

SFB 649 Discussion Paper 2017-010

# DATA SCIENCE & DIGITAL SOCIETY

Cathy Yi-Hsuan Chen\*  
Wolfgang Karl Härdle\*



\*Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>  
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin  
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN



## DATA SCIENCE & DIGITAL SOCIETY

**Cathy Yi-Hsuan Chen**  
**Wolfgang Karl Härdle**

[chencath@hu-berlin.de](mailto:chencath@hu-berlin.de)  
[HAERDLE@HU-BERLIN.DE](mailto:HAERDLE@HU-BERLIN.DE)

*Ladislaus von Bortkiewicz Chair of Statistics*  
*Institute for Statistics und Econometrics*  
*School of Business and Economics*  
*Humboldt-Universität zu Berlin*  
*Unter den Linden 6*  
*10099 Berlin, Germany*

*Sim See Boon Institute*  
*Singapore Management University*  
*50 Stamford Road, Singapore 178899*

### **Abstract**

Data Science looks at raw numbers and informational objects created by different disciplines. The Digital Society creates information and numbers from many scientific disciplines. The amassment of data though makes it hard to find structures and requires a skill full analysis of this massive raw material. The thoughts presented here on DS2 - Data Science & Digital Society analyze these challenges and offers ways to handle the questions arising in this evolving context. We propose three levels of analysis and lay out how one can react to the challenges that come about. Concrete examples concern Credit default swaps, Dynamic Topic modeling, Crypto currencies and above all the quantitative analysis of real data in a DS2 context.

**Keywords:** Data Science, Digital Society, social networks, herding, sentiments

**Acknowledgements:** We acknowledge financial support for this project from the Deutsche Forschungsgemeinschaft (DFG) through IRTG 1792.

# 1. Introduction

The Michelangelo painting in the sistine chapel, the image of the near touching hands of God and Adam has become iconic as it symbolizes the act of creation of life by sending an invisible spark to the human on the left. Imagine the original scene where on the left we see a young helpless, passive creature, that barely can raise the finger (in latin: digitus). With digitization we humans have created such a creature ourselves, not like just discretizing the act of life into a puzzle but by having created a new era of life, or even a digital life style. Today we have cheap computing power and unlimited access to information. We have digital products and creatures that still, so many hope, are a little helpless and barely can raise the digitus.

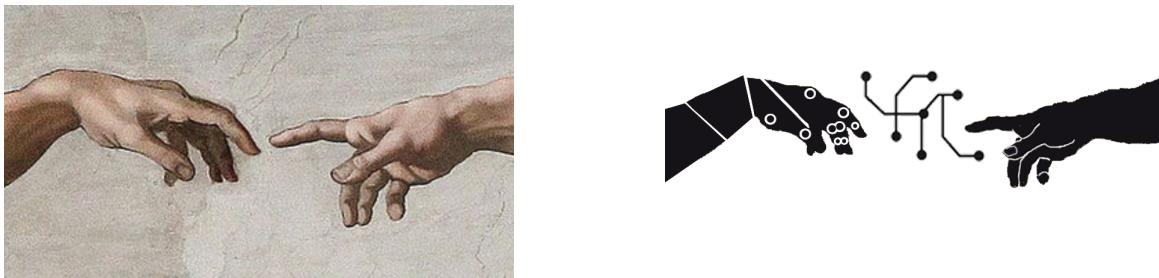
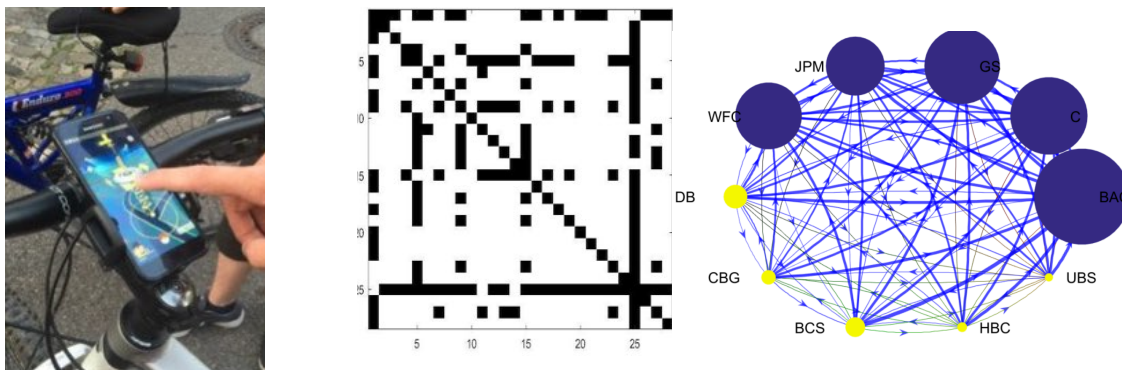


Figure 1 The touch of god to men and finally to digitus.

The combination of computing power and effortless data access though have uprooted numerous conventions regarding human relations. They have altered the way people perform tasks and how they interact with one another. How do digital technologies



change the fundamentals of economic relations? Digital goods tend to be easy to copy and difficult to appropriate; developing them, bringing them to market, and protecting their ownership pose great challenges to both businesses and legal institutions. More fundamentally, digital technologies challenge basic microeconomics tenets.

The thoughts presented here on DS2 - Data Science & Digital Society analyze these challenges and offers ways to handle the questions arising in a DS2 context. In the

next section we propose three levels of analysis and lay out how one can react to the challenges that come about.

*Figure 2 How Pokemon games enter networks and become actors and risk emitters.*

## **2. Digital Society & Data Science**

Imagine a mobile Pokemon GO situation. Clearly the touch of a finger by an individual connects people but exposes them more than ever. The players get connected and may start social neighborhood like activities. If one projects such a gaming connection into an adjacency matrix, also used for describing the interconnection of SIFIs (systemically important Financial Institutions) then one needs to ask, understand and characterise the evolution of such high dimensional data objects. Such an evolution is also observable in systemic risk stress scenarios where Credits Default Swaps (CDS) moves and Conditional Value at Risk (CoVaR) tail dependencies, play an important role.

These three examples are from the apparently disjoint disciplines computer science, statistics and economics but their very identical DS2 nature makes it possible to create cross discipline insights and deliverables for the acquisition of knowledge and the evolution of not only social networks in today's digital economy. References are but not limited to Chen S, Härdle WK, Liang C, Schienle M (2017), Chen CYH, Härdle WK, Okhrin Y (2017).

Let us boil it down in one sentence: A simple tap of a finger impacts the perceptions of many and at a higher level it affects the reality of production relations, finance, the law, the distribution of wealth and in general economic activity.

The DS2 key research foci are therefore:

*Digital technologies connect, empower and expose individuals*

*Dynamic interactions require smart data science*

*Digital society creates new business relations*

*Decision processes evolve in time*

*Data science impacts societies*

Partial analysis to problems related to these observations and questions have been given at different granulation levels in many disciplines but no overarching DS2 view yet has been established. Such an answer requires cross-pollination of e.g. economics, computer science, management, mathematics, statistics, and jurisprudence, but also

must involve views on digitisation of agriculture, energy science, geography and psychology. All these disciplines use data science and meet at these key research foci, must therefore create new structures of education, and must satisfy the DS2 research



demands. The above focal research questions may be organised into three areas and levels of complexity:

- A Micro      Decision making at the individual level
- B Meso        A dynamic network view
- C Macro      Aggregate consequences for DS2

*Figure 3 How Pokemon meets the bitcoins and CRIX the Crypto Index*

### **Micro level**

How are we advancing to answer the key questions? Area A studies behavioral actions and reactions of individual agents in processing digital signals and information at different levels of decision-making. Research here focuses on three cornerstones: social media, cognitive processing, and the optimization to best react to this information. One necessarily looks at very high dimensional multivariate objects, combining analyses of movements with behavioral analyses and e.g. neuro-scientific methods. Technology input is needed from area B Data Science. One may address how business models can be adjusted to fit consumer preferences in digital contexts. *Pokemon Go: download and play for free.* Better though: buy additional resources with bitcoins! Data science investigates the design of such freemium models and is neatly interwoven in a data science platform that will be described a little later.

A pure data science platform alone is not sufficiently endowed though to answer the research questions. Block chains have led us to a new era of net work based trust building. In these we observe herding effects, the power law and tail event dynamics. We need partners in academics and industry to create a trustful data science platform.

### **Meso level**

The meso level B is motivated by the confluence of individual decisions and collective allocations. This includes sentiment-driven stock-market reactions lifetime and

liquidity of crypto-currencies, the volatility of high-frequency financial data, see Chen S, Chen CYH, Härdle WKH, Lee TM, Ong B (2017), Härdle WK, Yu L, Wang W (2016), Härdle, W., Hautsch, N. and Mihoci, A. (2012).

This requires complex methods for modelling, aggregating and treating a large quantity of information. The main challenge being Network Emergence and Evolution. It is this dynamical aspect that we need to understand. The difficulty arising here is the relative proportion of size to dimension. Therefore dynamic information aggregation becomes inevitable for mere data processing. Machine learning and structured forecasting are essential co-domains. The huge data size of the objects studied requires large-scale automatization of operations and decisions. This requires though fair evaluation for risk aware decision making. A tilted evaluation will create a spill over, measurable at macro level which we study in area C.

### **Macro level**

By creating and destroying jobs, the digital economy affects all production factors. In the limit, the potential for maximal efficiency induces allocative inefficiency, and raises legal issues on antitrust and intellectual property rights. Larger networks are susceptible to security breaches and block chains become increasingly important. A second point is the reliability. What is the value of information, what is the veracity of content and what is the external effect of trust and truthfulness? In future digital times, images, movies blogs and tweets may move popular opinion faster than ever, herding effects create sentiments and affect prices. At the same time, crowds can be wise at times, and that trust-building is a central economic, psychological and institutional feature.

Finally, inequality is a great concern because access to public goods, public data, and public networks is unequal to begin with and can aggravate the inequality of income and wealth. Uneven access will threaten the equality of opportunity and social cohesion. Access to information, education and jobs are hallmarks of good governance in democracies with well-informed electorates. A great source of inequality is climate change, because mitigation strategies require rich information. At the same time, linking workflows of the digital economy with analyses of space data creates opportunities to overcome such inequalities.

A research data centre (RDC) plays a key role. First, it will store, organize, and provide data services to the involved disciplines. It will provide a hub for massive data processing. Second, the RDC hosts the Quantnet ([www.quantlet.com](http://www.quantlet.com)) infrastructure, which is an innovative mechanism to allow cooperating researchers all over the world, to access data, codes in different languages, and results developed in the DEDA scope using the GitHub platform. Such a non central centre like RDC is a great way to achieve transparency and replicability of research output and hence disseminates DS2 findings.

### **Concrete Situations**

DS2 is not just a bag or a combination of ideas. The invisible spike from the above Michelangelo painting that creates its life also invokes organic links. Data Science guarantees the stability of the research on the DS2 questions.

Here is an example: High frequency trading lets us identify liquidity problems and a network of stock flow. One may with data science examine the interfaces between regulation and innovation in digital goods and the share economy: Platforms such as Uber and AirBnB raise regulatory questions. Labour relations are impinged, on the other hand unprecedented connectedness enables workers to organize and mobilize to improve working conditions. Actions through social media channels can effectively sabotage the reputation of large firms. The coalescence of business practice and decision analytics leads to sophisticated algorithms to exploit market opportunities. Deep learning techniques lead to an increasing number of decisions being taken fully automatically and, accordingly, to the exclusion of human beings.

The next example is a dynamic network of loans of a banking system. Its flow or adjacency matrix may be analyzed for its centrality in order to check for systemic risk. The flow of risk emitters and receivers is given above in the first Figure 1. One sees the dominant US risk emitting banking sector, Xiu X, Chen CYH, Härdle WK (2017)

**Dynamic Topic Modeling.** DTM is a text mining tool that uses a bag of words model for the flow of topics over time. It may be employed to check for upcoming topics, forecasting events, analyzing flow of docs, Linton et al (2017). It is based on scraping via data science articles from e.g. NASDAQ news pages and distill then sentiments that may create additional information about future stock moves. Recent references are Linton M, Teo EGS, Bommers E, Chen CYH, Härdle WK (2017), SEC ruling, (2017), Zhu X, Wang W, Wang HS, Härdle WK (2017)

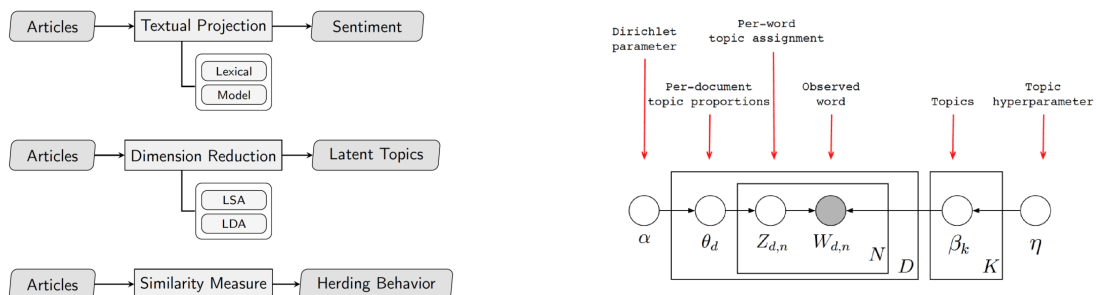


Figure 4 The DTM mechanism

**Face values of companies.** The face expressions are correlated with the stock of here Deutsche Bank.

Cryptocurrencies, [hu.berlin/crix](http://hu.berlin/crix) are yet another DS2 example where the above mentioned disciplines come together naturally to check their vola and trends. Here

data science in the big picture comes into play in terms of:

- Dynamic Risk Structuring
- High Frequency Markets
- Dynamic Topic Modelling
- Forecasting Volatility
- Transparency of Quant Data
- Crypto currencies acceptance
- Cyber security insurance
- Herding behavior



Figure 5 How Faces correspond to stock values

## Structural Goals

DS2 has several structural goals. One needs to create and induce an MDE Master in Digital Economy. The MDE will expose students to the full range of concepts and challenges in the modern world. This requires combined knowledge elements and joint research efforts from business, economics, computer science, law, and environmental sciences in other words those that come together here. MDE students will be familiar with economic theory and how international contract law governs real-life economic transactions. They will have a firm understanding of digital business strategies and how their implementation requires smart software systems. They will possess the skill to develop and estimate empirical models for decision analytics. At the same time, a firm understanding of psychological fundamentals will allow them to understand the cognitive process of decision making and how it is moderated by technologies. This diverse set of skills will allow MDE alumni to find answers to the most pressing questions of the digital economy.

## Conclusion



Data Science looks at raw material created by e.g. the Digital Society. We are all interested and highly dependent on a skill full analysis of the massive data material. The thoughts presented here on DS2 analyze these challenges and offers ways to handle the questions arising in this evolving context. We propose three levels of analysis and lay out how one can react to the challenges that come about. Concrete examples concern Credit default swaps, Dynamic Topic modeling, Crypto currencies and above all the quantitative analysis of real data in a DS2 context.

## References

Chen S, Härdle WK, Liang C, Schienle M (2017) Network dynamics of high-frequency trading data, J Econometrics submitted.

Chen CYH, Härdle WK, Okhrin Y (2017) Tail event driven networks of SIFIs. SFB DP2017-004 , J Econometrics, submitted

Chen S, Chen CYH, Härdle WKH, Lee TM, Ong B (2017) A first econometric analysis of the CRIX family, in Handbook of Blockchain, Digital Finance and Inclusion, Vol 1, Cryptocurrency, FinTech, InsurTech, and Regulation, David LEE Kuo Chuen Robert Deng, eds. ISBN: 9780128104415, Academic Press, Elsevier

Härdle WK, Yu L, Wang W (2016) TENET - Tail Event driven NETwork risk. J Econometrics, 192, 2, 499-513 DOI: 10.1016/j.jeconom.2016.02.013

Härdle, W., Hautsch, N. and Mihoci, A. (2012) Modelling and Forecasting Liquidity Supply Using Semiparametric Factor Dynamics, J. Empirical Finance. 19, 610-625, DOI <http://dx.doi.org/10.1016/j.jempfin.2012.04.002>

Linton M, Teo EGS, Bommers E, Chen CYH, Härdle WK (2017) Dynamic Topic Modelling for Cryptocurrency Community Forums. in Applied Quantitative Finance (Härdle, Chen, Overbeck eds) Springer Verlag, ISBN 978-3-662-54486-0

SEC ruling, (2017), <https://www.sec.gov/rules/sro/batsbzx/2017/34-80206.pdf>

Xiu X, Chen CYH, Härdle WK (2017) Dynamic credit default swaps curves in a network topology SFB 649 DP 2016-059

Zhu X, Wang W, Wang HS, Härdle WK (2017) Network Quantile Autoregression J of Econometrics, submitted SFB DP 2016-050

# SFB 649 Discussion Paper Series 2017

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Fake Alpha" by Marcel Müller, Tobias Rosenberger and Marliese Uhrig-Homburg, January 2017.
- 002 "Estimating location values of agricultural land" by Georg Helbing, Zhiwei Shen, Martin Odening and Matthias Ritter, January 2017.
- 003 "FRM: a Financial Risk Meter based on penalizing tail events occurrence" by Lining Yu, Wolfgang Karl Härdle, Lukas Borke and Thijs Benschop, January 2017.
- 004 "Tail event driven networks of SIFIs" by Cathy Yi-Hsuan Chen, Wolfgang Karl Härdle and Yarema Okhrin, January 2017.
- 005 "Dynamic Valuation of Weather Derivatives under Default Risk" by Wolfgang Karl Härdle and Maria Osipenko, February 2017.
- 006 "RiskAnalytics: an R package for real time processing of Nasdaq and Yahoo finance data and parallelized quantile lasso regression methods" by Lukas Borke, February 2017.
- 007 "Testing Missing at Random using Instrumental Variables" by Christoph Breunig, February 2017.
- 008 "GitHub API based QuantNet Mining infrastructure in R" by Lukas Borke and Wolfgang K. Härdle, February 2017.
- 009 "The Economics of German Unification after Twenty-five Years: Lessons for Korea" by Michael C. Burda and Mark Weder, April 2017.
- 010 "DATA SCIENCE & DIGITAL SOCIETY" by Cathy Yi-Hsuan Chen and Wolfgang Karl Härdle, April 2017.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
**<http://sfb649.wiwi.hu-berlin.de>**

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

