Doctoral School of Regional Sciences and Business Administration

Attila Sóti

Complexity Theory in Social Studies: A Portfolio of Applications for Understanding Real-World Phenomena

Doctoral dissertation

Supervisors: Gábor Vattay, DSc, Professor
Eötvös Loránd University, Budapest
Department of Physics of Complex Systems
Ancarani Alessandro DSc, Professor
University of Catania
Department of Civil Engineering and Architecture

Győr, 2021

Attila Sóti

Complexity Theory in Social Studies: A Portfolio of Applications for Understanding Real-World Phenomena

Supervisors: Gábor Vattay, DSc, Professor
Eötvös Loránd University, Budapest
Department of Physics of Complex Systems
Ancarani Alessandro DSc, Professor
University of Catania
Department of Civil Engineering and Architecture

Győr, 2021.
Széchenyi István University

Author's Declaration

No portion of the work referred to in this dissertation has been submitted in support of an application for another degree or qualification of this or any other university or any other institution of learning.

Furthermore, this dissertation contains no material previously written and/or published by another person, except where an appropriate acknowledgement is made in the form of bibliographical references, etc.

Abstract of the dissertation.
submitted by Attila Sóti
For the degree of Doctor of Philosophy,
Entitled: Complexity Theory in Social Studies: A Portfolio of Applications for Understanding Real-World Phenomena

Month and year of submission: MARC 2021

Abstract

Our world is complex. This complexity is the result of the interactions among socio-political-economic-environmental factors, the multitude of potential relationships between the individual components and the multiplicity of elements involved. It can be grasped by the "amount " of information needed to describe a system: the more complex a system is, the more information is required to describe it. On the other hand, we cannot know all the components of the system or how the components are connected to each other. Any given system with high complexity at the microlevel needs many variables to describe it. But it is no longer certain that at the macro level the complexity of the same system remains the same, so it is possible that it can be described with much less information.

Complex systems are observable in our everyday lives in many ways, for example in the living cells, the Earth's climate, the communication systems, the human relationships, stock markets, or economic markets**.** These examples cover several scientific disciplines. An interdisciplinary approach can be achieved when we work across disciplines to solve problems. The methods and concepts of complex systems can be applied to a wide range of fields in society and economics, biology or robotics. Up to nowadays, the lack of computing capacity and the inability to acquire and process large sets of data prevented the detailed exploration and simulation of such systems. Recently, the availability of digitized data has increased. This opens the possibility of studying areas such as the behaviour of individuals within a group, the analysis of web traffic, user feedback, stock market transactions etc. As a result, new models can be built and tested for emerging collective phenomena. In my dissertation, I researched human behavioural patterns with the help of the investigation methods of complex systems. There are now various sources of data available to support such research. I decided to use geo-referenced posts from the billion wide Twitter's database, stock market data, custom reports, and user

complaint messages from these data sources. With the help of the Internet, billions of people use various online community platforms to maintain relationships. Users send messages, comments, join specific groups, read and respond to the news, leave behind digital footprints. Collecting this information is critical for building successful models of complex systems. Digital human footprints often contain sensitive data. After careful anonymization, encryption and aggregation, meaningful insight can be gained into human behaviour without violating anyone's rights. However, the understanding of socio-economic phenomena does not merely stem from data collection, as it lacks structural insight and context and does not explain patterns that have been identified. The existing models that describe patterns of human behaviour often rely on earlier concepts of complexity theory. Therefore, in a new approach, methods should be developed that link microscopic data to macroscopic observations. Then, the results of these methods need to be evaluated, scaled, and compared. In this thesis, I would like to contribute to the evaluation of human footprints in digital data with the ultimate goal to learn more about the forces that are moving humanity.

At the beginning of the twenty-first century, the business and economic environment in the world changed significantly. The world is organized into a single complex, global market. With the expansion of online markets, capital can be immediately transferred from one place to another. Small changes can multiply very quickly in this market and lead to dramatic changes. Technical development is swift, so the competitive position of companies is continually changing. Market movement is inherent in everyday life. In such economic circumstances, we can understand what is happening around us better through chaos and complexity theory, than through traditional economic theories. Theories of chaos and complexity act as a real challenge to conventional economic theories and question traditional interpretations of economic equilibrium.

The study can be summarized in the identification of **one comprehensive and homogeneous approach that has been applied to five subareas**.

In general, the developed models rely on methods that link microscopic data with macroscopic observations. I evaluated, scaled, and compared the results of these methods, hypothesizing that, based on the methods used in the papers, we understand social reality better through chaos theory.

With respect to the five subareas:

- I have observed the development of a football team based on changes in order to study phases of a complex system. I hypothesize that understanding the motivations of smaller groups in the team and successfully integrating this knowledge to understand higher-level elements will facilitate understanding the complex transitions in a more extensive system.

- One way to understand stock market movements is to interpret the stock market as a complex system. I have highlighted the features of complex systems in the two different articles to show how small changes in the input can cause significant changes in the output. In my predictions, the goal was not to give an accurate estimate of what would happen in the future. The aim was to outline realistic scenarios, alternatives that could point out the way for the future. I hypothesize that there is a group of individual investors whose decisions about investing in a football club are driven not only by considerations of their long-term well-being but also by their daily emotional state related to the clubs.

- One of the basic assumptions of economic theory is that economic actors always act rationally However, there are significant problems with applying this assumption as we observe that some events have a greater impact on the market than we would expect. Because of this fact, one needs to pay more and more attention to nonlinearity issues. For example, today the big innovative companies have taken advantage of the web. When innovative products are created, the users of these products have to adapt to new challenges. They need to be familiar with the digital world. I hypothesize that there is a link between the frequency of the appearance of these companies in social media and the volatility of stock prices of these companies.

- The network of football fans also forms a complex system. Complexity is referred not only to the background but also to the methodology. Using the power of computing, I can prove relationships among the components of the system that help to understand better the whole system. By using modern technical equipment, significant events related to the biggest football clubs can be tracked from anywhere. In this direction, mapping the background of these clubs in order to understand the motivations of their supporters plays an important role. By using the urban scaling theory, I measured how

the size of a city relates to the number of football fans, hypothesizing that the lower is the GDP of a country, the more fans live in cities. The results of the study can be supplemented by the smart device usage habits of the people living in the given area and the internet coverage may differ between towns and villages in countries with lower GDP.

- The IT infrastructure also forms a complex system. In enterprise environments, the analysis of failure phenomena connected to this infrastructure is paramount. Complexity usually assumes some sort of hierarchy. The components of the system at a certain level of the hierarchy can interact with each other. What makes complex systems interesting is that as a result of the interactions between their parts, the behaviour of the parts changes in such a way that the whole system follows a qualitatively new pattern of behaviour. Analysing bug reports can help to find broader relationships across systems. I hypothesize that usage of semantic analysis with other data mining techniques can help to find the focus, patterns and trends in the texts connected to user feedback.

Acknowledgements

I would like to thank first of all to my family: my mother Etelka and my father Antal, who, although they can no longer be with me, have raised me to be aware of the importance of knowledge since my childhood. To my cousin, Karolina Koza, for her perseverance, precision, and never-ending optimism, who heroically endured with me during the final phases of the work and helped to explicate the texts. To my niece, Réka Sóti, for her helpful advice and huge energy, my brother Zsolt Sóti for setting a good example, for his faith and every word of encouragement, to my daughters Csenge and Dorottya for their youthful impetus and special thanks to Anikó Fehér, who endured and encouraged me when I was tired and grumpy.

I would like to thank my supervisors, Dr Gábor Vattay and Dr Alessandro Ancarani, for their support during my doctoral years. I am grateful that they have encouraged me to develop my skills and ideas and that I was trusted and provided with great independence. It meant a lot to me that I could turn to them with all sorts of questions and always count on their advice. Thanks to Dr. József Stéger and Dr. Eszter Bokányi for their benevolent and persistent help and expert advice for more than a year at the Department for Physics of Complex Systems at Eötvös Loránd University.

I would also like to thank the special attention of each member of the SEEDS team; Dr. Zoltán Baracskai, who, with his vast experience and tempered humour, has always helped me to get back on track. The former and current leaders of the school, Professor László Komlósi and Professor Márta Konczosné Szombathelyi, who have provided me with this fantastic opportunity to enter the world of science, as well as Tihana Vasic and Dr. Katalin Feketéné Czakó for their consistently precise and helpful work. Last but not least, I would like to thank my fellow students Laci Buics, Laci Szendrődi, Zoli Dobra, and Máté Dr. Tóth, for our cooperative work, unforgettable experiences and everything else.

List of Tables

LIST OF FIGURES

Key to Abbreviations

API	Application Programming Interface
BM	Bayern München
CCDF	Complementary Cumulative Distribution Functions
COVID 19	Coronavirus Disease
CRISP-DM	Cross-industry Data Collection Process – Data Mining
G7	The Group of Seven
GDP	Gross Domestic Product
GPS	Global Positioning System
FA	Football Association
FIFA	Fédération Internationale de Football Association
IT	Information Technology
JUVE	Juventus
LA Galaxy	Los Angeles Galaxy
NYSE	New York Stock Exchange
NLP	Natural Language Processing
MU	Manchester United
OLS	Ordinary Least Square
OSN	Online Social Networks
RCA	Root Cause Analysis
RM	Real Madrid
TV	Television
UEFA	Union of European Football Associations
UK	United Kingdom
USA	United States of America

Table of Contents

Introduction

This thesis has developed quantitative studies related to complex local and global systems that interweave our everyday lives. Most of the case studies analyzed are referred to as examples in the sports industry. This choice is motivated by several reasons. Firstly, nowadays sports are part of the global business world, and in the last two decades, it became a big business. Secondly, sport has been impacted by the globalisation that defines today's social mechanisms. Globalisation started earlier in the field of football than in the other areas of sport because it quickly became an international cultural phenomenon. Further, it is an industry with unparalleled global scope and power. Globally, sport-related turn-over accounts for three per cent of the world total economic activity. Significant events such as the football World Cup are watched all around the world.

Background and motivation

In recent years, almost every field of science has been dramatically transformed by the technological explosion. High-capacity computers have enabled data collection and analysis in several disciplines, significantly refining our picture of the world and laying the groundwork for many new applications. Handling a huge quantity of data requires new approaches and new models. In this aspect, the models I have developed rely on the results of the application of complexity science that describes patterns of human behaviour. The aim was to apply methods that make sense of macroscopic observations using a large number of essential data. Through the complexity theory used in the papers, we can better understand the given social environment. In light of the results, the correlation within the components of the environment becomes more transparent.

In the first for instance, in studies insights into the operation of complex systems has been gained by studying the link between popularity and development of football teams and big companies. Second for example several aspects of our lives are influenced by the possibilities offered by the environment in which one lives. As such, complex spatial structures and the dynamics of changes in them have been the focus of the next study.

Finally, in the corporate environment, analyzing the description of incidents related to this infrastructure is of paramount importance. Analyzing bug reports can help to find greater connections between systems.

Structure

In the first study, I consider the development of a football team as a complex system over time. In the second study, the stock market movements of football teams are analyzed. In the third study, I investigated the stock market movements of large international companies, whereas in the fourth study, the relationship between cities as complex systems and the most popular football teams. In the fifth study, I analyzed larger text files that appeared in large corporate environments. To understand the motivation behind the analyzes, in the first three chapters I give a brief overview of:

1. The economic development of today's global world
2. On the theory of complex systems and
3. About the methodological aspects I used in writing the following studies.

These summary studies help us to understand the relationships between the studies described in the dissertation such as the appearance and impact of today's global thinking, the appearance of common basic features of complex systems, and the use of common research methodologies in different areas of life. There is also a brief explanation in the studies of dissertation how the information in each section relates to the world of "big data".

1. Globalisation

1.1 Globalisation before the 1970s

In this dissertation, I researched human behaviour patterns using methods of complex systems. As a first step, I consider it essential to present a historical overview of our current socio-economic environment. Today's human social environment can be understood if we understand the process of globalisation. (Pretty & Ward, 2001). Globalisation is a process, by which the world is becoming increasingly interconnected because of massively increased trade and cultural exchange (Haivas, 2003). A better understanding of the past, the motivations behind, and the effects of globalisation are more urgent than ever before (Lamb, 2004). The two main forces for sustained cultural interaction before this century have been wars and religious conversations (Appadurai, 1990). The invention of printing enabled the spread of ideas (Clanchy, 1983). By the early 1900s, it was seldom to come across a town that was not influenced by foreign markets. All significant innovations, such as the automobile, electricity, chemicals and cinema emerged in most of the western countries at the same time. Mass production and assembly lines appeared in some corporate giants, so the appropriate quantity of goods became available (Boysen, Fliedner, & Scholl, 2007). Globalisation leads to the expansion of the production of goods and services. With progression in ship and rail transport and electronic communications, trade with other parts of the world became much more manageable. Towns were no longer restricted to what they alone could manufacture and what nearby towns could trade with them. The industrial potential of the world started to grow fast. (Figure 1) These developments in economic globalisation were disrupted by World War I (Christian, 2011a).

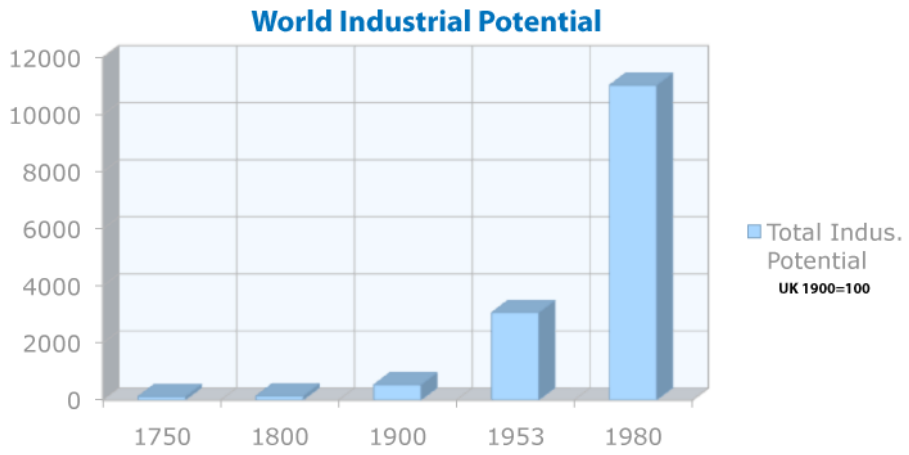

Figure 1 World Total Industrial potential - 1750 -1980

Source: (Christian, 2011b) (https://archive.org/details/B-001-015-437/page/n433/mode/2up)

It is noteworthy that 50% of the shares of global income were produced by the seven wealthiest countries in 1910, and we can see in Figure 2 how rapidly the USA, Europe, and Japan increased their Industrial potential in 1880. This proportion has not changed significantly later (1980) (Skaff, 2001).

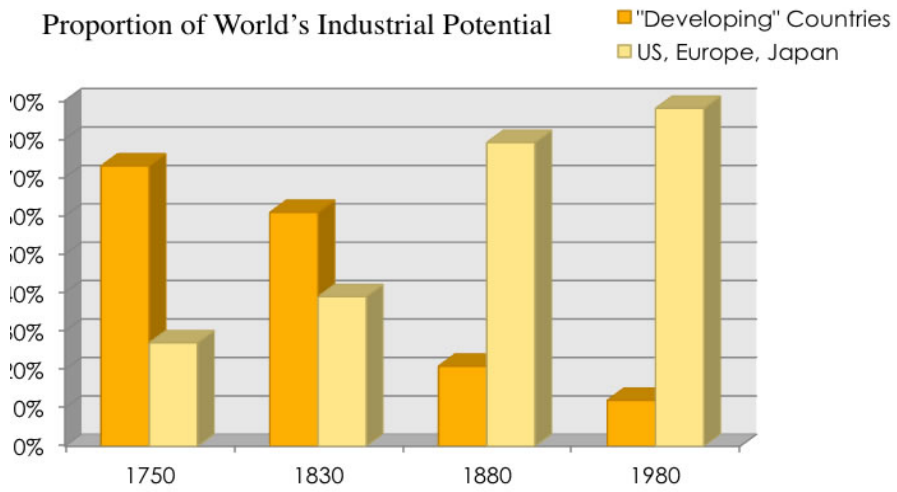

Figure 2 Proportion of World's Industrial Potential (11750 -1980)

Source: (Skaff, 2001) (http://webspace.ship.edu/jkskaf/INT%20200/6-Egypt%20&%20Economic%20Globalisation.html)

Subsequently, in the second era of globalisation culture, new technologies, and ideas appeared all over the world (Latham, 2001). Human knowledge, expertise and experience were globalized (Spring, 2008). Cultural globalisation increased cross-cultural contacts which helped the more developed societies to affect the other parts of the world. (like Hollywood**,** Jazz, Hamburger) (De Zoysa & Newman, 2002). We should not forget that the term globalisation implies transformations (Kellner, 2002). It has greatly influenced the culture, as this is the most basic cause for a person's wants and consumer behaviour. When people are changed, it generates different new needs.

Through discoveries and economic advancement (transport, telegraph, etc.) implementation cost of the first two phases of globalisation fell dramatically, contributing to the unprecedented growth of international trade. After World War I., most of the global economic powers constructed protectionist economic policies and introduced trade barriers (Keohane, 2005). This caused slowing down of worldwide trade and even led to some countries introducing immigration limits. The idea of the nation-state dominated political life and had a broad scope of well-known outcomes. In 1944, all in all, 44 nations attended the Bretton Woods Conference (James, 2013) with the aim of stabilizing the world's currencies and establishing credit for international trade in the post-World War II era. This conference commenced organisations such as the World Bank, the International Monetary Fund, and the International Trade Organisation, with main aims that became the fundaments of the global economy and global financial systems (Woods, 2014).

1.2 Globalisation after the 1970s

Globalisation had not recovered until the 1970s when all governments began to emphasise the benefits of trade (Braun & Raddatz, 2008). There were reforms in the socialist countries in the 70s (Csanády, 2012). The market-driven economy appeared gradually after the plan-driven economy. In other countries, such as India and Egypt, the market-driven economy started taking over the original ways of production ('Cambridge Econ. Hist. India', 1983). The industrialised states, except Japan, struggled with an economic recession due to the oil crisis that was caused by oil embargoes by the Organization of Arab Petroleum Exporting Countries (Cooper & Karl, 1998). The crisis saw the first instance of stagflation (a tighter monetary policy that did not reduce inflation, but it did lead to a smaller recession that increased unemployment).

That began a political and economic trend that has replaced Keynesian economic view with neoliberal economic theory (in the Keynesian view, aggregate demand does not necessarily equal to the productive capacity of the economy; but it is rather influenced by a host of factors such as production, employment, and inflation) (Cogan, Cwik, Taylor, & Wieland, 2010). This change included extensive economic liberalization policies such as privatisation, fiscal rigours, free trade, and reductions in government's spending, in order to increase the role of the private sector in the economy (Hazlitt, 1960). A combination of different elements, including the continued mass mobilisation of capital markets through neo-liberalism, the end of the decades-long Cold War (Lightbody, 2005), the beginning of a rapid increase of new media such as the Internet, and the end of the Soviet Union led to a reorganisation of economic and political authority across the world (Machowski, 2019). In the 2000s, we can see a significant downturn in the value of dot-com shares. However, the Internet keeps growing as a business and advertising medium, with permanent increases in online shopping and banking activities. We also need to know that in recent years, the characteristics of the Global trade are the developed countries purchasing the raw material from developing countries and the poorer countries buying products manufactured in the richer (more developed) countries. This process is increasing because the more prosperous countries have more money to spend on development, as well as to support a well-functioning economy (McKinsey Global Institute, 2019). It is easier to take a loan because there will be lower interest rates (the size of interest rates is in direct proportion to investment risks) (Appadurai, 2018).

1.3 Effects of globalisation

Globalisation has dramatic and varied effects on world economies, culture, political views and people's lives (Appadurai, 2014). For example, inward investments help countries by providing new jobs and upskill local people (Yen & Yen, 2016). Multinational companies can contribute to local economies by increasing wealth and bringing in foreign currency when they buy local resources, products and services. The additional money created by these investments can be spent on education (Sunthonkanokpong, 2011) health and infrastructure. In the field of culture: the sharing of ideas, experiences, and lifestyles is easier (Berger, 2000). However, we must take into consideration that globalisation benefits mainly the interests of the richest countries, that continue to dominate world trade at the expense of the developing

countries. The role of less economically developed countries in the world market is mainly to supply the North and West with the low-cost workforce as well as raw materials (Mosley & Friedman, 2006). It is not certain that the wealth from the inward investment will benefit the local community (Wade, 2004). Often, profits are sent back to the country where the multinational company is based. It can also happen that big companies, with their massive scale of economies, may drive local companies out of business (Bhattacharya & Michael, 2008). Due to an absence of strictly enforced international regulations, they can cause pollution, poor working conditions or low wages.

Cultural globalisation is viewed by many as a threat to the world's cultural diversity (Tong & Cheung, 2011). It is feared that it might drown out local traditions and languages. For example, a film made in Hollywood is far more likely to be successful worldwide, than one made in Hungary or Indonesia.

1.4 New reasons for globalisation in the 2000s

Labour availability, skills and improvement of communications became essential factors in the 2000s. Countries, such as India, have cheap labour costs and high skill levels (MacCarthy & Atthirawong, 2003). The Internet and mobile technology have enabled more excellent and faster communication between people in different parts of the world. The workers can work remotely, and therefore, the movement of services is more frequent than previously (Sauvé, 2003). It is also essential that not only work must be transferred but work conditions, ideas and way of thinking also need to be relocated. However, it is evident that the major starting points for involving employees in such monumental changes are communication and education (Lupton, 2014). One of the critical factors of organizational success is to bring new processes or technologies across borders and involve co-workers in innovations in order to support them to understand the whole process in which they should be productive. Three factors can accelerate economic globalisation in developing countries (Garrett, 2000):

- Progress of science and technology
- Market-oriented economic reforms
- Contributions of multinational corporations

However, we should know that these changes have an impact not only on the economy but also on cultural and political life, such for example, the Arab Spring (Khondker, 2011). In a world of growing attention to the economy, culture has become one of the key drivers of change (Holton, 2013). It can lead to two outcomes; homogenization or heterogeneity (Nacar & Uray, 2016). Homogenization is the situation when the richer culture has invaded the local culture, and it has become the dominant culture in the local area, that aims to eliminate the local culture. Society, therefore, becomes homogenous. Everyone conforms to the western ideal (Boli & Lechner, 2015). It also results in the loss of individual culture and religion. There is higher market competition as well (Kuo & Ye, 2009). Cultural models move through time and space where they come into contact with other cultural forms and settings and therefore influence each other (Pellegrini, 2010). New forms and changed cultural settings are produced (P. G., 2018). On the other hand, cultural heterogenization or multicultural society means that regional culture is widely disseminated (Choi & Park, 2014) and welcomed by other cultures/societies. In the meanwhile, the cultural diversity in local society is enhanced.

It can result in richer countries giving various incentives to poorer countries in order to protect their natural environment, as well as adopting more sustainable work practices (Annepu, 2012). We can observe that global criteria gradually replace local perspectives and reigniting old debates on individuality versus universality, modernity versus tradition etc. (Popescu, 2014). Globalisation can bring good ideas to the communities when it helps renovation (Appadurai, 2000).

1.4.1 Global market

If the goal of the study is the patterns of human behaviour, it cannot avoid observing the market and the most significant market is the global market. So, if the industry begins to be developed in emerging countries at the expense of jobs in wealthier countries, in the next years, we may face some surprises from the voters in developed countries, just like in the cases of Brexit (Irwin, 2015) and presidential elections in the USA (Rodríguez-Andrés, 2018). We know that the losers of globalisation are mostly lower-middle-class workers from developed countries (Baldwin, 2006). We also know that the global share of the total income of the G7 countries has fallen dramatically in just a few decades. If this trend continues, developed countries will clearly need to do something. (Global Market Insights, 2016) However, the questions are: will this be the end of the neoliberal economy and political culture in these countries, or will they

find alternative solutions? What will happen with the spread of cultures? If the economic links are changing, will the cultural preferences also change?

1.5 Common cores of our behaviour: risk, uncertainty, trust, innovation, failure and creativity.

The literature on the concepts developed in this chapter is vast. It goes beyond the scope of the dissertation, but I consider it important to describe the basic concepts for an easier understanding of my research.

Social capital includes the relationship between trust, reciprocity, standard rules, norms and sanctions (Pretty & Ward, 2001). In the next chapter, I look for the answer to: Where do these core values come from?

1.5.1 Risk and Uncertainty

Risk and uncertainty are an essential part of any innovation. Companies that are launching new and innovative products on the market are taking some kind of risk and uncertainty (Knight, Of, & Classics, 1921).

In the terminology of economics risk and uncertainty are defined as two entirely different terms. The difference between the two concepts is much more emphasised than in everyday speech. These two terms were in economics defined by Knight (Knight et al., 1921).

In the case of risk, the future is unknown, but the probability distribution of the future is known. Uncertainty, on the other hand, is characterized by both an unknown future and an unknown probability distribution.

In other words, the risk is present when future events have a measurable probability of occurrence; and uncertainty occurs when the probability of future events is unpredictable.

Key differences between risk and uncertainty

1. Risk is defined as the gain or loss of something valuable. Uncertainty is a state where there is no knowledge of future events.
2. Risk can be measured and quantified through theoretical models. Uncertainty is not quantifiable because future events are unpredictable.
3. The possible outcomes are known in terms of risk, while in the case of uncertainty the results are not known.

4. The risk can be controlled if appropriate measures are taken to control it. On the other hand, uncertainty is not under the control of that person or business either, as the future is uncertain.
5. The risk can be minimized by taking the necessary precautions. In the case of uncertainty, this cannot be minimized either.
6. In risk, probabilities are assigned to circumstances, whereas in uncertainty they are not possible. (Knight et al., 1921)

Knight aimed to research the processes in business and explore the nature of profits.
His primary goal was to classify the decision-making problems faced by the entrepreneurs: that the events can be considered certain, possible and uncertain. To the latter, he formed and estimation and attached a probability whenever that was possible.

Risk can be calculated by all companies right from the start of innovation. The method of calculation is fundamental in this case. In mathematics, the risk is the measurement of the likelihood of satisfactory or unsatisfactory results. Determining the probability of a combination of hazardous events for every business is useful. This can be, however, very complicated. We can easily understand that many executives are wary when it comes to the potential implementation of new ideas (Kahneman & Tversky, 1979). They prefer renovation rather than innovation.
Advances in science and technology can bring significant social benefits and help economic growth. They support the creation of new products and services. We know that risk can be actuarial, objective, subjective, intersubjective and perceived (Kelle & Miller, 2001). For us, however, the most critical question is: How risky innovation depends on the people who use our new product. As we cannot predict the behaviour of a particular individual, however, we can predict the behaviour of the mass, so in the majority of cases, uncertainty transforms into risk.
Global companies are complex units held together by a control network. Loosening these controls may entail unacceptable risks and costs regarding the activities of innovation teams. When thinking about the consequences of innovation, managers need to keep in mind the constraints of the models on which people base their decisions when applying innovation. Some models are only suitable for specific applications. The more complex the system that innovation

introduces, the more likely and serious the consequences. In general, many of the risks associated with innovation do not stem from innovation itself but from the infrastructure in which it is implemented. A new way of thinking is needed (C. Wang, Cheng, Yue, & McAleer, 2020). A systematic risk management approach that can encourage a controlled, consistent and flexible decision-making process within the organization. (Edwards & Bowen, 1998) described risk management as a combination of the following processes.

- Creating the right context
- Recognizing the risks of an innovation project involving stakeholders
- Estimation of identified risks
- Developing responses to recognized risks.
- Monitor and follow up risks throughout the project life cycle.
- Enable risk analysis after project completion.

There are areas where pioneers face more risks than other products or services. The most important benefit of innovation is that it has a monopoly on the innovative product before any of the other competitors. This is a short-term benefit. The long-term benefit is that when other competitors enter the market, the pioneer has the advantage of being the first to know the market movements. The pioneering brand has more loyal customers; they save on switching costs and have a wide range of products that anticipate competition. Innovation cycles follow non-linear sequences. These development cycles must be well understood. If this does not happen the uncertainty will grow.

1.5.2 Trust

Despite being childless at an old age, Abraham had the promise from God to become the ancestor of a great nation. Abraham believed in this promise, and he was willing to make a sacrifice for his faith.

Our current world is based on these ethics, we trust when something has been promised to us. All contracts are based on this convention. Max Weber explains in his book “The Protestant Ethic and the Spirit of Capitalism” (Weber, 2005) that the uncertainty of salvation under the Calvinist dispensation later became the very defining characteristic of the financialized

capitalism of our times. Weber's approach helps us to unpack the logic of contemporary finance (Appadurai, 2014).

If we examine the meaning of the word trust, we will find out that the word trust means in many languages' **belief** as well.

We can understand trust to be some form of a personal expectation, anticipation, belief in the positive outcome of some particular event.

This expectation (although generally cognitive-rational in essence) arises not exclusively from a rational basis but contains emotional elements as well.

In economics, we examine trust as an element that influences economical transactional costs.

In the economic theory based on the analysis of transactional costs defined by Williamson (Williamson, 1993) trust appears as an element that is suitable for lowering these costs.

In this sense, this means the expectation that the other person will behave in accordance with his obligations. That he will be trustworthy in the negotiations and will not take advantage of his partners. The positive experiences amongst the participants of the business world increase credibility and trust. This credibility changes the uncertainty into risk in business.

1.5.3 Derivatives - when one buys a promise.

The financial expression derivative is a contract that obtains its value from the performance of an asset, index, or an interest rate, and is often simply called the "underlying entity". Derivatives can be used for many purposes, including insurance against price movements, speculations related to the increasing exhibition to the price movements, or getting access to otherwise difficult-to-trade assets or markets. The most common derivatives include forwards, futures, options, and variations of these. Derivatives belong to one of the three main categories of financial instruments, (other two being stocks and debt-based instruments, such as bonds) can be roughly classified as "lock" or "option" products. Lock products, for example, swaps, futures, or forwards, obligate the contractual parties to the terms throughout the validity of a contract. The option products, for instance, interest rate swaps, provide the buyer with the right, but not the obligation to enter the contract according to the specified terms. From the economic point of view, we can say that financial derivatives are cash flows, which are conditioned stochastically and discounted to the actual value. The market risk inherent in the underlying asset is attached to the financial derivative through contractual agreements, and hence it can be traded separately. That is, the underlying asset does not have to be necessarily

acquired. Derivatives, for that reason, allow the breakup of ownership and participation in the market value of an asset. This also provides a considerable extent of freedom regarding the content of the contract.

That contractual freedom enables the modification of the participation in the performance of the underlying asset almost arbitrarily. Hence, specifically, the risk related to the market price of the underlying asset can be controlled in almost every situation (figure 3).

Figure 3 What is the use of derivatives.

Source: (https://www.kotaksecurities.com/ksweb/Research/Investment-Knowledge-Bank/what-is-derivative-trading)

1.5.4 Innovation

Product innovation involves either creating a new product or making changes to an existing one. Innovation means employing new knowledge to provide customers with a new product or service they desire. In other words, innovation can be referred to as invention and commercialization. On the other hand, we can define innovation as a new idea that also may be a recombination of old ideas, a plan that challenges the present system, a formula, or an exclusive method that is perceived as new by the individuals involved. Innovation became a critical factor for social, economic and business development, and it is the key component for creating wealth. Innovation cannot exist without uncertainty and risk (Ethiraj & Levinthal, 2004; Kahneman & Tversky, 1979). Creating and offering something new is always risky. As innovation involves the development of a new product, service or modification to an existing one, so it cannot be predicted whether it will be received with appreciation, or not. For example,

a car company produces cars for women with innovative design, to attract more customers. Here there are two possibilities; the innovation may be a great success, as many women like it, and it becomes part of a new trend, or fashion, which everyone would like to follow. The second possibility is that it may not attract customers, and the car company may be unable to recover its production cost so that the innovation will cause financial loss. So, it cannot be predicted whether an innovation will be received with an appreciation or not. Innovation can be a company's most powerful tool of operation, a key driver of value and a key source of economic growth. It can provide new opportunities for employment because it offers a possibility to realize environmental benefits and exploit change as an opportunity for a different business or a different service.

1.5.5 Failure

We must remember that we learn from failure and not from success. When we combine failure with innovation, we must not forget that anyone who has never made a mistake, has never tried anything new. The criteria for failure are highly dependent on the circumstances of use. They may refer to a single observer or belief system. You may consider one situation to be unsuccessful, whereas someone else may see it differently. This is especially true in direct competition. Similarly, the extent to which a situation succeeds or fails can vary by individual participants. There are situations in which one considers a significant failure and another a small success or neutral situation (Figure 4):

Figure 4 What if?

Source: (https://parentsfriend.files.wordpress.com/2015/10/failing-to-try.jpg)

The key problem is that failure can be a great teacher, but also not necessarily unambiguously. To learn from a failure, we need to decode the " moments of teaching " that are hidden within them. We need a method to determine precisely what these lessons are and how they can improve our chances of success in the future.

The truth is that we always have more control over things than we think. We need to understand precisely the circumstances under which we may change. (Bowen, Govender, & Edwards, 2014)

1.5.6 “Creative destruction" in economics.

The term 'creative destruction' describes innovative entry by entrepreneurs as the force which sustains long-term economic growth, even though it destroys the value of established companies that have enjoyed some degree of monopoly power (Schumpeter, 2006). Due to the significant barriers to entry that monopolies enjoy, new entrants must be radically different: ensuring that fundamental improvement is achieved, not only a mere difference of packaging. The threat of market entry is that it forces monopolists and competitors to continuously reinvest their profits in new products and ideas in order to avoid becoming the next dinosaurs. In the past few centuries, the motivation behind technological innovation has been the desire to secure profits in a “money is king world” which has served to concentrate ownership and wealth towards large corporations. With all the upheaval associated with innovation over the centuries, a pressing question is: What happens with a complex system like stock trading, whose basic operating principle questioned, and a new concept appears in customer motivation: the passion. In the following study, we measure the behaviour of the number of investors who are emotionally attached to a football club. This approach is in stark contrast to the owners who have moved the money market so far and have only owned companies through derivatives, for example.

1.6 Football as a global phenomenon

For modern social scientists dealing with cultural phenomena like a sport is a specific area. Sport is a part of globalisation, which defines the social mechanisms of today, which started earlier than elsewhere because sport, more than others, is an international cultural phenomenon. Football is directly or indirectly in people's lives. This is the most popular TV show, and we talk a lot about it. Everyone understands politics and coaching. Not just our

culture puts sport, especially football, on a pedestal (Biti, 2017)and our sporting heroes are models. We watch them on TV, we follow them on the Internet, we know their stories, and we tell our own stories when we comment about them in a stadium or this new age on the Internet. On the other hand, geography plays a vital role in many social phenomena, like building a social network. Many aspects of life are influenced by the possibilities offered by the environment in which one lives. In line with this, market researchers and some brand representatives invest efforts and resources into creation and maintenance of databases of census data, including several variables that describe the local population and economic activity on the regional scale. These data collection and monitoring activities are commonly limited by the significant efforts required to collect and process data. There are many approaches in the social sciences to harness the potential of large amounts of data generated by online social networks (OSNs). Access to information about such a large-scale user opens up a wide range of opportunities. The question that remains to be answered satisfactorily, however, is how demographics are represented in the content of OSN - are there relevant representations across society, or in this case are we only observing a few distinct levels.

To support this research, I will first present some data on the age distribution and geographic location of Twitter users and football fans. Top 10 Twitter countries, per capita use - Kuwait, Netherlands, Brunei, UK, USA, Chile, Ireland, Canada, Sweden, Puerto Rico, but we also know that Top 10 countries, number of Twitter users: USA, UK, Canada, Australia, Brazil, Germany, Netherlands, France, India, South Africa. Twitter age demographics - each age group is shown as its percentage of overall Twitter users: Ages 15-19 (31%), 20-24 (35%), 25-29 (15%), 30-34 (7%), 35-39 (4%), 40-44 (3%), 45-49 (2%), 50-54 (2%) and 55-60 (1%). It should be noted that China would be high on this list but does not allow Twitter. List of the 8 cities with the most Twitter users: Jakarta, New York, Tokyo and London, Sao Paulo, Paris and Los Angeles, and Bandung Indonesia. (Lipman, 2014) It is worth comparing these data with those who follow the football issues. However, only limited data are available for this, according to an analysing Premier league fans from the resource published by the global web index company at 2015 Q1. (Lipman, 2014)

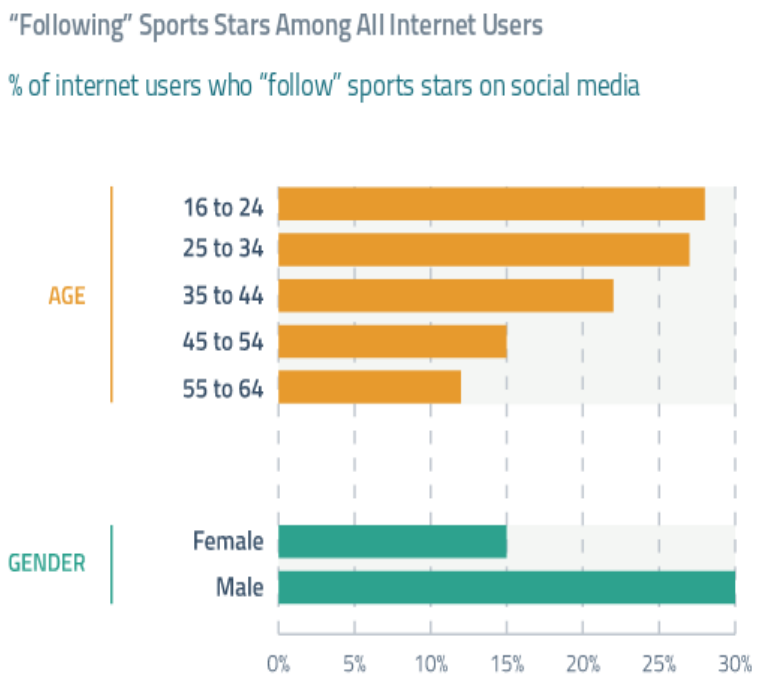

Figure 5 Age distribution of internet users who "follow" sports stars on social media

Source: (Mander & McGrath, 2015) (http://insight.globalwebindex.net/hs-fs/hub/304927/file-2593818997-)

The data shows that Twitter users are significantly younger on average than Premier league fans. However, we know from the same resource that 30% of Premier League Fans are using Twitter, but unfortunately, we do not know about the age distribution of this 30%. (Mander & McGrath, 2015) (Figure 5)

Figure 6 Geographical location of internet users who "follow" sports stars on social media

Source: (Mander & McGrath, 2015) (http://insight.globalwebindex.net/hs-fs/hub/304927/file-2593818997-)

The geographical distribution of Premier League Fans shows a lot of similarities with data that we collected from the geolocation of Twitter users (Figure 6). The primary fan bases are in Western Europe, North and Latin America as well as in the pacific region of Asia. (Mander & McGrath, 2015)

This research uses only one source; therefore, its strength is not to give an accurate picture of the reality but to offer a cost-effective solution with a relatively good estimation that can be applied also in other markets. In this case, I use the Twitter dataset and some computer programs written on my own with minimal resources, using the universities' database for data collection and some data analysis.

2. The complex system

2.1 Definition

As an area of study, the complex system is a subset of system theory. A complex system is a system that consists of many components that can interact with each other. Complex systems include the human brain, the global climate of the Earth, infrastructure such as the electrical grid, living organisms, transport or communication systems, and social organizations (such as cities), living cells, and finally the entire universe (Ramos-Villagrasa, Marques-Quinteiro, Navarro, & Rico, 2018). According to Hayek, complexity is nothing more than the minimum number of elements that a sample instance must contain in order to display all the characteristic properties of its class (Hayek, 1994). Zadeh divides complex systems into two categories, namely systems that are complex in time and systems that are complex in size (Zadeh, 2013). Arthur sees the economy as a complex, evolving system in several ways (Arthur, 2018). Ecosystems also are good examples of complex adaptive systems in which higher-level samples come from localized interactions and selection processes at lower levels. Simon argues that the degree of complexity is often naive or overly pessimistic; this may be due to a deterioration of the underlying assumptions, i.e. a too specific analysis of the selected sub-cases (Simon, 1977). According to Tamás Vicsek (Vicsek, 2002), although there is no unified complexity theory, there are some key concepts with which it is usually associated. Complex systems are systems whose behaviour is fundamentally difficult to model because of dependencies, races, relationships, or other types of interactions between their parts or between the system and its environment. The complex systems have various characteristics that arise from these relationships, such as nonlinearity, appearance, spontaneous order, adaptation and feedback loop. As such systems appear in a wide variety of areas, the common features between them have become the subject of their own research.

The application of the complex system theory in social science is limited by the fact that it has evolved from physical systems without considering the fundamental differences between physical and social science. In the world of social science, results often reflect very complex underlying relationships that involve the interaction of several potentially chaotic systems (Kiel

& Elliott, 2009). The price of wheat, for example, is influenced by the interaction of economic and weather systems. Finding a simple set of equations to explain a complex phenomenon seems like a futile experiment. Social and physical systems also differ in their unpredictability. In the physical world, unpredictability is the result of many iterations, non-linearities, and inaccuracies in its initial conditions. On the other hand in social science, there is no need for a precise definition of starting conditions (McDaniel Jr. & Driebe, 2005). The ultimate difference between the two is that physical systems are shaped by unchanged natural laws, whereas social systems are shaped by the intervention of individuals and organizations.

Warren Weaver did an evident and in-depth analysis of complexity, distinguishing between disorganized complexity and organized complexity (Waren Weaver, 1948). Disorganized complexity can be analysed and described with the help of statistics (probability, relative frequency). This is the area of modern physics, quantum mechanics and other systems, where the mathematical theory of probability can be adequately used due to disorganization, randomness and a large number of elements or events (for example, insurable risk). In the case of organized complexity, the parts of a system have lots of interrelated connections, and the part of the system creates an organization, where the problem is the following: there are a sizable number of factors that are interrelated into an organic whole. In this case, statistics (probability) cannot help, because the particles of the system do not behave randomly. Weaver describes many examples of the organized complexity in the field of biology, healthcare, society, economics and management. (Warren Weaver, 1949). Disorganized complexity is typical in physics, organized complexity in biology and social sciences.

2.1.1 Self-organizing structures

One of the most provocative and controversial elements of complex theory is that complex systems can spontaneously self-organize into more complex structures (Allen, 2001). This concept has been applied to the biological evolution, and economic systems likewise (Rasmussen, Mosekilde, & Sterman, 1985). Complex theory suggests that new, more complex forms of organization appear more often than it would be expected as a result of purely random mutations.

2.1.2 Significant changes within the system

Complex systems can generate large fluctuations internally. For example, sudden, large-scale changes in population levels are caused by internal changes in the system and not due to external influences (Radzicki, 2009). The same phenomenon can be observed when examining economic systems. We do not have to look for wars or natural disasters as the cause of economic depression or stock market crash. In complex systems, fluctuations between periods have a characteristic probability distribution (Bak & Chen, 1991). In this distribution, large fluctuations occur more frequently than in the normal distribution.

2.1.3 Short-term forecasts

A simulation model of a complex system, with well-defined basic conditions, can provide useful predictions (See Lyapunov exponent in chapter 2.3.1) for at least a short period of time. Short-term forecasting is possible so that we can calculate the conditions in the system at a time "t + 1", considering the time conditions "t". Complex systems can also follow repetitive patterns that provide useful information. For example, the economy is moving through recession and recovery. However, we cannot accurately predict the depth or the duration of a particular recession. An exciting feature of the patterns that are followed by complex systems is that they are independent of scale; in other words, similar patterns can be tracked regardless of the distance we use to view them. Economic time series often show this property. Smaller patterns within larger patterns are called fractals. In nature, fractals are found in a wide variety of phenomena, for example, the shapes of shorelines or in ice crystals.

2.2 The tipping points.

The Tipping point (turning point) is the point where the series of small changes or events become significant enough to make a bigger, more critical change and drastically affect the outcome. The term allegedly comes from the field of epidemiology when an infectious disease reaches a point that goes beyond the possibility to be prevented from spreading locally. The term was further promoted in 2000 by Malcolm Gladwell in “The Tipping Point: Little Things Can Make a Big Change”(Gladwell, 2000). Gladwell's ideas extend the definition of the turning point to a particular branch of mathematics, namely, a part of complexity theory called bifurcation analysis. In mathematics, bifurcation is the point where small changes begin to proliferate through positive feedback. He examined whether systems could evolve during periods of rapid change as a result of the dynamics of internal systems. The classic form of development of complex systems is evolution. The evolutionary process takes place unhindered until a turning point arrives, where the fact and time of the next step depend on the resilience of the system.

2.2.1 Examples for tipping points

- The straw that breaks the camel’s back

There is a well-known statement that a straw can break the camel's back. In this example, a linear increase in the weight of the straw placed on the camel's back leads, beyond a certain point, to a sudden change in the spine of the abovementioned camel. This is a classic example of a small change, in this case adding a straw, crossing a turning point, and having a disproportionate effect.

- World Population

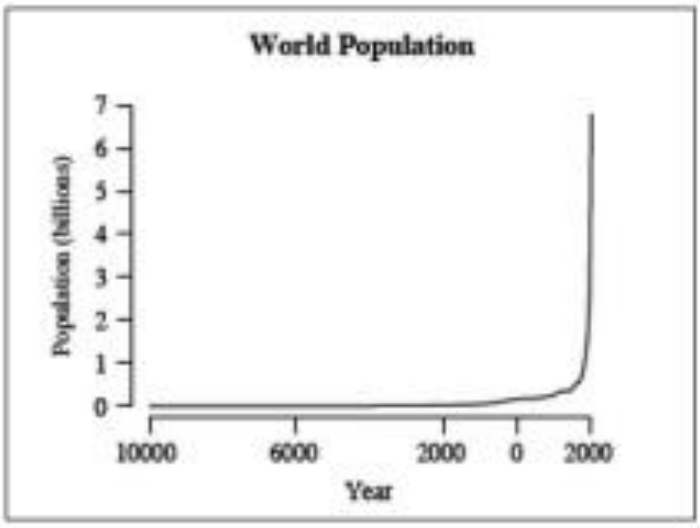

Figure 7 World population charted on a long-time scale

Source :('Critical transitions in nature and society', 2009) (https://www.theautomaticearth.com/2012/03/the-nature-of-tipping-points/)

This almost L-shaped curve is familiar to those studying exponential growth (Figure 7). It can be concluded that we can see a distinct tipping point on the curve beyond which the growth of the population has suddenly accelerated.

- Boiling water

100 degrees Celsius is the tipping point of water when it turns from liquid to steam.

- New York crime rate

The crime rate in New York was very high between 1975 and 1992, with more than 600,000 crimes per year, including about 2,000 murders. In 1993, a tipping point came, and the crime rate dropped dramatically. Within five years, serious crimes have declined by half, and murders have decreased to a third.

There was a five-year project that focused on improving the quality of the neighbourhood, such as removing all the graffiti and repairing all the broken windows in the street.

The restoration of the underground travel system was also part of the project. 170,000 people travelled daily without buying their tickets. Losing revenue was not as important as the effect of losing respect for law and order. The travel fare enforcement officers forced the payment of fares. The non-payers were handcuffed and put on the platform for a while. The remarkable result of the new policy was not only that people paid their fares, but also that serious crimes such as robbery, rape and murder fell more than 50%. A firm stance against minor crimes has also reduced major crimes, making New York one of the safest cities in the world.

Scientists have observed tipping points in many real-world systems, such as marine fisheries, lake water quality and the world's climate change**,** financial crisis or spread of viruses. (Meadows-Klue, 2004)

2.2.2 Foreseeing tipping points

The decline of the resistance near the turning points will be more readily understood from the stability diagrams presented in Figure 8. In the case of off-peak conditions, a system is resilient: the attractiveness of the pool is high. The external or internal influences do not quickly drive the system toward an altered state. In the case of a system close to a turning point, a disturbing effect may result in the system being moved to a subsequent pool. The state of the system does not in itself support such "fragility." However, the dynamics of the system around the equilibrium are typically different from what we see when the attraction of the pool is large (as in (a)). In the risky state (b), the rate of recovery from small disturbances is slower (arrow). In the stochastic environment, fluctuations are larger and more correlated over time, as shown by the side maps. Such dynamic changes are general indicators of the proximity of turning points (Scheffer, 2010). The motivation for writing this study was that the prediction of critical transitions in socio-ecological systems is crucial for sustainable development and ecosystem management, where we need to avoid catastrophic changes in the lives and social well-being of people (Tušak, 1997).

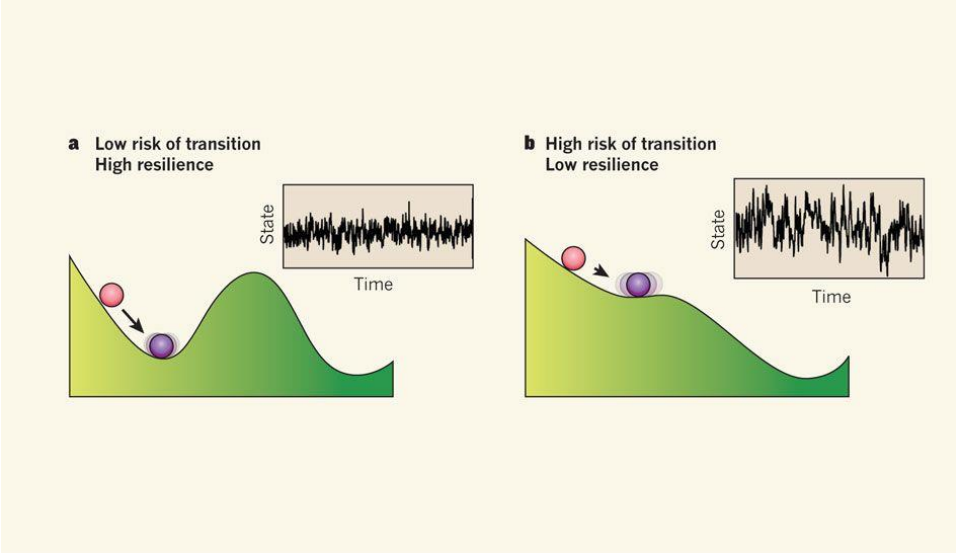

Figure 8 Foreseeing tipping points.

Source (Scheffer, 2010): (https://www.nature.com/articles/467411a/figures/1)

2.2.3 Transformations

Several researchers (Hampson, Held, McGrew, Goldblatt, & Perraton, 1999; Kuziemsky, 2016; Xiaolong Wang, Farhadi, & Gupta, 2016) - following their path independently from each other - investigated what was happening during transformations. The behaviour of substances has been studied near the transformation points where one substance moves from one state to another - from liquid to gas or from magnetized to non-magnetized. Phase transitions - as the singular limit between the two empires of existence - exhibited their strong non-linear character in mathematical terms. The predictable behaviour within the phases does not help much in understanding transitions. Researchers of complexity (Ramos-Villagrasa et al., 2018) believe that if the world is complex, making it more straightforward until it is manageable means omitting essential elements. Abstract theoretical models focus on scientific goals, while policy-oriented models focus on policy-oriented recommendations. In order to substantiate the usability of the scaling-based thoughts, many mathematical analyses and experiences are required about real-world systems in order to understand that these seemingly independent transitions follow the same rules.

As data became increasingly aware of the variability of dynamical systems, he realized that less complicated systems, than quadratic functions, can give other types of unexpected patterns. (Feigenbaum, 1978) In some systems, more stable solutions can be hidden. For a very long time, the observer may experience only one kind of behaviour, although different types of

behaviour also exist within the system. There are times when we think we understand how it works. For example, how prices change in the stock market. Then a crisis comes, like this worldwide epidemic of the COVID 19 virus, and all our existing experiences are gone. The system enters a different state, whose rules must be rediscovered.

2.3 Chaos – historical overview

Gleick's famous sentence "Chaos begins where classical science ends" reveals much (Gleick & Hilborn, 1988). Concerning the scientific concept of chaos, which has been widespread since the 1980s, it should first be noted that it does not refer to a current situation or state, but the behaviour of the system over time. Namely, considering the time, the evolution of any quantity can be considered as motion in a general sense.

I did not find an exact definition of the concept of chaos. Some examples among the definitions proposed (Sardar, 1994):

- A kind of order without periodicity.
- Seemingly randomly repeating behaviour within a simple deterministic system.
- A qualitative study of unstable, aperiodic behaviour within deterministic non-linear dynamical systems.
- The ability of simple models without built-in random features to exhibit very irregular behaviour.

By its very nature, chaos is a dynamic phenomenon that occurs when something changes. It connects our everyday experiences with the laws of nature by exploring the deep-seated connections between simplicity and complexity and order and disorder. Chaos in this modern sense refers to the nature of motion and dynamics. In other words, chaos is a constant movement that does not repeat itself. (Baumol & Benhabib, 1989) By non-repetitive motion, we mean that motion is not periodic in time, not even approximately. Constant movement refers to a long-lasting, non-damping movement. Such irregular, unpredictable dynamics can be observed in many everyday phenomena, such as the drop of falling leaves. It is generally true that the study of chaos involves the study of elementary non-linear systems that lead to extremely complicated

behaviour, and complexity is generally about the (simple) interactions of many things (often repeated) leading to higher-level patterns.

Practitioners of the science of chaos have diverted science from its reductionist pursuit: from studying systems only through their components. They are looking for the whole thing. The revolution of chaos directly affects the world of visible and tangible human-scale things.

Chaos theory is essential to modern science because it connects our everyday experiences and the laws of nature by exploring the connections between simplicity and complexity and order and disorder. It presents a universe that obeys the fundamental laws of physics but is able to become disordered and unpredictable. Predictability is a rare phenomenon and operates only within the limits that science has filtered out of our complex world. It provides an opportunity to simplify complex phenomena. It combines math with the impressive computing power of modern computers. Chaos theory questions the traditional model-building methods of science. There are limitations to understanding and predicting future events at all levels of complexity. (Gould, 2008) In chaos systems, small disturbances become more frequent over time due to the non-linear relationships and the dynamic, repetitive nature of the chaotic systems. As a result, such systems are extremely sensitive to initial conditions, and this makes forecasting difficult. The fundamental problem is the use of finite measurements in an infinite world. As systems develop dynamically, they are exposed to several tiny random effects that cannot be integrated into the model. The traditional understanding of linear models and the influence of random errors cause us to make better predictions (See Lyapunov exponent in chapter 2.3.1) of better models and baseline conditions. Long-term planning of complex systems is not only difficult but also fundamentally impossible; it has severe consequences for organizations that are looking to develop a strategy based on future planning. Rather than investing large amounts of resources in forecasting, strategic planning must consider several possible scenarios.

Fundamentally, chaos is a narrow part of the field of the “complexity theory”, in which “chaos” is a particular mode of behaviour.

2.3.1 The Lyapunov - exponent

The Lyapunov exponent's simple formula helps to understand why it is so difficult to predict anything in a complex system. Lyapunov’s exponent (λ) is used to characterize unpredictability. In complex systems, the distance between two close points increases exponentially over time.

$$(r'_t - r''_t) = e^{\lambda}(r'_0 - r''_0) \quad (3)$$

where r 'and r'' are the two selected close starting points. The distance of the deviation after n, time steps is $e^{\lambda(t)}$ where λ (t) is called the local Lyapunov exponent that depends on where the two starting points were chosen and how many steps we took. (Lyapunov, 1992) The mean of local Lyapunov exponential is the Lyapunov exponent (λ), which describes the typical departure rate of nearby starting points. In the chaotic systems, when Lyapunov exponential is positive, the exponential orbits move away from each other. If the system is not chaotic, then the distance is usually slower than exponential; hence the Lyapunov exponent is not positive. (Nguyen, 2018)

2.3.2 The financial market as a complex system

Nowadays, the business and economic environment in the world has changed dramatically. The world is organized into a single complex, global market. With the spread of electronic markets, capital can be immediately transferred from one place to another. Market turbulence is inherent in everyday life. Under such economic circumstances, through chaos and complexity theory, we can understand better what is happening around us, than through traditional economic theories. Chaos and complexity theory questions old economic theories. Feedback plays a vital role in the development of new ideas. Analysts are increasingly interested in the impacts with different speeds to the financial markets. In addition to their process orientation, complex models are also able to deal with the time dimension of economic policy. This means that all their results will be temporary and are going to change. It is known from the chaos and complexity theory that non-linear equations can describe very complex motions and are in many respects closer to real-time series than to linear models. Another interesting feature of such models is that they illustrate the importance of expectations in economics. One of the essential drivers of dynamic phenomena in social science is that people's expectations are diverse and changing.

This feature has significantly influenced the development of social sciences in recent years. An excellent example is the appearance and success of the neoclassical school in economics. One of the critical assumptions of the neoclassical school is the assumption of rational expectations.

The neoclassic assume that people do not make systematic mistakes in their expectations. With the information available, their expectations are the best estimates for the future. This assumption is the assumption of efficient markets, which says that all past information has already appeared in the market price. Everything that has happened so far is embedded in the price. Only an event that happens unexpectedly can change the price. This cannot be inferred from past data. If this would have been possible, investors would have already deduced it and incorporated it into the price.

A vital feature of complex systems is long-term unpredictability. This means that if we examine the behaviour of a given system starting from two points close to each other, these two points will move away from each other at an expected exponential rate.

City as a complex system.

Nowadays, more than 50% of the world's population lives and works in cities. The proportion and the total number of urban population will further increase in the coming decades (Bokányi et al., 2016). Understanding the creation of sustainable cities is an essential aspect of using data and geographic information. Such data include population growth, resource utilization, infrastructure analysis and modelling, and other economic indicators. Each city is a complex system of many subdivisions, where different urban infrastructures, energy and human flows act as a holistic whole and are influenced by society (Batty, 2009). Cities have become the primary medium of economic and social exchanges in today's increasingly interconnected world. Connections are interlinked insights and naturally lead to networks, which pervade complex spatial systems. (Jacobs & Lees, 2013)

If we represent human relationships as a graph, where the node of the graph are humans, and the edge of the graph is the relation between two persons. In that case, it can be seen from the nature of this graph that if the nodes of the graph increase linearly, then the number of possible edges increases exponentially. The social position of a person in a given society can increase factors such as for example, valuable knowledge, efficient usage of resources or social capital that may influence the person's income. Modern cities have become the primary medium of economic and social exchanges in today's increasingly interconnected world. This connection naturally leads to networks, which pervade complex spatial systems.

2.3.3 The team as a complex system

In the early twenty-first century, Arrow and his colleagues depicted teams as complex adaptive systems. (Arrow, McGrath, & Berdahl, 2000) The team, according to the classical definition, is a group with a common goal, whose members work closely and synergistically to be successful together, knowing the goal and mobilizing their intentions and abilities to achieve it. (Agile Alliance, 2015) Complex, ever-changing tasks (such as developing a software package, providing social services, or playing in a football team) require agile communication through a wider range of channels, responding quickly and emphatically to information, and allowing for trial and error. The team cannot do without the essential elements of feedback in the area of psychological safety and risk-taking. (DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004) Team members need to know the goals of their organization, as well as their values and interests, in order to continue their internal and external communication accordingly. (O'Daniel & Rosenstein, 2008) The team works as one, but it ultimately consists of individuals. According to Michel Jordan: Talent wins the competition, but teamwork and intellect win the championship. It is a misconception that successful teams can only be made up of heroes. Big teams are created with the right goal system, awareness, shared consciousness, operating culture of decision-making, and delegating the responsibility of the right level. (Borck, 2016) The properties listed above are the properties of a complex system with feedback loops, nonlinear connections, and unexpected turning points.

3. Methodological considerations

3.1 Identification of research stages

In the research presented below, a standard cross-industry data collection process (CRISP-DM) was used. This model is designed as a generic model that can be used in many industries and applied to diverse business problems. The general CRISP-DM process model includes six phases that address the significant issues of data mining. The six phases fit together in a cyclic process (Chapman et al., 2000). These six phases cover the entire data mining process, including integrating data mining into larger business practices. These phases are shown in the figure below. The diagram illustrates the iterative nature of the data mining project (Figure 9.)

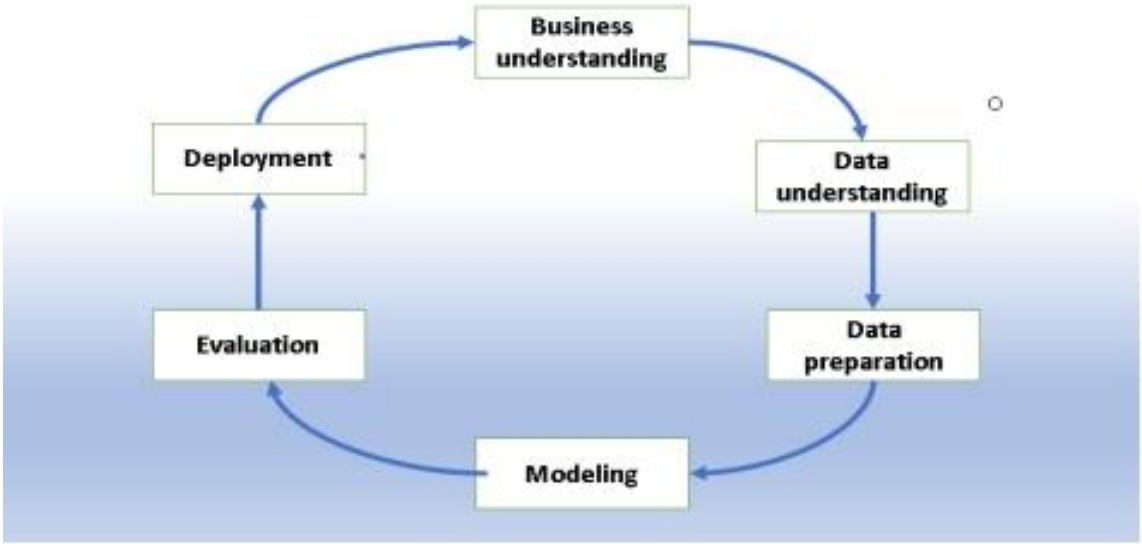

Figure 9 CRISP-DM Diagram

Source: Self-prepared chart using *Chapman et al., 2000*

3.1.1 Business understanding

Understanding a business includes determining the objectives of the business, assessing the situation, determining data-mining goals and producing a project plan. For data mining, business understanding includes determining what data are available, and crucially, thinking

about what information can be gained from the data. Therefore, sometimes only the data analysis can discover how much important information the data contains.

3.1.2 Data understanding

"Data provides the "raw material" of data mining" (Shearer et al., 2000). Data are usually representative (e.g. age, height, weight, colour, blood pressure, habits of a person) however it may be derived (e.g. data generated from some other data, such as for example percentage change over time) (Kitchin, 2014). The data understanding phase addresses the need to understand resources and the characteristics of these resources. It includes gathering initial data, describing data, exploring data, and verifying the quality of data. During the initial data collection, it should be checked whether or not sufficient data can be obtained (Salganik., 2017). Besides, I had to load the data into any devices that I wanted to use for data mining to verify that the devices are compatible with the data. It was a good idea to make a list of the types of data needed to achieve my data mining goals. The list should be supplemented with details such as the data format. If in some cases of the data is not available, I decided how to solve the problem. I considered alternatives such as:

- Replacement with an alternative data
- Narrowing the scope of the project

I prepared a report which described my needs and explain in detail exactly what data I collected and from what sources. After the data was available, a data descriptor report should be prepared. It describes the format and source of the data, the number of cases, the number and description of the fields, and any other important information. It also briefly evaluates the suitability of the data for its data mining purposes. During the data analysis process: for each variable, the range of values and their distributions must be examined.

A data exploration report is a place where I documented hypotheses or initial results formed during the data exploration. The report included a detailed description of the data, including allocation, aggregations, and any signs of data quality issues.

During the verification of the quality of the data, I needed to determine if the data is good enough to support my aims. The data quality report summarizes the data available to it, the minor and major quality issues found, and possible solutions to the quality issues or alternatives.

At this stage, I had to determine what concepts and categories the data belong, and which patterns of these concepts are expected to be discovered in the data.

3.1.3 Data preparation

The phases of data preparation include selecting, cleaning, constructing, integrating, and formatting data. These tasks are critical for the success of the data-mining project. All data mining projects are adversely affected by problems of missing data or any errors in the data.

- Missing data: Some records do not have all entries. In the first step, they were missed from the selection.
- Errors in data: Errors are usually more numerous in some types of data: anything created directly by humans with minimal editing will include errors. This type of data also commonly includes abbreviations or different formats for the same information, such as dates.

3.1.4 Modelling

At this stage, the data were already in good condition and it was worth looking for samples in them. (For example, in studies 2 and 3: Under what circumstances has the number of tweets increased on a given day). This phase included not only the selection of modelling techniques, but also the creation of test designs and the construction and evaluation of models.

Similar to building standard statistical models, model development is a repetitive process, (Shearer et al., 2000) therefore, during the writing of studies, I used a number of models and modelling techniques before finding the best one.

Another basic task of this phase was also to create a test plan: How will I be able to re-measure my models based on the past data that are available to me.

After creating the models and setting the parameters adequately, I had to evaluate my models. It was important to interpret the relationships revealed by the models: E.g. The relationship between the number of tweets and the number of shares sold.

The question was: What do these connections mean in business? How can I explain this to potential users?

It was also important to test under what circumstances the models work well and when they fail to do so? In those cases when they did not work well, several questions arose: Is it possible to create additional variables based on these experiences? Can I combine different models?

An important aspect of model evaluation is, whether a model meets the business criteria which I have defined for the business understanding and data comprehension that I established in the phases of CRISP-DM. It has happened several times that the model was technically acceptable, but it still could not be applied in the business environment.

3.1.5 Evaluation

At this stage, I assessed whether our model met the business criteria formulated in the first stage. If not, what can it be used for? It has happened several times that a model was, after all, useful from another research perspective, not the one I would have wished to prove.

It is no coincidence that the CRISP-DM methodology is iterative: if the model could not be used in the form in which it was created, I returned to one of the previous phases. In this case: I need to incorporate new data, take new data preparation steps, or to select new modelling techniques.

3.1.6 Deployment

As the models were not installed in the classical sense on the users' systems, there was no need to create an installation or maintenance plan.

3.2 The Monte Carlo method

The Monte Carlo method essentially solves the problem by directly simulating the underlying processes and then calculating the (average) result of those processes. In the studies I detailed here, this process was the number of tweets generated in a single day.

During the elaboration of the two papers, I used computer tools to produce the final result of the particular experiment, after which I evaluated the numerical characteristics obtained.

The mean distribution and standard deviations were calculated to determine the possible error of the result. However, it is important to note that the Monte Carlo method includes a broad group of computational algorithms that are based on repeated random sampling in order to achieve numerical results. The basic concept is, to use randomness to solve problems that can be deterministic in principle. The Monte Carlo method essentially solves the problem by directly simulating the underlying process and then calculating the (average) result of the process.

The main steps of the Monte Carlo method are:

- To specify the range of possible inputs
- To generate inputs randomly from the range of probability of distribution
- Perform a deterministic calculation of the inputs.
- Summarize the results.

In the cases elaborated in the following chapters, I have used computer tools in order to produce the final result of a given experiment, after which I evaluated the resulting numerical characteristics. In order to determine the possible error of the result, I carried out the calculation of the standard deviation.

3.3 About the models used in the papers

The technical development also renewed the research methodology. In the age of computing, future managers run simulations on machines and look for patterns to predict events. According to the science of forecasting, the main goal is not to give an accurate estimate of what will happen in the future but to outline realistic scenarios. Correct prediction can estimate the direction of a change and the possible effects of the change taking into account the factors that can be changed. Therefore it can provide decision-makers with a set of tools that support them to make choices by providing information on the possible consequences of decisions. (Roubelat, 2000)

With the rapid development of data collection and storage technologies, organizations have been able to accumulate large amounts of data. However, obtaining useful information has already proved to be a major challenge. Tools of conventional data analysis and methods are often unavailable due to the enormous amount of data. Occasionally, the data is non-traditional, which means that traditional approaches cannot be applied even in case of a relatively small data set. In other cases, questions to be answered cannot be addressed by methods of existing data analysis, and new methods must be developed.

3.3.1 Data mining

Data mining is a technology for processing large amounts of data that blends methods of traditional data analysis and sophisticated algorithms. This technology has opened up exciting opportunities for exploring and analysing new types of data and for exploring old types of data in new ways. Data mining techniques can be used to support a wide range of business

intelligence applications such as customer profiling, targeted marketing, business process management, business premises fitting and fraud detection. (Chapman et al., 2000) These methods also help the analyst answer critical business questions such as "Who are the most profitable customers?", "What products or services can be cross-sold or devalued?", or "How can I improve the quality of my service, in a smart but economical way?" The frequent occurrence of these questions motivated the development of a new data analysis method: The association rule analysis.

Learning the automatic lexical and structural preferences from corpora is one of the foundations of a statistical NLP approach. So first, I create a model (rules, types, synonyms) to be used for more prominent data analysis (Figures 10 and 11).

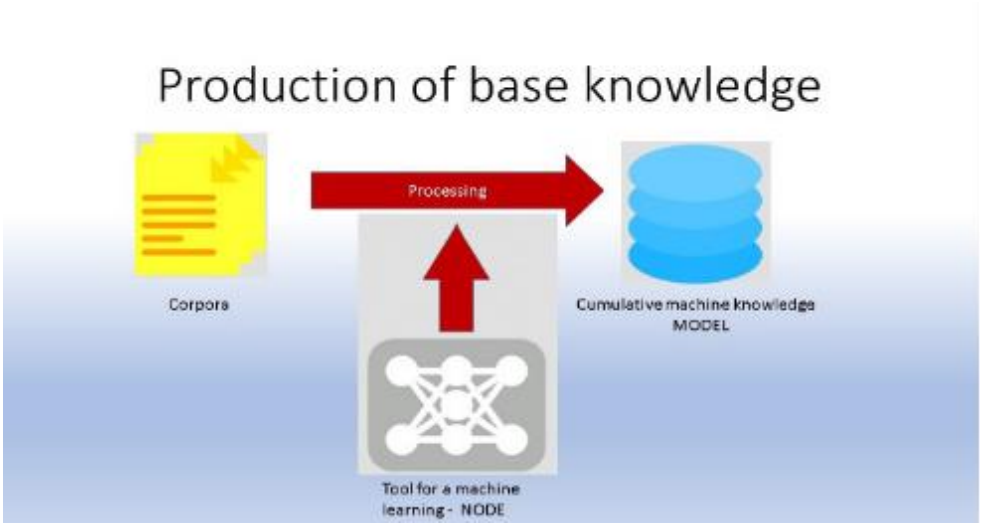

Figure 10 Statistical NLP approach phase I

Source: Self-prepared chart

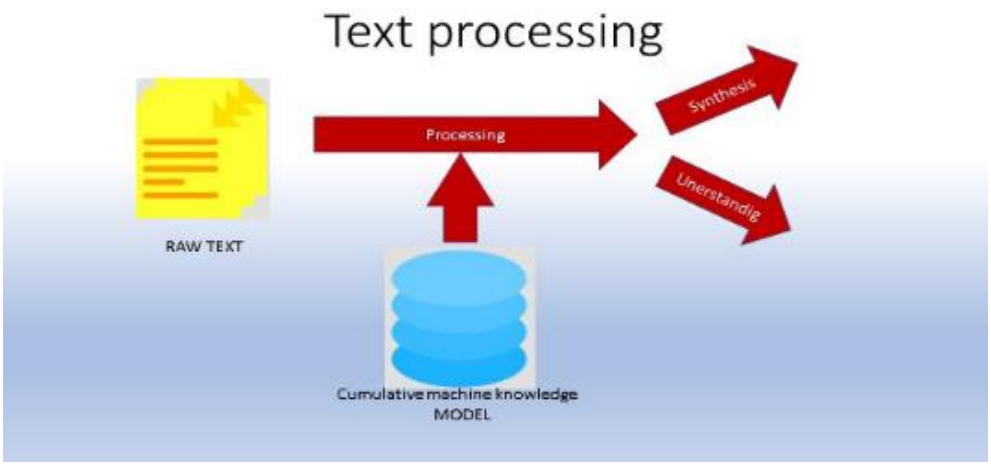

Figure 11 Statistical NLP approach phase II

Source: Self-prepared chart

3.3.2 Data analytics

Statistics, as one of the fast-growing disciplines of our time, faces many challenges today and will continue even more in the future. Nowadays, data are browsed not only by a small group of professionals, but useful information is seen as a value in every area of life. Any prestigious newspapers or educational publications are expected to back up the results of an article with data. They must do this in a scientific, but clear and spectacular way, such as

- Do data mining (discover data to find new patterns and relationships)
- Use full statistical analysis and quantitative analysis to explain why specific results occur.
- Use predictive modelling and predictive analysis to predict future results.

Big business has a mission. Most companies train newly hired and long-term employees to the core values that drive the company to success. However, many companies fail to quantify these values. Data mining techniques help companies measure whether these values are expressed in corporate communications.

With more information available to businesses, it is easier to empower the team to make quick decisions. Rapid movement and development are essential for a business if it wants to remain ahead of the competition. It is important to carefully consider individual decisions, and this applies not only to senior managers but to other employees too. That is why all departments should have access to the analysis. It has to be built into the corporate culture.

Using visual data can make businesses more agile (Sóti, 2020). It can help decision-makers to extract information from the data quicker. It helps them make decisions without spending too much time understanding what is going on in the market.

Today's world is moving faster than ever. The way you reach your customers is also changing. Way too many changes are happening fast. Using data mining technique and data analysis can help companies adapt quickly. Using analytic tools, managers can be able to make forecasts.

3.3.3 Text mining

Text mining is the process of extracting knowledge and information from natural language texts (Y. Li, Zhou, Bruza, Xu, & Lau, 2008). Usually, segmentation of text descriptions (by creating clusters) – also used in this research - can help to identify the problem area, and we can detect which parts of the service have the most significant impact. This is the most challenging task for text mining, and it involves the use of dictionaries, glossaries, lexicons, typologies, and so forth as part of it. (Briscoe, Copestake, & Boguraev, 1990) The following steps were taken during text processing (Yessenov et al., 2013):

Text pre-processing - Replacement of some special characters with spaces, and determination at the end of sentences and the end of paragraphs.

The extraction step begins by attaching a label to each word and making the first attempt at the identification of candidate terms. It is registered that terms are the building blocks of text analysis, corresponding to single or tricky words that are considered relevant or interesting. The following actions were taken during this step:

Text tokenization: A process that identifies character strings (tokens) from the input text, based on delimiters. Examples of delimiters are spaces, tabs, carriage returns, and punctuation marks.

Text normalization: Helps to manage poor punctuation in the text, such as improper use of a period, comma, semi-colon, colon, forward-slash, etc. The input text is "corrected" internally to place spaces around improper punctuation.

Candidate term extraction: Identifies relevant words and compound words from the input text.

Labelling the part of the speech: Each token in the text stream is labelled by a Part of Speech tag, which comes from the base of the dictionary. Traditional English grammar divides words based on eight parts of speech: verb, noun, pronoun, adjective, adverb, preposition, conjunction, and interjection.

Forcing / Excluding step: When the expert reviewed the text data after extraction, he should discover that some words or phrases were not extracted. While usually, these words are verbs or adjectives that we are not interested in. These may be terms that are specific to an industry as product names, acronyms, etc. If an expert would like to have these words and phrases extracted, he can force a term into the dictionary.

Typing step: A type is a higher-level concept that contains one or more terms. There are default types for Organization, Product, Person. After extracting and modifying terms as described above, the extractor next assigns each concept to a type, wherever possible.

Creation of categories step (grouping): The next step in text meaning is usually the creation of categories, which represent the information that we consider to be important in the responses. There are three categorization methods:

- concept inclusion
- concept derivation
- semantic networks

The category refers to a group of closely related concepts, opinions, or attitudes used in text mining. Each category is defined by one or more descriptors that are concepts, types, patterns, or conditional rules. Categories can be created from a single concept or type, but it is common to combine multiple descriptors. The three language methods provide a natural language-based approach to categorization. For example, Figure 12 presents the overlap for the categories to which the type (Performance) belongs according to the selection. (Figure 12)

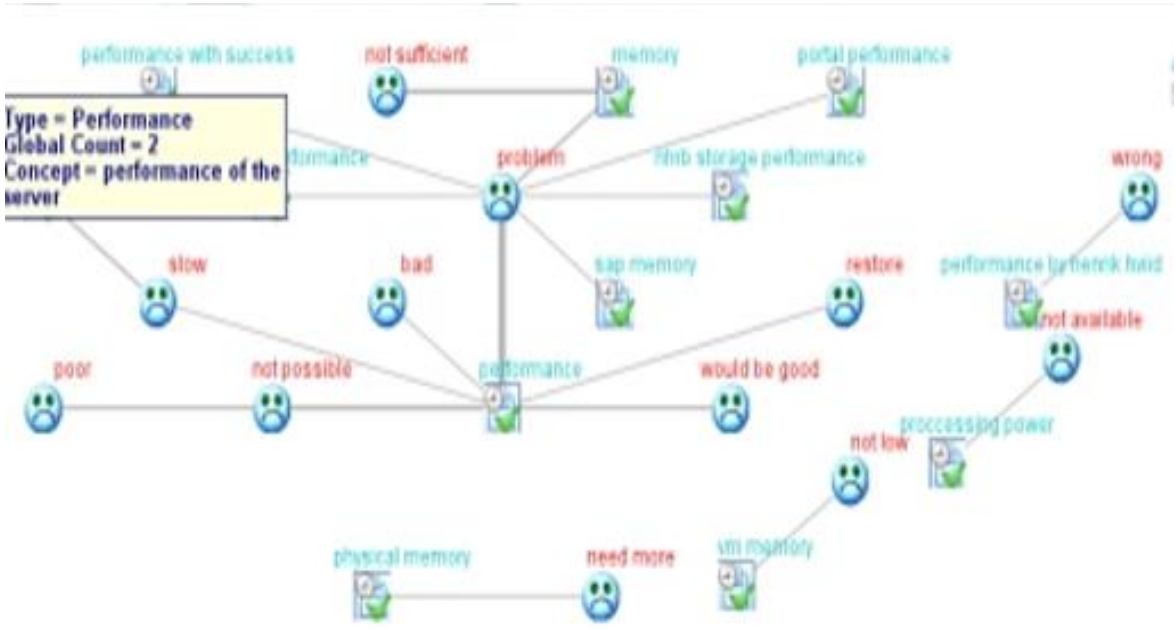

Figure 12 The overlaps for the categories (Performance)

Source: Self-prepared chart

Indexing step: The types are indexed by establishing a pointer between a text position and the representative term for each equivalent class.

Matching patterns and events extraction step: The text mining can discover not only types and concepts in the text but also relationships among them.

3.4 Limitation of models in papers

In the first paper, I used a model whose parameters were suggested by experts in the field and were based on their observations. These were however subjective opinions and unfortunately, due to lack of data, only one team was included in the described case study. By analyzing the model, I discovered the following limitations: I defined the football team as a complex system with elements such as its players and its coach. When a player who I assumed to be an atomic (elementary) entity left the team, it should have been examined whether the resulting system was comparable to the original. Is it possible to state, that if only one player leaves the team it will still remain the same team? If not, this would mean that the player is merely an attribute of the system.

My other concern was that it was unlikely that I could draw appropriate global conclusions from the experience gathered by observing a team, even if observed for five years.

The experiences of all the teams in the league explored and analyzed at the same time would have been much more exciting and thorough. On the other hand, I consider that the phenomenon described in the study, to be a sketch of a realistic scenario, and the principles of the group interaction may apply to other social collectives as well.

The above conclusion is valid for the groups that have the following attributes: the number of the members of the group falls between 10-20, voluntary based, carried out as a hobby, has regular activities and the group participates in events where performance plays a crucial role. A further area of research could be the observation of the elements of football as a complex process as well as increasingly analyse the observed attributes of the players such as for example their skills, tricking techniques or discipline.

In the second paper: exploring and investigating the causal relationship between two phenomena requires several considerations, often of philosophical depth. By definition, the variable X is considered to be the cause of Y if it can be used to give a better estimate of Y than without it. Two measurements may be different for the same mean due to different standard deviations. The F-test reveals whether this difference is significant. An F-test can be understood as any statistical method that compares the standard deviations of two or more samples. In the first step, it had to be proven that the data generated by the two models (described in chapter 5.6) are significantly different. The F-test is extremely sensitive to the abnormal distribution, but since this was not the case in our study, we used the F-test with confidence. In the next step, we wanted to see when the impact of the model we created was greatest. This was measured by generating two thousand random numbers within the given interval, taking their average effect and comparing. We chose two thousand samples because, in the case of a larger quantity of samples, the distribution of the data did not change significantly. Also considering the existing computer capacity, and the time for generating the test, the quality of data was acceptable. A weak point in the calculation of the abnormal return described in Equation (9) (Chapter 5.8) is that the extreme function values at the beginning and at the end of the row arranged in the formula can skew the results. Therefore, in each case, 50 values were omitted from the beginning of the sorted row and 10 from the end. We could have discarded more from the beginning of the line because there the relatively small values occurred much more frequently than the large values. In this study, I have not examined behavioural patterns of football fans

related to their place of living and geographical position. It is possible, for example, to compare the opinion of a norther English fan to the opinion of a fan from South Africa in case of an identical socio-cultural context. By the description of these opinions, we would need to examine the given socio-cultural environments. This beyond the area of competence of this study. It can be established based on the study, that the quantitative analysis of tweet entries with negative content can help in the automatization of the trade-in stock market.

In the third paper, I used a similar model as in the second paper because the goal was the same as in the second paper just in this paper, I examine innovative companies, not soccer clubs.

In this research, distance correlation is used to measure the dependency between two vectors, not necessarily of the same dimension. The distance correlation measures both the linear and non-linear association between the two vectors. It evaluates whether vectors tend to change together in terms of their distance from all other points. The distance correlation between any two variables ranges from zero to one. Zero means that the variables are independent, while a point closer to one indicates a dependent relationship. The distance correlation can be applied to perform a statistical test of dependence with a permutation test. I first calculated the distance correlation between two random vectors and then compared this value with the distance correlation of other measured data. Thus, the distance correlation data described in Section 6.3 must be viewed relative to each other. This reveals on which day there was a strong relationship between the number of tweets and stock market movements. The weakness of this method is that it does not necessarily show a causal relationship. There may exist a third event that triggers both of our observed events.

In this study, I have not examined the disposition of the tweets and I have not observed the effect of the positive and negative spirit of the registered tweets on the movements of the stock market. The more precisely determined analytical aspects would have provided more detailed results. The analysis of the occurrence discussed in this study clearly shows the relationship between the number of comments appearing in tweets and the movement of the exchange rate. We can establish based on this that the method is suitable for planning the more successful stock market investments in case of innovative companies.

In the fourth paper, I used a model of urban scaling. About 1-2% of Twitter data is free to download via the API. For those who enable this option on their smartphone, the exact GPS coordinates are attached to the messages. The weak point of the method described in Section 7.3 is that we do not know the internet coverage of the given country at the measured time. In

terms of results, the method provides a promising approach for developed countries. On the other hand, in the case of emerging countries, the development of the economy of each region may be interesting. In the diagrams of this section, we fit a line to the log Y - log N pairs. The estimated magnitude of the error was printed everywhere. Error calculation assumes that larger follower numbers carry fewer errors when fitting scaling curves.

The result of the study shows that tweet entries of a given topic (soccer fan) show different distribution in the geolocational aspect between the village and the city in terms of the frequency of occurrence. These data are suitable for the targeted marketing of products. The results of the study can be completed by observations related to the habit to use smart devices of the residents of the given area.

In the fifth paper, we use semantic analysis which can detect connection among different tags. One of the weaknesses of the analysis of this type is that it can also highlight relationships that are not relevant to us. However, this can only be judged by an expert in the field. So, such an analysis cannot be automated. Another weak point of the analysis is that it can only show focus points and trends. It lacks the evaluation of the quality of the given service.

In short, we do not know whether the change is pointing in the right or wrong direction. The connection of the expressions related to the size of the mistake to the trends analysed so far, would give more information about the operator of the system.

3.5 Hypotheses

General system analysis is usually satisfied with a simple exploration of the characteristics of the system. In this case, it is sufficient to examine what output response is given by the system to the input characteristics. By the description of complex systems lack of knowledge is a common issue. Although we know some parameters of the input data of the system, we cannot accurately describe the general behavior of the elements of the system (individuals, stock traders or even users). The possible reason for this that we have neither an infinite quantity of data nor infinite computational capacity to calculate this behaviour. I resolve this contradiction by building a model from the sphere of complex systems on a specific example. The model is significantly simpler than reality. The model contains only the most necessary properties of the system (and thus only the most important components building the system).

Modelling enables features that are important from the aspect of operation to come to the fore. These characteristics would be obscured by the complex nature of the actual system during the study. The model, therefore, contains only the most important elements.

Benefits of modelling:

- Based on the modelling, we have a better understanding of what system-level behaviour we can describe well. However, we need to take into account which part of the observed system cannot be described by the current model.
- Creation of a model does not tear the system into parts, so you see it as a whole, with its relationships and effects. Modelling gives you a chance to find effective solutions, as this method is able to move systematically towards the optimal solution.
- Inspection of the model is usually less expensive and safer than the analysis of the actual system.
- Modelling makes possible abstract introspections to be performed, therefore it provides the possibility to use mathematical and other abstract, theoretical tools.

The disadvantage of modelling is that the model is necessarily always less accurate than reality. One of the most important questions of model-creation in this dissertation is if the approach I have chosen under what circumstances, with what limits and how well it describes the part of the observed system that I would like to examine in particular. Human behaviour is much more complex than I can describe. I capture only some characteristic of the behaviour (e.g., willingness to buy stock or become a soccer fan) in order to examine. The limitations of the methods I use have been described in detail in Section 3.1.

In this study, the focus is on understanding the relationship between the patterns of human behaviour and complex systems. The following hypotheses were raised:

$H1_1$: The methods of complex systems can be used to describe the behavioural patterns of a football team as a complex system.

$H1_0$: The methods of complex systems cannot be used to describe the behavioural patterns of a football team as a complex system.

$H2_1$: There is a group of individual investors whose decisions about investing in a football club are driven not only by the consideration of their long-term wellbeing but also by their daily emotional state concerning the club.

$H2_0$: All decisions about investing in a football club is driven by investors' long-term wellbeing. There is no influence on their daily emotional state concerning the club.

$H3_1$: There is a link between the frequency of appearance of innovative companies in social media and the volatility of stock prices of these companies.

$H3_0$: There is no link between the frequency of appearance of innovative companies in social media and the volatility of stock prices of these companies.

$H4_1$: The lower is the GDP of a country, the more football fans from this country live in cities.

$H4_0$: The country's GDP does not affect where the football fans from in the country live, in a village or a town.

$H5_1$: The usage of semantic analysis with other data mining techniques can help to find focus, patterns and trends in texts connected to user feedback.

$H5_0$: The usage of semantic analysis with other data mining techniques cannot help in finding focus, patterns and trends in texts connected to user feedback.

3.6 Parts of this study

In **Chapter 4**, I present my conclusions by analysing the components of micro and macro relationships of complex system components. The observer can easily describe the set of system states. Internal and external forces can induce the transition between these states. I hypothesise, that understanding the motivations of smaller groups and successfully integrating this knowledge to understand higher-level elements of the system, facilitates the understanding of complex transitions.

In **Chapter 5,** my dissertation aims to fill the gap by focusing on the trading of stocks of football clubs or big companies. In both cases, the mood is an essential factor. For example, the fans are

emotionally attached to the teams; therefore, the number of their referrals to their favourite team in tweets fluctuates continuously. I hypothesise that their decisions about investment in the football club are not only driven by the desire of long-term prosperity, but also by their daily emotional state towards the clubs.

In **Chapter 6,** I hypothesize a strong economic connection in the relationship between the usage of social media and the volatility of the share prices of large companies. The results of the research provide measurable evidence for the assumption that in the case of large companies' financial market activities and the number of tweets is related. The novelty of the research is that it shows, which daily tweet rate has the most significant impact on the stock market.

In **Chapter 7,** we analyse followers of prominent football clubs by obtaining their home locations from Twitter messages. We then measure how the size of the city is connected to the number of followers by using the theory of urban scaling. I hypothesise that the scaling exponents of the club followers depend on the income of their country. These findings could be used to understand the structure and potential growth areas of global football audiences.

Chapter 8 is an empirical approach to the linguistic analysis of customer feedbacks using statistical natural language processing (NLP). When considering the customer feedback, there is a need to use text mining techniques in order to gain insight or perceive the focus area of the text because they are mainly written in an unstructured way. The case is more complicated when the description of the events or incidents is associated with a large organization providing a wide range of IT services. I hypothesise that usage of semantic analysis with other data mining techniques can help in finding focus, patterns, and trends in the texts connected to the users' feedback.

4. Paper 1: Micro and macro interactions in social neurobiological systems

Abstract: The approach of complex systems as a form of new approach provides an opportunity to replace the dominant, mechanistic view of sport-related phenomena. By analysing the micro and macro relationships of complex system components, the observer can describe well the set of states of the system. The transition between these states can be induced by internal and external forces. Understanding the motivations of smaller groups and successfully integrating this knowledge to understand higher-level elements of the system, facilitates the understanding of complex transitions. The novelty in this research is that micro and macro changes are integrated as a common driver, and they generate some entirely new property for the whole system. In the following case study, a small football team is presented as a complex system. There are statuses examined and elaborated that this small team and its members go through during the observed period (2005-2009) which can be similar to other complex systems.

The case study successfully mirrored the behavioural dynamics of agents in a social neurobiological system, exemplified by interactions of statuses in a team sport.

4.1 Introduction

Complex systems (teams, tactics) observed in sports consist of structurally and functionally heterogeneous components that interact (usually informatively and/or mechanically) with different intensities and spanning different spatial-temporal scales (Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006). These systems are purposefully changing their behaviour and adjust to the constraints that arise (Schuster, 2005). This feature significantly increases their level of complexity, and this is a major challenge for modelling techniques. In such systems, new forms of behaviour are constantly appearing under various constraints, without being designed or enforced previously. This is one of the main characteristics of the sport-related phenomena.

However, even complex systems can behave as simple ones, because their interacting components can form coalitions of cooperative elements that reduce the dimension of the behaviour. (Glazier & Davids, 2009). This way, a complex system displays simple behaviour

and can be treated as a simple system at the macroscopic level (Asur & Huberman, 2010). On the other hand, the complex system regulates itself. It is created in a selfless organization without control, patterns or structures can emerge, and then something unique happens, and a completely new quality of the system is revealed (Anderson, 1999). This new quality is called the emergent feature of the complex system. More than 35 years ago, the science of complexity influenced some trends in the sport. Nowadays, systematic research has been established to observe the coordination dynamics (Balague, Torrents, Hristovski, Davids, & Araújo, 2013). Coordination dynamics are defined as the science of coordination that describes, explains, and predicts how patterns of coordination form, adapt, persist, and change in living organisms. This field of work aims to understand the principles and laws that lead to the formation of dynamics of behavioural patterns under changing constraints (i.e., boundary conditions). (Preiser, 2019) These constraints may be classified into three sub-classes: task constraints, personal constraints, and environmental constraints (Glazier & Davids, 2009). These constraints are also reflected in the case study.

4.2 Literature review

The complex system has many components that can be complex systems themselves, (Liu & Barabási, 2016) and components can belong to multiple systems at the same time (S. A. Levin, 1998). There are dynamic links, or interactions between the elements, which reinforce or weaken each other (positive and negative feedback) (Pomerening, Sun, & Ferrell, 2005). The nature of interaction can be energy, material or information exchange, or a combination of these (Gros, 2008). Emergence plays a central role in theories of integrative levels and of ***complex systems.*** In short, it means that the system creates a new quality of its characteristics. The emerging property is somewhat expedient, but this global objective is not present in the components that make up the system: the components have local interactions, but they all make up something entirely new, which is the specific feature of the whole system (Filo, Lock, & Karg, 2015). Another important concept in complex system theory is that of swarm intelligence (X. Li & Clerc, 2019). Ants and bees are social insects that as individuals do not have special intelligence, but using interpersonal relationships (communication by pheromones), a "super-intelligence" is formed (super-intelligence meaning that it is above the individual's level) (Karaboga & Akay, 2009; Rajasekhar, Lynn, Das, & Suganthan, 2017). This is typical for big

cities as well. Therefore, the economic efficiency of big cities, if we look at it at the citizens' level, is far better than in the villages (Grimm et al., 2005). Another simple definition of a complex system is that if we change some input parameters of the system slightly, then the output parameters will change significantly. We can see from this simple definition that the modelling of a complex system is a very difficult task (Balague et al., 2013).

4.3 Methodology

A good example of complex systems is the social and economic organizations of people; for example, cities (Barankai, Fekete, & Vattay, 2012; Grimm et al., 2005). In this research, a football team has been observed as an example of a social and economic organization of people. The chosen methodology used for studying complex systems is the analysis of the collective behaviour within the team. In this case, the integrated levels interact at different levels during the performance. (Ethiraj & Levinthal, 2004). Different techniques, tactics, physical abilities, decisions, thinking or physiological processes, creativity or social dynamics are no longer seen as isolated or independent aspects, but as interdependencies and commonalities (Filo et al., 2015). New proposals make it possible to generate both multi-personal and individual learning and coaching strategies (Davids et al., 2013). The behavioural variability of the system acquires a functional value. It can provide information about the state~~s~~ of the system (its resilience and adaptability to changes, or, conversely, its inability or inflexibility) (Liu & Barabási, 2016; Liu, Slotine, & Barabási, 2013).

Previous studies have shown, that different systems in a sports environment are very sensitive to the limitations and small changes of variables in critical areas (Balague et al., 2013; Davids et al., 2013; Torrents & Balagué, 2006). The benefit of these investigations was that they showed some universal properties of the dynamics of complex systems, which were originally unchanged at the organizational level of the case. Various authors have studied the impact of new methods on the training of the practitioners in sport and the new role of athletes and coaches (Malina, 2010; Torrents & Balagué, 2006). The differentiated learning approach suggests that teams tend to find optimal performance patterns (Sanchez-Segura, Hadzikadic, Dugarte-Peña, & Medina-Dominguez, 2018) by adding noise during exercise (Passos et al., 2011). The coaches or athletes explore the state of the space until they find the best solution. This research has focused on decision making and creativity in sport. For example, new

activities may arise if a coach or a team has the opportunity to explore the new region of the action workspace.

The stochastic movements in the environment may give rise to the invention of new and functional actions (Bolotin, Tur, & Yanovsky, 2009). Crucial elements of this approach are the various independent agents of the system with defined relations between them (players, coach, and so on).

4.4 Case study: a small football team as a complex system

This case study describes and evaluates novel applicability of network methods in understanding human interpersonal interactions in social neurobiological systems (Passos et al., 2011) such as a small-town football team in the period 2005-2009 ("Electro – Vojvodina" – Szabadka – Serbia). It will be presented how collective system networks are supported by the sum of interpersonal interactions that come into view from the activity of system agents (such as players or coach of a football team). For example, it was observed how the interactions between the coach and the team members influenced the team's mood or tactics and the response effect of these changes.

The aim of this research is to describe the well distinguishable states of the system. The transition from one state to the next was always time-consuming (usually a half of a year). Driving forces behind each transition have been described as simply as possible.

First, the system agents were listed, then the list of the attributes associated with them. In cases where it is not clear, it has also been specified what value can an attribute take. Finally, the stages of the system and the transitions are listed in chronological order.

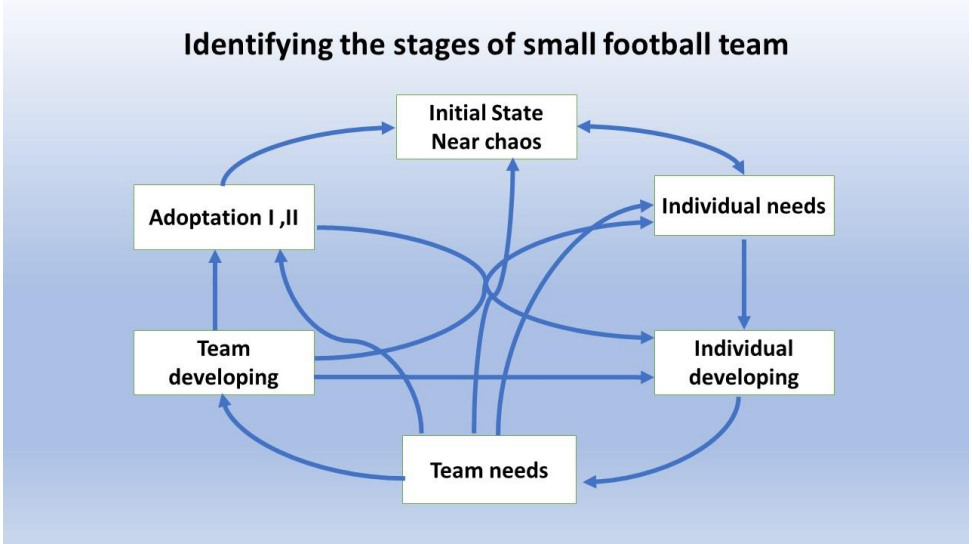

Figure 13 Stage diagram of the small football team

Source: Self-prepared chart

The creation of a model aims to contribute to the clarity of the observed systems. (Figure 13) The key elements (Agents) of the system are **players, coach, mood** and **tactics on the team level**.

Each agent has certain attributes that can change over time.

Shortlist of the attributes of the agents:

The attributes of Players: age, experience, position in the team, motivation (Tušak, 1997)

The attributes of Coach: experience, sport diplomatic influence, motivation

The attributes of the mood of the team:

- Good (the players attend the training, they meet each other in private life, there are regular out-of-work activities with family members, there are constant fans)
- Medium (the attendance of the training is sufficient, there are no common activities, a few fans)
- Poor (the attendance of training is unsatisfying, there are no common activities, there are no fans)

The attributes of tactics of the team:

- Good (technical elements exist and take the opponent's peculiarities into consideration)
- Medium (Technical elements exist, but there are no adaptations to the opponent yet)
- Poor (Technical elements do not exist – only individual solutions)

Identifying the complex system stages and transactions between these stages

In the initial stage, all attributes of the system agents will be listed separately. The following sections show only the possible changes.

Stage: **Initial state -Near to chaos (**season 2005/06 autumn Figure 6)

At the beginning of the observation, the team is on the edge of disruption. Coach motivation is very low.

Players:

- age: the average age of players is 26 years
- experience – several years
- motivation – medium
- position in the team – there are no defined positions for individual players

Coach:

- experience – high.
- influence of sports diplomacy – high
- motivation – low

Team mood: Medium

Tactics: Poor

The driving force behind the transition: The old coach leaves the club. The most experienced player takes over the coach's role.

The new coach:

- experience – limited
- influence of sports diplomacy – limited

- motivation – extra high: the aim is to win the championship

Stage: **Individual needs** (season 2005/06 – spring Figure 6): The new coach transforms the team into a new state. The championship begins. The coach starts the training. He improves the general mood by increasing the extent of running and starts to map the players' motivation. He tries to develop smaller groups within the team (forwards, defenders, midfielders, goalkeeper – according to the players' wishes). Therefore, the players' motivation increases. The team performance is good in the championship, and it is a surprise for their opponents.

Stage: **Individual development - (**season 2006/07- autumn)

During the season, the coach tries to change the motivation of the players. Based on a unique survey, he makes a personal workout plan. As a result, the motivation of some players improves, but the team's events are missing.

Stage: **Team needs, Individual needs** and **Initial state (**season 2006/07 **–** spring)

The coach's experience increases. The average age of players is growing (no new players). The coach tries to alter the roles within the team according to the team's needs, so the motivation of the players is decreasing, but the team's tactical repertoire increases. Some players leave the team. The team is close to chaos again.

Stage: **Team development and Individual needs and Individual development (**season 2007/08 **–** autumn)

Recruitment of young, talented players for specific roles within the team. The average age of players decreases. Because of the new players, the level of motivation in the team is declining. The team score is moderate.

Stage: **Adaptation I and Individual development (**season 2007/08 **–** spring)

Tactics level increases - adaptation to the opponent. The atmosphere is excellent. Motivation is appropriate. The team leads throughout the championship, but another team is eliminated, and therefore the points scored are recalculated. The coach has a low degree of sports diplomacy, so after the recalculation, they finish in the second place.

Stage: **Adaptation II** and **Individual development (**season 2008/09 **–** spring and autumn)

Level of team tactics increases - adaptation to the opponent. The atmosphere is excellent. Motivation is appropriate. The team is leading the championship, change in the rule

of sending the referees does not affect the team negatively because the coach has a high degree of sports diplomacy, the team wins the championship.

Stage: **Near to chaos (again) - Season 2009/10**

There is a significant change in the coach's life: he has a new job, and the old team members are no longer motivated. Some key players take advantage of the success achieved and transfer to other teams. This results in a drop at the tactical level. The coach leaves the team. No new young players arrive, a new coach is to be chosen. Return to the initial state.

4.5 Conclusion

Sport is not only a social phenomenon in our world but also a true depot of experiments of human behaviour. It provides an opportunity to study effectively and efficiently the effects of intensive change in complex life systems at many levels (psychological, social) (Vega, 2015). On the one hand, the study of complex systems in sport might lead to a better understanding of evolutionary processes, optimization of resource extraction/allocation, and economic transactions; strategies for economic and ecosystem resilience and sustainability (Sasou & Reason, 1999). Analysis of factors influencing performance, motivation, and determination can often be understood through interactions (synergies) (Liu & Barabási, 2016). For example, a player's performance depends not only on his knowledge, his role in the team but also on the state of the receiving environment.

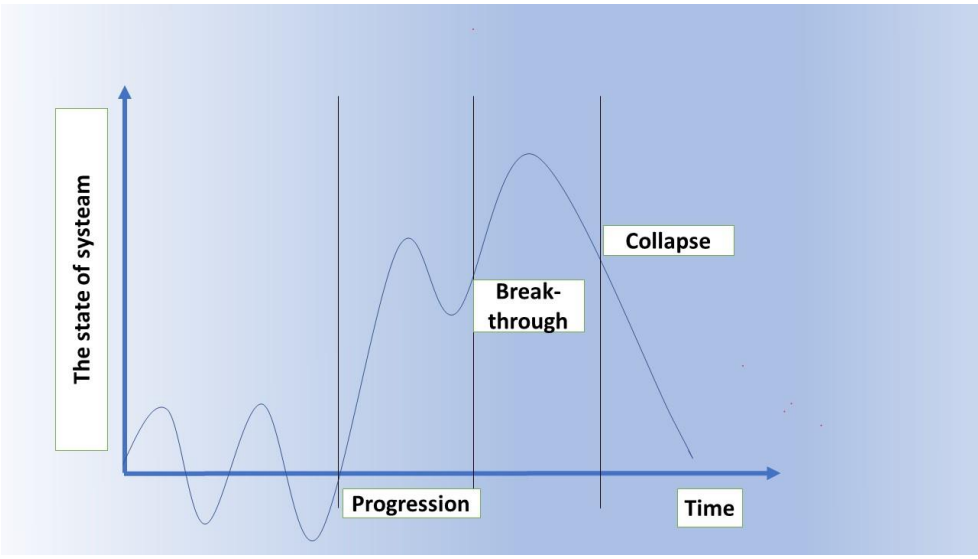

Figure 14 System macro changes (decreasing or increasing its performance)

Source: Self-prepared chart

The observed system undergoes many macro changes (Figure 14). Due to inefficiency, increasing tension forced the participants to change roles: from a player to coach, as well as from individual player to a team member role. Despite the initial stumbling block, the coach's experience grew, and the quality of the player's motivation has improved. However, they did not consider that the opponents have also learned their playing style since they were built on simple elements of individual solutions. The breakthrough was finally achieved by the evolution of team attributes. They were able already to adhere to the circumstances at the team level, to adapt to their opponents (Memmert, Lemmink, & Sampaio, 2017). The collapse began on the one hand with the loss of the motivation and became complete by an external factor. This could have been avoided so that a resource must be maintained in order to maintain development. For example, by developing an assistant-coach, the team would not be so much dependent on one person.

In summary, studying the complex systems of sport can have some specific and general benefits. On the one hand, it can help to improve human performance in sports and create new strategies for teams and coaches. Improved understanding of human social networks, including player psychological drivers and group participation; developing a policy of participation in sport and education of young people (Holt, Tamminen, Black, Sehn, & Wall, 2008). The principles governing these group interactions are expected to apply to other biological collectives, where dynamics derive from a mixture of competition and physical principles (Boccaletti et al., 2006; Cao, 2015; Steels, 2000). These include superorganisms such as ant colonies, birds and penguins (Karaboga & Akay, 2009). Because of the potential for obtaining rapid empirical feedback on formal models, the modelling of sport-related phenomena requires particular attention from the point of view of science and complex systems theory (Cao, 2015). Sophisticated sciences can contribute to changing the mechanical view prevalent in sport and can usefully contribute to understanding complex systems.

4.6 Limitations and future work

In this research, the development of only one team has been monitored during a particular period. However, were carefully selected the parameters of the coach, the players and

so on. Some important attributes may have been missed, depending on the special circumstances. In the future, this research could involve more teams, other sports and include additional parameters.

The scope of some future research may include the analysis of the influence of key control parameters on the non-linear behaviour of the systems. It could be extended to the team-environment and the possible relationships between dynamics and constraints that affect team sports on different spatial-temporal scales.

5. Paper 2: Influence of twitter activity on the stock price of football clubs

Abstract: The interrelation between the fluctuations of stock prices of football clubs and the social media activity of their fans is of major economic interest. The results in this research provide measurable support for the suggestion that financial market activities and the number of tweets are linked in the case of big football clubs. The novelty of the research is to show which daily ratio of tweet counts and their mood can predict the stock market. Based on these findings, sport managers could use social media strategically to build better relationships with consumers and stakeholders.
KEYWORDS: Twitter, social media; football, stock market

5.1 Introduction

Research has investigated daily fluctuations of stock prices and found that they are influenced by the emotional state of the trading individuals (Vivek, Beatty, & Morgan, 2012). In particular, the emotional state can be affected by economic trends and major political events but according to several studies also by sports events, even if with a smaller impact (Ashton, Gerrard, & Hudson, 2003). Among these events are the main football games (FIFA World Cup, UEFA Champions League, major matches of South American and European football leagues) that attract millions of people around the world (Yang, Wang, & Billings, 2016).

Nowadays, football fanatics are following their favourite clubs much closer than ever before thanks to the availability of many communication channels (Nielsen, Storm, & Jakobsen, 2019; Sanderson J., 2011; Scholtens & Peenstra, 2010). These channels are not limited any more to the one-way distribution of messages and images. They include social media that allow two-way online interactions (Hromic & Hayes, 2019), in which thoughts and opinions can be disseminated and accessed from all over the world (Dellarocas, 2003; Sóti, Bokányi, & Vattay, 2018). In this light, the emergence of social media has profoundly impacted the delivery and consumption of sports events (Filo et al., 2015; Park & Dittmore Stephen, 2014).

Previous studies have shown that victory or defeat of a sports team affects the stock market (Renneboog & Vanbrabant, 2000; Vicentini & Graziano, 2016), however, to the best of our knowledge, no study has proposed a method to evaluate the relative importance of different events related to a football club, e.g. not only the impact on the stock market of victories and defeats but also of events like the change of a coach or winning a championship. This study tries to fill this gap and considers that the fans are emotionally attached to the teams while their moods are constantly fluctuating (Dodds, Harris, Kloumann, Bliss, & Danforth, 2011). The hypothesis is that the decisions about investment in a football club are not only driven by the feeling of long-term prosperity, but also by their daily emotional state towards the club (Mael & Ashforth, 1992). Based on this approach, the paper investigates whether public sentiment, as expressed in large-scale collections of daily Twitter posts (Ji, Chun, Wei, & Geller, 2015) - can be used to refine prediction in the stock market.

The analysis was carried out by creating a linear model to estimate whether the fluctuation of a given stock after a given event matches the fluctuation of the twits exchanged by fans in a few days after the event. The results of this research show that the accuracy of the prediction of conventional stock exchange forecasting models is greatly improved by incorporating the mood dimension, measured as the how many times fans refer to their favourite team in their on-line messages on Twitter. Findings confirm that when an event occurs the number of tweets increase in the days that follow (Guille & Favre, 2015), and they predict in most cases the stock price deviation from the model that has been created.

The methodology has been tested using three famous football clubs, namely Manchester United Juventus, and Ajax. These clubs are suitable case studies because they are among the most important football clubs in Europe, they are listed in the stock market, and Twitter is one of the preferred platforms for interaction among their fans.

The study is relevant both from the scientific and the managerial points of view. From the scientific point of view, it provides a methodology to refine the prediction of the trading stock fluctuations following main events in the life of the sports teams. From the managerial perspective, findings support the importance for managers to closely monitoring media and the need to develop appropriate strategies to influence conversations among fans (Hinz, Skiera, Barrot, & Becker, 2011; van der Lans, van Bruggen, Eliashberg, & Wierenga, 2009).

The paper is divided into 6 sections. The second section provides another view of the previous results that have motivated the current research. After that, data collection and handling

procedures are proposed. Then, the results section presents the application of the methodology of the three case studies. The conclusion chapter shows the importance of the result, whilst the limitations and future work section list possible directions for further research.

5.2 Literature review

Several studies have investigated the complex fluctuations of the listed companies on the stock market, but no one has so far provided a completely reliable method to forecast stock market evolution (Fama, 2002; Schredelseker & Fidahic, 2016). Stock trading is a complex process, dependent on interactions between companies and customers, where stock prices often show unpredictable behaviour (Roy & Sarkar, 2010).

Due to the potential benefits of reliable forecasting, the stock market prediction has attracted much attention from science as well as from businesses. Predictions used to be based on the interpretation of market conditions, speculation, and available information (Schredelseker & Fidahic, 2016). Early research on stock market prediction was an outcome of the efficient market hypothesis (Fama, 2002), and claimed that stock market prices are largely driven by new information, especially news, rather than present and past prices.

Recent literature on behavioural economics tells us that emotions can profoundly affect individual behaviour (Topal, Koyutürk, & Özsoyoğlu, 2017) and decision-making (Weimar & Schauberger, 2018). Several studies have confirmed that public mood is correlated and can even predict economic indicators (Bollen, Mao, & Zeng, 2011). Psychological research has shown that although news certainly influence~~s~~ stock market prices, public mood, states or sentiment may play an equally important role, as they play a significant role in human decision-making (Kahneman & Tversky, 1979; Stieglitz & Dang-Xuan, 2013). Therefore, according to findings in the literature, there is proof that financial decisions are significantly driven by emotion and mood and that it can be reasonable to assume that the public mood and sentiment can drive stock market values as much as news (Nofsinger, 2005).

Since it was not possible to collect a huge quantity of data inexpensively in the past, the impact of events was evaluated by using official odds (Scholtens & Peenstra, 2010). However, recently the prediction of stock market fluctuations based on on-line activities has been proposed in literature at the national macro-level. Some contributions have shown how online activity in chat predicts sales of different products like books (Gruhl, Guha, Kumar, Novak, &

Tomkins, 2005) and movies (Mishne & De Rijke, 2006; Rui, Liu, & Whinston, 2013). Recently, public sentiment related to movies, as expressed on Twitter, has been connected to the prediction of box office tickets sales (Asur & Huberman, 2010).

A few models of stock market fluctuations have been proposed based on football games' results (Edmans, García, & Norli, 2007), and several papers have investigated the link between the performance of the football or basketball teams with the stock returns (Scholtens & Peenstra, 2010; Vicentini & Graziano, 2016) and volatility movements (Brown & Hartzell, 2001).

The majority of global companies use social media (i.e. 97%), and many of them are increasing their social media marketing budgets (Kumar, Bezawada, Rishika, Janakiraman, & Kannan, 2015). Among them, also large football clubs and football leagues largely invest money in establishing official social media channels (Bonchi, Castillo, Gionis, & Jaimes, 2011) to engage with their fan base (Filo et al., 2015; Price, Farrington, & Hall, 2013). In this direction, many football clubs try to exploit the possibilities of influencing the latest news through immediate messaging (Bruns, Weller, & Harrington, 2014).

The online social network Twitter has more than 300 million active users monthly and 67% of sports fans are likely to use Twitter to increase their experience of television viewing compared to non-sport fans (Mancuso & Stuth, 2013). Using the possibility given by this platform, the paper has analysed the possibility to identify a relationship between Twitter-based analysis of football clubs and the short-term market performance of the clubs (Roberts & Emmons, 2016) using large scale tweet data.

5.3 Case study selection and Twitter data collection

5.3.1 Twitter data

This study relies on Twitter data originally collected using the Twitter API and maintained in the Virtual Observatory of Eötvös Loránd University (Dobos et al., 2013). Twitter API cannot access the data stream of the entire Twitter platform, as only a randomly sampled fraction is available. In order to compensate for the sampling, the number of the football-related Twitter data had to be related to the overall stream rate of the background Twitter data. From the Virtual Observatory database, tweet records generated from January 1st, 2012 to December 28th, 2014 were downloaded. This time span is called the "observation

period". Data collected contain information about the unique identifier of the Twitter user, the timestamp of the creation of the tweet, and the text content. In the first step, data were filtered against the text content, and records related to the most tweeted football clubs. Due to the characteristics of the information necessary to perform the analysis, multiple case-study approaches were selected. In the selection of clubs, three conditions needed to happen together. The first was that the club is popular enough to inspire a large number of tweets. Next, the club needed to be a public company listed in a stock market, to have the information on market fluctuation. Finally, it was necessary that historical data on tweets are available for a period in which there are a sufficient number of events to be analysed. These three conditions were fully satisfied only by a few clubs, namely Manchester United, Juventus and Ajax Amsterdam, which are among the most influential clubs in Europe (Kantar Media, 2013). In fact, two other very popular clubs listed in the stock market, namely Borussia Dortmund and Arsenal, were discarded because they did not satisfy the selection criteria. In particular, Borussia Dortmund does not have enough tweets daily to use for analysis, and the share price of this team does not change significantly to appreciate the variations due to the sports events. Next, the English team Arsenal could not be considered because of the extensive use in English of the word "arsenal" not associated with the name of the football team, and if #arsenal was to be used there were not enough tweets daily. Finally, other teams that are listed on the stock market, such as Porto, Benfica, Roma, or Celtic, are not popular enough on Twitter, and it was not possible to collect enough daily tweets.

The selection of the three case-studies (Manchester United, Juventus, and Ajax Amsterdam) was further supported by the fact that all the other major clubs had far smaller tweet records than the three teams selected. Data on other top clubs like Real Madrid, Barcelona, and Bayern Munich (all of them not listed in the stock exchange market) were used to compare the evolution of tweets during the observation period.

In a second step, tweets related to the case-studies were filtered based on word matches in the content of the tweeted text. The filtered words (besides the denominations of the clubs) are reported in Table 1, along with the sample size and the number of followers. It can be estimated that in the given period, roughly 2-3% of the reported followers of each club were captured via their tweets.

Table 1 Summary of Twitter data

Football club	Words	Followers (No.)	Followers in database (No.)
Manchester United	manunated glorymanunated	17.3M	436515
Juventus	juventus, juve	6.46M	122345
Ajax	Ajax, ajaxamsterdam	1,3M	40234

In what follows, the three football clubs selected are described regarding their sportive performance during the observation period.

5.3.2 Juventus Football Club

Juventus stocks are traded on the Borsa Italiana, in Italy, Juventus's home country. Juventus Football Club S.p.A. (Juve for short, "La Vecchia Signora" (The Old Lady) to its fans) is an Italian football club playing in Serie A. The club, founded in 1897, is the best-supported football club in Italy. It has over 12 million fans, and they represent approximately 34% of all Italian football fans. According to research published in September 2016 by an Italian research agency (Demos & Pi, 2017), Juventus is one of the most supported football clubs in the world, with over 300 million fans whereof 41 million live in Europe (Juventus Football Club S.p.A., 2017), especially in the Mediterranean countries.

The club has seen great successes in recent times as well as during the observation period **(Table 2**), winning over and over again the Italian Championships at 2012/13, 2013/14 and 2014/15. The club was ranked 9th in Europe (1st in Italy) in 2013 in the Football Money League (Admiral Market, 2019).

Table 2 Results of Juventus in the observation period.

Season	Serie A	UEFA Champions League	Italian Cup
2012/2013	First place	Quarter Final - lost against Bayer Munich	Semi-finals
2013/2014	First place	Semi-Final - lost against Benfica	Quarterfinals

2014/2015	First place	Final - lost against Barcelona	Win a Cup

Tweets from 01.01.2012 to 29.12.2015 show that the most important events for the fans were firstly those related to the international cup competition, rather than those related to the national championship, and only lastly the replacement of the coach (**Table 3**).

Table 3 Events connected to Juventus in the observation period.

Date	Event	Tweets in the database (No)
2012-09-19	Draw against Chelsea	3439
2012-11-20	Win against Chelsea (3-0)	4378
2013-02-12	The first leg against Celtic	6476
2013-03-06	The second leg against Celtic	3784
2013-04-02	The first leg against Bayer Munich	3789
2013-04-21	Juventus-Milan 1-0	1816
2014-06-16	M. Allegri new head coach of Juventus	546

5.3.3 Manchester United Football Club

Manchester United is the largest football club traded on the stock market. It is an English football club playing in the Premiership. The Club was founded in 1878 named Newton Heath Lancashire & Yorkshire Railway Football Club and has now 659 million fans. About half of the Manchester fans (325 million) live in the Asia-Pacific region, 173 million in Africa and the Middle East, 71 million in America and 90 million in Europe. More than 108 million fans come from China. More than 30% of South Korea's 49 million inhabitants are fans of Manchester United. Nigeria, with about 22 million fans, has the most Manchester United fans in Africa and the Middle East (Kantar Media, 2013). Manchester United stocks are traded on the NYSE (New York Stock Exchange) (Admiral Market, 2019).

Tweets from 20.08.2012 to 29.12.2014 show that the club has seen some success and some failures in recent times (**Table 4**). It won the Championships in 2012/13, but then in 2013/14 it had a very disappointing result in the Premier League and did not qualify in 2014/15 for UEFA

Champions League. Nevertheless, the club was ranked 1st in Europe in the Football Money League (Admiral Market, 2019).

Table 4 Results connected to Manchester United in the observation period.

Season	Premier league	UEFA Champions League	FA Cup
2012/2013	First place	Round of 16 - lost against Real	Round 6 lost against Chelsea
2013/2014	7th place	Quarter Final - lost against Bayer Munich	Round 3 lost against Swansea City
2014/2015	4th place	Did not compete	Round 6 lost against Arsenal

The number of tweets shows (**Table 5**) that it is more important for fans that the team can pass competitions successfully through several turns in an international cup, than the replacement of a coach or negative results in the national championships.

Table 5 Events connected to Manchester United in the observation period.

Date	Event	Tweets in the database (No)
On 20 August 2012	The first game of the Premier League campaign 2012/2013	1523
On 9 December 2012	Win against Man City	2525
On 20 January 2013	Draw Against Tottenham	2185
On 13 February 2013	The first leg against Real Madrid	5425
On 5 March 2013	The second leg against Real Madrid	7756
On 22 April 2013	United won an unprecedented 20th English top-flight title	3247
On 8 May 2013,	United's manager, Sir Alex Ferguson announced that he would retire	1063

On 5 January 2014	Lost against Swansea City FA Cup	390

5.3.4 Ajax Amsterdam

Amsterdamsche Football Club Ajax also known as **AFC Ajax**, **Ajax Amsterdam**, or simply **Ajax**, is a Dutch professional football club. AFC Ajax is the only Dutch club with an Initial public offering. The club is registered as Naamloze vennootschap (N.V.) and is listed on the stock exchange Euronext Amsterdam, since 17 May 1998. According to a study by Football International (Holland's biggest football magazine), Ajax has more supporters than any club in the country, that is around 3.7 million. If someone looks at the history of Ajax, can only really judge why this club is so popular all over the world. There was also a player on the team, such as Johan Cruyff, who would become the greatest Dutch footballer of all time. Tweets from 1 January 2012 to 29 December 2014 show that the club achieved some success and some failures during this time. The club has achieved great successes in recent times **(Table 6**), winning over and over again the Dutch Championships in 2012/13 and 2013/14. However, the team is rapidly eliminated from international cups in all three years.

Table 6 Results connected to Ajax in the observation period.

Season	Eredivisie	UEFA Champions League	KNVB Cup
2012/2013	First place	3rd Place in Group D	Semi-finals
2013/2014	First place	3rd Place in Group H	2nd place
2014/2015	2nd place	3rd Place in Group C	Fourth Round

The number of tweets shows (**Table 7**) that the highest number of tweets was measured when the team played with a prestigious opponent in the international cup.

Table 7 Events connected to Manchester United in the observation period.

Date	Event	Number of tweets in our database
On 24 October 2012	Ajax – Manchester City (3–1)	3432
On 4 December 2012	Ajax - Real Madrid (1–4)	2525
On 18 September 2013	Barcelona – Ajax (4–0)	3156
On 26 November 3013	Ajax – Barcelona (2–1)	2934
On 17 September 2014	Ajax – PSG (1–1)	3027
On 21 October 2014	FC Barcelona – Ajax (3–1)	3067
On 25 November 2014	Paris Saint-Germain – Ajax (3–1)	2016

5.4 Semantical analysis

In order to understand stock market movements, it was important to determine the mood of the text of tweets on a given day. This information was needed to distinguish the stock market rise from the fall through the mood of the tweets as well. In order to perform the semantic analysis of tweets, results reported in Dobbs et al. (Dodds et al., 2011) and the average happiness value of the words have been evaluated from the given website ('http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0026752.s001'. The value was calculated for each meaningful word in all tweets. This value was attributed on a daily basis to each team and multiplied by the number of tweets per day. Then, the time series of the tweet mood intensity have been analysed. The tweet mood intensity $\mathbf{v_j}$ of club j on a given day is defined as the ratio of the daily mood intensity of that club $\boldsymbol{T_j}$ and the overall tweet emotional state of the same day $\boldsymbol{T_m}$:

$$\boldsymbol{v_j} = \frac{\boldsymbol{T_j}}{\boldsymbol{T_m}} \qquad (4)$$

5.5 Stock data

The second source of this study is stock market data collected from https://finance.yahoo.com, providing daily aggregates of each stock product. Figures (1,2,3) report for the three case-studies the closing prices.

In what follows, the fluctuation of the stock market prices for each of the three case studies in the observation period is reported along with the identification of the main sports events that occurred to each club.

5.5.1 Juventus

Figure 15 shows the evolution of the closing price of Juventus in the observation period. Values oscillate during the time with a few abrupt changes, likely caused by the outcomes of matches. For example, peak "1" between May 2nd and 5th, 2012 appears right after the draw between Juventus and Lecce (1:1), when for Juventus this point was enough to win the Italian championship "Serie A". Similarly, peak "2" on October 15th, 2013 comes right after the draw between Italian and Armenia national football teams (2:2), when Italy qualified for the FIFA World Cup in the first place from their group, with 7 Juventus players listed in the team.

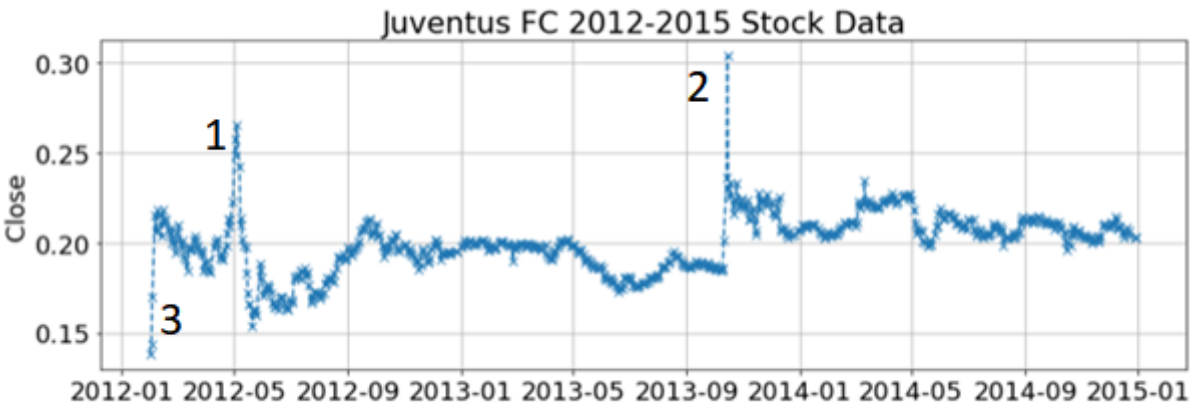

Figure 15 The closing price evolution of the Juventus stocks

Source: Self-prepared chart

The counterexample is the drop marked as "3" when Juventus was defeated in a series: January 6th, 2012 against Parma (1: 4), January 9th, against Napoli (0:3) and February 2nd, against Palermo (1:2). Taking these observations into account, one can see that Juventus is a typical stock exchange company, whose closing price values tend to be sensitive to important events.

5.5.2 Manchester United

The series of closing prices of the Manchester United stocks shows a more stable temporal evolution than Juventus, regardless of whether or not the team performed somewhat worse at the end of the investigated period from a sportive performance point of view (Figure 16). Therefore, it seems that as a market-leading stock exchange company, it has a more stable investor background. Then Manchester United's stocks are less sensitive to the events related to the football club's performance. Fluctuation peaks and drops do not show a strong correlation with the team's performance. At the beginning of 2013, the exchange rate increased because the team won the championship (Figure 16, "1"), and the peak point "2" in July 2014 can be linked to the friendly match which they won against LA Galaxy. (7:0).

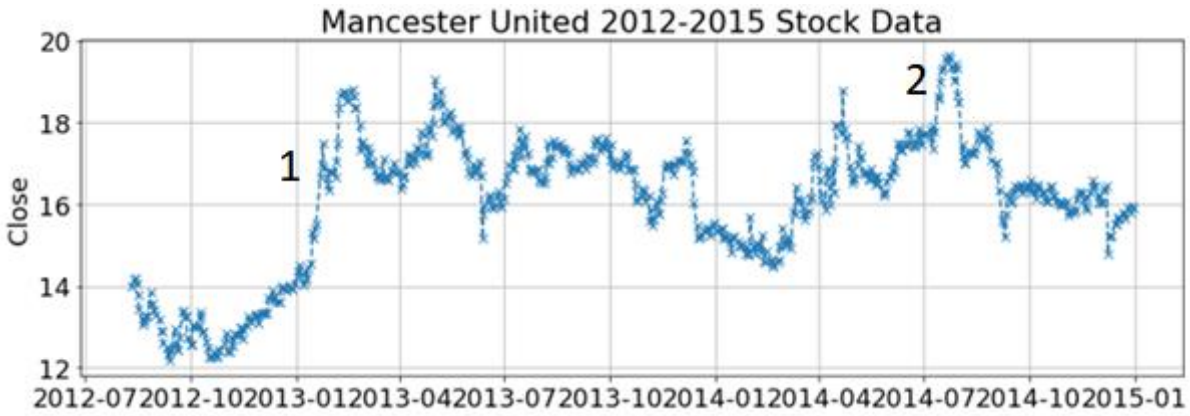

Figure 16 The closing price evolution of Manchester United stocks.

Source: Self-prepared chart

5.5.3 Ajax Amsterdam

Figure 11 shows the evolution of the closing price of Ajax in the observation period. Figure 17 shows that shares of Ajax Amsterdam rose steadily until almost the end of the period. At the end of the observed period, there is a larger increase, followed by a slight correction. Two exchange rate increases are indicated in Figure 11. The first increase is reported in Figure 11 ("1"), when Ajax won the national championship in the last match in the 2012/13 league against the NEC (ending with a 2: 2 draw). It was not a bright win, yet it did mean that AJAX will start in the champions 'league next season, so its stock price rose. The second increase is in point "2" (Figure 11), a peak point occurs at "17" September 2014. It can also be linked to a draw result, (promising start in the Champions League) when AJAX plays 1: 1 against Paris Saint Germain in the first round of the Champions League.

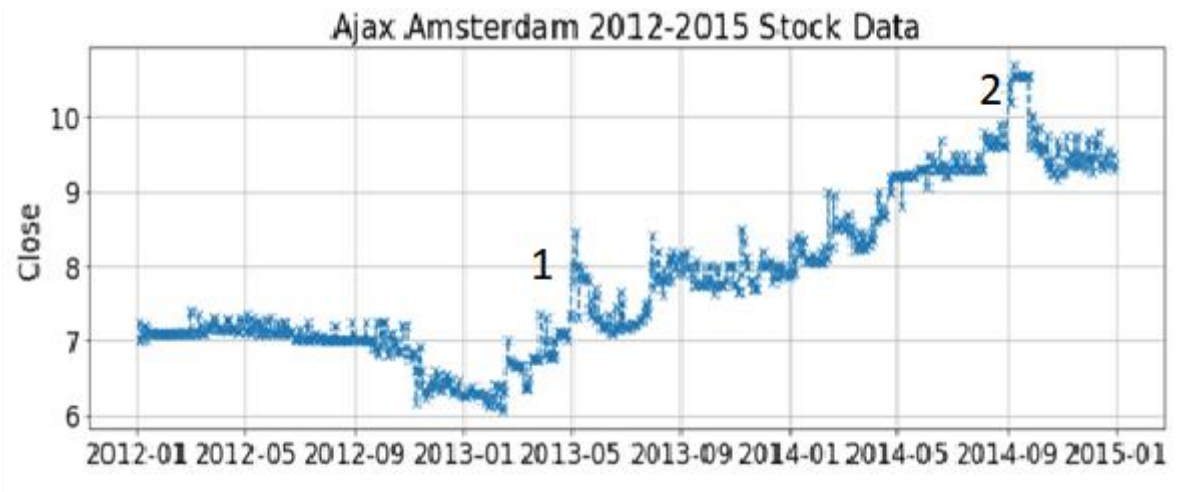

Figure 17 The closing price evolution of Ajax stocks

Source: Self-prepared chart

5.6 The estimation models.

From the closing price series $\boldsymbol{C_{j,t}}$ related to club j at a given day t - the magnitude of the stock's normal rate of return (1) - is calculated as follows:

$$R_{j,t} = C_{j,t} - C_{j,t-1^*} \tag{5}$$

where $t - 1^*$ indicates the previous trading day, i.e. when t stands for a Monday and $t - 1^*$ points to the Friday the week before.

The Dow Jones market index served as the basis for modelling the background processes of the stock market. The marginal normal rate of return $R_{j,t}$ was derived with the same formula (5). A linear relationship to the background is proposed as the first approximation of the average return rate (Model 1).

$$\textbf{MODEL 1:}\ R_{j,t} = b_j + \sum_{i=1}^{4} a_j R_{i(m,t-1)} + AR_{j,t} \tag{6}$$

where parameters a_j and b_j for football clubs are estimated by ordinary least squares method. Parameter a_j is the risk indicator of the club j in the stock market.

The market model defines the abnormal return of the stock.

$$AR_{j,t} = R_{j,t} - (a_j R_{(m,t)} + b_j) \quad (7)$$

Another model was considered for taking into account the mood of the tweets related to a given team to estimate the stock change (Model 2).

$$\textbf{MODEL 2:}\ R_{j,t} = b_j + \sum_{i=1}^{4} a_j R_{i(j,t)} + \sum_{i=1}^{4} cX_{i(j,t)} + AR_{j,t} \quad (8)$$

Where X_i is the daily sum of the mood of the tweets calculated for the given team.

The Granger causality analysis was performed according to MODEL-1 and MODEL-2 shown in Eq. 3 and 5 for the period between Jan 01, 2012, to Dec 31, 2014. The first model (M1) uses only lagged values of $R_{j,t}$, i.e. ($R_{j,t-1}$,·…., $R_{j,t-n}$) for prediction, while the second model (M2) uses lagged values of both $R_{j,t}$, and the results of calculated tweet mood of time-series denoted X_{t-1},⋯, X_{t-n}.

The Granger causality test is a statistical hypothesis test, that determines if time series are useful for predicting other time series. A time series X is said to Granger-cause Y if it can be shown, usually through a series F-tests on lagged values of X (and with lagged values of Y also included), that those X values provide statistically important information about future values of Y. Usually, the regression reflects “mere” correlations, but economic causality can be examined by measuring the possibility to predict future values for a particular time series using values of previous time series. In the analysis, it was investigated the hypothesis that public mood, measured by the mood in tweets, is predictive of future share values. In the Granger causality analysis, the share price values were correlated to the mood of tweets of the past n days.

5.7 Results

The two data sources are joined together based on time, which has the granularity of one day, in order to track tendencies. It can be noted that while stock data is not available on the weekend, as the markets are closed, Twitter data still exist during the weekend. To verify whether tweeting activity impacts trading decisions, a lag τ in days where $\tau \in \{1,2,3\}$ is allowed when joining the two datasets. A lag $\tau = 1$ means Twitter data of a day $t + 1$ and the stock data of the next day $t + \tau$ are moving together.

Based on the results of Granger causality (shown in Table 8), the null hypothesis can reject that the tweet-s mood time-series do not predict share price change, with a high level of confidence. It can be observed that on the first day, all soccer clubs have the highest Granger causality relation (p-values<0.05). On the second day, the Granger causality relation is smaller (0.0.5<p-values<0.1). On the third day, for all three clubs, the p-value is above the significance level (p-value>0.1)

Table 8 Statistical significance (p-value) of bivariate Granger-causality correlation between moods and stock price change (F-test)

	Juventus	Manchester	Ajax
1 day	**F=3.8742, p=0.0497***	**F=4.9879, p=*0.0260***	**F=5.0543, p=0.0251***
2 days	**F=2.8582, p=0.0586 ** **	F=2.4109, p=0.1080	**F=2.6631, p=0.0910** **
3 days	F=1.7605, p=0.1360	F=1.1632, p=0.3265	F=1.7298, p=0.1426

(p-value<0.05*, p-value<0.1**)

The Monte Carlo method essentially solves the problem by directly simulating the underlying process and then calculating the (average) result of the process. (number of tweets)

In the elaboration in the following chapter, has been using the computer tools in order to produce the final result of a given experiment, after which the resulting numerical characteristics were evaluated. In order to determine the possible error of the result, the calculation of the average distribution and standard deviations were performed.

In each case (Juventus, Manchester United and Ajax), three days were always observed: the next three days when tweets were released.

Joint samples are classified as the following. A threshold $\boldsymbol{v}^*$ is introduced. Samples of club j are said to belong to the positive group if $\boldsymbol{v_{j,t}}$ exceeds the threshold value. Discarding to differentiate between profitable or loss movement in the stock, the mean value of the absolute abnormal return is calculated as follows:

$$AR_j(v^*, \tau) = < \{ \left|AR_{j,t^*+\tau}\right| \left| t^* : v_{j,t^*} > v^* \right\} \qquad (9)$$

Dropping explicit notation of parameters j and τ. Let AR^c denote the value of group $AR_j(\boldsymbol{v}^*, \boldsymbol{\tau})$ and $\overline{AR^c}$ is the average of a complementary group. In Figures (4,5,6) the abnormal return of the statistics of $\frac{AR^c}{\overline{AR^c}}$ are presented for all clubs. The figures (12,13,14) show the following values: The impact of an event was measured by calculating the average response of the stock market to a set of relative tweet numbers. Twothousand sets of random relative tweet numbers were generated and the average effect on the stock market. was calculated This average distribution and one and two standard deviations from the average are shown in the upper part of these Figures (12,13,14) (green and orange stripes). The blue line represents the calculated effect of the relative number of tweets on the given day. The bottom part shows the difference between the calculated effect and the mean effect, measured in standard deviation (yield as σ confidence) (Figures 12,13,14).

Juventus

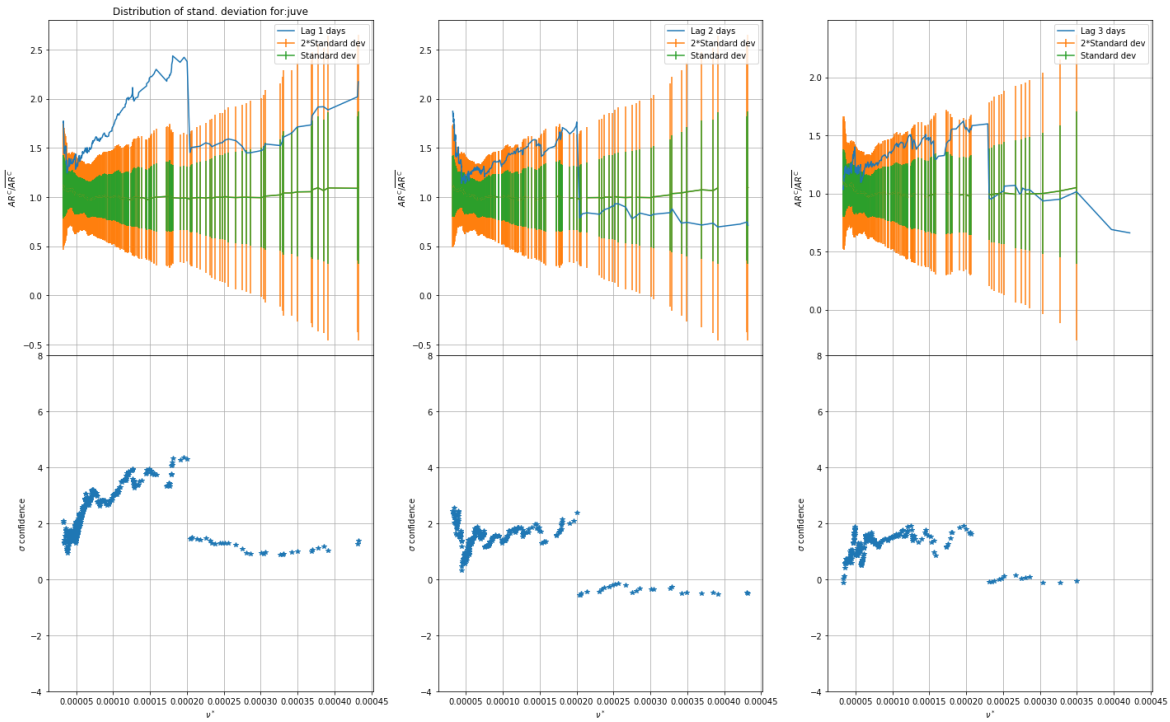

Figure 18 Abnormal return upon relative tweet intensity 1-3 days and the standard deviation of that point from the mean (Juventus FC).

Figure 18 above shows that one day after the relative number of tweets (V *) increases to 0.0002, the deviation from the expected stock price also increases. This means that the market will react the next day to the increased relative number of tweets up to 0.0002. Above this value, there is no longer a measurable relationship between the measured values. After 1 day, the peak from the mean 4 standard deviations can be measured from the mean for the 0.0002 values of relative tweet numbers. On the second day, the value measured is close to the two standard deviations. Figure 18 on the right shows that after 3 days the effects of the event on the increased tweets number are not any more detectable on the stock market.

Manchester United

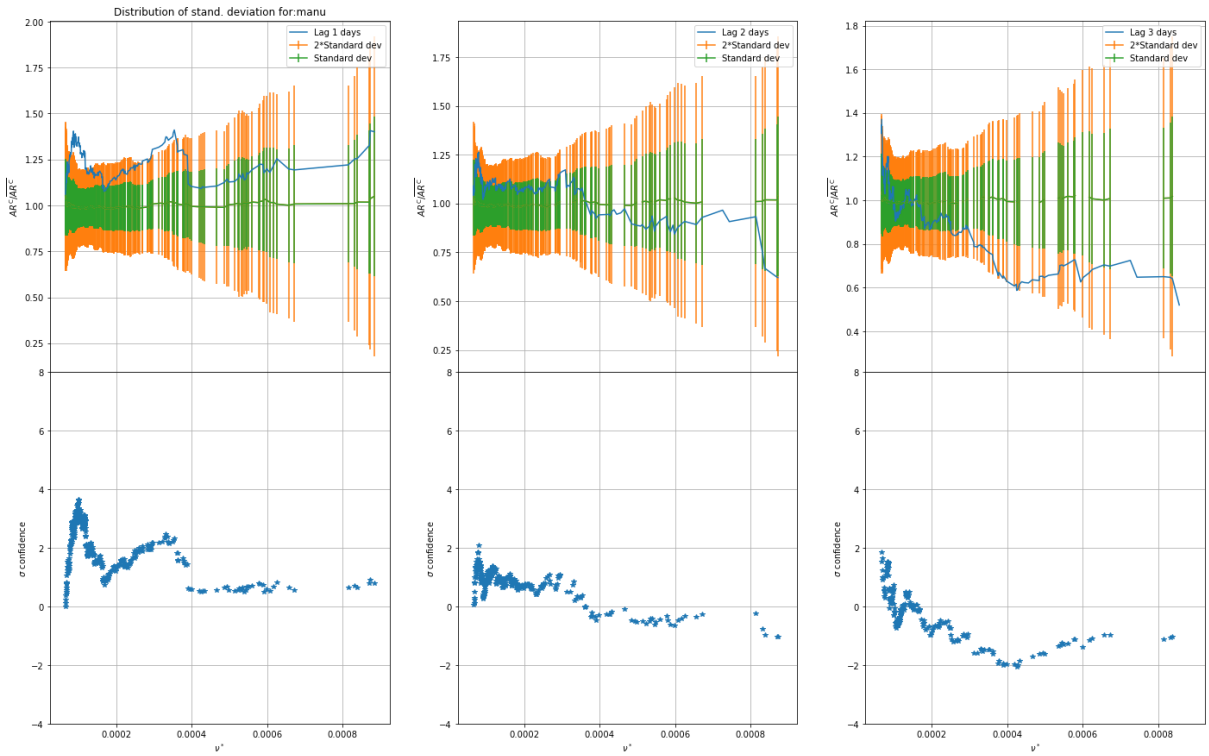

Figure 19 Abnormal return upon relative tweet intensity and the standard deviation of that point from the mean (Manchester United).

The Manchester United data shown in Figure 19 is slightly different from that of Juventus. Figure 19 above shows that one day after the relative number of tweets (V *) increases to 0.0001, the deviation from the expected stock price also increases. This means that the market will react the next day to the increased relative number of tweets up to 0.0001. After this value, it decreases for a short period but then reaches a significant level again. When a relative number of tweets reach the value 0.00035, the peak is reached. Above this value, there is no longer a measurable relationship between the measured values. After 1 day, it can be measured the peak from the mean 3,8 standard deviations from the mean for the 0.0001 value of relative tweet numbers and the peak from the mean 2,2 standard deviations from the mean for the 0.00035 value. Figure 5 on the right shows that the event that led to an increase in the number of tweets will have a minor impact on the stock market over the next two days.

Ajax Amsterdam

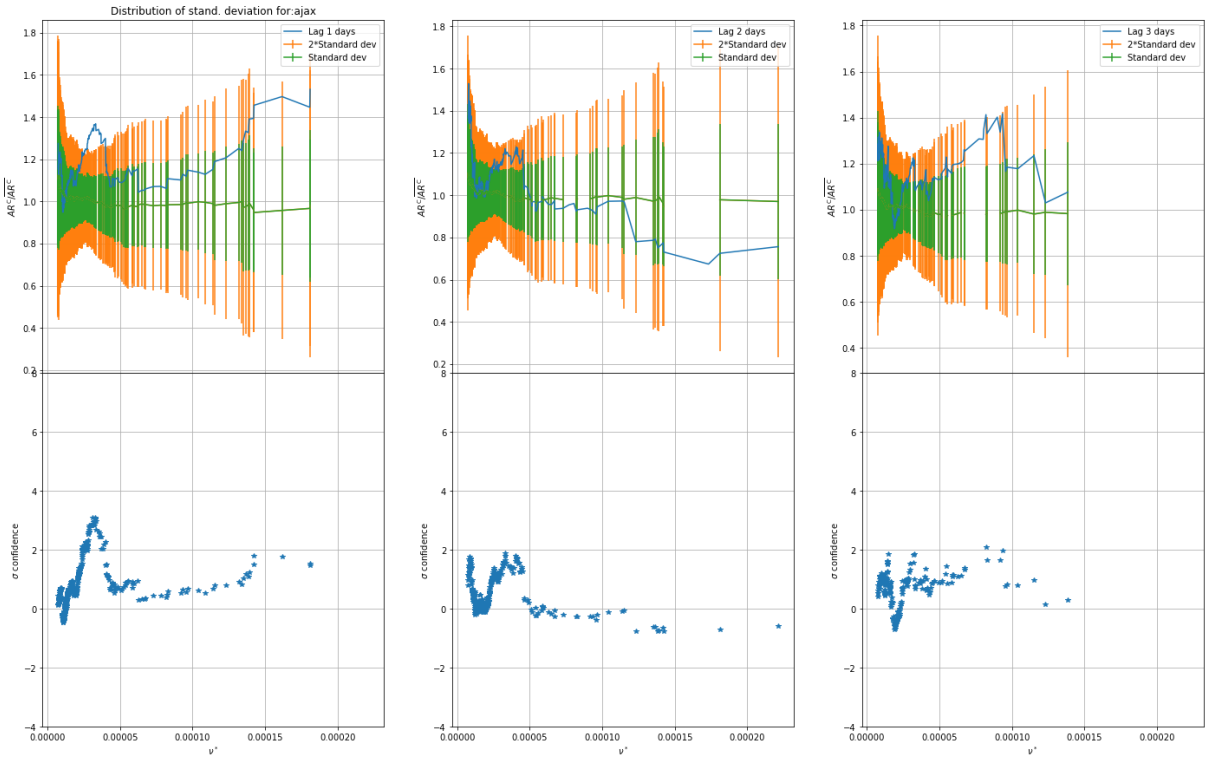

Figure 20 Abnormal return upon relative tweet intensity and the standard deviation of that point from the mean (Ajax Amsterdam).

The Ajax Amsterdam data shown in Figure 20 are similar to Juventus Figure 20 shows that one day after the relative number of tweets (V *) increases to 0.00003, the deviation from the expected stock price also increases. This means that the market will react the next day to the increased relative number of tweets up to 0.00003. Above this value, there is no longer a measurable relationship between the measured values. After 1 day, can be measured the peak from the mean three standard deviations from the mean for the 0.00003 value of relative tweet numbers. On the second day, the value measured is close to the two standard deviations. Figure 5 on the right shows that after 3 days the effects of the event, which increased the tweets number, are not any more detectable on the stock market.

5.8 Discussion and conclusions

This article gives a new perspective to the analysis of sports events, taking into account individual investors in football clubs' stocks. The behaviour of individual investors is influenced by public information, and their mood is affected by different events related to their preferred football club. As a result, individual investors develop a variety of opinions about positive and negative sports events involving the club (Roshanaei & Mishra, 2015). In order to find out how investors respond to different events, tweets exchanged immediately after a sporting event was examined regarding three major football clubs and correlated with the changes in their stock prices. In the life of a football club, the result of the match is the most important event, and the study hypothesises that this result can also be measured through the number of tweets associated with the match - regardless of the fact which team wins or loses and how much the outcome was unexpected. For example: in July 2014 Manchester United sumptuously won in a friendly match against LA Galaxy (7:0). This happened in the U.S.A. where Manchester United has a lot of fans and listed in the New York exchange market. The victory was overwhelming, and the victory of this magnitude had instantly swept over the stock market and the tweets. It is worth noting here that the relationship between the number of tweets and stock market fluctuation is impacted both by direct march involving the club and by the result of other matches related to the club, for instance when players of the club are involved in national team games. For example: on October 15th, 2013 after the draw between the national football teams of Italy and Armenia (2:2), Italy qualified for the FIFA World Cup as first from their group with 7 Juventus players listed in the team. In that case, both Juventus stock prices and the number of related tweets increased. The significance of the mood of the tweet in the given days was compared with the value of the price change in the following three days. When our results were related to the previous research, our observations show that matches and big events indeed have a direct impact on the stock market (Scholtens & Peenstra, 2010). We also found in line with (Zuber, Yiu, Lamb, & Gandar, 2005) that investor–fan trades are based on temporary emotions, and for fans, stock ownership is a value by itself. It can be stated that football investors include a significant portion of football fans.

The mood quantification method developed in this study has been applied to three top football clubs by measuring how many times fans and in what mood refer to their favourite team in tweets in a certain period. Findings confirm that these two indicators move together as a result of the public mood. The novelty of our method is that this method can quantify the mood with tweet intensity. The results also proved that this link is present only up to a certain level above which the increase in the relative number of tweets no longer affects the stock market. This is an important sign for managers to take some action to exploit the potential of on-line interactions but also warns them not to overuse tweets as they no longer affect the stock market and may have even negative effects. This means that managers should carefully consider where the fan groups on Twitter are located and use different languages and motivations to impact on the mood of fans in different countries. Therefore, a manager can consider how to affect the news through well-placed tweet messages and influencing the number of tweets, increase rumours, and also shape the stock market, maybe counterbalancing the effects of a negative event.

Findings are also interesting as they outline that certain events, such as changing coaches (e.g., when United's manager, Sir Alex Ferguson disclosed that he wants to retire, or M. Allegri was named in an announcement as the new head coach of Juventus) do affect the number of tweets and the stock market movements, although to a lesser extent and rather in the long run.

Several limitations of this study must be highlighted. First, the results are limited by the relatively small number of active Twitter users in the population. The demographics of active users are skewed towards the young and male population. This means that the study of tweets does not represent the mood of the entire public. Also, in this paper, could be tracked only the stock movements of teams whose data are public and listed on the stock exchanges and at the same time also have many supporters on Twitter. The semantically analysed was done only on the tweets written in English. In the future, the relation between twits and stock market prices should be observed over a longer period. Further, the identification of the country from which the tweets are originated could be of interest to understand differences in impacts on the stock market fluctuations by fans in a different location.

6. Paper 3: Impact of Twitter activity on stock prices of innovative companies

Abstract: For innovative companies, demonstrating the relationship between the use of social media and stock price volatility is an important economic request. The results of the research provide measurable support for the assumption that the performances of the most rapidly developing companies of our time are related to the number of tweets that mention them. The novelty of the research is to show where are the tweet-number limits that still affect the stock market. Using the results of research, the conscious use of social media can help to fine-tune the estimation of stock market movements.

Keywords: Innovative companies. Twitter, share price.

6.1 Introduction and literature review

The world is often considered "linear" and, therefore, it is assumed that it is conducted by very clear "causal relationships" (Alagidede, Panagiotidis, & Zhang, 2011). Majority of studies of economic science are based idea of equilibrium systems: for example, the symmetry between supply and requirement (Wu, Li, Gou, & Gu, 2017), risk and benefit,(D. Levin & Smith, 1994) price and quantity (Kelly, 2005). This view originates from the idea that economics is a scientific discipline like Newtonian physics, which implies consequence and explicit predictability. When a system in equilibrium is hit by an exogenous shock, it absorbs the shock and yields to equilibrium. (Salati & Vose, 1984). For most of the complex systems, (Soti, 2019), the causal link is very weak and does not end with "deliberate consequences of correlations". For example, predictions such as: "If Trump becomes president, markets will collapse" or, "loose monetary insurance policy does not lead to hyperinflation as recommended by some economists" and much more.

Why were these forecasts not fulfilled? Because stakeholders in the stock markets use their version of market status, speculation, and information available to them to shuffle their trading decisions (Vivek et al., 2012). This results in diverse, complicated interactions between the stakeholders, and - eventually -the stock markets form a complex system, where the stock price

of various firms shows emergent behaviour that is difficult to predict. (Fama, 2002; Schredelseker & Fidahic, 2016). Normally, the analysis of the financial markets is a discipline using time -serial publication data and linear extrapolation. (Rai, Kasturi, & Huang, 2018). The analysis of individual time-series data becomes very complicated (as several stocks increase value in the market). Several studies have investigated this complex fluctuation of the stock market, but no one can provide a completely reliable method of how the stock market will move in the future (Alagidede et al., 2011; Fama, 2002; Vicentini & Graziano, 2016). Research has come to the conclusion that daily fluctuation of stock prices is also influenced by the emotional state of the trading individuals (Bollen et al., 2011; Vivek et al., 2012). Their emotional state can be affected not only by economic trends and major political events, but other news connected to big innovative companies (Ashton et al., 2003).

The benefits of forecasting reliable stock markets have attracted much attention not only in the field of science but also in business. Such forecasts are based on the interpretation of market conditions and speculation (Schredelseker & Fidahic, 2016). Early research on stock market forecasts was based on an efficient market hypothesis. They argued that stock prices are largely shaped by new information, not by current and past values. (Dockery, Vergari, & Vergari, 2001; Schumaker & Chen, 2006).

Recent literature, based on behavioural economics, tells us that emotions can profoundly affect individual behaviour and decision-making (Weimar & Schauberger, 2018). It is confirmed that public mood is correlated with economic indicators (Bollen et al., 2011). Psychological research literature has shown that individual emotions can play an important role in investor decisions. (Kahneman & Tversky, 1979)

We hypothesize that investors attach importance to the image of innovative companies in social media. The results of this research show that the prediction accuracy of traditional stock exchange forecasting models can be greatly improved by incorporating certain mood dimensions. This study examined whether public sentiment - in this case, relative tweet intensity - expressed in large collections of daily Twitter posts - could indeed be used to refine stock market forecasts. In this research, we first created a linear model to estimate the future movement of a given stock. It shows that if an event occurs, that increases the number of tweets, it could affect changes in stock price and volume from this forecast analysis model onwards.

This analysis can be important from a scientific and managerial point of view. On the one hand, it can provide a methodology for predicting the volatility of trading stocks following major

events in the life of innovative companies. On the other hand, from a managerial perspective, it can help to develop a better strategy that can be used to influence conversations between customers and the company. (Hinz et al., 2011). Social media can serve as a promotional tool in companies' integrated marketing plans (Strahinja, Golob, & Subašić, 2017)to increase users' brand loyalty (McClung, Eveland, Sweeney, & James, 2012).

The article is divided into 5 sections. The second section outlines the data collection procedures. In the third section, we present the application of the methodology in three case studies. The fourth section summarizes the importance of the result and the possible future research directions are listed in the fifth section.

6.2 Datasets and preparation

This study has two main sources of data. It relies on Twitter data and stock market data. The steps for collecting and preparing the data are detailed in the next section.

6.2.1 Twitter data

Twitter data was originally collected via the Twitter API and maintained at the Virtual Observatory of Eötvös Loránd University (Dobos et al., 2013) The Twitter API does not have access to the entire Twitter feed. It is estimated that we managed to collect about 2-3% of the tweets that contain the names of any of these three companies during the observed period. We downloaded tweet records created from January 1, 2012, to December 28, 2014, from the Virtual Observatory database. The information collected includes the unique identifier of the Twitter user, the timestamp of the tweet, and its textual content. Table 5 summarizes the filter words, with the companies selected in the first column.

Table 9 Summary of Twitter data- innovative companies.

Football club	Word	Total number of followers	Tweet numbers in our database
Sony	Sony	4.4M	0,4M
Amazon	Amazon	3.1M	0,4M
Google	Google	21.7M	1M

Source: own elaboration

We then analyzed the time series of the tweet intensity. The tweet intensity ν_j of innovative company jon a given day is defined as the ratio of the daily tweet count of that company $\boldsymbol{T_j}$ and the overall tweet count of the same day in our database $\boldsymbol{T_m}$:

$$\nu_j = \frac{T_j}{T_m} \tag{11}$$

6.2.2 Stock data

The second source of data for this study was stock market data collected from https://finance.yahoo.com. These records contain daily aggregated data. The data includes the closing prices of shares and volume.

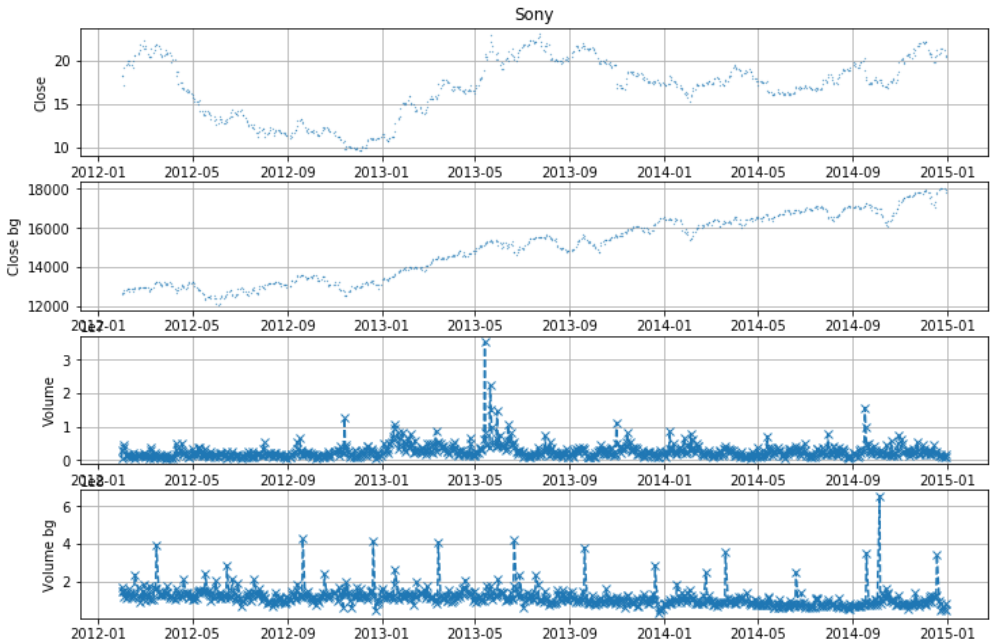

Figure 21 The closing price and volume evolution of the Sony stocks

Source: Self-prepared chart from stock data

It seems from Figure 21 that neither the closing price of the stock nor the volume of the stock does not follow the Dow Jones index (background process).

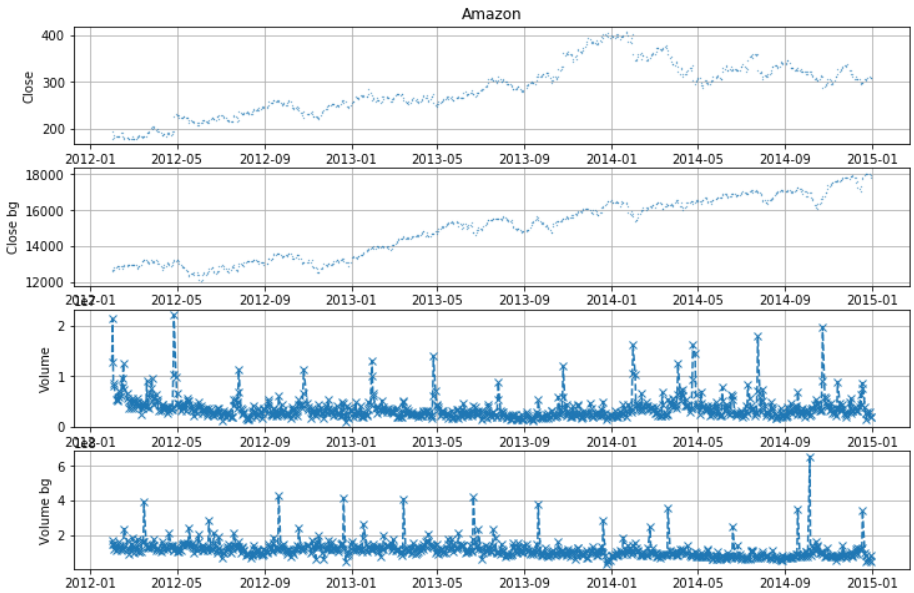

Figure 22 The closing price and volume evolution of Amazon stocks

Source: Self-prepared chart from stock data

In contrast to this observation, the series of closing prices and volume of the amazon stocks in Figure 22 shows more simultaneity with a background process.

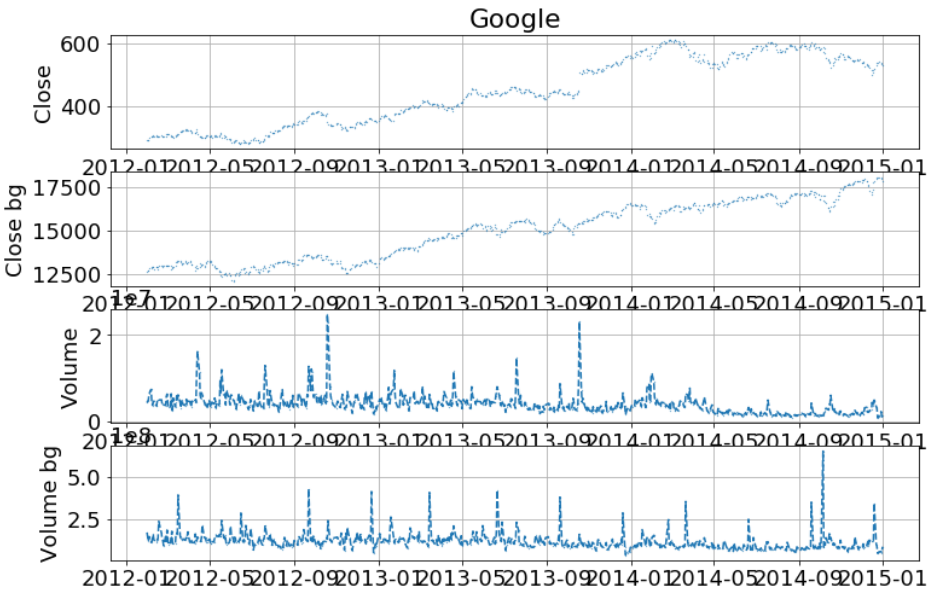

Figure 23 The closing price and volume evolution of Google stocks

Source: Self-prepared chart from stock data

Finally, the third observed company, the google stock closing prices and volume changing shows also some simultaneity with a Dow Jones closing prices and volume. (Figure 23)

6.2.3 Methods

From the series of closing prices for companies on a given day, the standard rate of return on a share (2) is calculated as follows:

$$R_{j,t} = abs(\, C_{j,t} - C_{j,t-1^{*})}) \qquad (12)$$

where $t - 1^*$ indicates the previous trading day, i.e. when t stands for a Monday $t - 1^*$ points to the Friday the week before.

The Dow Jones USA market index was used to model the stock's background. The normal marginal rate of return $R_{j,t}$ shall be calculated by the formula (12).

$$R_{j,t} \approx \sum_{i=1}^{4} (a_j R_{i(j,t)} + b_j) \qquad (13)$$

where parameters a_j and b_j for companies are estimated by ordinary least squares method.

This relatively simple model defines the abnormal return of the stock.

$$AR_{J,T} = R_{j,t} - (A_J * R_{M,T} + B_J) \qquad (14)$$

The other data obtained from stock exchange data is the volume of Vj shares of the three companies, which were divided over the same period by the volume of the Dow Jones Index. The following section sets out applicable case studies for Sony, Amazon and Google to analyze the impact of the tweets on the exchange rate and volume. In order to keep track of trends, the two data sources are connected on a time basis on a daily basis. Note that stock data is not available on weekends as markets are closed while twitter data is still available on weekends. In order to check whether tweeting affects trading decisions, a τ delay in days where $\tau \in \{0,1,2,3\}$ is allowed when the two data sets are merged. Delay $\tau = 0$ represents twitter data for day t and stock data for $t + \tau$ for the next day move together. The pooled samples are classified as follows. A threshold $\nu *$ is introduced, and samples of the club j are said to belong to the positive group if $\boldsymbol{\nu_{j,t}}$ exceeds the threshold. If we exclude the distinction between profitable or

loss-making movements of stocks, the average of the absolute abnormal yields shall be calculated as follows:

1

$$AR_j(v^*, \tau) = < \{ \left|AR_{j,t^*+\tau}\right| \left| t^*: v_{j,t^*} > v^* \right\} \qquad (15)$$

Dropping explicit notation of parameters j and τ. Let AR^c denote the value of group $AR_j(\mathbf{v}^*, \boldsymbol{\tau})$ and $\overline{AR^c}$ is the average of a complementary group.

In Figures connected to abnormal return the statistics of $\frac{AR^c}{\overline{AR^c}}$ are presented for all three companies.

The relative daily volumes of the stock data were also investigated:

$$RhO = \frac{V_j}{V_m} \qquad (16)$$

where v- is the volume of given stocks and v_m is the volume of background index. Similarly, to AR samples, a positive sample group is defined RhO_j:

$$RhO_j(v^*, \tau) = < \{ |RhO(j, t\hat{}* + \tau)| \left| t^*: v_{j,t^*} > v^* \right\} \qquad (17)$$

Let $\mathbf{RhO_j}$ denote the average of the samples in the group (14). In Figures connected to volume, the statistics of $\frac{RhO}{\overline{\mathrm{RhO}}}$ are presented for all three companies.

In this research, distance correlation is used to measure the dependency between two vectors, not necessarily of the same dimension. The correlation coefficient of the distance between the two data sets is zero if, and only if the vectors are independent. The distance correlation measures both the linear and non-linear association between the two vectors. Distance correlation can be used to perform a statistical dependency test with a permutation test. In this analysis one first computes the distance correlation (including the ring of the most recent Euclidean distance matrices) between the two vectors and then compares this value to the distance correlation of many shufflers in the data. (Székely & Rizzo, 2009, 2013)

Figures connected to distance correlation presented distance correlation between a close price change and tweet number, volume change and tweet number, and close price change and volume change for all three companies.

6.3 Results - case studies

6.3.1 Sony

In all three cases, we always took four days into account: the day on which we counted the number of tweets and the following three days. The quotient of AR^c and $\overline{AR^c}$.

Distance correlation and relative volume quotient are defined in the previous section.

Abnormal return - Sony

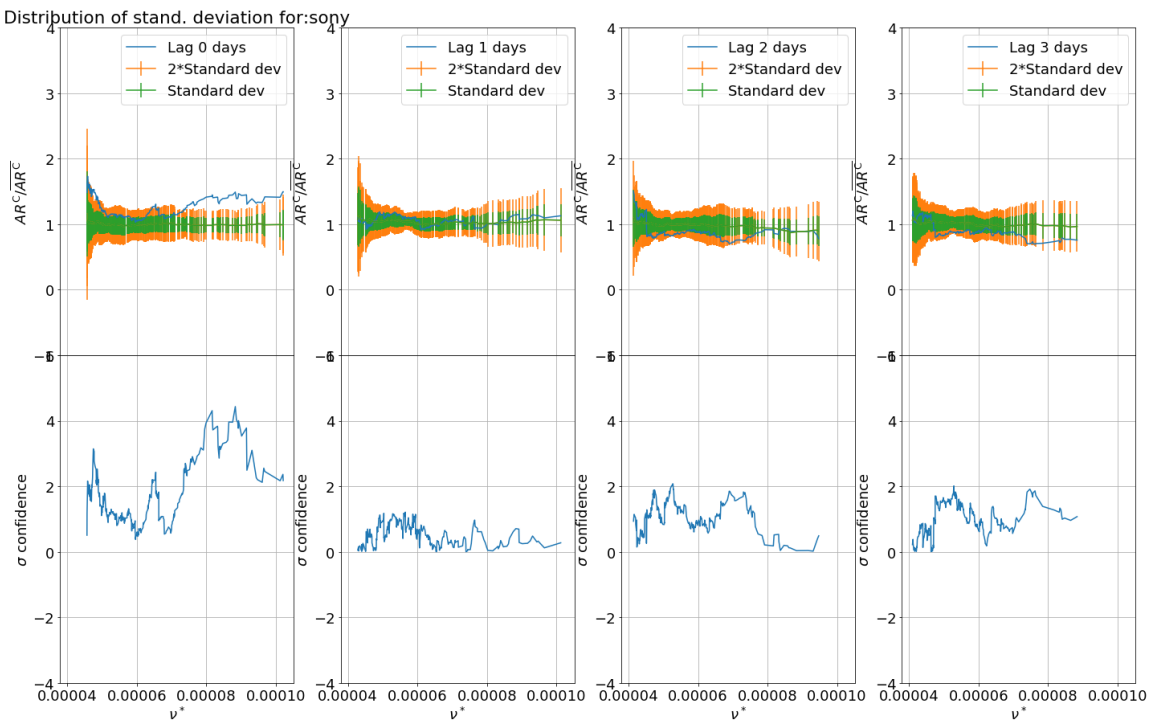

Figure 24 Abnormal return upon relative tweet intensity 0-3 days and the standard deviation of that point from the mean - (Sony)

Source: Self-prepared chart

The following values are shown on this Figures (24,26,27,29,30,32) type: we measure the impact of an event by first calculating the stock market's average response, (In one case, Abnormal return in the other case Volume) to a set of relative tweet numbers. (We generate

2000 sets of random relative tweet numbers and calculate the average effect on the stock market). This average distribution and one and two standard deviations distance from this are shown in the upper part of this type of Figures (24,26,27,29,30,32) (green and orange stripes). The blue line represents the calculated effect of the relative number of tweets on the given day. Figure 24 bottom part shows the difference between the calculated effect and the mean effect, measured in standard deviation (yield as σ confidence).

At the beginning of Figure 24, bottom part shows that a little more than the daily average number of tweets has the biggest impact on the same day on the stock market. The highest value is reached when the relative tweet value is 0.00008, and then its value is 4 standard deviation from the mean.

Figure 25 middle and last part show that after 1, 2 or 3 days, the effect of the event, which increased the tweet numbers is not any more detectable on the stock market.

Distance correlation - Sony

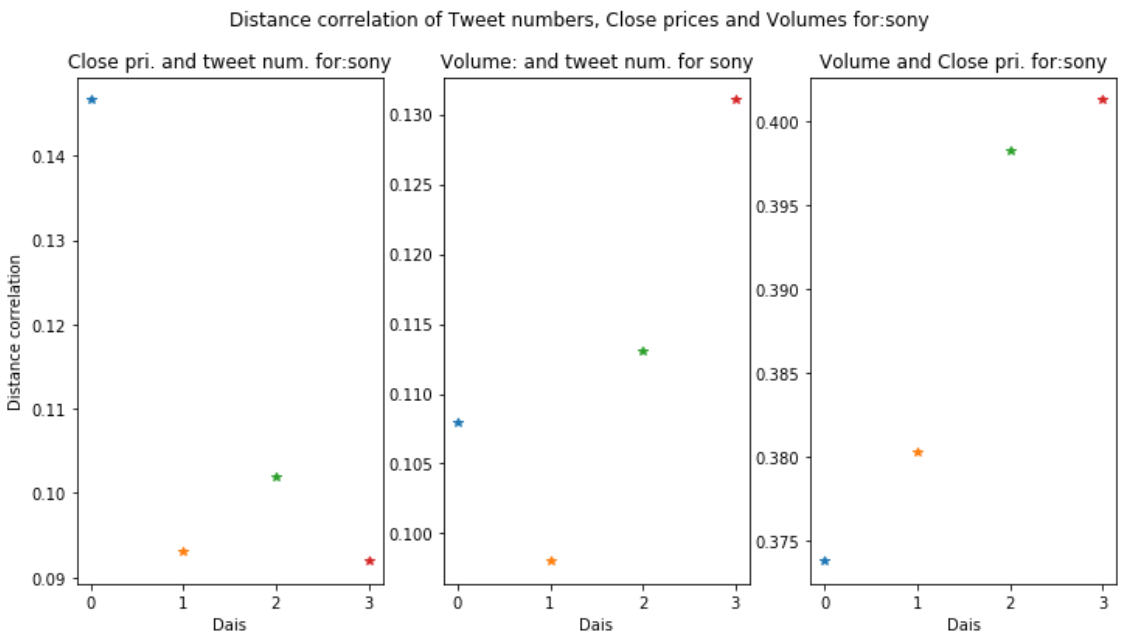

Figure 25 Distance correlation on 0-3 days between tweet number, close price change and volume for Sony

Source: Self-prepared chart from stock data

This graph type shows the distance correlation between the calculated values. First, these values are close price change and relative tweet number, then volume change and relative tween number and last volume change and close price change. The X-axis shows the days of shifting.

In the first part of Figure 25 shows that on the same day when we calculated the relative number of tweets, this value and close price change has 0,15 distance correlation (dc). For the next three days, this value is negligible. (Following the values shown in Figure 24). The middle of Figure 25 shows that between relative tweet numbers and volume on the third day, the connection is the strongest (0.13 dc). The last part of Figure 25 shows that between a close price and volume change, the strongest connection also falls on the third day (0.4 dc).

The volume of sold stocks- Sony

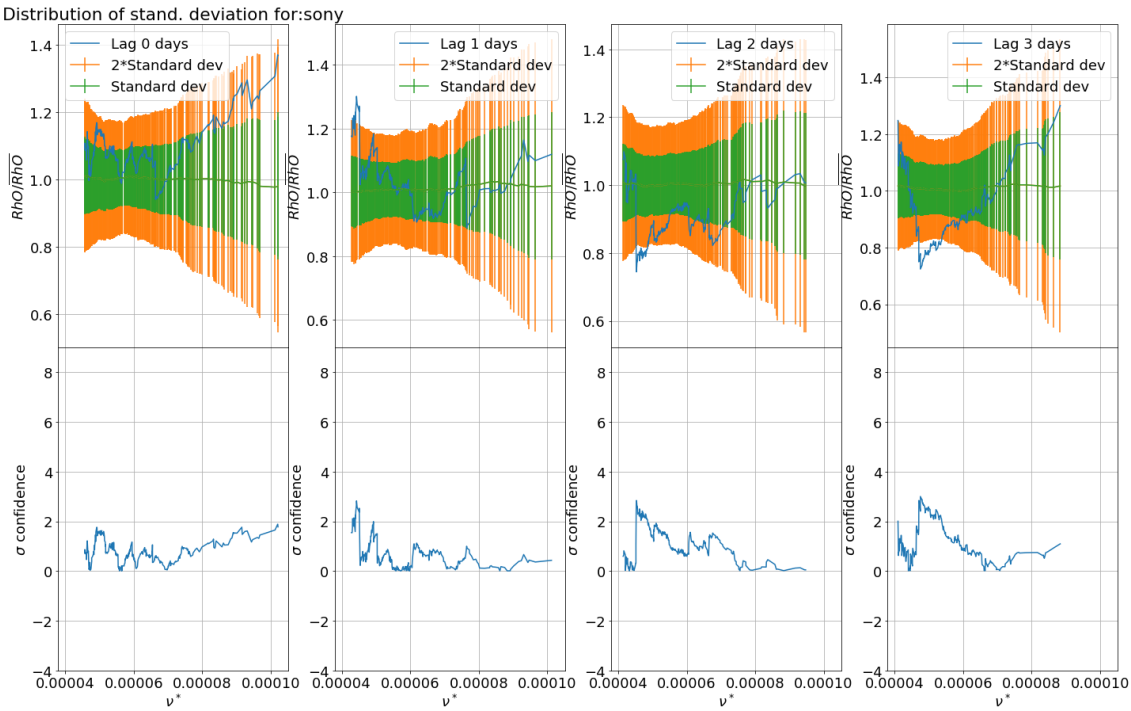

Figure 26 Stock volume upon tweet intensity and the standard deviation of that point from the mean – (Sony)

Source: Self-prepared chart from stock data

The first part of Figure 26 also shows that for all numbers of tweets, the deviation is less than two standard deviations from the average. This is in line with the middle Figure 25 (0 day - is 0.1 dc).

Last part of Figure 26 shows that the effects of the event that increased the tweet numbers will be increased after 3 days. This is also in line with the middle Figure 25 (3 day - 0.134 dc).

Summary for case Sony: The relationship between the numbers of tweets and abnormal return is measurable on day 0, and on day 3. The relationship between the Volume is measurable with both: with the number of tweets and with an abnormal return of stock price.

6.3.2 Google

In the case of Google, we also took four days into account, and the same metrics were examined as in the case of Sony.

Abnormal return - Google

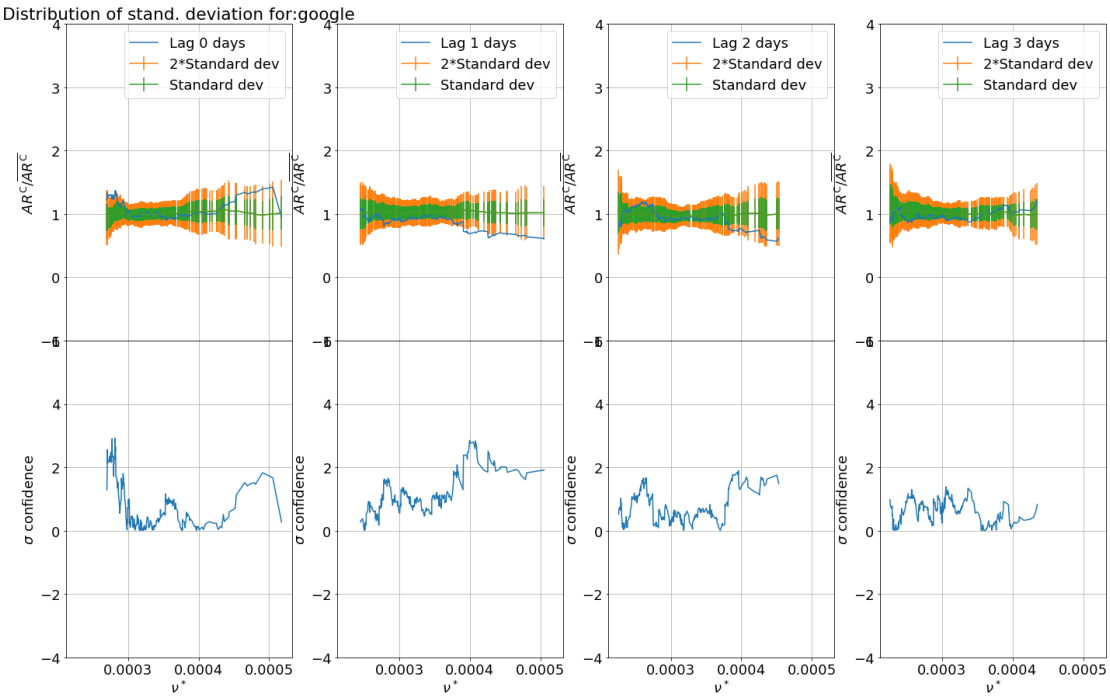

Figure 27 Abnormal return upon relative tweet intensity and the standard deviation of that point from the mean - (Google)

Source: Self-prepared chart from stock data

The Google case shown in Figure 27 is somewhat different from that of Sony. The reaction of stocks takes a longer time to respond to tweet intensity. The stock market reacts within 1 day to an increased number of tweets. The figure below shows that the highest value is reached when the relative tweet value is relatively high, and its value is 2.1 standard deviation from the mean.

Middle and las part of Figure 27 shows that after 2 or 3 days the effect of the event that increased the tweet numbers are not any more detectable on the stock market concerning Google.

Distance correlation - Google

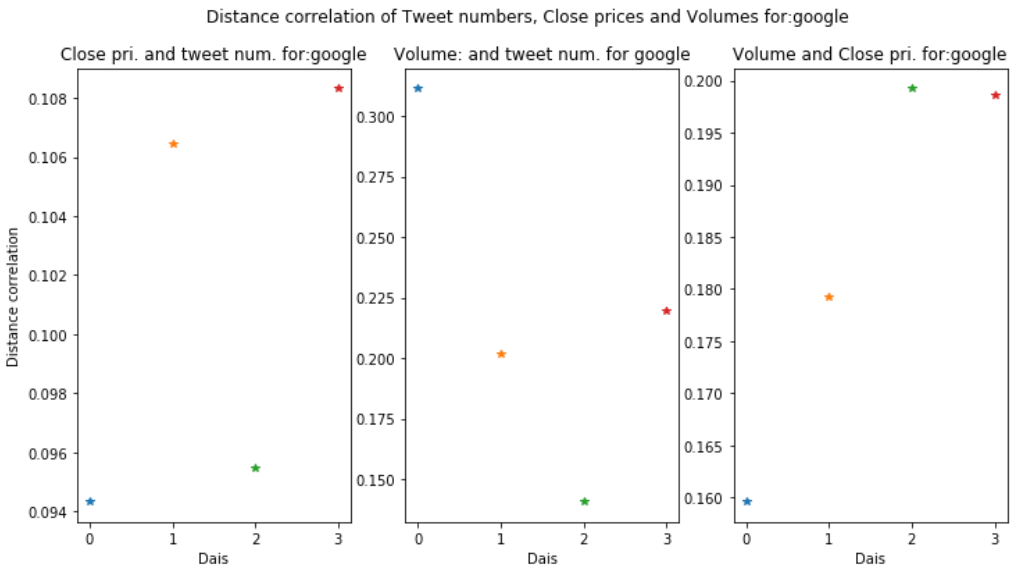

Figure 28 Distance correlation on 0-3 days between tweet number, close price change and volume for Sony

Source: Self-prepared chart from stock data

Figure 28 shows that on all days when we calculate the distance correlation (dc) between the relative number of tweets and close price change is small (0.1 dc). The middle of Figure 28 shows that the connection between a relative tweet number and volume on the 0 days is the strongest (0.34 dc) and the last part of Figure 28 shows, that between a close price and volume change connection is the strongest also on the second and third day. (0.2 dc)

The volume of sold stocks- Google

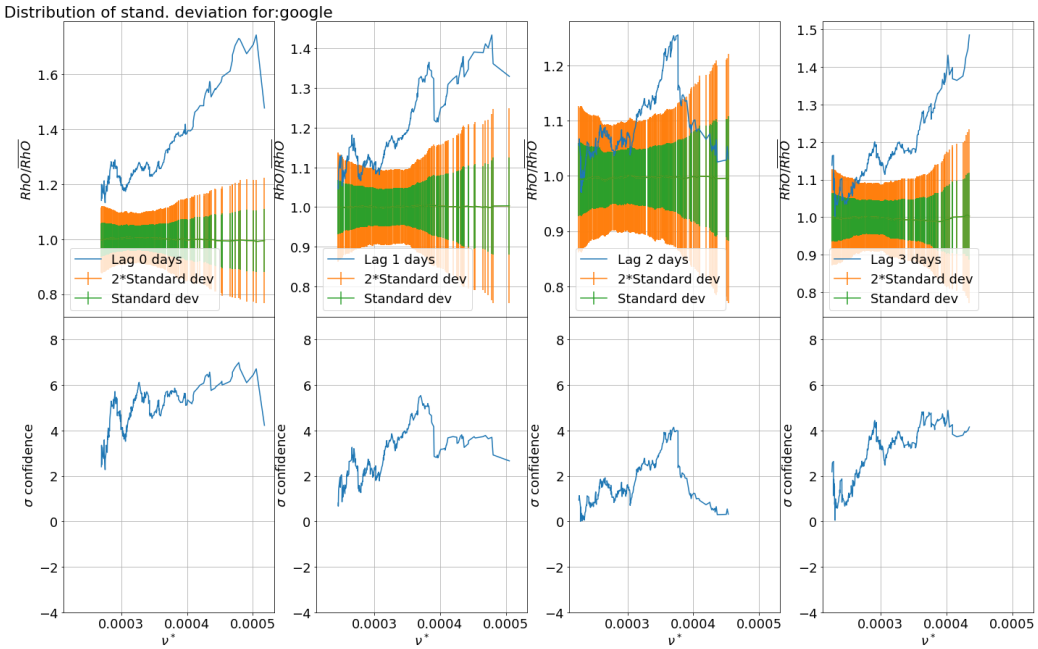

Figure 29 Stock volume upon tweet intensity and the standard deviation of that point from the mean – (Google)

Source: Self-prepared chart from stock data

Figure 29 shows a very strong connection between volume and the number of tweets. On the first day, this effect is high. Its value is more than 6 standard deviation from the mean. Less effect can be observed in the next days, and the effect continues to decline. These observations are in line with the intermediate figure of Figure 28.

Summarized for case Google: The relationship between the numbers of tweets and abnormal return is very poorly measurable; however, the relationship between volume and the number of tweets is impressive. The relation between a close price and volume change is solid.

6.3.3 Amazon

In the case of Amazon, we also took four days into account, and the same metrics were examined as in the case of Sony.

Abnormal return - Amazon

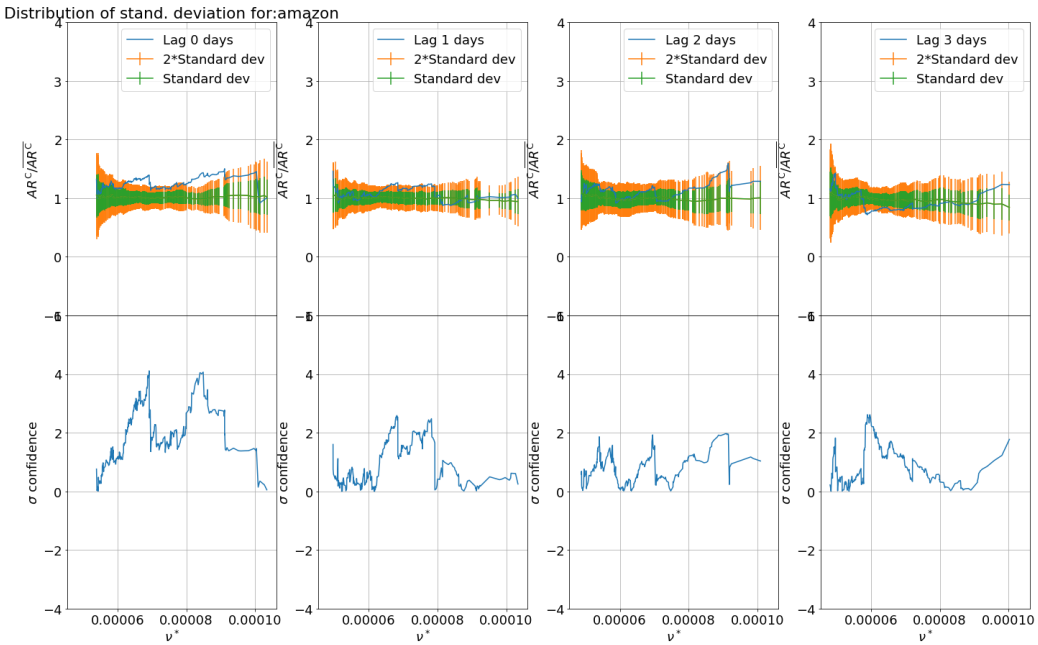

Figure 30 Abnormal return upon relative tweet intensity and the standard deviation of that point from the mean - Amazon

Source: Self-prepared chart from stock data

The Amazon case shown in Figure 30 is a little different from that of Sony. The first part of the figure also shows that after a certain number of tweets, the deviation is more than 4 standard deviations from the average. The figure below shows that the highest value is reached when the relative tweet value is 0.00009, and then its value is 4 standard deviation from the mean.

The Figure 30 middle and las part shows that after 1, 2 or 3 days the effects of the event which increased the tweet numbers are detectable to a much lesser extent on the stock market connected to Google.

Distance correlation - Amazon

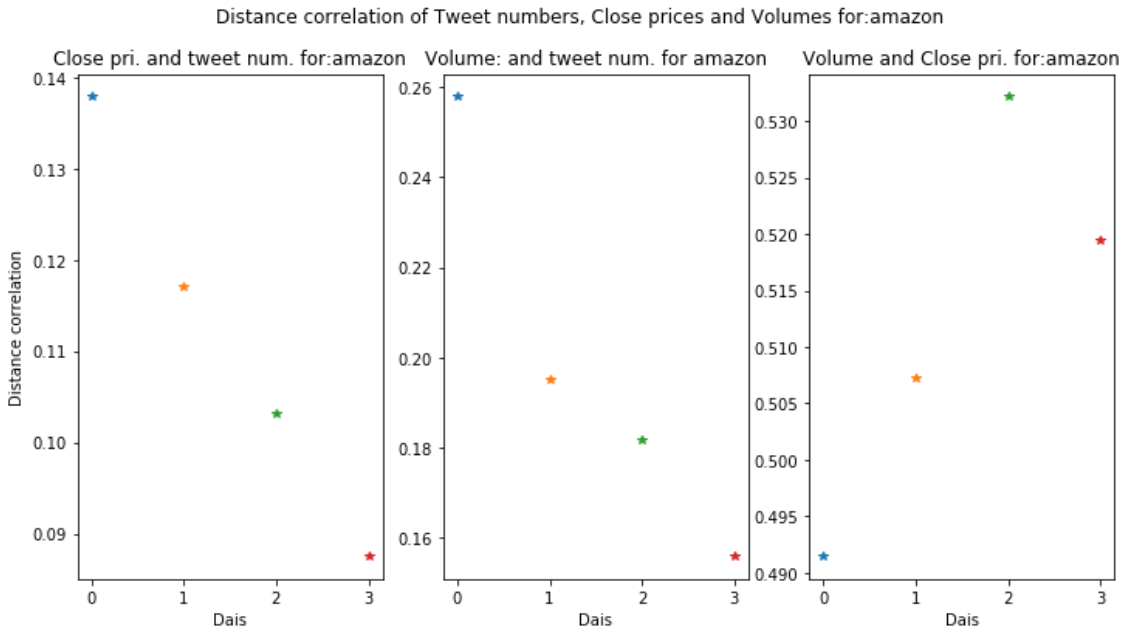

Figure 31 Distance correlation on 0-3 days between tweet number, close price change and volume for Amazon

In the first part of Figure 31 shows that on the same day when we calculated the relative number of tweets, then this value and close price change has 0,14 distance correlation. For the next three days, this value is negligible. (Following the values shown in Figure 30). The middle of Figure 31 shows that between a relative tweet number and volume on the same day when we calculate the relative number, the connection is the strongest (0.26 dc). The last part of Figure 31 shows that the connection between a close price and the volume change is the strongest also on the second day (0.53 dc).

The volume of sold stocks – Amazon

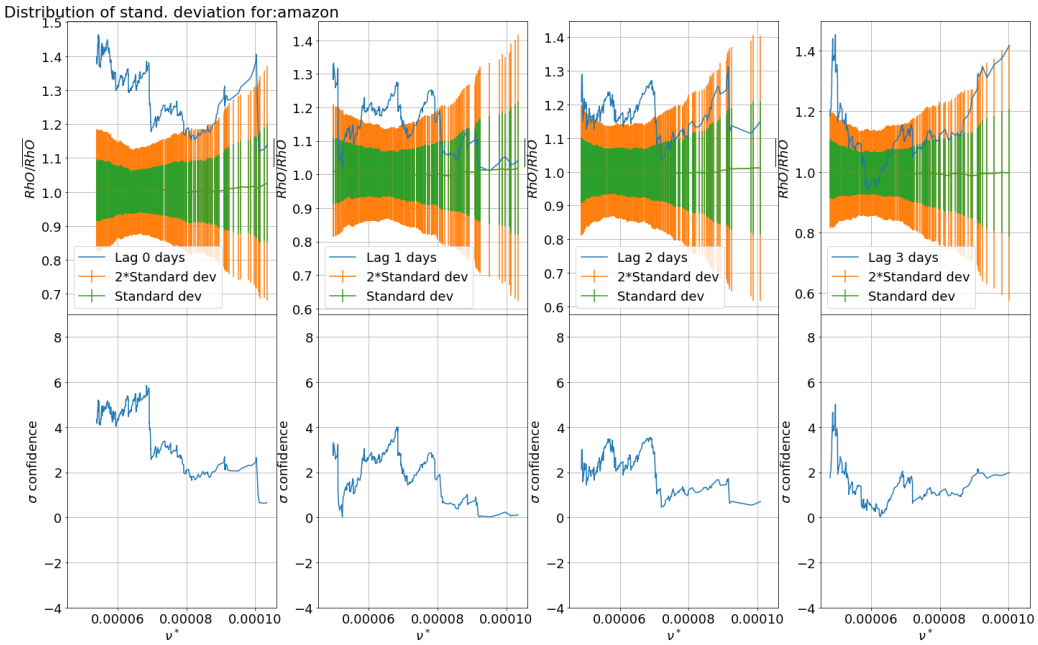

Figure 32 Stock volume upon tweet intensity and the standard deviation of that point from the mean – Amazon

Figure 32 shows a very strong connection between volume and the number of tweets. On the first day, this effect is high. Its value is 6 standard deviations from the mean. During the next days, less effect can be observed, and the effect continues to decline. These observations are in line with the intermediate figure of Figure 31.

Summary for case Amazon: The relationship between the numbers of tweets and abnormal returns is measurable, whereas the relationship between volume and the number of tweets is very impressive. The relation between a close price and volume change is high.

6.4 Conclusion

This study provides a new perspective on the analysis of events, taking into account a specific category of stakeholders, namely individual investors. The behaviour of individual investors is influenced by public information and sentiment. Some investors may have completely different opinions about events. To understand how investors are responding to different events, we have examined several tweets about three large companies. The significance of the events was related to the relative number of tweets on a given day. The

change in the absolute value of the price or volume in these cases changes after one, two or three days. We have developed a mood quantification method for large companies - the number of occasions customers and users refer to a given company in tweets - and realized that there are relative numbers of tweets when absolute stock price movements are relatively high. In some cases, these two indicators move together. We have also proven that once a relative number of tweets reaches the threshold, the increase in relative tweets no longer affects the stock market. This ensures that if tweets are created artificially, they will not affect the stock market after a while. The novelty of our method is that we can quantify the mood with a Tweet intensity and assign a number to it. This is an important signal to managers that too many tweets are no longer affecting the stock market.

6.5 Limitations and future work

This study is limited by the particular proportion of active Twitter users in the population, which, for example, is about 15% of the US population. Active user demographics point to young people and male users. This means that studying the tweet does not represent the full public mood (Zhang et al., 2013). In addition, in this article, we could only track the movement of companies whose data is public and listed, while having many users on Twitter.

7. Paper 4: Urban scaling of football followership on Twitter

Abstract

Social sciences have an important challenge today to take advantage of new research opportunities provided by large amounts of data generated by online social networks. Because of its marketing value, sports clubs are also motivated in creating and maintaining a stable audience in social media. In this paper, we analyse followers of prominent football clubs on Twitter by obtaining their home locations. We then measure how city size is connected to the number of followers using the theory of urban scaling. The results show that the scaling exponents of club followers depend on the income of a country. These findings could be used to understand the structure and potential growth areas of global football audiences.

Keywords: urban scaling, Twitter, social media, football1

7.1 Introduction

Today the online social network Twitter has more than 300 million monthly active users, with many of them actively following sports events, stars or clubs to exploit the possibilities of obtaining the latest news through instantaneous messaging (van der Lans et al., 2009). Large football clubs and football leagues invest money in establishing official social media channels to engage with their fan basis (Price et al., 2013) and seek to purchase players who bring them a massive number of Twitter followers. Social media presence is especially important for clubs that rely more heavily on broadcasting and commercial revenues than on matchday revenues, such as the global top 20 clubs from a recent analysis of the Deloitte Football Money League (Boor, Hanson, & Ross, 2020). Because global fans have limited options to be present at matchday events, popularity on Facebook together with Twitter is a good indicator to judge the global follower success of a football club.

On the other hand, the geographic and socio-economic environment of a user still plays an important role in determining the probability of engaging with a globalized phenomenon. As such, complex spatial structures and the dynamics of changes in them have for some time been a focus of the scientific community as well as marketing experts. Recently, there has been

growing literature on the concept of urban scaling, which connects measurable outputs of cities to their size. (Bettencourt, 2013; Cottineau, Hatna, Arcaute, & Batty, 2017; O. A. dos Santos, 2014; Yakubo, Saijo, & Korošak, 2014) Urban scaling laws have been detected for various quantities with respect to the city size, such as GDP (Bettencourt, Lobo, Strumsky, & West, 2010), urban economic diversification (Strumsky & Lobo, 2015), touristic attractiveness (Bojic, Belyi, Ratti, & Sobolevsky, 2016), crime concentration (Hanley, Khatun, Yosef, & Dyer, 2014; Oliveira, Bastos-Filho, & Menezes, 2017), human interactions (Schläpfer et al., 2014), or election data (Bokányi, Szállási, & Vattay, 2018). Some of these measures follow a super-linear relationship with urban size, which means that the quantities are disproportionally overrepresented in larger cities. These measures include GDP, number of patents or certain business types, where larger cities facilitate more the accumulation of wealth and resources needed for such phenomena. On the other hand, infrastructural - like quantities have sublinear scaling laws reflecting efficiency due to urban agglomeration effects.

In this paper, we investigate urban scaling laws for geolocated Twitter football club followers for three majors widely acknowledged clubs: Real Madrid, Manchester United and Bayern Munich. We calculate the scaling exponents for the number of followers of each club in the urban systems of five different countries. While the scaling exponents of clubs differ significantly within countries as well, the variations in the exponents across countries suggest, that the wealthier a country is, the more sublinear its follower scaling exponent, and vice versa.

7.2 Materials and methods

Twitter freely provides approximately 1.2% of its data for download through its API. For those users that allow this option on their smartphones, the exact GPS coordinates are attached to their messages, the so-called tweets. By focusing the data collection on these geolocated tweets, we could determine the home location for the most active users selected from the database, using the friend-of-friend algorithm clustering on their coordinated messages. This left us with a total of 26.3 million Twitter users that have home coordinates associated with them. We constructed a geographically indexed database of these users, permitting the efficient analysis of regional features (Dobos et al., 2013). Using the Hierarchical Triangular Mesh scheme for practical geographic indexing, we assigned cities to each user. City locations were obtained from **http://geonames.org** , city bounding boxes via the Google Places API. We

downloaded the Twitter user identifiers of the followers of three selected football clubs: Real Madrid, Manchester United and Bayern Munich. Table 1 shows the number of followers (people who follow at least one of the three teams, later referred as overall follower count) that are also in our geolocated user database, which meant roughly 2-3% of all followers in all three cases.

Table 10 Number of total followers for each football club on Twitter and the number of followers from the geolocated user database used in our analysis.

Team name	Total number of followers	Geolocated followers
Real Madrid	28.7M	808,427
Manchester United	17.3M	436,515
Bayern Munich	4.3M	119,056

Source: own elaboration

The theory of urban scaling suggests that there is a power-law relationship between a socio-economic indicator measured in a city and its size (O. A. dos Santos, 2014). We can formulate this power-law relationship with the following equation:

$$Y = Y_0 * N^{\beta} \tag{18}$$

Where Y denotes the investigated quantity, N is the number of inhabitants in a city, Y_0 is a normalization constant, and ß is the so-called exponent that characterizes the behaviour of the quantity in connection to changing city size. In the literature, it has been observed that this ß parameter differs only slightly from 1. Most urban socio-economic indicators have a superlinear ß > 1 exponent, which is caused by larger cities being the centres of wealth, innovation and creative processes. Sublinear scaling ß < 1 characterizes material quantities associated with infrastructure, where the agglomeration into cities is more economic, which manifests in fewer overall road length, or overall cable need etc. (Bettencourt et al., 2010).

If we take the logarithm of both sides, the equation becomes a linear relationship:

$$\log Y = \log Y0 + ß * \log N \tag{19}$$

It is then enough to fit a line onto the log Y -- log N pairs. We used binning of the data, where we took the mean of log N and log Y in each bin, and then fitted a line onto them using an OLS fit with weighting the bins by $\mathbf{1} / \sqrt{\boldsymbol{N}}$. This error calculation assumes that higher follower numbers carry less error when fitting the scaling curves (Bettencourt et al., 2010).

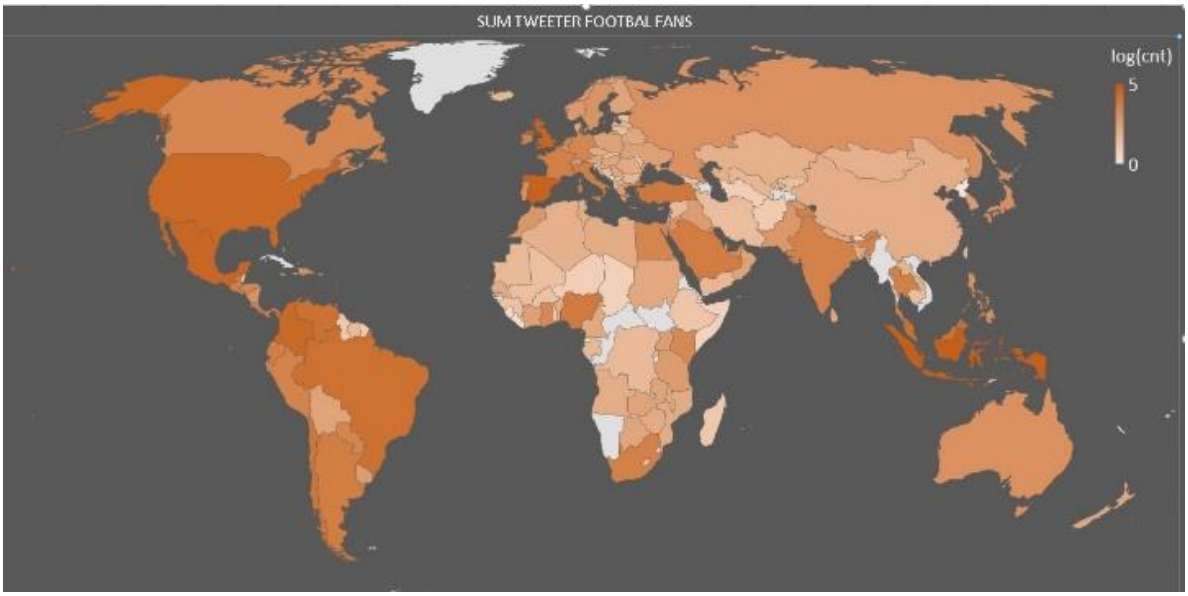

Figure 33 Distribution of geolocated Twitter users that follow at least one of the three selected clubs. Countries are coloured according to the logarithm of the number of users

.

7.3 Results and discussion

The geographical distribution of users that follow at least one of the three clubs can be seen in Figure 33. A major fan base is in Western Europe, North and Latin America as well as in the Pacific Region. Because Spanish and English teams are among the investigated clubs, the number of followers is high in Spain and in Great Britain.

As analysed countries, we chose the home countries of two of the teams, Spain and the UK, and we included traditional football supporter countries such as Mexico. We chose Indonesia from the Pacific Region and Columbia from South America. We also analyse the USA since it is a country with high Twitter penetration.

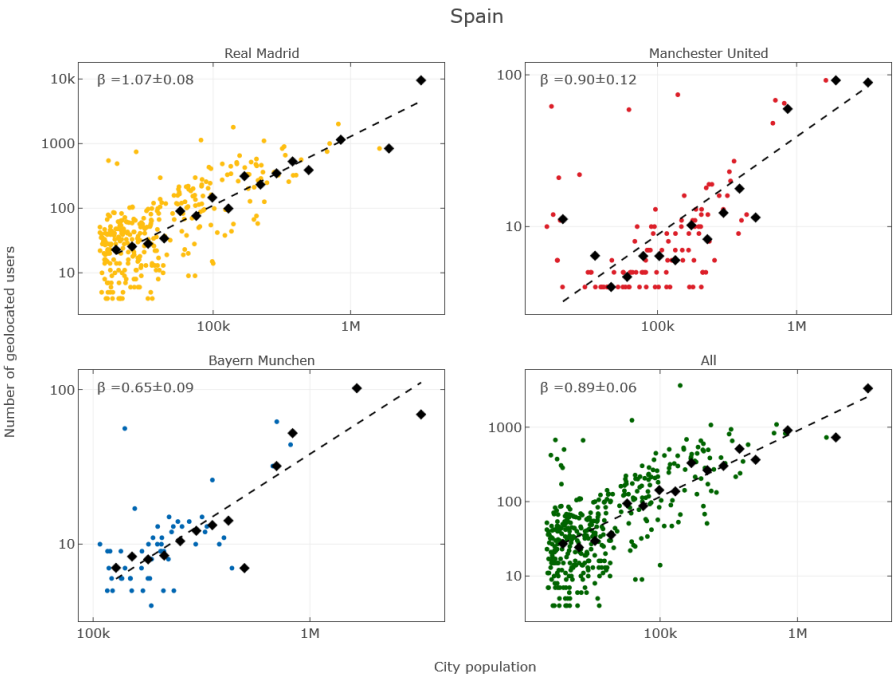

Figure 34 Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in the Spain

Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents β_{RM}=1.07±0.08; β_{MU}= 0.90±0.12; β_{BM} = 0.65±0.09 and β_{All} = 0.89\±0.06, respectively.

Source: Self-prepared chart

In the top left corner of Figure 34, we can see the urban scaling relationships of Spain for the three clubs (Real Madrid in the top left, Manchester United in the top right and Bayern Munich in the bottom left corner), and for the number of overall followers (bottom right corner). The exponent of Real Madrid, the "home" team, is super-linear β_{RM} = 1.07±0.08, while the exponent of the other two teams is sublinear with β_{MU} = 0.90 ± 0.12 for the Manchester United, and β_{BM} = 0.65 ± 0.09 for the Bayern Munich, respectively. It is spectacular how the second biggest city in Spain, Barcelona is a clear outlier in the Real Madrid urban scaling curve, with having much fewer followers than the size of the city would predict. The overall follower numbers in Spain also has a sublinear scaling.

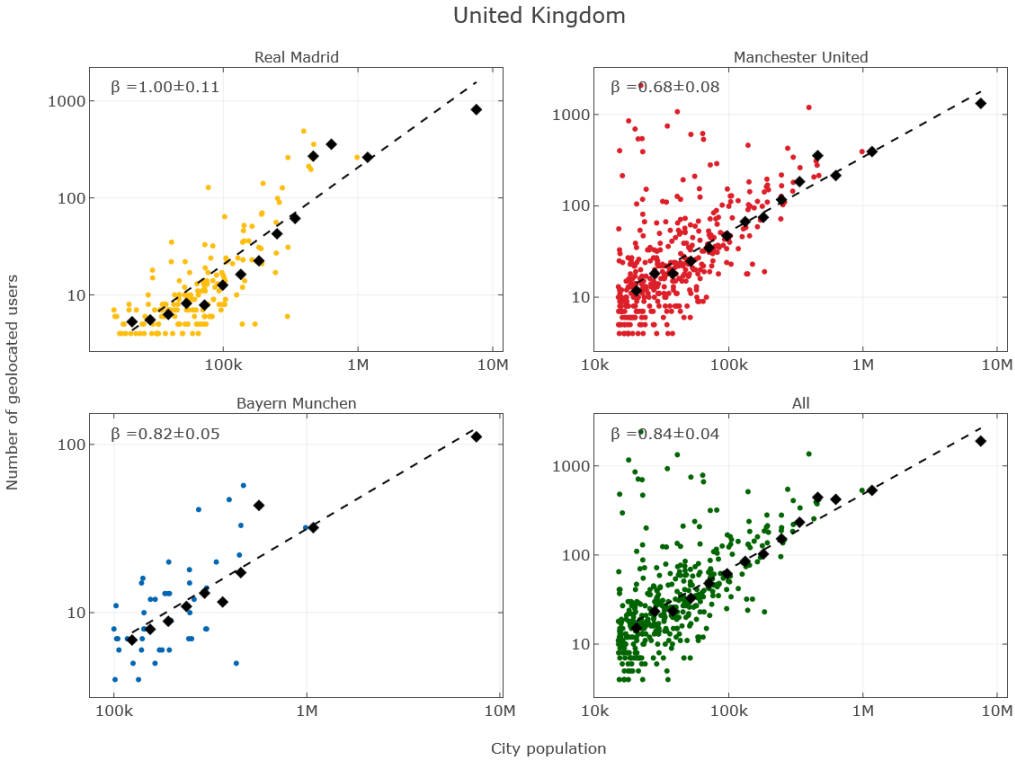

Figure 35 Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in the UK.

Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents β_{RM}*=1.00±0.11,* β_{MU}*= 0.68±0. 08,* β_{BM} *= 0.82±0.05 and* β_{All} *= 0.84±0.04, respectively.*

Source: Self-prepared chart

In Figure 35 when we look at scaling curves in the UK, which has the longest football traditions of all of the countries, we again see a similar picture of the exponents, with that of Real Madrid being higher than the other two. However, it is only around the linear regime with $ß_{RM}$ =1.00 ± 0.11. However, Manchester United apart from the outlier points of Manchester and its surroundings has an astoundingly low sublinear exponent $ß_{MU}$ = 0.68 ± 0.08 that suggests a strong relative decline of interest for this team with the city size. The overall follower trend is also strongly sublinear in the UK.

In case of the USA Figure 38 is similar to that of Spain, where Real Madrid followers scale super linearly, but the other two clubs have a sublinear relationship with city size.

Figure 36 Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in the US.

Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents $\beta_{RM} = 1.08\pm0.06$, $\beta_{MU} = 0.78\pm0.06$, ; $\beta_{BM} = 0.74\pm0.06$ and $\beta_{All} = 0.91\pm0.06$, respectively.

Source: Self-prepared chart from stock data

A very different effect takes place in Indonesia, according to Figure 37. Here, all four scaling relationships are in the highly super linear range, which means that club followership is a measure that is driven by urban factors. Though less pronounced because of slightly smaller, but still super linear exponents, this is also the case for Columbia in Figure 38

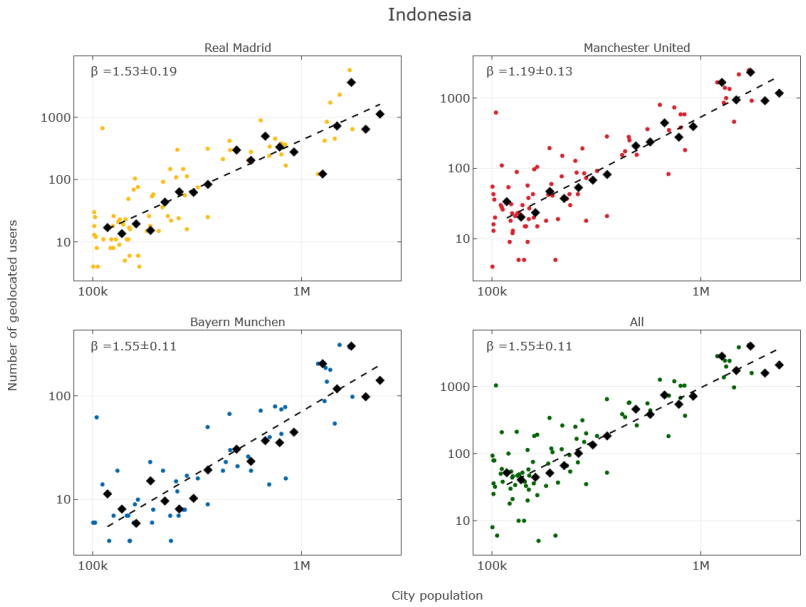

Figure 37 Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in Indonesia

Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents β_{RM} = 1.53±0.19, β_{MU} =1.19±0.13,; β_{BM} = 1.55±0.11 and β_{All} = 1.55±0.11, respectively.

Source: Self-prepared chart from stock data

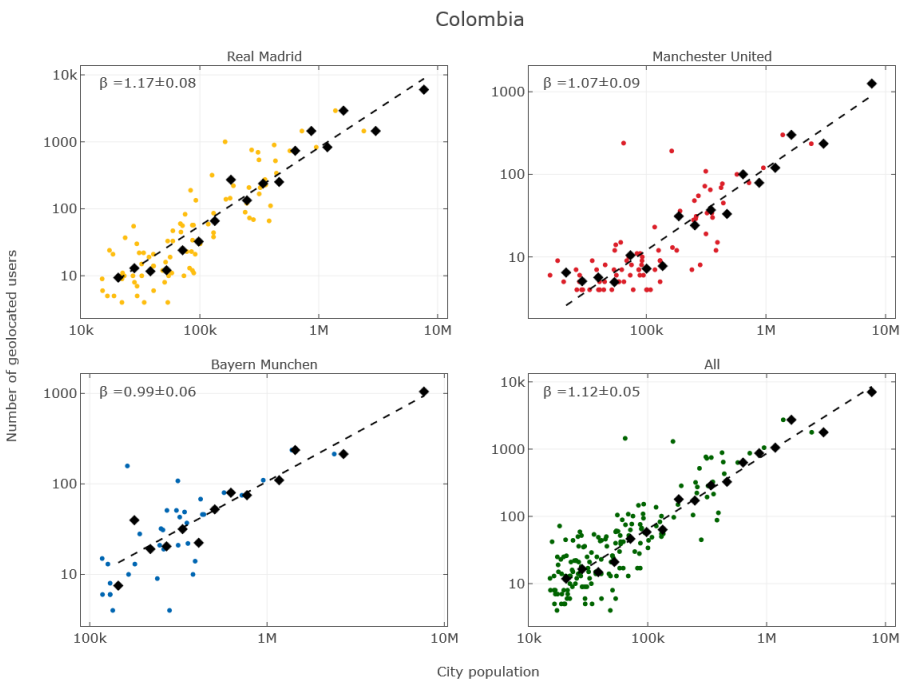

Figure 38 Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in Columbia.

Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents β_{RM} = 1.17±0.08, β_{MU}= 1.07±0.09, ; β_{BM} = 0.99±0.06 and β_{All} =1.12±0.05, respectively.

Source: Self-prepared chart from stock data

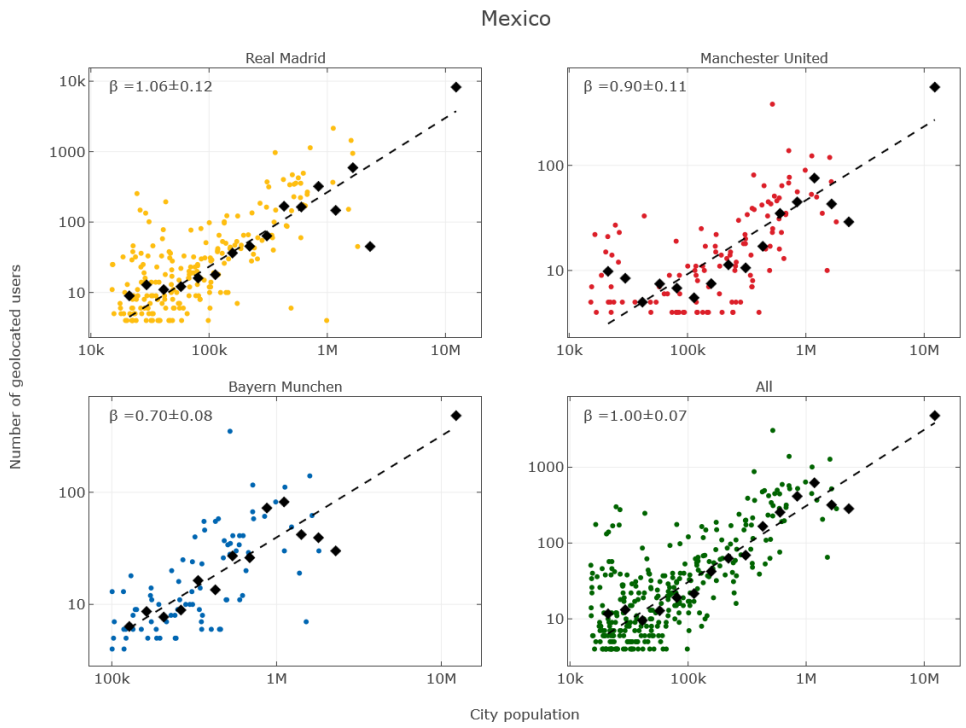

Figure 39 Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in Mexico.

Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents β_{RM} = 1.06±0.18, β_{MU}= 0.90±0.11, ; β_{BM} = 0.70±0.08 and β_{All} =1.0±0.07, respectively.

Source: Self-prepared chart from stock data

The summary Figure 40 shows that Columbia, Indonesia and Mexico, are the countries whose exponents for the overall supporter count are super-linear. This means that in these countries that globalized football tracking is an increasingly urban phenomenon. In countries where football culture is older, and/or general income is higher, sublinear exponents may signal a relative attention shift for football to smaller settlements and a change in the composition of

consumers of football-related content. This may be an important message for marketers trying to increase social media attention and responsiveness because people from different environments may need quite different targeting messages.

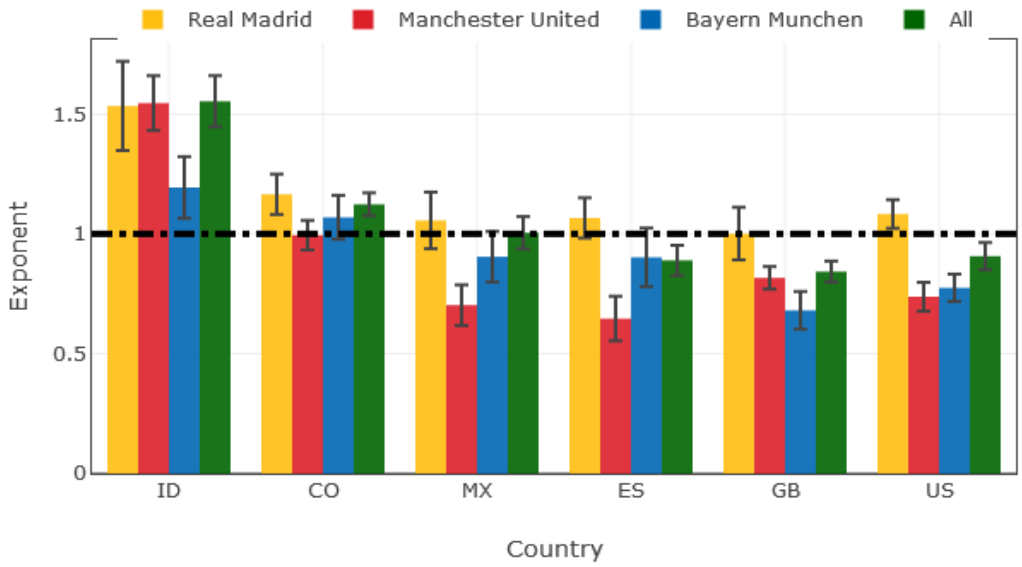

Figure 40 Summary figure showing the exponents per team according to countries. The vertical line at β=1 corresponds to linear scaling.

Source: Self-prepared chart from stock data

The difference between the club exponents in the same country suggests that Real Madrid followers are relatively more represented in bigger cities, and the other two teams have the same exponents. This suggests that even within a country, different clubs may have different follower audiences and fans from different cultural backgrounds.

7.4 Conclusion

In this paper, we analysed urban scaling in the follower numbers of three football clubs, Real Madrid, Manchester United and Bayern Munich. We determined user geolocation from Twitter messages that had GPS coordinates attached to them and fitted scaling relationships using population data for cities of six different countries. While for higher-income countries, urban scaling exponents tended to be in the sublinear, linear, or in a few cases, a slightly super-linear range, exponents for lower-income countries are almost exclusively super-linear. This

suggests that in globalized football fandom, followership is driven by very different factors. Exponents also exhibited variations between clubs, which suggests that the followers of different football clubs are embedded in different socio-economical environments, and that is related to the degree of urbanization as well.

8. Paper 5: NLP analysis of the incident and problem descriptions

ABSTRACT: This article has an empirical approach to the linguistic analysis of customer feedback, using statistical natural language processing (NLP). Considering customer feedback, as texts are mainly written in an unstructured way, there is a need to use text meaning techniques to gain insight or perceive the focusing area of the text. The case is more complicated when the description of events or incidents is associated with a large organization providing a wide range of IT services. This article describes how to use semantic analysis with other data mining techniques which can help to find focus, patterns and trends in texts connected to user feedback.
Keywords: Statistical Natural Language Processing, User feedback, IT services

8.1 Introduction

During everyday work, one can be faced with an incident and problem management, and this is now an everyday activity in large companies. Besides, it is in the IT service management process area as well. The goal of the incident management process is to reconstruct a normal service operation as quickly as possible, and to minimize its effect on the business operations, therefore ensuring that the highest levels of service quality and availability are maintained. A problem record should be created when multiple occurrences of related incidents are observed. The management of a problem is different from the process of managing an incident, and it is typically executed by different staff and controlled by the problem management process. (Figure 41) Root cause analysis (RCA) is part of problem resolution.

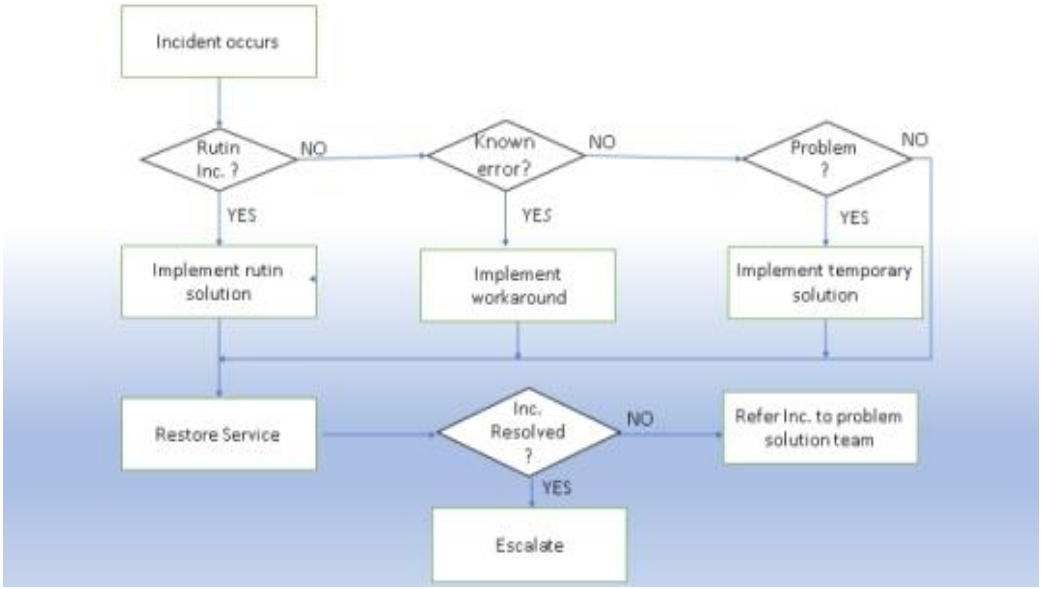

Figure 41 Process of Incident and Problem management

Source: Self-prepared chart

It is estimated that 80% of business-relevant information is recorded in unstructured and semi-structured text data. In other words, without using textual analysis, to discover that 80% of data, all business information and customer's behaviour data would be lost (Halevy, Norvig, & Pereira, 2009). Incident descriptions are usually written by non-professionals and, many times, they are not formulated in competent English, but they still contain a lot of information. (Pan, Liu, Xiang, & Yang, 2011) In order to use this information thoroughly and systematically, it is increasingly necessary to understand customer behaviour and attitudes (Song & He, 2010). For example, if the cause of most incidents is known, measures can be taken to decrease their occurrences. There are no standard rules for writing customer's feedback so that the computer can understand them. The language and the meaning for every piece of text vary depending on the purpose. The only way to include accurately unstructured data in a data-meaning project is to understand the language and the context within which the text is created.

8.2 Literature review

Understanding human language is based on linguistics, commonly referred to as Natural Language Processing (NLP) (Nadkarni, Ohno-Machado, & Chapman, 2011). NLP is a

methodology for computers to analyse, understand, and derive meaning from human language in a smart and useful way (Turney & Pantel, 2010). When developers utilize NLP, they can organize and structure the knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, and topic segmentation. Algorithms of NLP are typically based on machine learning algorithms (Reddy et al., 2015). Instead of hand-coding a large set of rules, NLP can rely on machine learning. A Statistical NLP approach seeks to solve problems by automatic learning of lexical and structural preferences (Cohen & Dolbey, 2007) from corpora, and if the existing dictionary is extended (rules, types, synonyms) to support this process, we might be able to analyse the non-formal descriptions (Jordan & Mitchell, 2015). A system that incorporates NLP can intelligently extract terms, including compound phrases, and permit classification of terms into related groups (Le & David Jeong, 2017). Linguistic systems are knowledge-sensitive: the more information is contained in the linguistic resources (dictionaries), the higher is the quality of results (Yu, Li, Merigó, & Fang, 2016). Modification of the dictionary content, such as synonym definitions, can simplify the resulting information and focus attention on the most relevant concepts.

Text meaning must consider the universal fact that languages contain ambiguities (Kiss & Strunk, 2006). The same words can be different parts of speech (nouns, pronouns, verbs, adjectives, adverbs, etc.), and therefore play different roles in meaning (Goldberg, 2018). The same word, even when used as the same part of speech, can have different meanings depending on how it is used and the context within which it appears. The linguistic analysis involves the study of the elements, structure, and mining of language.

8.3 Methodology

Text mining is the process of extracting knowledge and information from natural language texts. Text mining proceeds in two stages.

Stage 1: Key concepts/terms are extracted from the text that represents the essence of information the text contains.

Stage 2: These concepts/terms are grouped into categories that represent the higher-level ideas contained in the text.

Root Cause Analysis (RCA) is focusing on identifying root causes -, while the incident description is simply describing the circumstance or symptom of an issue. During the analysis, the problem description can provide insight into the problematic area on which the enterprise needs to focus. Also, RCA focuses on the problem, which is typically not a one-time incident. However, a repetitive issue, which means being able to identify a problematic area on the enterprise level, can bring extensive benefits. There are several ways how NLP can help to gather insights out of any text description. Segmentation of text description (by creating clusters) can help to identify the problem area, where we can detect which parts of the service have the biggest impact on the operation. Another possibility is to analyse the problem description based on semantic analysis (stage 1.), the analysis of the meaning of the words, phrases, sentences, and texts. The best examples are synonyms and homonyms. For example, to ring (call or phone) versus ring (some noise in my ear) versus ring (on my finger) or ring (sports ground for boxing). Semantic analysis is the most difficult task for text mining, and it involves the usage of dictionaries, thesauruses, glossaries, lexicons, typologies, and so forth.
Traditional English grammar divides words based on eight parts of speech: verb, noun, pronoun, adjective, adverb, preposition, conjunction, and interjection. However, part-of-speech (PoS) tagging in Text Analytics uses the following tags:

N: **Noun**

A word used to name a person, place, thing, quality, or action and can function as the subject or object of a verb.

V: **Verb**

A word that expresses existence, action, or occurrence such as be, shall, and happen.

A: **Adjective**

A word that is used to modify a noun by limiting, qualifying, or specifying it.

B: **Adverb**

A word that modifies a verb, an adjective, or another adverb.

O: **Coordination**

Conjunction such as "and" and "or".

D: **Determiner**

A noun modifier including articles, demonstratives, possessive adjectives, and words such as any, both, or whose.

G: **Gerund**

A noun derived from a verb. In English ending in the suffix "ing", as smoking is harmful

P: **Participle**

A verb used as an adjective, most often ending in "ing" (present) or "ed" (past), as in “returning home”.

C: **Preposition**

A word placed before a noun indicates the relation of that noun to a verb, an adjective, or another noun. For example, the word "of" is a preposition. Most prepositions are tagged as S or Stop words.

X: **Auxiliary**

A verb such as is, have, can, could, or will usually accompany the main verb in a clause.

S: **Stop word**

A broad category of words is excluded from extraction. It contains all pronouns, particles, and prepositions (except the word "of")

8.4 Results

Semantic analysis has several advantages; it can detect connection among different tags. There are several combinations, which have a powerful meaning, like verb-noun connections. In the case of the nouns, we can increase the significance if we identify not a single term, but a sequence. A typical noun sequence is a bi-gram, but it can be any number of nouns (n-grams) (I. Santos, Penya, Devesa, & Bringas, 2015). Advantage of the n-grams is, that compared with the daily-speech environment, occurrences of n-grams are relatively low (Gries & Mukherjee, 2010), while in the technical environments these occurrences are high (typically identifying technical terms or expressions) and have higher business meaning value, which we can use for analysis (Xuerui Wang, McCallum, & Wei, 2007). One example of how Bi-grams can help to identify areas that we can see here is, that often customer feedback refers to hardware problems, issues related to file systems, or problems describing cases related to performance (Figure 42). In such cases, we need to look at the relationship between different IT environments where the problem occurs. We discovered, for example, that some types of databases could not work

together with old operating systems (performance problems). In other cases, we have discovered that a poorly written application program in different IT environments has always caused a storage problem. The third case refers to the vulnerability of a hardware element.

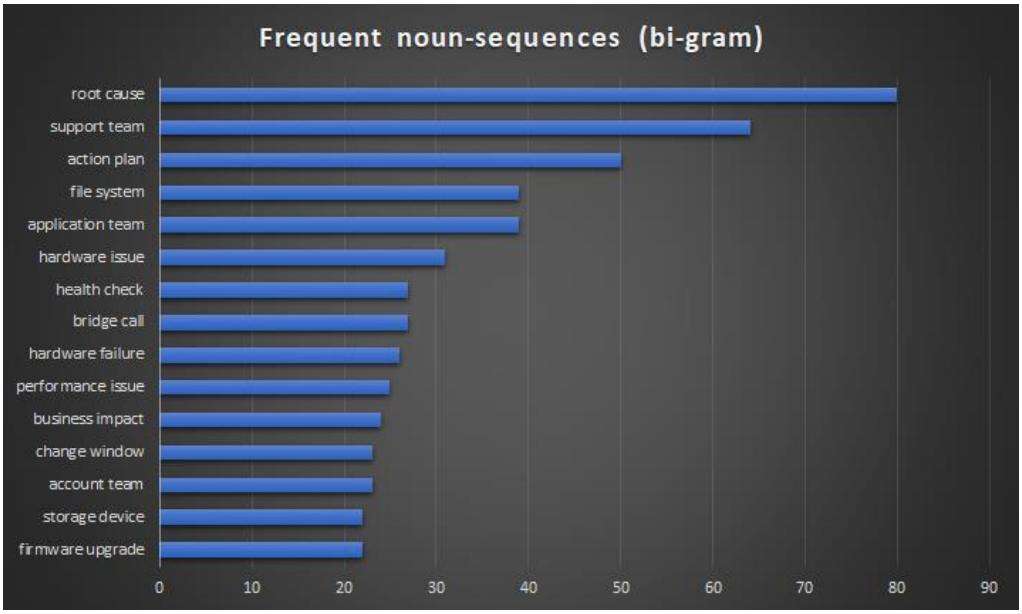

Figure 42 Frequent noun-sequence

Source: Self-prepared chart

Since problems usually do not occur at the same time, enriching our data with time scales can easily help us to identify trends related to some common grammar tags. For example, the usage of some IT environments is often not linear, but there are some frequent periods. In such cases, for example, periodical usage of the term "slow response time" can be observed. In another case, however, we can follow the elimination of the problem. For example, comments mentioning the "file system" bi-gram were reduced at the beginning of the year because some applications have begun automatically to save the data. (Figure 43)

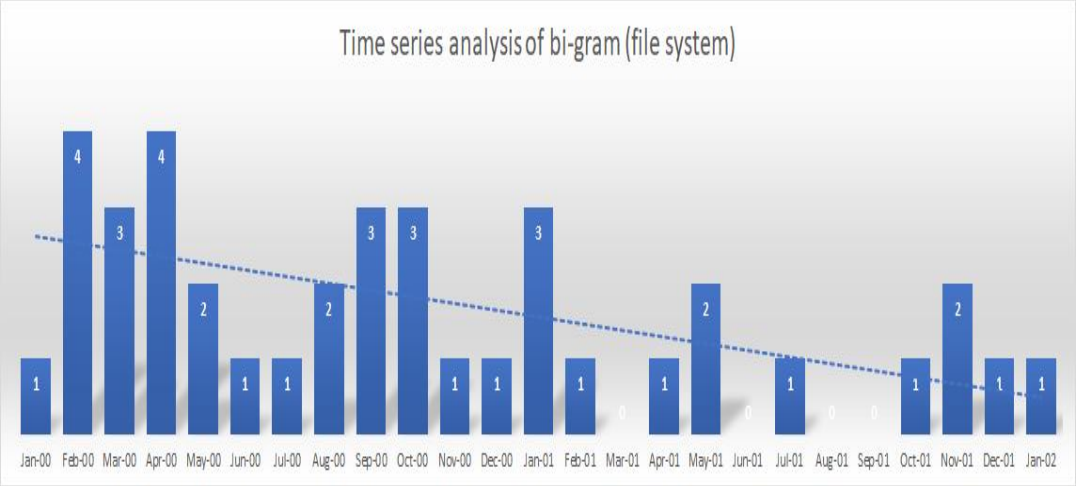

Figure 43 Time series analysis of bi-gram (file system)

Source: Self-prepared chart

At the same time series can be analysed based on deviations in points of view. That helps us to identify those points in the time, that behave differently than we would predict based on previous data points (Figure 44). We could use this method as well in case of problems with smaller parts of the system. For example, a poorly formatted SQL query occasionally causes high memory usage. It is easy to find out which of the thousands of queries are working badly with this method since only the running time of queries had to be compared with the time of deviation.

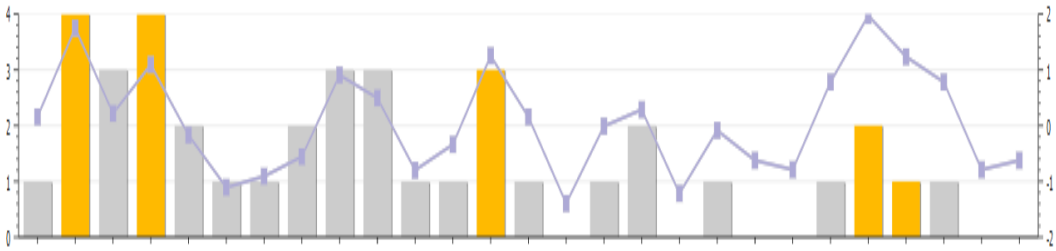

Figure 44 Deviation detection

Source: Self-prepared chart

Semantic analysis can help to detect meaningful connections among different grammatical parts of the customer feedback.

Let us see two examples of connection by detecting a correlation between noun sequences (bi-gram) and modified nouns (Chen, Seymore, & Rosenfeld, 1998). For example, we gain more insight by seeing that a larger part of the file system issues is connected to the Unix team (Figure 45). So, we can identify that this problem occurs mainly around the Unix environment. In such cases, the parts of a text (“same issue”, “file system”, “Unix team”) are not necessarily mentioned sequentially. It may also happen that in different customer feedback, only some synonyms of terms appear, nevertheless, we discover the relationships among the text elements.

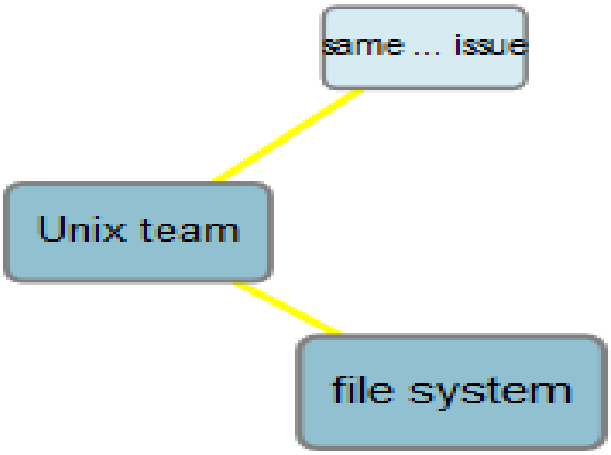

Figure 45 Connection network sample I

Source: Self-prepared chart

Another example shows that there is a correlation between hardware failure and firmware upgrade, (Figure 46) which can help us to identify common issues on the firmware level. In this case, a correlation occurs between the four elements of the text. The incorrect writing of the "hardware" word does not cause difficulty in recognizing the connection among the text elements.

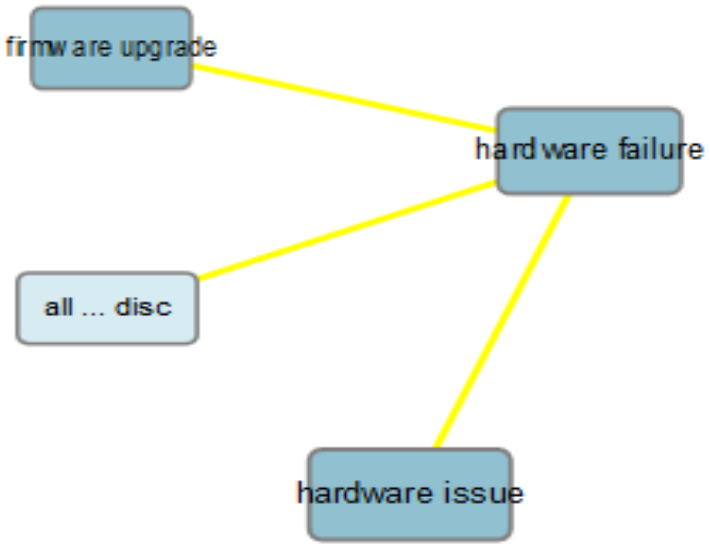

Figure 46 Connection network sample II

Source: Self-prepared chart

Beyond the forms mentioned so far, we can identify structures such as noun sequences, modified nouns or nouns with predicates, which help us to identify more relevant text structures in describing our problem space (Figure 47).

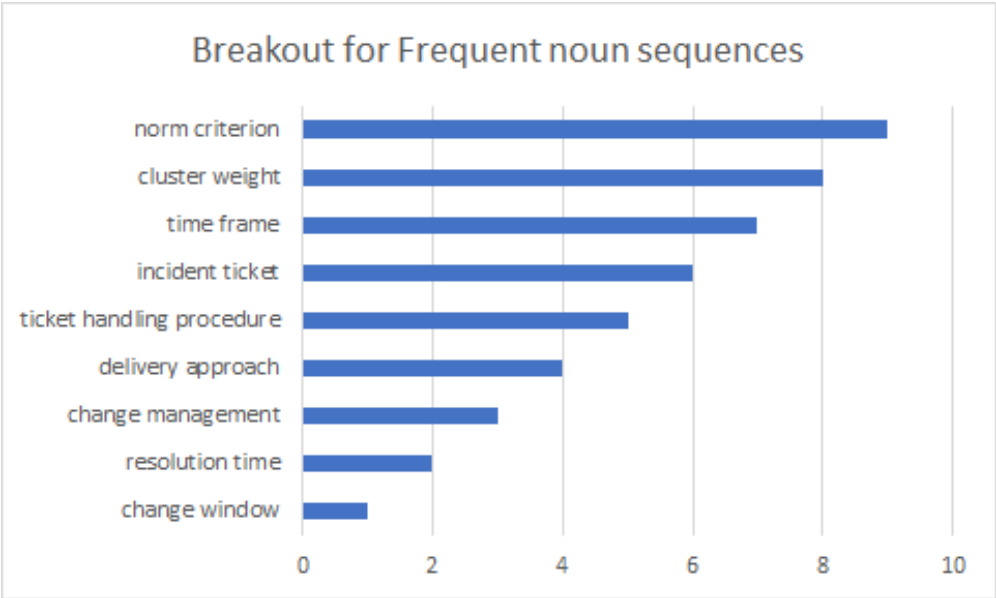

Figure 47 Frequent noun sequences

Source: Self-prepared chart

The frequency of certain terms and text structures may reflect the severity of any problematic area. Based on that, we need to consider what solution shall we choose. For example, suppose many posts contain the expression "time frame", which refers to a certain period, in which something planned is to happen in the IT system. In that case, we need to pay special attention to planning the introduction of changes in the future, because clearly, this is an important area for our customers.

8.5 Conclusion

NLP is an efficient way to discover patterns of large sets of textual feedback. Simply identifying grammatical structures helps to detect trends and deviations related to the components of provided service or the applied service. There are grammatical structures, which are more valuable for this purpose than nouns and verbs that are the key language components. Customers' feedback and the incident descriptions describe certain elements of the customers' opinion; however, they also provide further information for the complex enterprise environment. These details are usually highly unstructured, and we need to apply techniques to extract information out of them. By combining NLP procedures with other methods, such as time series analysis, deviation detecting and connection detecting, we can identify trends and deviations by focusing on business-relevant expressions and terms.

8.6 Limitation and future work

In the future, I would like to develop textual analysis procedures that perform semantic analysis of a given text and define a positive or negative opinion within that text. Using these options, we would like to analyse the texts in the future and show how the opinion of users affects the movement of the company's stocks.

9. Conclusion

9.1 Hypotheses conclusions

The systems studied above are complex systems whose elements and the relationships between the elements, are difficult to recognize. Therefore, I examined these systems through the models I developed. Characteristics of models: they are *simpler than reality* because they contain only the most important elements and processes. This way features that are important for operation come to the fore. The advantage of modelling was that it made possible the mathematical examination of models.

The systems I study are often chaotic. The reason for the chaotic behaviour is the nonlinear nature of systems. This means that the response of the system to a change in an input parameter is not proportional to the change in the output. Nonlinear systems are sensitive to initial conditions. This sensitivity is well felt in the so-called butterfly effect. (Term butterfly effect can be illustrated by the fact that the delicate wing blows of a butterfly can cause massive windstorms and hurricanes in a more remote region.) That is, changes of microscopic size are transformed into macroscopic size. This phenomenon is also called the phenomenon of small causes that generate large effects. The elaborated papers are also linked by the fact that, at all levels of complexity, there are limitations to understanding a critical event and predicting future events. Within these limits, however, chaos theory helps to simplify the world around us. Based on the above studies, and the corresponding five hypotheses here is what we now understand better about the social reality through chaos theory.

H1$_1$: The methods of complex systems can be used to describe the behavioural patterns of a football team as a complex system. - True

I studied the development of a football team. Systems operating in a sports environment are susceptible to small changes in variables in critical areas. A complex methodology that focuses on processes helped me to elaborate processes between stable states. This study demonstrated that the club showed some universal properties of the dynamics of complex systems. The

differentiated learning approach suggested that teams tend to find optimal performance patterns. In contrast, understanding the motivations of smaller groups in the team and successfully integrating this knowledge into understanding higher-level elements of the system, contributed more to the understanding of complex transitions. Micro- and macro-changes have been integrated as a common driving force and generated entirely new features of the whole system.

H2$_1$: There is a group of individual investors whose decisions about investing in a football club are driven not only by the consideration of their long-term well-being but also by their daily emotional state related to the club. – True

One way to understand stock market movements is to interpret the stock market as a complex system. I have highlighted the feature of complex systems, that small changes in the input can cause big changes in the output. Individual investors' behaviour is influenced by public information, and their mood is affected by different events related to football clubs. In predictions, the goal was not to give an accurate estimate of what would happen in the future. The aim was to outline realistic scenarios, alternatives that could point the way for the future. As a result, individual investors may have completely different opinions about positive and negative sporting events. The aim of the study of Twitter messages can help to find out how investors respond to different events. In the life of a football club, the result of the match is the most critical event. However, how unexpected the outcome of a match is, can be well measured by the number of tweets connected to the match. I investigated several tweet messages connected to two major football clubs and compared them with changes in their stock price and in trading volume. In some cases, these two indicators move together.

H3$_1$: There is a link between the frequency of appearance of innovative companies in social media and the volatility of stock prices of these companies. - Partly True

The big innovative companies of our time have taken advantage of the web. Products have been created that are intricately linked to innovation. The users of these products have to adapt to new challenges. They need to be familiar with the digital world. One of the basic assumptions of economic theory is rationality: that is, economic actors always act rationally. However, there are significant problems with the application of the assumption as more and more attention is

paid to the non-linearity. This kind of nonlinearity can lead to significant temporary changes. To understand how investors are responding to different events, I have examined several tweets about three large companies. The significance of the events was related to the relative number of tweets on a given day. The change in the absolute value of the price or volume in these cases changes after one, two or three days. I have developed a mood quantification method for large companies - the number of occasions customers and users refer to a given company in tweets - and realized that there are relatively high numbers of tweets when absolute stock price movements are relatively high. In some cases, these two indicators move together.

H4$_1$: The lower is the GDP of a country; the more football fans from this country live in cities. - True

The network of football fans also forms a complex system. With the use of modern technical equipment, significant events related to the biggest football clubs can now be tracked from anywhere. Complexity renewed not only the thought background but also the methodology. In the age of computing, I can prove relationships that help to understand the whole system better. I determined user geolocation from Twitter messages that had GPS coordinates attached to them and fitted scaling relationships using population data for cities of six different countries. Using urban scaling theory, I have measured how the size of a city relates to the number of football fans. While for higher-income countries, urban scaling exponents tended to be in the sublinear, linear, or in a few cases, a slightly super-linear range, exponents for lower-income countries are almost exclusively super-linear.

H5$_1$: The usage of semantic analysis with other data mining techniques can help to find focus, patterns and trends in texts connected to user feedback. - True

Complexity usually assumes some sort of hierarchy. The system is complicated by the fact that elements at a certain level of the hierarchy can influence each other. What makes complex systems interesting is that as the result of the interaction between their parts, the behaviour of the parts changes in such a way, that the whole system follows a qualitatively new pattern. The IT infrastructure forms a complex system. In enterprise environments, the analysis of failure phenomena connected to this infrastructure is paramount. Analysing bug reports can help to

find broader relationships across systems. Identifying grammatical structures in these reports helps to detect trends and deviations related to the components of the provided service or the applied service. There are grammatical structures, which are more valuable for this purpose than nouns and verbs, that are the critical language components. Customers' feedback and the incident descriptions determine some aspects of the customer's opinion; however, they also provide further information for the complex enterprise environment. By combining NLP procedures with other methods, such as time series analysis, deviation detecting and connection detecting, we can identify trends and deviations by focusing on business-relevant expressions and terms.

9.2 Methodology

An essential element of complex systems is the schema to decipher the behaviour of complicated phenomena. The schema summarizes the things we have noticed about our environment. This makes it easier to deal with complex phenomena. The design of the schemes is influenced by public perception as well as by the observer himself. More complex relationships can only be schematically arranged by a professional.

I studied schemes of human behaviour with methods characteristic for complex systems. After that collected a lot of data using the internet. Moreover, I used the computing power of the computer that allowed me to do up to millions of complicated calculations in seconds. However, understanding socio-economic phenomena is not an easy task. Models had to be used that combine microscopic data with macroscopic observations. During my work, I built networks. I was looking for connections and focus points. After that, the results of these methods had to be evaluated, scaled and compared. I trust that this work provides valuable insight for future research of any community trying to understand the complex systems around us.

9.3 Limitation and future work

I have tried to use the potential of large data sources to understand the behaviour of a particular group of people. I used the spatial, temporal and textual characteristics of the data sets. I developed models to explain the collective phenomena measured in the data. Some of the limitations I mentioned in the previous section. Numerous experiments and experiences support the methods used. Variables, boundary conditions, and characteristics were described

separately during problem-solving. Due to the diversity of the topics, I was often able to describe the limitations of the chosen method only to a limited extent. The same applies to future research directions, the access to data was limited in many cases; research should be conducted with multiple data sources in the future. Thus, the models will become increasingly accurate, which may help to explore newer aspects of human relationships.

10. Results and summary

Due to the spread of the crown virus, tournaments have largely stopped. Significant events, such as the European Football Championship have been postponed. In such times, one can only remember the old days of the football, and therefore, it is time for analysis. This study fulfils that second need. In the first part, I gained insight through the development of a football team into the operation of complex systems. The results show that the interactions of the elements of the system generate completely new functions for the observed structure. The results of the second part show that the prediction accuracy of the conventional stock exchange forecasting models is greatly improved by incorporating specific mood dimensions. The methodology has been tested using two famous football clubs, namely Manchester United and Juventus and some big innovative companies, namely Google, Amazon, and Sony. From the managerial perspective, this method can help managers to develop integrated marketing plans in order to increase fans' brand loyalty.

In the next part, I was looking for an answer to the question: what is the force which persuades fans to watch football games or follow news connected to football in social media? Determining the geographic location of football supporters helped me to understand the motivation of fans. The phenomenon that I found during my research is that there are football fans in a higher ratio in the villages and smaller towns in developed countries, than in developing countries. It is also true that fans of some clubs tend to live in big cities, while fans of other teams rather live in small towns. Why is this important? Because different products can be advertised through the clubs that have supporters from big cities, and different ones-can be promoted by clubs whose fans are more small-town people.

In the last section, I studied customer feedback that describes the opinions of individual customers separately. However, they contain more information, not only an opinion. These

details are very unstructured, and we need to use different techniques to extract information from them. Using previously known language analysis methods and NLP can help the analysis of non-structured texts from corporate environments. Why is this important? Because it helps to identify trends and deviations in different services. Once we become aware of this, we can improve the quality of our service.

10.1 Publications supporting the thesis:

In the order of Chapters 4-8.

Soti, A. (2019). Understanding Micro and Macro Interactions in Social Neurobiological Systems. *Interdisciplinary Description of Complex Systems*. https://doi.org/10.7906/indecs.17.4.1

Soti A, Acarani A, Steiger J, Vattay G.
Influence of Twitter activity on the stock price of soccer clubs: Social Network Analysis and Mining **https://link.springer.com/article/10.1007/s13278-020-00691-2**

Soti A, Impact of Twitter activity on stock prices of innovative companies
Submitted to journal: Management and production engineering review

Sóti, A., Bokányi, E., & Vattay, G. (2018). Urban scaling of football followership on Twitter. *Acta Polytechnica Hungarica*, *15*(5). https://doi.org/10.12700/APH.15.5.2018.5.14

Sóti, A. Dobos Z. (2018)., *NLP analysis of incidents descriptions in case of IT services*
In proceedings: Proceedings of 25th EurOMA Conference 2018-06-24 [Budapest, Hungary_ http://euroma2018.org/wp-content/uploads/proceeding/570_654_4.pdf

Sóti, A. (2020). A Python programozási nyelvről statisztikusoknak. *Statisztikai Szemle*, *4*(1). https://doi.org/10.20311/stat2020.4.hu0324

11. References

[1.] Admiral Market. (2019). General football's data Retrieved. Retrieved from https://admiralmarkets.com/?regulator=fca

[2.] Agile Alliance. (2015). Team Definition. *Glossary of Management*.

[3.] Alagidede, P., Panagiotidis, T., & Zhang, X. (2011). Causal relationship between stock prices and exchange rates. *Journal of International Trade and Economic Development*. https://doi.org/10.1080/09638199.2011.538186

[4.] Allen, P. M. (2001). A complex systems aproch to learning in adaptive networks. *International Journal of Innovation Management*. https://doi.org/10.1142/s136391960100035x

[5.] Anderson, P. (1999). Complexity Theory and Organization Science. *Organization Science*. https://doi.org/10.1287/orsc.10.3.216

[6.] Annepu, R. K. (2012). Sustainable Solid Waste Management in India. https://doi.org/10.1007/978-981-4451-73-4

[7.] Appadurai, A. (1990). Disjuncture and Difference in the Global Cultural Economy. *Theory, Culture & Society*. https://doi.org/10.1177/026327690007002017

[8.] Appadurai, A. (2000). Grassroots globalization and the research imagination. *Public Culture*. https://doi.org/10.1215/08992363-12-1-1

[9.] Appadurai, A. (2014). Arjun Appadurai. *Globalizations*. https://doi.org/10.1080/14747731.2014.951209

[10.] Appadurai, A. (2018). The production of locality. In *Sociology of Globalization: Cultures, Economies, and Politics*. https://doi.org/10.4324/9780429493089

[11.] Arrow, H., McGrath, J. E., & Berdahl, J. L. (2000). Groups as Complex System. In *Small Groups as Complex Systems: Formation, Coordination, Development, and Adaptation*.

[12.] Arthur, W. B. (2018). The Economy as an Evolving Complex System II. In *The Economy as an Evolving Complex System II*. https://doi.org/10.1201/9780429496639

[13.] Ashton, J. K., Gerrard, B., & Hudson, R. (2003). Economic impact of national sporting success: Evidence from the London stock exchange. *Applied Economics Letters*, *10*(12), 783–785. https://doi.org/10.1080/1350485032000126712

[14.] Asur, S., & Huberman, B. A. (2010). Predicting the future with social media.

Proceedings - 2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010. https://doi.org/10.1109/WI-IAT.2010.63

[15.] Bak, P., & Chen, K. (1991). Self-organized criticality. *Scientific American*. https://doi.org/10.1038/scientificamerican0191-46

[16.] Balague, N., Torrents, C., Hristovski, R., Davids, K., & Araújo, D. (2013). Overview of complex systems in sport. *Journal of Systems Science and Complexity*. https://doi.org/10.1007/s11424-013-2285-0

[17.] Baldwin, R. (2006). Globalisation: the great unbundling(s). *Economic Council of Finland*.

[18.] Barankai, N., Fekete, A., & Vattay, G. (2012). Effect of network structure on phase transitions in queuing networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. https://doi.org/10.1103/PhysRevE.86.066111

[19.] Batty, M. (2009). Cities as Complex Systems: Scaling, Interaction, Networks, Dynamics and Urban Morphologies. In *Encyclopedia of Complexity and Systems Science*. https://doi.org/10.1007/978-0-387-30440-3_69

[20.] Baumol, W. J., & Benhabib, J. (1989). Chaos: Significance, Mechanism, and Economic Applications. *Journal of Economic Perspectives*. https://doi.org/10.1257/jep.3.1.77

[21.] Berger, S. (2000). Globalization and politics. *Annual Review of Political Science*. https://doi.org/10.1146/annurev.polisci.3.1.43

[22.] Bettencourt, L. M. A. (2013). The origins of scaling in cities. *Science*. https://doi.org/10.1126/science.1235823

[23.] Bettencourt, L. M. A., Lobo, J., Strumsky, D., & West, G. B. (2010). Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0013541

[24.] Bhattacharya, A. K., & Michael, D. C. (2008). How local companies keep multinationals at bay. *Harvard Business Review*.

[25.] Biti, O. (2017). Hrvatska fanovska scena: nogomet u televizijskim reklamama za pivo. *Glasnik Etnografskog Instituta SANU*.

[26.] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics Reports*. https://doi.org/10.1016/j.physrep.2005.10.009

[27.] Bojic, I., Belyi, A., Ratti, C., & Sobolevsky, S. (2016). Scaling of foreign attractiveness

for countries and states. *Applied Geography*. https://doi.org/10.1016/j.apgeog.2016.06.006

[28.] Bokányi, E., Kondor, D., Dobos, L., Sebők, T., Stéger, J., Csabai, I., & Vattay, G. (2016). Race, religion and the city: Twitter word frequency patterns reveal dominant demographic dimensions in the United States. *Palgrave Communications*. https://doi.org/10.1057/palcomms.2016.10

[29.] Bokányi, E., Szállási, Z., & Vattay, G. (2018). Universal scaling laws in metro area election results. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0192913

[30.] Boli, J., & Lechner, F. J. (2015). Globalization and World Culture. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*. https://doi.org/10.1016/B978-0-08-097086-8.10409-X

[31.] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*. https://doi.org/10.1016/j.jocs.2010.12.007

[32.] Bolotin, Y., Tur, A., & Yanovsky, V. (2009). Stochastic resonance. *Understanding Complex Systems*. https://doi.org/10.1007/978-3-642-00937-2_7

[33.] Bonchi, F., Castillo, C., Gionis, A., & Jaimes, A. (2011). Social network analysis and mining for business applications. *ACM Transactions on Intelligent Systems and Technology*. https://doi.org/10.1145/1961189.1961194

[34.] Boor, S., Hanson, C., & Ross, C. (2020). *Deloitte Football Money League | Deloitte UK*. Retrieved from https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/deloitte-football-money-league.html

[35.] Borck, T. (2016). Team of Teams: New Rules of Engagement for a Complex World. *The RUSI Journal*. https://doi.org/10.1080/03071847.2016.1174487

[36.] Bowen, P., Govender, R., & Edwards, P. (2014). Structural equation modeling of occupational stress in the construction industry. *Journal of Construction Engineering and Management*. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000877

[37.] Boysen, N., Fliedner, M., & Scholl, A. (2007). A classification of assembly line balancing problems. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2006.10.010

[38.] Braun, M., & Raddatz, C. (2008). The Politics of Financial Development: Evidence from Trade Liberalization. *Journal of Finance*. https://doi.org/10.1111/j.1540-6261.2008.01363.x

[39.] Briscoe, T., Copestake, A., & Boguraev, B. (1990). Enjoy the paper: Lexical semantics via lexicology. *Proceedings of the 13th ...*. Retrieved from https://doi.org/10.3115/997939.997947

[40.] Brown, G. W., & Hartzell, J. C. (2001). Market reaction to public information: The atypical case of the Boston Celtics. *Journal of Financial Economics*. https://doi.org/10.1016/S0304-405X(01)00047-2

[41.] Bruns, A., Weller, K., & Harrington, S. (2014). Twitter and Sports. In *Twitter and Society*. https://doi.org/10.1017/CBO9781107415324.004

[42.] Cao, L. (2015). Complex systems. In *Advanced Information and Knowledge Processing*. https://doi.org/10.1007/978-1-4471-6551-4_1

[43.] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Daimlerchrysler, T. R., Shearer, C., & Daimlerchrysler, R. W. (2000). Step-by-step data mining guide. *SPSS Inc*. https://doi.org/10.1017/CBO9781107415324.004

[44.] Chen, S. F., Seymore, K., & Rosenfeld, R. (1998). Topic adaptation for language modeling using unnormalized exponential models. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. https://doi.org/10.1109/ICASSP.1998.675356

[45.] Choi, S., & Park, H. W. (2014). Flow of Online Content from Production to Consumption in the Context of Globalization Theory. *Globalizations*. https://doi.org/10.1080/14747731.2014.904172

[46.] Christian, D. (2011a). History and time. *Australian Journal of Politics and History*. https://doi.org/10.1111/j.1467-8497.2011.01601.x

[47.] Christian, D. (2011b). Maps of time: An introduction to big history. In *Maps of Time: An Introduction to Big History*. https://doi.org/10.1080/03612750409605157

[48.] Clanchy, M. T. (1983). Looking Back from the Invention of Printing. *Literacy in Historical Perspective*.

[49.] Cogan, J. F., Cwik, T., Taylor, J. B., & Wieland, V. (2010). New Keynesian versus old Keynesian government spending multipliers. *Journal of Economic Dynamics and Control*. https://doi.org/10.1016/j.jedc.2010.01.010

[50.] Cohen, K. B., & Dolbey, A. (2007). Foundations of Statistical Natural Language Processing (review). *Language*. https://doi.org/10.1353/lan.2002.0150

[51.] Cooper, R. N., & Karl, T. L. (1998). The Paradox of Plenty: Oil Booms and Petro-States.

Foreign Affairs. https://doi.org/10.2307/20048806

[52.] Cottineau, C., Hatna, E., Arcaute, E., & Batty, M. (2017). Diverse cities or the systematic paradox of Urban Scaling Laws. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2016.04.006

[53.] Critical transitions in nature and society. (2009). *Choice Reviews Online*. https://doi.org/10.5860/choice.47-1380

[54.] Csanády, A. (2012). Structure Change in the Economy of Romania – Part Two. *Társadalomkutatás*. https://doi.org/10.1556/tarskut.30.2012.3.3

[55.] Davids, K., Hristovski, R., Araújo, D., Serre, N. B., Button, C., & Passos, P. (2013). Complex systems in sport. In *Complex Systems in Sport*. https://doi.org/10.4324/9780203134610

[56.] De Zoysa, R., & Newman, O. (2002). Globalization, soft power and the challenge of hollywood. *International Journal of Phytoremediation*. https://doi.org/10.1080/1356977022000025678

[57.] Dellarocas, C. (2003). The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms. In *SSRN*. https://doi.org/10.2139/ssrn.393042

[58.] Demos & Pi. (2017). *Sul Capitale Sociale degli Italiani I giovani e le passioni (in italian)*. Retrieved from http://www.demos.it/

[59.] DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of Applied Psychology*. https://doi.org/10.1037/0021-9010.89.6.1035

[60.] Dobos, L., Szule, J., Bodnar, T., Hanyecz, T., Sebok, T., Kondor, D., … Vattay, G. (2013). A multi-terabyte relational database for geo-tagged social network data. *4th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2013 - Proceedings*. https://doi.org/10.1109/CogInfoCom.2013.6719259

[61.] Dockery, E., Vergari, D., & Vergari, F. (2001). Explaining the behaviour of stock prices in an emerging market: An empirical analysis of the greek stock market. *Managerial Finance*, *27*(1–2), 82–98. https://doi.org/10.1108/03074350110767510

[62.] Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE*, *6*(12).

https://doi.org/10.1371/journal.pone.0026752

[63.] Edmans, A., García, D., & Norli, Ø. (2007). Sports sentiment and stock returns. *Journal of Finance*. https://doi.org/10.1111/j.1540-6261.2007.01262.x

[64.] Edwards, P. J., & Bowen, P. A. (1998). Risk and risk management in construction: A review and future directions for research. *Engineering, Construction and Architectural Management*. https://doi.org/10.1108/eb021087

[65.] Ethiraj, S. K., & Levinthal, D. (2004). Modularity and Innovation in Complex Systems. *Management Science*. https://doi.org/10.1287/mnsc.1030.0145

[66.] Fama, E. F. (2002). The Behavior of Stock-Market Prices. *The Journal of Business*. https://doi.org/10.1086/294743

[67.] Feigenbaum, M. J. (1978). Quantitative universality for a class of nonlinear transformations. *Journal of Statistical Physics*. https://doi.org/10.1007/BF01020332

[68.] Filo, K., Lock, D., & Karg, A. (2015). Sport and social media research: A review. *Sport Management Review*. https://doi.org/10.1016/j.smr.2014.11.001

[69.] Garrett, G. (2000). The causes of globalization. *Comparative Political Studies*. https://doi.org/10.1177/001041400003300610

[70.] Gladwell, M. (2000). [hooking] The Stickiness Factor: Sesame Street, Blue's Clues, and the Educational Virus. In *The tipping point : how little things can make a big difference*. Retrieved from https://archive.org/details/tippingpointhowl00glad

[71.] Glazier, P. S., & Davids, K. (2009). Constraints on the complete optimization of human motion. *Sports Medicine*. https://doi.org/10.2165/00007256-200939010-00002

[72.] Gleick, J., & Hilborn, R. C. (1988). Chaos, Making a New Science . *American Journal of Physics*. https://doi.org/10.1119/1.15345

[73.] Global Market Insights. (2016). Global Market Insights. *Gmi*.

[74.] Goldberg, Y. (2018). Neural network methods for natural language processing. *Computational Linguistics*. https://doi.org/10.1162/COLI_r_00312

[75.] Gould, A. E. (2008). Leadership and the New Science: Discovering Order in a Chaotic World. *The Social Science Journal*. https://doi.org/10.1016/j.soscij.2008.09.005

[76.] Gries, S. T., & Mukherjee, J. (2010). Lexical gravity across varieties of English: An ICE-based study of n -grams in Asian Englishes . *International Journal of Corpus Linguistics*. https://doi.org/10.1075/ijcl.15.4.04gri

[77.] Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., …

DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*. https://doi.org/10.1126/science.1116681

[78.] Gros, C. (2008). Complex and adaptive dynamical systems: A primer. In *Complex and Adaptive Dynamical Systems: A Primer*. https://doi.org/10.1007/978-3-540-71874-1

[79.] Gruhl, D., Guha, R., Kumar, R., Novak, J., & Tomkins, A. (2005). *The predictive power of online chatter*. https://doi.org/10.1145/1081870.1081883

[80.] Haivas, I. (2003). Globalization: A Very Short Introduction. *BMJ*. https://doi.org/10.1136/sbmj.0307258

[81.] Halevy, A., Norvig, P., & Pereira, F. (2009). The Unreasonable Effectiveness of Data. *IEEE Intelligent Systems*. https://doi.org/10.1109/mis.2009.36

[82.] Hampson, F. O., Held, D., McGrew, A., Goldblatt, D., & Perraton, J. (1999). Global Transformations: Politics, Economics, and Culture. *International Journal*. https://doi.org/10.2307/40203424

[83.] Hanley, Q. S., Khatun, S., Yosef, A., & Dyer, R. M. (2014). Fluctuation scaling, Taylor's law, and crime. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0109004

[84.] Hayek, F. a. (1994). Chapter 4: The Theory of Complex Phenomena. *Readings in the Philosophy of Social Science*.

[85.] Hazlitt, H. (1960). The Critics of Keynesian Economics. In *New York*. https://doi.org/10.1111/j.1813-6982.1963.tb02952.x

[86.] Hinz, O., Skiera, B., Barrot, C., & Becker, J. U. (2011). Seeding Strategies for Viral Marketing: An Empirical Comparison. *Journal of Marketing*. https://doi.org/10.1509/jm.10.0088

[87.] Holt, N. L., Tamminen, K. A., Black, D. E., Sehn, Z. L., & Wall, M. P. (2008). Parental involvement in competitive youth sport settings. *Psychology of Sport and Exercise*. https://doi.org/10.1016/j.psychsport.2007.08.001

[88.] Holton, R. J. (2013). Economy and society. In *Economy and Society*. https://doi.org/10.4324/9780203388365

[89.] Hromic, H., & Hayes, C. (2019). Characterising and evaluating dynamic online communities from live microblogging user interactions. *Social Network Analysis and Mining*. https://doi.org/10.1007/s13278-019-0576-8

[90.] Irwin, G. (2015). BREXIT: the impact on the UK and the EU. *Global Council*.

[91.] Jacobs, J. M., & Lees, L. (2013). Defensible space on the move: Revisiting the urban

geography of alice coleman. *International Journal of Urban and Regional Research*. https://doi.org/10.1111/1468-2427.12047

[92.] James, H. (2013). The multiple contexts of Bretton woods. *Oxford Review of Economic Policy*. https://doi.org/10.1093/oxrep/grs028

[93.] Ji, X., Chun, S. A., Wei, Z., & Geller, J. (2015). Twitter sentiment classification for measuring public health concerns. *Social Network Analysis and Mining*. https://doi.org/10.1007/s13278-015-0253-5

[94.] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*. https://doi.org/10.1126/science.aaa8415

[95.] Juventus Football Club S.p.A. (2017). Bilancio di sostenibilità 2015–16" (PDF) (in Italian). Retrieved from https://www.juventus.com/it/club/

[96.] Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*. https://doi.org/10.2307/1914185

[97.] Kantar Media. (2013). European football social media report. Retrieved from https://www.kantarmedia.com/uk?=country-site

[98.] Karaboga, D., & Akay, B. (2009). A survey: Algorithms simulating bee swarm intelligence. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-009-9127-4

[99.] Kelle, P., & Miller, P. A. (2001). Stockout risk and order splitting. *International Journal of Production Economics*. https://doi.org/10.1016/S0925-5273(00)00137-7

[100.] Kellner, D. (2002). Theorizing globalization. *Sociological Theory*. https://doi.org/10.1111/0735-2751.00165

[101.] Kelly, D. L. (2005). Price and quantity regulation in general equilibrium. *Journal of Economic Theory*. https://doi.org/10.1016/j.jet.2004.07.006

[102.] Keohane, R. O. (2005). After hegemony: Cooperation and discord in the world political economy. In *After Hegemony: Cooperation and Discord in the World Political Economy*. https://doi.org/10.2307/40202461

[103.] Khondker, H. H. (2011). Role of the New Media in the Arab Spring. *Globalizations*. https://doi.org/10.1080/14747731.2011.621287

[104.] Kiel, L. D., & Elliott, E. (2009). Chaos theory in the social sciences foundations and applications. In *Chaos Theory in the Social Sciences*.

[105.] Kiss, T., & Strunk, J. (2006). Unsupervised multilingual sentence boundary detection. *Computational Linguistics*. https://doi.org/10.1162/coli.2006.32.4.485

[106.] Kitchin, R. (2014). The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences. In *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. https://doi.org/10.4135/9781473909472

[107.] Knight, F., Of, R., & Classics, E. (1921). Risk, Uncertainty and Profit. *Climate Change 2013 - The Physical Science Basis*. https://doi.org/10.1017/CBO9781107415324.004

[108.] Kumar, A., Bezawada, R., Rishika, R., Janakiraman, R., & Kannan, P. K. (2015). From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior. *Journal of Marketing*. https://doi.org/10.1509/jm.14.0249

[109.] Kuo, Y.-K., & Ye, K.-D. (2009). The causal relationship between service quality, corporate image and adults' learning satisfaction and loyalty: A study of professional training programmes in a Taiwanese vocational institute. *Total Quality Management & Business Excellence*. https://doi.org/10.1080/14783360903037085

[110.] Kuziemsky, C. (2016). Decision-making in healthcare as a complex adaptive system. *Healthcare Management Forum*. https://doi.org/10.1177/0840470415614842

[111.] Lamb, M. (2004). Integrative motivation in a globalizing world. *System*. https://doi.org/10.1016/j.system.2003.04.002

[112.] Latham, A. (2001). Globalization: The Human Consequences. *Political Geography*. https://doi.org/10.1016/s0962-6298(00)00045-7

[113.] Le, T., & David Jeong, H. (2017). NLP-Based Approach to Semantic Classification of Heterogeneous Transportation Asset Data Terminology. *Journal of Computing in Civil Engineering*. https://doi.org/10.1061/(asce)cp.1943-5487.0000701

[114.] Levin, D., & Smith, J. (1994). Equilibrium in auctions with entry. *American Economic Review*. https://doi.org/10.2307/2118069

[115.] Levin, S. A. (1998). Ecosystems and the biosphere as complex adaptive systems. *Ecosystems*. https://doi.org/10.1007/s100219900037

[116.] Li, X., & Clerc, M. (2019). Swarm Intelligence. In *International Series in Operations Research and Management Science*. https://doi.org/10.1007/978-3-319-91086-4_11

[117.] Li, Y., Zhou, X., Bruza, P., Xu, Y., & Lau, R. Y. K. (2008). A two-stage text mining model for information filtering. *International Conference on Information and Knowledge Management, Proceedings*. https://doi.org/10.1145/1458082.1458218

[118.] Lightbody, B. (2005). The cold war. In *The Cold War*. https://doi.org/10.4324/9780203979143

[119.] Lipman, V. (2014). Top Twitter Trends: What Countries Are Most Active? Who's Most Popular? *Forbes*. Retrieved from http://www.forbes.com/sites/victorlipman/2014/05/24/top-twitter-trends-what-countries-are-most-active-whos-most-popular/

[120.] Liu, Y. Y., & Barabási, A. L. (2016). Control principles of complex systems. *Reviews of Modern Physics*. https://doi.org/10.1103/RevModPhys.88.035006

[121.] Liu, Y. Y., Slotine, J. J., & Barabási, A. L. (2013). Observability of complex systems. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1215508110

[122.] Lupton, D. (2014). Digital Sociology. In *Digital Sociology*. https://doi.org/10.4324/9781315776880

[123.] Lyapunov, A. M. (1992). The general problem of the stability of motion. *International Journal of Control*. https://doi.org/10.1080/00207179208934253

[124.] MacCarthy, B. L., & Atthirawong, W. (2003). Factors affecting location decisions in international operations - A Delphi study. *International Journal of Operations and Production Management*. https://doi.org/10.1108/01443570310481568

[125.] Machowski, H. A. (2019). The soviet union. In *Economic Warfare or Detente: An Assessment of East-west Economic Relations in the 1980s*. https://doi.org/10.4324/9780429035197-19

[126.] Mael, F., & Ashforth, B. E. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of Organizational Behavior*. https://doi.org/10.1002/job.4030130202

[127.] Malina, R. M. (2010). Early sport specialization: Roots, effectiveness, risks. *Current Sports Medicine Reports*. https://doi.org/10.1249/JSR.0b013e3181fe3166

[128.] Mancuso, J., & Stuth, K. (2013). Social media and the second screen. *Marketing Research*, Vol. 25, pp. 18–19. Retrieved from https://www.researchgate.net/publication/283986324_Social_Media_and_the_Second_Screen

[129.] Mander, J., & McGrath, F. (2015). *Premier League Fans summary*. 1–8. Retrieved from https://thisisfootballx.files.wordpress.com/2015/03/summarypremierleaguefansq12015.pdf

[130.] McClung, S., Eveland, V., Sweeney, D., & James, J. D. (2012). Role of the Internet Site

in the Promotion Management of Sports Teams and Franchise Brands. *Journal of Promotion Management*. https://doi.org/10.1080/10496491.2012.668429

[131.] McDaniel Jr., R. R., & Driebe, D. J. (2005). Uncertainty and Surprise in Complex Systems: Questions on Working with the Unexpected. In *Media*. https://doi.org/10.1007/b13122

[132.] McKinsey Global Institute. (2019). Globalization in transition: the future of trade and value chains. *McKinsey & Company*.

[133.] Meadows-Klue, D. (2004). The Tipping Point: How Little Things Can Make a Big Difference. *Interactive Marketing*. https://doi.org/10.1057/palgrave.im.4340262

[134.] Memmert, D., Lemmink, K. A. P. M., & Sampaio, J. (2017). Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine*. https://doi.org/10.1007/s40279-016-0562-5

[135.] Mishne, G., & De Rijke, M. (2006). Capturing global mood levels using blog posts. *AAAI Spring Symposium - Technical Report*, *SS-06-03*, 145–152. Retrieved from https://www.aaai.org/Papers/Symposia/Spring/2006/SS-06-03/SS06-03-028.pdf

[136.] Mosley, L., & Friedman, T. L. (2006). The World Is Flat: A Brief History of the Twenty-First Century. *International Journal*. https://doi.org/10.2307/40204208

[137.] Nacar, R., & Uray, N. (2016). The challenge of international market segmentation in emerging markets. In *Handbook of Research on Impacts of International Business and Political Affairs on the Global Economy*. https://doi.org/10.4018/978-1-4666-9806-2.ch003

[138.] Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: An introduction. *Journal of the American Medical Informatics Association*. https://doi.org/10.1136/amiajnl-2011-000464

[139.] Nguyen, N. T. (2018). Lyapunov stability theory. In *Advanced Textbooks in Control and Signal Processing*. https://doi.org/10.1007/978-3-319-56393-0_4

[140.] Nielsen, C. G., Storm, R. K., & Jakobsen, T. G. (2019). The impact of English Premier League broadcasts on Danish spectator demand: a small league perspective. *Journal of Business Economics*. https://doi.org/10.1007/s11573-019-00932-7

[141.] Nofsinger, J. R. (2005). Social Mood and Financial Economics. *Journal of Behavioral Finance*. https://doi.org/10.1207/s15427579jpfm0603_4

[142.] O'Daniel, M., & Rosenstein, A. H. (2008). Professional Communication and Team

Collaboration. In *Patient Safety and Quality: An Evidence-Based Handbook for Nurses*.

[143.] Oliveira, M., Bastos-Filho, C., & Menezes, R. (2017). The scaling of crime concentration in cities. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0183110

[144.] P. G., M. (2018). Factors Influencing the Smartphone Addiction Among Students of the North Central Province in Sri Lanka. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2794735

[145.] Pan, W., Liu, N. N., Xiang, E. W., & Yang, Q. (2011). Transfer learning to predict missing ratings via heterogeneous user feedbacks. *IJCAI International Joint Conference on Artificial Intelligence*. https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-386

[146.] Park, J. A., & Dittmore Stephen, W. (2014). The relationship among social media consumption, Team identification, And behavioral intentions. *Journal of Physical Education and Sport*. https://doi.org/10.7752/jpes.2014.03050

[147.] Passos, P., Davids, K., Araújo, D., Paz, N., Minguéns, J., & Mendes, J. (2011). Networks as a novel tool for studying team ball sports as complex social systems. *Journal of Science and Medicine in Sport*. https://doi.org/10.1016/j.jsams.2010.10.459

[148.] Pellegrini, A. (2010). The Role of Play in Human Development. In *The Role of Play in Human Development*. https://doi.org/10.1093/acprof:oso/9780195367324.001.0001

[149.] Pomerening, J. R., Sun, Y. K., & Ferrell, J. E. (2005). Systems-level dissection of the cell-cycle oscillator: Bypassing positive feedback produces damped oscillations. *Cell*. https://doi.org/10.1016/j.cell.2005.06.016

[150.] Popescu, M. A. (2014). Community of singularities: the possibility of being-with in the work of Heidegger, Lévinas and Derrida.

[151.] Preiser, R. (2019). Identifying general trends and patterns in complex systems research: An overview of theoretical and practical implications. *Systems Research and Behavioral Science*. https://doi.org/10.1002/sres.2619

[152.] Pretty, J., & Ward, H. (2001). Social capital and the environment. *World Development*. https://doi.org/10.1016/S0305-750X(00)00098-X

[153.] Price, J., Farrington, N., & Hall, L. (2013). Changing the game? The impact of Twitter on relationships between football clubs, supporters and the sports media. *Soccer and Society*. https://doi.org/10.1080/14660970.2013.810431

[154.] Radzicki, M. J. (2009). System Dynamics and Its Contribution to Economics and Economic Modeling. In *Complex Systems in Finance and Econometrics*.

https://doi.org/10.1007/978-1-4419-7701-4_39

[155.] Rai, B., Kasturi, M., & Huang, C. yu. (2018). Analyzing stock market movements using news, tweets, stock prices and transactions volume data for APPLE (AAPL), GOOGLE (GOOG) and SONY (SNE). *ACM International Conference Proceeding Series*, *D*, 109–112. https://doi.org/10.1145/3243250.3243263

[156.] Rajasekhar, A., Lynn, N., Das, S., & Suganthan, P. N. (2017). Computing with the collective intelligence of honey bees – A survey. *Swarm and Evolutionary Computation*. https://doi.org/10.1016/j.swevo.2016.06.001

[157.] Ramos-Villagrasa, P. J., Marques-Quinteiro, P., Navarro, J., & Rico, R. (2018). Teams as Complex Adaptive Systems: Reviewing 17 Years of Research. *Small Group Research*. https://doi.org/10.1177/1046496417713849

[158.] Rasmussen, S., Mosekilde, E., & Sterman, J. D. (1985). Bifurcations and chaotic behavior in a simple model of the economic long wave. *System Dynamics Review*. https://doi.org/10.1002/sdr.4260010108

[159.] Reddy, S., Sharoff, S., Collobert, R., Weston, J., Bottou, L., Karlen, M., … Sangal, R. (2015). Deep Learning for NLP (without magic). *Arxiv*. https://doi.org/10.1017/CBO9781107415324.004

[160.] Renneboog, L., & Vanbrabant, P. (2000). Share Price Reactions To Sporty Performances Of Soccer Clubs Listed On The London Stock Exchange And The AIM. In *CentER Discussion Paper*. Retrieved from https://pure.uvt.nl/portal/files/534671/19.pdf

[161.] Roberts, C., & Emmons, B. (2016). Twitter in the Press Box: How a New Technology Affects Game-Day Routines of Print-Focused Sports Journalists. *International Journal of Sport Communication*. https://doi.org/10.1123/ijsc.2015-0113

[162.] Rodríguez-Andrés, R. (2018). Trump 2016: ¿president elected thanks to social media? *Palabra Clave*. https://doi.org/10.5294/pacla.2018.21.3.8

[163.] Roshanaei, M., & Mishra, S. (2015). Studying the attributes of users in Twitter considering their emotional states. *Social Network Analysis and Mining*. https://doi.org/10.1007/s13278-015-0278-9

[164.] Roubelat, F. (2000). Scenario Planning as a Networking Process. *Technological Forecasting and Social Change*. https://doi.org/10.1016/S0040-1625(99)00125-0

[165.] Roy, R. B., & Sarkar, U. K. (2010). Capturing Early Warning Signal for Financial Crisis

from the Dynamics of Stock Market Networks: Evidence from North American and Asian Stock Markets. *Society for Computational Economics 16th International Conference on Computing in Economics and Finance, London, UK*, 1–15. Retrieved from https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CEF2010&paper_id=89

[166.] Rui, H., Liu, Y., & Whinston, A. (2013). Whose and what chatter matters? the effect of tweets on movie sales. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2012.12.022

[167.] Salati, E., & Vose, P. B. (1984). Amazon Basin: A system in equilibrium. *Science*. https://doi.org/10.1126/science.225.4658.129

[168.] Salganik., M. J. (2017). *Bit by Bit: Social Research in the Digital Age*. Retrieved from princeton: Princeton University Press, 2017. 423 pp. ISBN: 978-0-691-15864-8

[169.] Sanchez-Segura, M. I., Hadzikadic, M., Dugarte-Peña, G. L., & Medina-Dominguez, F. (2018). Team Formation Using a Systems Thinking Approach. *Systems Research and Behavioral Science*. https://doi.org/10.1002/sres.2536

[170.] Sanderson J. (2011). *How social media is changing sports: It's a whole new ballgame* (Sport Mana). https://doi.org/10.1016/j.smr.2012.02.008

[171.] Santos, O. A. dos. (2014). Urban common space, heterotopia and the right to the city: Reflections on the ideas of Henri Lefebvre and David Harvey. *URBE - Revista Brasileira de Gestão Urbana*. https://doi.org/10.7213/urbe.06.002.se02

[172.] Santos, I., Penya, Y. K., Devesa, J., & Bringas, P. G. (2015). *N-GRAMS-BASED FILE SIGNATURES FOR MALWARE DETECTION*. https://doi.org/10.5220/0001863603170320

[173.] Sardar, Z. (1994). Conquests, chaos and complexity. The Other in modern and postmodern science. *Futures*. https://doi.org/10.1016/0016-3287(94)90036-1

[174.] Sasou, K., & Reason, J. (1999). Team errors: Definition and taxonomy. *Reliability Engineering and System Safety*. https://doi.org/10.1016/S0951-8320(98)00074-X

[175.] Sauvé, P. (2003). Services. In *Regionalism and the Multilateral Trading System*. https://doi.org/10.1787/9789264101371-en

[176.] Scheffer, M. (2010). Foreseeing tipping points. *Nature*. https://doi.org/10.1038/467411a

[177.] Schläpfer, M., Bettencourt, L. M. A., Grauwin, S., Raschke, M., Claxton, R., Smoreda,

Z., … Ratti, C. (2014). The scaling of human interactions with city size. *Journal of the Royal Society Interface*. https://doi.org/10.1098/rsif.2013.0789

[178.] Scholtens, B., & Peenstra, W. (2010). *Scoring on the stock exchange ? The effect of football matches on stock market returns : an event study Scoring on the stock exchange ? The effect of football matches on stock market returns : an event study*. *6846*. https://doi.org/10.1080/00036840701721406

[179.] Schredelseker, K., & Fidahic, F. (2016). Stock Market Reactions and Formula One Performance. *Journal of Sport Management*. https://doi.org/10.1123/jsm.25.4.305

[180.] Schumaker, R., & Chen, H. (2006). Textual Analysis of Stock Market Prediction Using Financial News Articles. *Proceedings of AMCIS 2006*. https://doi.org/10.1145/1462198.1462204

[181.] Schumpeter, J. (2006). The Process of Creative Destruction. In *Capitalism, Socialism and Democracy*. https://doi.org/10.4324/9780203202050.ch7

[182.] Schuster, H. G. (2005). Complex adaptive systems. In *Collective Dynamics of Nonlinear and Disordered Systems*. https://doi.org/10.1007/3-540-26869-3_16

[183.] Shearer, C., Watson, H. J., Grecich, D. G., Moss, L., Adelman, S., Hammer, K., & Herdlein, S. a. (2000). The CRISP-DM model: The New Blueprint for Data Mining. *Journal of Data Warehousing*.

[184.] Simon, H. A. (1977). *The Organization of Complex Systems*. https://doi.org/10.1007/978-94-010-9521-1_14

[185.] Skaff, J. (2001). Genghis Khan: Conqueror of the World, by Leo De Hartog. 230 pages, maps, endnotes, bibliography, index. London, New York: I. B. Taurus, 1999. $16.95 (Paperback) ISBN 1-86064-375-2. *Middle East Studies Association Bulletin*. https://doi.org/10.1017/s0026318400041808

[186.] Song, Y., & He, L. (2010). *Optimal rare query suggestion with implicit user feedback*. https://doi.org/10.1145/1772690.1772782

[187.] Soti, A. (2019). Understanding Micro and Macro Interactions in Social Neurobiological Systems. *Interdisciplinary Description of Complex Systems*. https://doi.org/10.7906/indecs.17.4.1

[188.] Sóti, A. (2020). A Python programozási nyelvről statisztikusoknak. *Statisztikai Szemle*, *4*(1). https://doi.org/10.20311/stat2020.4.hu0324

[189.] Sóti, A., Bokányi, E., & Vattay, G. (2018). Urban scaling of football followership on

Twitter. *Acta Polytechnica Hungarica*, *15*(5). https://doi.org/10.12700/APH.15.5.2018.5.14

[190.] Spring, J. (2008). Research on globalization and education. *Review of Educational Research*. https://doi.org/10.3102/0034654308317846

[191.] Steels, L. (2000). Language as a complex adaptive system. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

[192.] Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*. https://doi.org/10.1007/s13278-012-0079-3

[193.] Strahinja, R., Golob, M., & Subašić, T. (2017). Sportski marketing u Hrvatskom nogometnom klubu Rijeka. *Zbornik Veleučilišta u Rijeci*. https://doi.org/10.31784/zvr.5.1.6

[194.] Strumsky, D., & Lobo, J. (2015). Identifying the sources of technological novelty in the process of invention. *Research Policy*. https://doi.org/10.1016/j.respol.2015.05.008

[195.] Sunthonkanokpong, W. (2011). Future global visions of engineering education. *Procedia Engineering*. https://doi.org/10.1016/j.proeng.2011.03.029

[196.] Székely, G. J., & Rizzo, M. L. (2009). Brownian distance covariance. *Annals of Applied Statistics*. https://doi.org/10.1214/09-AOAS312

[197.] Székely, G. J., & Rizzo, M. L. (2013). The distance correlation t -test of independence in high dimension. *Journal of Multivariate Analysis*. https://doi.org/10.1016/j.jmva.2013.02.012

[198.] The Cambridge Economic History of India. (1983). In *The Cambridge Economic History of India*. https://doi.org/10.1017/chol9780521228022

[199.] Tong, H. K., & Cheung, L. H. (2011). Cultural identity and language: A proposed framework for cultural globalisation and glocalisation. *Journal of Multilingual and Multicultural Development*. https://doi.org/10.1080/01434632.2010.527344

[200.] Topal, K., Koyutürk, M., & Özsoyoğlu, G. (2017). Effects of emotion and topic area on topic shifts in social media discussions. *Social Network Analysis and Mining*. https://doi.org/10.1007/s13278-017-0465-y

[201.] Torrents, C., & Balagué, N. (2006). DYNAMIC SYSTEMS THEORY AND SPORTS TRAINING. *Baltic Journal of Sport and Health Sciences*.

https://doi.org/10.33607/bjshs.v1i60.609

[202.] Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*. https://doi.org/10.1613/jair.2934

[203.] Tušak, M. (1997). Razvoj motivacijskega sistema v športu. *Psihološka Obzorja*.

[204.] van der Lans, R., van Bruggen, G., Eliashberg, J., & Wierenga, B. (2009). A Viral Branching Model for Predicting the Spread of Electronic Word of Mouth. *Marketing Science*. https://doi.org/10.1287/mksc.1090.0520

[205.] Vega, H. (2015). Review of Systems Thinking for Social Change: A Practical Guide to Solving Complex Problems, Avoiding Unintended Consequences, and Achieving Lasting Results. *The Foundation Review*. https://doi.org/10.9707/1944-5660.1258

[206.] Vicentini, F., & Graziano, E. A. (2016). Football cultural events and stock market returns: the case of FIFA WORLD CUP. *International Journal of Environmental Policy and Decision Making*, *1*(1), 1. https://doi.org/10.1504/ijepdm.2016.10001600

[207.] Vicsek, T. (2002). Complexity: The bigger picture. *Nature*. https://doi.org/10.1038/418131a

[208.] Vivek, S. D., Beatty, S. E., & Morgan, R. M. (2012). Customer Engagement: Exploring Customer Relationships Beyond Purchase. *Journal of Marketing Theory and Practice*, *20*(2), 122–146. https://doi.org/10.2753/mtp1069-6679200201

[209.] Wade, R. H. (2004). Is globalization reducing poverty and inequality? *World Development*. https://doi.org/10.1016/j.worlddev.2003.10.007

[210.] Wang, C., Cheng, Z., Yue, X.-G., & McAleer, M. (2020). Risk Management of COVID-19 by Universities in China. *Journal of Risk and Financial Management*. https://doi.org/10.3390/jrfm13020036

[211.] Wang, Xiaolong, Farhadi, A., & Gupta, A. (2016). Actions ~ Transformations. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2016.291

[212.] Wang, Xuerui, McCallum, A., & Wei, X. (2007). Topical N-grams: Phrase and topic discovery, with an application to information retrieval. *Proceedings - IEEE International Conference on Data Mining, ICDM*. https://doi.org/10.1109/ICDM.2007.86

[213.] Weaver, Waren. (1948). Science and complexity. *American Scientist*.

https://doi.org/10.1007/978-1-4899-0718-9_30

[214.] Weaver, Warren. (1949). Recent contributions to the mathematical theory of communication. In *The mathematical theory of communication*.

[215.] Weber, M. (2005). The protestant ethic and the spirit of capitalism. In *The Protestant Ethic and the Spirit of Capitalism*. https://doi.org/10.4324/9780203995808

[216.] Weimar, D., & Schauberger, M. (2018). The impact of sporting success on student enrollment. *Journal of Business Economics*. https://doi.org/10.1007/s11573-017-0877-1

[217.] Williamson, O. E. (1993). Calculativeness, Trust, and Economic Organization. *The Journal of Law and Economics*. https://doi.org/10.1086/467284

[218.] Woods, N. (2014). The globalizers: The IMF, the world bank, and their borrowers. In *The Globalizers: The IMF, the World Bank, and their Borrowers*. https://doi.org/10.1080/03056240903086360

[219.] Wu, Y., Li, H., Gou, Q., & Gu, J. (2017). Supply chain models with corporate social responsibility. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2017.1346833

[220.] Yakubo, K., Saijo, Y., & Korošak, D. (2014). Superlinear and sublinear urban scaling in geographical networks modeling cities. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. https://doi.org/10.1103/PhysRevE.90.022803

[221.] Yang, Y., Wang, Y., & Billings, A. C. (2016). Online Chinese discussions about the 2014 World Cup. *Online Information Review*, *40*(6), 834–848. https://doi.org/10.1108/OIR-01-2016-0031

[222.] Yen, Y. Y., & Yen, W. T. M. (2016). Knowledge transfer of foreign and local employees in multinational companies. *International Business Management*.

[223.] Yessenov, K., Tulsiani, S., Menon, A., Miller, R. C., Gulwani, S., Lampson, B., & Kalai, A. (2013). A colorful approach to text processing by example. *UIST 2013 - Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*. https://doi.org/10.1145/2501988.2502040

[224.] Yu, D., Li, D. F., Merigó, J. M., & Fang, L. (2016). Mapping development of linguistic decision making studies. *Journal of Intelligent and Fuzzy Systems*. https://doi.org/10.3233/IFS-152026

[225.] Zadeh, L. A. (2013). Fuzzy logic. In *Computational Complexity: Theory, Techniques,*

and Applications. https://doi.org/10.1007/978-1-4614-1800-9_73

[226.] Zhang, N., Campo, S., Janz, K. F., Eckler, P., Yang, J., Snetselaar, L. G., & Signorini, A. (2013). Electronic word of mouth on twitter about physical activity in the United States: exploratory infodemiology study. *Journal of Medical Internet Research*. https://doi.org/10.2196/jmir.2870

[227.] Zuber, R. A., Yiu, P., Lamb, R. P., & Gandar, J. M. (2005). Investor-fans? An examination of the performance of publicly traded English Premier League teams. *Applied Financial Economics*. https://doi.org/10.1080/0960310042000338713