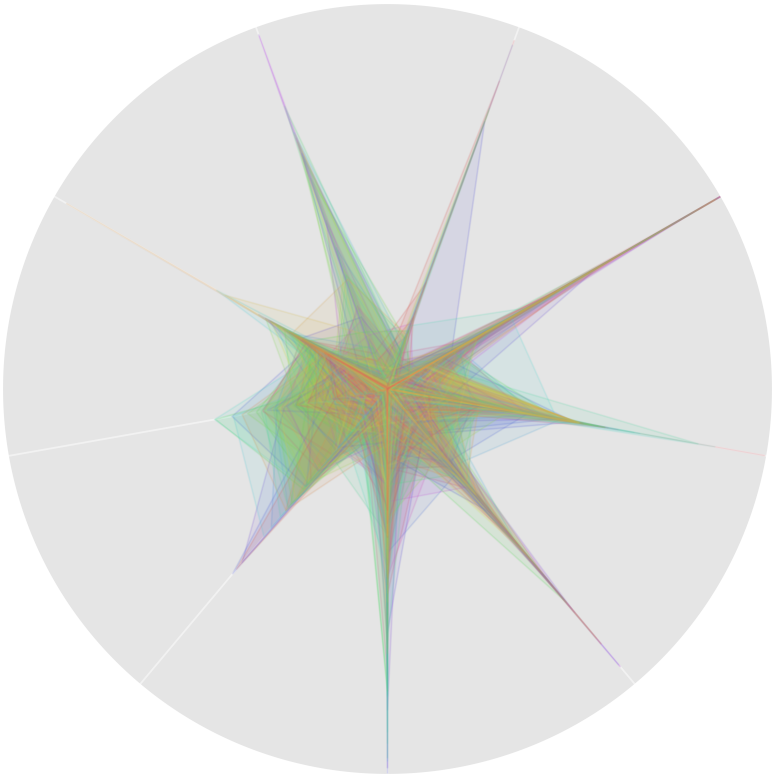
FACULTY OF SCIENCES

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

STATISTICAL PHYSICS OF SOCIAL-ECONOMIC SYSTEMS IN REAL AND VIRTUAL WORLDS



GHENT UNIVERSITY

Andrés M. Belaza Vallejo

FACULTY OF SCIENCES

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

STATISTICAL PHYSICS OF SOCIAL-ECONOMIC SYSTEMS IN REAL AND VIRTUAL WORLDS

Supervisor: Prof. Dr. Jan Ryckebusch
Co-supervisor: Prof. Dr. Koen Schoors

Dissertation submitted in fulfillment
of the requirements for the degree of:
Doctor of Science: physics and applied economic sciences

Ghent University
Faculty of Sciences:
Department of Physics and Astronomy
Faculty of Economics and Business Administration:
Department of General Economics
Academic year 2019-2020

The figure on the cover corresponds to the activity profiles of 372 players for a year in EVE-Online as calculated in chapter 7.

GHENT UNIVERSITY

Andrés M. Belaza Vallejo

Acknowledgments

Tl;dr, Thanks to all the people who help me to reach where I am now.

A bit more than four years ago, I was a young graduated student from Spain who never left their home, taking a flight to a new adventure in Belgium. Now, I am a bit more experience in science and life. It has been a long trip, and it would not have been possible without many people that, directly or indirectly, support me in this adventure.

My First thanks are for my supervisors: Jan and Koen. Thanks for giving me this opportunity to do a Ph.D. under your supervision. I know that I could be a bit stubborn and chaotic, so thanks for your patience and guidance. Jan, without you, this work would not have been possible. Thanks for helping me with your expertise in physics and humanity. And thanks to both for your fabulous ideas, I wish I would have had the time to apply them all.

Another person that was crucial for my being able to be able to finish my Ph.D. has been Benjamin. You help me a lot with their corrections and comment. And also, the rest of the LABM group: Milan, Aaron, Ken, Kevin. Thanks for your patience, your ideas, and I would have liked to have had more opportunities to work on more projects together. To Olivier and Corneel, I learned from you more than you from me. Thanks also to the members of the jury for their suggestions to improve the text and to bring many exciting ideas for discussion.

Raúl, gracias. Gracias por guiarme en las frías calles de Bélgica. Por enseñarme los trucos y recodos de la Universidad, la ciudad y de la vida. No hubiera sobrevivido sin tu ayuda. To Alexis, Kajetan, Martina, thanks for those fantastic times out of the university that give me a reason to move forward when life looks like it is going against me. Also, thanks to Marem and Prem, for many nights of board games and dinner, so necessary to recharge my battery. And thanks for letting me be part of your circle and be the godfather of one of your lovely daughters. I hope I could be an excellent example for her. Paula, gracias por tu paciencia y tus croquetas, pero que tu estancia en Gante fuera fantástica ;). To Nils, thanks for being so friendly and helpful (and thanks for the help with the language), I wish you the best in your new future. También a Laura, a la que te deseo el mejor de los comienzos con el doctorado y tu vida en Bélgica.

The INW looks at my home sometimes more than my house. That was also thanks to amazing people that always were there, like Alexis, Alessio, Sam, Christos, Ivan, Jannes, Ward, Ianthe, and the rest. Thanks.

I also would like to appreciate the fantastic time I have at Roosakker. Thanks, Nele, Tine, Shandra, and Javi, for your patience. I know I was not always the more social of the housemates, thanks.

I would not be in Ghent without my amazing times in Zaragoza. Gracias, Iñigo, Ainara, Andrea, Héctor, Rocío, Camón, Jorge, Juancho, Pablo y Aleta por hace la Universidad mas llevadera. Y por los chupitos en el ambiente. Aleta, ojalá hubiéramos podido trabajar mas juntos, gracias. +Profesors A Nora, por enseñarme a ver la vida con otra óptica, gracias por ayudarme a abrir los ojos. A Carlos Miguel: nos embarcamos a la vez en esto, pero sabia que terminarías antes, eres genial. También a Amelia y Rocio por nuestras meriendas durante la uni. Al grupo del instituto, Vera, Raúl, Cris, Xabi, os he perdido un poco la pista, pero esto también es gracias a vosotros. A mi grupo de rol, Aitor, Ros, Julio, Jesús, Miki, Taleb, gracias por ayudarme a desconectar y adaptaros para que pudiera seguir jugando con vosotros. A Luis, por ser una de las columnas que me ayudo a sobrevivir en Bélgica al principio: volveremos al Karaoke. A mi twitter-mundo, Amanda, Javi, Chus, Vir. Gracias por aguantar mis desvaríos y mis dudas estos años. Y por presentarme a Alba, aun os debo una ;). Alba, gracias por dejarme compartir mi vida contigo y por venirte por mi a la aventura. A todos aquellos que han gastado su tiempo y esfuerzo en venir a verme aquí: Edu, David, Helena, Juma... Gracias por ayudarme a salir de la rutina. A mi cuadrilla de Los Arcos, Ainara, Haritz, Cris, Diego, Lorena, Bea, Migue, Paula, Iosu... Gracias por acogerme cada vez que iba, aunque hubieran pasado 10 meses desde la ultima vez.

Thanks to my Family. A mis Padres por estar siempre allí apoyándome, pero dejándome tomar mis propias decisiones. A mi hermano Fer por siempre estar ahí para escucharme. Fer, te admiro mucho por todo lo que haces. A los Belaza, por estar ahi en lo bueno y en lo malo. A mis primos Vallejo, sobretodo a Fran y a Manu que me sirven de ejemplo a seguir. A todos los Vallejo Ramos en general, por ser una fuente de risa y alegría que me ayuda a llevar el frio.

Thanks to everyone that help me to reach where I am, even If I forgot to mention you. Thanks.

Gent, January 2020
Andres Maria Belaza Vallejo

Table of Contents

1 Introduction

The work presented in this thesis can be subdivided into two parts that are connected by the source of the data that they are based on. The first part deals with the development and application of a Statistical-Physics inspired model for balance in signed social networks, thereby focusing on political networks as for example encountered in international relationship networks between countries. The second part contains a series of studies of some specific economic and social aspects of a Virtual World, from a more statistical perspective. The different chapters in both parts address a variety of topics and methodologies. The common thread of these chapters is the use of detailed data from virtual worlds in combination with data from the real world.

In the forthcoming section of this introduction, we present the background to frame this work. In section 1.2, we outline the structure and content of the chapters. Finally, in section 1.3, we sketch the software and hardware tools that have been used in order to obtain the outcomes detailed in this thesis.

1.1 Introduction

1.1.1 Physics in Social and Economical Systems

In modern-day science, there is a consensus that many laws of nature arise from local interactions, thanks to statistics. As the laws of thermodynamics

are macro laws that emerge from the interactions of the "smaller" parts in the system, the idea to describe global behavior starting from statistical models for micro-interactions have been around for a long time [1, 2]. The success of Statistical physics in explaining the origin of these laws was the reason for many physicists to try the application in interdisciplinary fields like the description of selected social [3–5] and economic phenomena [6–8]. For these phenomena, the fundamental constituents are human or groups of humans, and they typically interact with a given set of rules with a small number of other constituents. Even humans are more complicated than gas molecules; human societies present regularities as time cycles in transport [9–11], order/disorder phases for segregation [12], culture [13] or language [14, 15]; and city structure and size [16, 17].

The last decades have seen the availability of new databases and of hardware/software combinations to properly study them. In addition the appearance of a new social paradigm thanks to the internet, has boosted the interest of physicists for the social sciences. The available computational power allows one to simulate large systems with many interacting agents and makes the systematic study of these new models possible like in [18, 19]. A conceptual difficulty emerges, however, when the methodology of statistical physics approaches social dynamics. Statistical physics usually deals with components like atoms or molecules, whose properties are reasonably well known. Many macroscopic phenomena represent emergent effects of the interaction of a large number of these components. Humans are more much complex and the individual behavior of each person is complicated. It is hard to predict the action of an individual, so substantial simplifications are usually required to develop these models as in [12, 20].

Statistical physics has been applied to many topics of interest in the social sciences [4, 5, 21],"sociophysics" or social physics. Castellano et al. in [4] expose a deep review of models based on statistical physics for social systems, from opinion dynamics, over crowd behavior to language dynamics. Barthelemy in [5] focuses on dynamics and presents a variety of models inspired by statistical physics. Barthelemy presents models for spacial organization to city mobility. Connected with the topic of mobility, Galleoti in [21] presents a study of human mobility thereby making use of GPS data for car movement in Italy.

Economics is another field where methods of statistical physics have been applied. One often refers to this field as "Econophysics"- roughly defined as a research field in which models and methodologies gleaned from physics are used to study selected topics in economics. Examples of works that are based on this perspective are [6, 7, 22]. In their work on complex systems, Stanley et al. [6, 7] build on the idea that nontrivial long-range correlations

can be used to explain the behavior of the measures that emerge from large groups of humans including for example the common economic indexes. In [22] Sornette reviewed the connection between Physics and Economics.

As an example of a system that can be modeled using statistical physics is the distribution of money in a closed system. In "Statistical mechanics of money" [23], Drăgulescu and Yakovenko propose a system where agents interchange random amounts of money. Similar to the energy of molecules of a gas, the stationary probability distribution of the money resembles a Boltzmann-Gibbs law. In [24] one finds evidence for this model in the income distribution in the USA. In [20], Yakovenko digs into the differences of money, wealth, and energy in the model. Yakovenko concludes that they behave similarly, and that conservation laws are at work in both physical and monetary systems.

1.1.2 Networks in Social and Economical Systems

A characteristic of many socio-economic systems (among others) is that they are composed of many interacting agents. Three possible viewpoints are possible. Firstly, one can focus on the study of a particular agents, for example the individual behavior of humans. Secondly one can study how agents interact with each other, for example what kind of language is used for communication purposes. Thirdly the focus of the study can be on the patterns that emerge from the connections. These efforts are often conducted within the context of “Network science” [25]. A network is a collection of points (nodes or vertices) connected with lines (edges or links). The use of statistical physics in networks has been tested [26,27]. For example, in [27] Cimini et al., present a framework for modeling networks using the rewiring algorithm based on entropy and likelihood maximization. They propose this framework to perform network reconstruction from incomplete data and to determine whether patterns found in networks are statistically significant.

Many fields use network science to reveal essential pieces of information that would be hidden otherwise. The use of transport networks plays an essential role in global disease propagation [28, 29]. For diseases [29] to rumors and ideas [30,31], the topology of the network on which the diffusion takes place is key [32]. Other studies for which networks have proven their usefulness are studies of the stability vulnerabilities of electrical grids [33] and of the human brain [34]. Thereby, the structure of the network can help to pinpoint the points where the propagation of the undesired effects can be stopped.

In the field of social science, network theory has become a vital resource [35] for example to test hypotheses and address highly relevant ques-

tions. For example, how connected are human societies. How many connections are (on average) between two randomly chosen persons. An old experiment has set this number around six [36]. The use of social networks, such as Facebook or Twitter, has allowed one to test this hypothesis [37]. This kind of networks are called "small worlds" and are prevalent in the nature [38, 39]. Even with a more connected society thanks to technology, it is suggested that the number of real connections is limited [40, 41]. The Dundar's number [40] has been suggested as a cognitive limit to the number of people with which a person can maintain stable social interactions. Dunbar proposes that this number lies in the range 100-250, and later works support this hypothesis [41].

Another example of the application of network science to social systems is the study of international relationships [42–45]. Thereby the countries play the role of nodes, and between them, different kind of edges get established. These edges can refer to alliances, to trade, to war and other things. Galam in [43] develops a model inspired by spin glasses. There are historical connections between countries that vary in strength and sign (friendly or negative). Each country can choose to align itself to one alliance or another thereby seeking to maximize the number of positive interactions inside the alliance and minimize negative interactions between alliances. This model has been extended by Vinogradova in [46]. Among others, one of the more common theories that have been applied to the international network is Social Balance theory.

1.1.2.1 Social Balance Theory

Balance theory (also called "structural balance theory") was initially proposed by Heider [47] as an explanation for attitude changes in psychology. It was generalized for networks by Cartwright and Harary [48]. Balance theory has been refined and applied to a variety of scenarios but in this thesis the focus is on its application to international relationships [44, 45]. In balance theory, three agents remain in a balanced situation if the product of their connection is positive. As a result "a friend of a friend is also a friend" and "an enemy of my enemy is my friend" represent the balanced situation. The idea is that if the system is unbalanced, the agents in the network will modify the status of their connections so as to achieve a more balanced status. In the limit, the system reaches full balance when it contains only positive links, or when it is divided into two adversarial groups. Different ways to measure balance have been proposed [49, 50].

Some works have pointed out that not all systems appear to be evolving toward the balanced state [44, 51]. It has been suggested that other forces [52] could inhibit the fully balanced state as "homophily" (similar

nodes prefer to have a positive connection) and "heterophobia" (nodes with different properties prefer to be enemies).

With the help of statistical physics, Marvel et al. in [51] propose an interesting approximation for social balance. They reinterpret social balance theory in terms of an energy function, thereby associating the existence of unbalanced status to the occurrence of an excitation energy. With this transformation, they investigated the energy landscape and found the occurrence of so-called jammed states: network configurations with a finite number of unbalanced triads that cannot evolve to a fully balanced situation. Another thermodynamic formulation of balance theory inspired by the Ising model is introduced in [53] by Lee and collaborators.

1.1.3 Virtual worlds as a source of Data

In 2005, an error in the code of the most popular virtual world triggered a virtual pandemic [54]. In World of Warcraft [55], the players can control a virtual avatar to fight each other and challenges placed by the developers. The death of these avatars is not permanent, but is very annoying as it prevents the players to continue for a few minutes. An error in the code created annoying effects namely it infected the characters in the game and made the avatars to die slowly. The mere fact of two characters or avatars moving in each other's proximity was transmitting an infection, and for a week, the virtual world became pandemic. This effect quickly killed lower-level characters and drastically changed normal game play, as players did what they could do to avoid infection. The facts that players were abandoning densely populated cities or even just not playing the game for the week, caught the attention of epidemiologists [56,57]. In the last decades, we have seen a rise in the study of these virtual worlds as a highly controlled environment for testing human behavior [58,59].

A virtual world that has been studied profoundly [58,60] is Pardus [61]. In Pardus, the players compete for space in a futuristic universe where they can interact in many ways, from fighting to trade. They start with a low-level spacecraft, and over time they increase their wealth, rank and skills in the game. Optionally, they can join factions, develop trade routes, and construct building and starbases to produce other items. They can collaborate in this process, or they can attack each other to steal eachother's wealth. There are three different servers, or "universes", on which a player can create a single character per universe. Using both the friendship and enmity networks from Pardus, researchers have found strong evidence of social balance theory [58].

The use of virtual worlds as a source of data presents some special characteristics. The sample present in those virtual worlds could show age

and gender bias. Those biases, however, can also be present in real-world datasets. For example, data extracted from cell-phone or public-transport usage have a certain bias with regard to age and/or economic-status distributions. Also, one could think that the motivations and risks of players operating in virtual worlds are different from those of humans in the real world. For these reasons, one must always keep in mind elements from the real world whilst studying virtual environments. The virtual behavior is fascinating by itself as it can reveal social dynamics, psychological profiles, and economic behavior. In its comparison with phenomena in the real world the use of virtual-word data reaches its full potential as a proxy . For this dissertation, we compare both real and virtual world data. Even with differing motivations, some studies have suggested that the virtual world and the real world can present, qualitatively, similar human patterns [56, 58]

1.1.3.1 EVE Online

A common thread in this dissertation is the use of data from a virtual world called "EVE Online" (or, in brief EVE) [62]. EVE is a Massively Multiplayer Online Role-Playing Game (MMORPG) developed by CCP Games. The game was released in 2003, and it is well known in the community for its complexity and scale in the player interactions. All EVE events occur in real-time, in the universe composed for all players. Each day the universe stops functioning for 30 minutes for maintenance. Five thousand (5000) different solar systems compose the EVE universe, some of them can be under control of a group of players (alliances). Planets, Asteroids, Stars, pirated NPC (Non-Player Characters), star-facilities among others stay in this universe even when the players are not in-game. EVE online, thanks to its sandbox qualities, is a strong candidate as a source of data. Even it is populated mainly by males (85% according to statistics of the developers), the age distribution is, on average, older than other video games and closer to the real world [63].

A crucial point for chapters 2 and 3 of this work is the study of the network of alliances. An alliance is a group of players that operates as a group. In-game, they can formalize the creation of this alliance and extend their ability to modify the universe. An alliance can go to war with another alliance, collect tax from their members, control territory, and exploit their resources. They have a variety of internal political systems: some alliances work democratically, but others are despotic and strongly hierarchical. Big alliances also have spies in other alliances to reveal the size of the resources, the production, and supply line and other secrets. Some big alliances also have a social security system that allows getting compensation if a player dies helping the alliance. In many respects, the operation of an alliance

resembles the operation of countries.

The players can spend their time doing a large variety of actions. From piracy to mining natural resources or explore the space, players interact between them or with the environment or for curiosity or looking for action or for wealth. In chapter 4, we will go more in detail about these activities. Another essential key to EVE is the production and the market. In EVE, almost every item is produced, mined, transported and traded by players. Imagine players buying in the market a new Raptor (a spaceship). First, they would go to one of the few solar systems that work as a trading hub. Here, they will face the double auction open market as they can set a ticket asking of buying the ship they want to the price or they can fulfill a selling ticket that other player set. The seller could be a producer of the ship or buy the ship from someone that can produce. In any case, probably the ship should have been transported from a site which production facility. For producing, the producer needs to have a set of skill, a blueprint, and the materials. This blueprint and material need to also buy from other players. The material can be produced items also o be row materials, extracted by another player in the asteroid and also sell in the market. The entirely player-based market and production line catch the attention of many fanatics of realism and researchers looking to study economic theories. In chapter 4 and 5 will use market data to analyze our research questions.

1.2 Outline

This Ph.D. dissertation consists of two parts. The first part introduces a new statistical-physics based model for international relationships. The second part is devoted to studies of selected socio-economic aspects in EVE online. The first part covers the model we develop (Chapters 2 and 3), which operate as a generalization of several spin models systems as Ising [64, 65] and Blue-Came model [66, 67]. This first section ends with the chapter 4 that digs into this model. The second part covers a set of works dealing with the economy in EVE Online. The chapter 5 covers the connection of in-game activity with real-world macro measures for aggression and economic data, while chapter 6 focuses on the price formation in EVE online. These sections end with chapter 7 that deals with the effect of real-world cycles in the activity and character classification. An overview of the structure of this work is displayed in Fig 1.1.

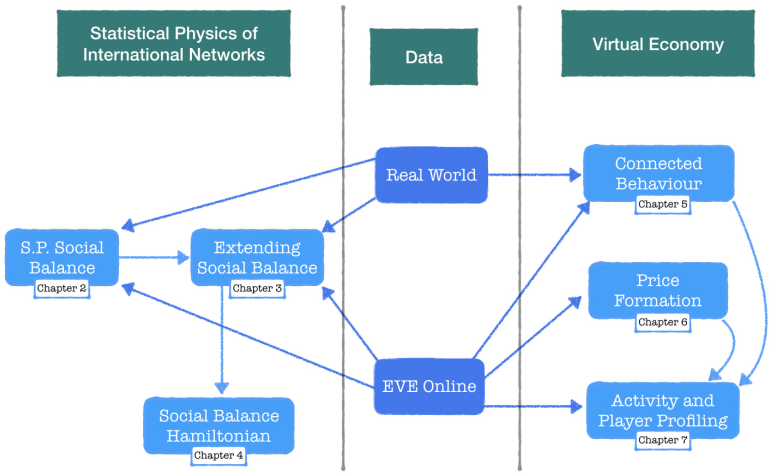


Figure 1.1: Structure of this dissertation. At the center, we have the two sources of data used in this work. On the left side, we sketch the different components in the first part of this work that deals with the statistical physics of international networks. The arrows represent the connections between the different chapters. On the right side, we summarize the content of the second part that deals with the economy in a virtual world. Also for the second part, the arrows are indicative for the connections between the different chapters.

1.2.1 International social networks from the perspective of statistical physics

Chapter 2: Statistical Physics of Balance Theory. This chapter exhibits the work published in *PLOS One* on August 28 of 2017 [68]. Based on the ideas of Social Balance theory, we develop a model from Boltzmann-Gibbs statistical physics. Central to our model are the triads' incidence rates and the idea that those can be modeled by assigning a specific triadic energy to each type of triad. We emphasize the role of the degeneracy of the different triads and how it impacts the degree of frustration in the political network. We go beyond the original formulation of social balance (the strict rule, whereby the triads can be either "balanced" or "unbalanced") and include information of recent studies that include more detailed information. Thereby, four different types of triads are considered with strong and weak versions of the balanced and unbalanced triads.. In order to learn about the dynamics and motives, we propose a generic Hamiltonian with three terms to model the triadic energies. One term is connected with a three-body interaction that captures balance theory. For a reader with a background in spin physics, it could be interesting to remark that in this model, the spin

information is carried by the edges of the political network. In the context of physics it is commonplace for the models to have that the spin information is carried by the nodes in the network. This may seem as conceptually very different, but it will be shown that the concepts of social balance can be mathematically mapped on selected spin models that have been studied in great detail in physics. The other terms take into account the impact of heterogeneity and of negative edges in the triads. The validity of our model is tested on four datasets, two from a virtual world and other two from the Real World. From a massively multiplayer online game (MMOG), we use the time series of triadic relationships for the standings between two classes of alliances. From the real world, we analyze real-world data for the relationships between the "agents" involved in the Syrian civil war, and the relations between countries during the Cold War era. We observe emerging properties in the triadic relationships in a political network, reflecting itself in a persistent hierarchy between the four triadic energies, and in the consistency of the extracted parameters from comparing the model Hamiltonian to the data. The work contained in this chapter is meant to be a proof-of-concept of the Hamiltonian-based model for social balance. The model gets further developed and put to the test in chapters 3 and 4.

Chapter 3: Social Stability and Extended Social Balance. This chapter exhibits the work published in *Physica A: Statistical Mechanics and their applications* on March 15 of 2019 [69]. Extending the previous model, we include the creation and destruction of relationships. The inclusion of a new kind of relationship brings the possibility of having three possible connection between agents (friendly, hostile and inactive). This addition results in ten types of triads, bringing the advantage that the network analysis can be done with fully connected networks. To each type of triadic relationship, we assign an energy that is a measure for its average occupation probability as in the previous model. We propose a novel extended Hamiltonian with three interaction terms and a chemical potential (capturing the cost of edge activation) as an underlying model for the triadic energy levels. Our model is suitable for empirical analysis of political networks. We test it with the same dataset from EVE Online used in the previous chapter and the real-world data for the relationships between countries during the Cold War era. The addition of the non-active edges, does not affect the persistent hierarchy between the ten triadic energy levels across time and networks that we found in the previous analysis. Also, the analysis reveals consistency in the extracted model parameters and a universal data collapse of a derived combination of global properties of the networks. We illustrate that the model has predictive power for the transition probabilities between the different

triadic states.

Chapter 4: Social Balance Theory, Supplementary Material. This chapter contains not-previously published work based on the three-body Hamiltonian of Social Balance introduced in this work. As an extension of the work discussed in the previous chapters, we develop a mean-field approximation and compare its results with Monte-Carlo simulations with the Hamiltonian that includes the chemical potential. For all possible combinations of the Hamiltonian's parameter we extract the position of the critical point. As a sanity check, both the Monte-Carlo and the mean-field calculations retrieve the results of the Ising system for the proper limit of the three-body interaction and the chemical potential.

1.2.2 Economy in a virtual world

Chapter 5: Connections between real-world and virtual-world human behavior. With the data of the activity of the players in the virtual world "EVE Online", we study the connection between in-game behavior and the real-world nation-level context. Tracking the actions conducted by players from different countries, we create country-profiles that capture all actions performed. Some neighboring countries present similar profiles (like Germany and Austria) and other present high differences (like Ukraine and Canada). Using different coordinates of these profiles, we correlate real-world aggressiveness (as quantified by the Global Peace Index and Global Terrorist Index) with virtual-world aggressiveness. We found a negative connection between real-world aggression level and in-game, indicate that the more aggressive countries have more pacific and friendly players. Also, we test whether in-game trading behavior can be linked to a player's real-world socioeconomic environment. We observe a positive relationship between the level of in-game trading activity and both the unemployment rate and the local dollar exchange rate. With our analysis, we present evidence that the virtual world can be used to infer the real-world context.

Chapter 6: Trade and production in an Online World: Case of EVE-Online. In this chapter, we investigate mechanisms related to price formation ("pricing") in the virtual market in "EVE Online". More specifically, we test which conditions of the market and the production chain affect the decision of the players to sell and buy the manufactured items to specific prices. We found strong correlations with many other market measures, as material cost, the average size of the transaction, but also with properties fixed by the developers as production time and proprieties of the

production network as centrality and position in the chain. We found evidence that pricing is a multivariable process, and other factors apart of the cost are dynamically taking into account. Besides, We test how the market responds to changes in the production conditions. For the period we studied, EVE Online has every month expansions that can have an impact on the production as changing the distribution of materials in the universe, the times of production, introducing new items, among others. Around this expansion dates, we perform our regressions to measure the predictability of the model. We found that there is a heterogeneous market response to these changes, some becoming more uncertain before the expansion and others do not be affected by it. We conclude that this could be a consequence of a lack of homogeneity about the information pre-expansion.

Chapter 7: Virtual Economy, Supplementary Materials. This chapter contains some extra material for activity in EVE online. We use EVE data to explore some initial research questions not connected between them. In the first part, we test the effect of weekly time in the market activity and motion of the players in-game. Using the daily frequency of market transaction and Jump between solar systems, we calculate their total probability distribution and for each of the days of the week. We found that the weekday and weekend day distributions do not have an underlying distribution, being the player more active in the weekend than weekdays. The second part is composed of player profiling. We calculate the player profile based on their action in-game, and we use a clustering algorithm to finding groups of players that have similar behavior. Depend on the sensibility of the algorithm; we found a general cluster where we can extract smaller and more specialist clusters.

1.3 Software and Computational setup

All scripts for the extensive data processing and analyses conducted in the context of this work, have been coded in Python [70] thereby making extensive use of its libraries. Python is a high-level interpreted general-purpose programming language, that is widely used for Data Analysis thanks to its simplicity and readability. The numerical calculations have been done using the library NumPy [71] that contains many functions written in C/C++, speeding up many numerical calculations. For scientific calculations we have frequently made use of the SciPy [72] library. For some of statistical analyses use was made of Statmodels [73] and Sci-kit [74]. The figures have been generated with the aid of Matplotlib [75] that is a plotting library for Python of general propose. To manage the big tables of data use has

been made of the PANDAS library [76], as it is a library for easy-to-use data structuring and data analysis for Python. For the network analysis and corresponding plots, use has been made of the libraries NETWORKX [77] and GRAPH-TOOL [78]. They are similar libraries for network analysis but have complementary functions. For the symbolic manipulation in Chapter 4, we used the library SYMPY [79]. The processing of EVE's data has been done on our group's server, which has 32 cores and 256 GB of RAM. The Monte-Carlo simulations in Chapter 4 were conducted using the computational resources and services provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government.

Part I

International social networks from the perspective of Statistical Physics.

2

Statistical Physics of Balance Theory

The content of this chapter reflects the work published in *PLOS One* on August 28 of 2017 [68]. The format of the document has been adapted to match the rest of the exposition. In addition, the references are unified with the remaining chapters. My role in this work cover Conceptualization, Data curation, Formal analysis, Software, Visualization, Writing - original draft.

2.1 Abstract

Triadic relationships are accepted to play a key role in the dynamics of social and political networks. Building on insights gleaned from balance theory in social network studies and from Boltzmann-Gibbs statistical physics, we propose a model to quantitatively capture the dynamics of the four types of triadic relationships in a network. Central to our model are the triads' incidence rates and the idea that those can be modeled by assigning a specific triadic energy to each type of triadic relation. We emphasize the role of the degeneracy of the different triads and how it impacts the degree of frustration in the political network. In order to account for a persistent form of disorder in the formation of the triadic relationships, we introduce the systemic variable temperature. In order to learn about the dynamics and motives, we propose a generic Hamiltonian with three terms to model the triadic energies. One term is connected with a three-body interaction that captures balance theory. The other terms take into account the impact of

heterogeneity and of negative edges in the triads. The validity of our model is tested on four datasets including the time series of triadic relationships for the standings between two classes of alliances in a massively multiplayer online game (MMOG). We also analyze real-world data for the relationships between the "agents" involved in the Syrian civil war, and in the relations between countries during the Cold War era. We find emerging properties in the triadic relationships in a political network, for example reflecting itself in a persistent hierarchy between the four triadic energies, and in the consistency of the extracted parameters from comparing the model Hamiltonian to the data.

2.2 Introduction

Signed social networks are those with both positive and negative edge weights used to capture the valence as well as the strength of dyadic relationships, such as friendship/ally and animosity/enemy. By far the dominant method to analyze such networks is balance theory (also called "structural balance theory"). Originally proposed by Heider [47] as an explanation for attitude change, and formally generalized for graphs by Cartwright and Harary [48], balance theory has been refined and applied to a variety of social, economic, ecologic and political scenarios [44, 45, 51, 80–84]. Balance theory provides a means to capture a system of such relationships and measure the degree to which it is balanced/stable or frustrated/unstable. Here we provide a further expansion of balance theory utilizing methods from Boltzmann-Gibbs statistical physics to show that after assigning energy levels to unique configurations of triads in signed graphs, we can extract the characteristic strength parameters of their interactions.

The original formulation was concerned only with whether graphs were balanced or not, where "balanced" meant that all cycles among all nodes contained only an even number of negative edges [48]. Later work [85] argued that only triads of nodes (i.e., connected triples, 3-cycles) were relevant for most social science applications of balance theory, and also showed that the proportion of unbalanced n-cycles and 3-cycles increase monotonically with each other. Thus the triadic version has become dominant (although see [49] and [50] for alternate measures of balance). Balanced configurations are still those with an even number of negative edges; specifically we capture the ideas of "The enemy of my enemy is my friend" $[+--]$ and "The friend of my friend is also my friend" $[+++]$ as balanced/stable situations. The two other types of signed triadic configurations ($[++-]$ and $[---]$) are considered unstable and give rise to frustration in the network. Because this was developed as a theory of attitude change, the frustrated triads are

considered to be posed to change to increase systemic balance.

As a further refinement, social scientists since [86] have observed that the two types of balanced triads are not equally balanced, and the two types of frustration are not equally frustrated. If we consider the classic version above the "strict rule" for balance, the "loose rule" for balance rates $[+\,+\,+]$ as more strongly balanced than $[+\,-\,-]$, and $[+\,+\,-]$ is more strongly frustrated than $[-\,-\,-]$. This breakdown reflects the observation that while triple negative triads are not stable, they do not actually inject much frustration into the system. Likewise, although $[+\,-\,-]$ is balanced, a system containing entirely triple positive triads is more stable.

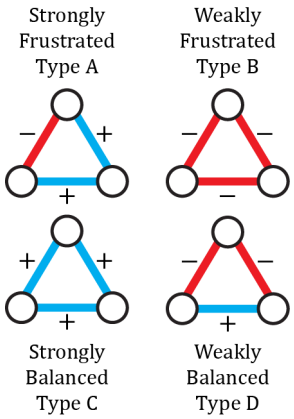


Figure 2.1: **Different types of triadic relations and their classification according to the strict and loose rule.** *In both classifications, C and D are balanced. According to the strict rule, the triads A and B are equally unbalanced. According to the loose rule, A is more unbalanced than B.*

Heider's original hypothesis was that a network of signed relations would tend to evolve towards a more balanced situation, eventually having solely triple positive balanced triadic relationships (often referred to as "utopia"). Several simulation models explore the formal conditions for this outcome [48, 87–90], but empirical studies reveal significant deviations. The validity of SBT has been tested on a variety of social and political relationships, such as political relations between countries [44,45,82]. For example, in the time span between 1946 and 1999 the fraction of unbalanced triads in the political network studied by [44] fluctuates over time and is consistently in the 5-15% range. Furthermore, there was no decreasing trend in the fraction of unbalanced triads. Although the rationale behind Heider's theory is compelling, and there may be intervening and exogenous factors (such as those explored by [82]) that explain the discrepancy, the power of balance theory

to explain political dynamics remains an open question.

One recurring problem in the study of social and political systems is that the data are incomplete and/or subject to privacy-related restrictions. Considering this, the virtual worlds of massively multiplayer online games (MMOGs) are fascinating social laboratories that can serve as sources for high-quality data in a controlled environment over extended periods of time. For example, the authors of [58] infer the signed social network in the online virtual world PARDUS using information available such as the formation of alliances, trade, attacks, the exchange of private messages, etc. They found different structural properties in the positive and negative networks with respect to clustering coefficients, reciprocity, and degree distributions. They also found that the two types of unbalanced triads have a different incidence rate; specifically the $[- - -]$ (weakly frustrated) triads are more frequent than the $[+ + -]$ (strongly frustrated) ones. The two types of balanced triads also have a markedly different incidence rate; specifically the $[+ + +]$ (strongly balanced) triads have a higher frequency than the $[+ - -]$ (weakly balanced) ones. Aside from implying that cooperation between the players is an important incentive in PARDUS, this provides support for using the loose version of balance theory.

For our main analysis we use data on the ally/enemy relationships between two different classes of alliances from the virtual world of EVE Online (another MMOG). These data are aggregated daily and cover a time period of over a year. Because these datasets come from a virtual world, complete and accurate records of all alliance standings across time are available. We also analyze two datasets from the real world: (1) the relations among fourteen political participants in the military intervention against ISIS (DAESH), and (2) the relations between states during the Cold War era.

Pioneering work in applying physics-based models inspired by the Ising model to political networks can be found in [43] and [46]. In these works, the spin variable is used to assign a political agent to a peculiar global alliance. Thereby the spins are also subject to history-based interactions. We develop an alternate approach based on balance theory.

Our method applies principles of Boltzmann-Gibbs statistical physics to balance theory. First, we test the accuracy of the traditional strict and loose rules of balance theory. Next, we argue that the incidence rate (or, occupation probabilities) for the four types of triadic relationships are good measures to gain a better understanding of balance theory. We associate a specific energy value to each type of triad ("triadic energies") and introduce the concept of *temperature* as a measure of the total and persistent systemic frustration. The information entropy corresponding to the occupation prob-

abilities of triads is introduced. In order to unravel the dynamics behind structural balance, we introduce a Hamiltonian with three parameters that can be extracted from model-data comparisons. We find a high degree of consistency among the model parameters extracted from the four datasets analyzed in this article, thus implying that our statistical physics approach to balance theory may be broadly useful.

2.3 Materials and Methods

In this section, we capture social balance in a statistical physics framework. Previous formal approaches to analyzing systemic balance and dynamics often focus on changing the link values or structure to understand tendencies towards or away from balance. Antal et al. [87] proposed a model in which randomly selected links change with the aim of balancing unbalanced triads. In this model, dubbed "local triad dynamics", they found that a finite network relaxes into an equilibrium state with balanced triads. Abell and Ludwig explore the tradeoff between increasing the number of positive links and the nodes' tolerance to imbalance. Variations in these parameters provided evidence for three behavioral phases, including features resembling self-organized criticality [88]. Gawronski et al. use continuous link values to reformulate balance theory in terms of dynamical equations. They show that, given certain constraints, in a fully connected network the system converges to Heider's balanced state (or "utopia") after a finite number of interactions [89].

A formulation of balance theory in terms of energy levels using only the strong rule was proposed by [51], but its focus was on explaining why the systems do not necessarily evolve to lower frustration. They did this by investigating the landscape of possible networks and found so-called "jammed states" that impede the total relaxation of the system. Although balance theory includes a tendency toward reduced frustration, we are also interested in various systems' tolerance levels of systemic frustration.

We propose a framework allowing one to determine the proportions of triad types a given system adopts. To do this we assign a characteristic energy to each of the four triad types in Fig. 2.1, which produces a ranking of energy levels. Those so-called "triadic energies" determine how many of each type of triad persists in the system. In this framework, low-energy triads will be more common than what would be expected based on their rate of occurrence in a random network. The opposite is true for high-energy triads. By comparing various systems one can infer whether some robust properties emerge with respect to the frequencies of each type of triad.

2.3.1 Energies and Entropies of Political Networks.

Political networks are dynamic, so the proportion of each type of triad and the density of the triads will fluctuate over time. This includes triads among particular triplets of nodes that change type, an effect that can be modeled by considering that it changes energy level. In our applications the system never resides in its energetically most favored state ("utopia" or the"ground state"), as is commonplace in physical system with a non-vanishing temperature. The social network dynamics of our systems can be attributed to several sources (conflict, political bargaining, access to resources, trading partners, etc.), but we will not discriminate between these different drivers of change. We model their overall (averaged) effect through the concept of "temperature"[1]. This can be seen through fluctuations in quantities such as the average number of triads of a particular kind around the mean.

Now, we are going to apply the methodology of classical statistical physics to the study of social networks, focusing on the triadic relations as the level where the preferences of the different nodes play the main role. First, we need to define the variables that characterize the system. A social network of N nodes is composed of $\frac{1}{2}N(N-1)$ possible symmetric edges s_{ij} between nodes i and j, each of which must have one of three values: $s_{ij} \in \{+, -, \times\}$ (positive, negative or nonexistent). Thus a microstate μ of such a network is uniquely specified by the complete set of s_{ij} values (together with any relevant node attributes).

Our unit of analysis here is triads, specifically the four types of triads presented in Fig. 2.1. We define a triad micro-state as the set of $\frac{1}{6}N(N-1)(N-2)$ triads T_{ijk} between nodes i, j and k, each of which must have one of five states: $T_{ijk} \in \{A, B, C, D, \times\}$ (The four kinds of Fig. 2.1 or nonexistent). The map from network micro-states μ to triad micro-states η is not bijective: all micro-states μ correspond to a unique triad micro-state η, but not vice-versa. The states of two triads that share a link are not actually completely independent, but taking the potential correlations into account is notoriously challenging. For example, although there are no shared links between the $[+++]$ and $[---]$ triads, a negative edge may be part of triads of type A, B, or D in various permutations. For sparse networks the number of triads sharing links is low and so, to a first-order approximation, the triads can be considered as independent. From the perspective of the

[1] Intuitively, the social temperature acts as the element that allows the system to go out of the ground state and move for unbalance systems. The temperature aggregates many factors, like greed or resource distributions, that are exogenous of the network and hard to quantify. For example, country A that have resources that two of its allies B and C do not have can tolerate a negative link between B and C if it can still sell its resources to both.

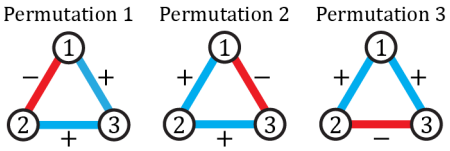


Figure 2.2: **Geometrical degeneracy of triads.** *The* $[++-]$ *triad (Type A) has 3 different micro-states defined by the permutations of edge valences shown here. These micro-states constitute the geometrical degeneracy of the triads.*

triads, edge permutations can be seen as the geometrical degeneracy of the triads, and the number of equivalent edge microstates differs for the various types of triads. Type B and C triads have three identical edges, so there is only one edge combination that can generate those. Triad types A and D can be generated in three different ways depending on which edge is the odd one out (see Fig. 2.2).

We must first estimate the probabilities of each type of triad: p_A, p_B, p_C, p_D. Because these types constitute a partition of triadic micro-states, we can divide the number of each type m_i by the total number of triadic relations in the network M to produce proportions. We make the simplifying assumption that the probabilities can be estimated accurately by these observed proportions:

$$p_{i\in\{A,B,C,D\}} = \frac{m_i}{M} \,. \tag{2.1}$$

Unlike the case of the network in which each unique configuration of the edges counted as a distinct micro-state, for the triads we are considering a micro-state as defined by the *proportion* of each type of triad. This means that the micro-states μ are uniquely determined by the variables $\{p_A, p_B, p_C, p_D\}$. In this procedure, we aggregate both the geometrical degeneracy within each type as well as which particular triads are in which particular configuration. We do this because we are interested in the total average energy and entropy of the system, and such macro-variables are determined by a weighted sum over all possible micro-states.

We assign an energy E_i to each type i of triad. The corresponding degeneracy $g(E_i)$ is defined as the total number of different ways of creating a triad of type i. Apart from the geometrical degeneracy (see Fig. 2.2), there can be other sources contributing to $g(E_i)$[2]. The triadic entropy S_T generated by the probabilities $\{p_A, p_B, p_C, p_D\}$ of the energy levels of each

[2]As the degeneracy is defined as the number of micro-states that belong to the same state, each system can have different degeneracies. Each step between the particular state to the undirected signed graph where we merge two o more states equivalent will add another source to the degeneracies. In our case, geometrical and directional sources are the sources we found.

type of triad can be calculated as

$$S_T = -K \sum_{i \in \{A,B,C,D\}} p_i \left(\ln p_i - \ln g(E_i)\right) \quad , \tag{2.2}$$

where K is a constant that connects the units of entropy and energy. This constant is the Boltzmann's constant (k_B) in statistical physics.

We now derive an expression for the probabilities $p_{i \in \{A,B,C,D\}}$. If we assume that each type of triad can be connected to a specific energy, the principle of maximum entropy applied to the S_T under constraints of normalization of the probabilities and a finite average energy leads to the Boltzmann distribution [91]. Thereby, each energy level $i \in \{A, B, C, D\}$ has a probability of occupation p_i given by

$$p_{i \in \{A,B,C,D\}} = \frac{g(E_i)\ \exp(-\frac{E_i}{KT})}{Z} \ , \tag{2.3}$$

whereby the partition function

$$Z = \sum_{i \in \{A,B,C,D\}} g(E_i)\ \exp(-\frac{E_i}{KT}) \ , \tag{2.4}$$

acts as a normalization constant to ensure the total probability $\sum_{i \in \{A,B,C,D\}} p_i = 1$. Further, (KT) has the units of energy and is the Lagrange multiplier associated with the constraint that the system has an average total energy. In the remainder of this work, we also use the notation $\beta \equiv \frac{1}{KT}$. From this formulation we create a model in which, for a "high" temperature (defined as $KT >> |E_i - E_j| \, \forall i, j$), the triads will be stochastically occupied according to their degeneracy $g(E_i)$, because $\exp(-\frac{E_i}{KT}) \approx \exp(-\frac{E_j}{KT}) \forall i, j$. For "small" temperatures (defined as $KT << |E_i - E_j| \, \forall i, j$) all triads will be in the lowest possible energy state E_g, because $\exp(-\frac{E_g}{KT}) >> \exp(-\frac{E_j}{KT})$. In the parlance of balance theory, this specific small-temperature situation corresponds with utopia, a network of exclusively $[+ + +]$ triads.

The expression of Eq. 2.3 connects the occupation probability of a certain triadic state to the ratio of its energy and the temperature (so it is invariant under translations of the energy scale). Our methodology determines the triadic energies from the occupation probabilities; so although it is fundamentally impossible to determine the temperature, without any loss of generality we can set KT as the unit of energy.

We study the time evolution of social networks, which may be considered out-of-equilibrium systems. However, because our data are collected at time intervals that are short relative to the social dynamics, we interpret them as quasi-static. Consider that political networks that typically evolve over

the time scales of years may be interpreted as quasi-static if the data are recorded on a daily basis. Thus Eq. 2.3 provides the triadic occupation probabilities for such quasi-static out-of-equilibrium scenario.

In a political network with agents not displaying preferences, the relative occupation probabilities are determined solely by the degeneracies. The rate of incidence of the different types of triads is then a reflection of the degeneracy. When the agents display preferences –which is reflected in a certain hierarchy of energy levels –, the degeneracy still plays a major role. A level with a relatively high energy, which means that it is not the preferred state of the agents, could still have a large incidence if its degeneracy is large enough. Random networks can be used as a reference for the preferences among the different energy levels. In a study with random networks, Szell et al. [58] report relative occupation probabilities in their random network for the triads B:A:D:C close to 1:3:3:1 which reflects their geometrical degeneracy. Using triadic relationships produced by random signed networks as a control, we can detect which triads are conceived of as more unbalanced (the ones underrepresented in the data relative to random networks) or more balanced (overrepresented relative to random networks).

Given data for the occupation probabilities p_i of the different types of triads, the relative energies $(E_i - E_j)$ can be extracted from

$$\beta E_i - \beta E_j = \left[-\ln \frac{p_i}{g(E_i)} - \ln Z \right] - \left[-\ln \frac{p_j}{g(E_j)} - \ln Z \right] . \tag{2.5}$$

Throughout this article we systematically refer to the quantity "$-\ln \frac{p_i}{g(E_i)}$" as the "extracted triadic energy" $\beta E_i + \ln Z$ for triad type i. Using the β as a reference of the energy scale and the inferred degeneracy of each type of triad, we have developed a methodology that allows us to solve for the relative energies between the triad types based on their frequency of occurrence in data of a political network. As we will see shortly, the energies can be connected with the underlying dynamics.

2.3.2 Hamiltonian for the Dynamics of Triadic Relationships

The use of the combination of degeneracy and energy for each specific triad will allow us to embark on quantitative studies of the underlying dynamics of political networks. Up to this point we have connected the rate of incidence of a specific triad $i \in \{A, B, C, D\}$ to its energy E_i (different for each state, and displaying the preferences) and a temperature KT (which is a systemic variable). The authors of [51] showed that strict-rule balance theory can be mapped into a genuine three-body interaction among the edges because the product of the three edge valences discriminates between balanced and

Type of triad	Symbol	Associated energy from Hamiltonian (2.6)	Geometrical degeneracy
Highly frustrated	A: $[++-]$	$H_A = \alpha + \gamma + \omega$	$g_G(E_A)$ =3
Lowly frustrated	B: $[---]$	$H_B = \alpha - 3\gamma - 3\omega$	$g_G(E_B)$ =1
Lowly balanced	D: $[+--]$	$H_D = -\alpha + \gamma - \omega$	$g_G(E_D)$ =3
Highly balanced	C: $[+++]$	$H_C = -\alpha - 3\gamma + 3\omega$	$g_G(E_C)$ =1

The corresponding Hamiltonian in the space of triadic relationships is given by the Eq (2.6). The geometrical degeneracy is explained in Fig 2.2.

Table 2.1: ***Energies and degeneracies associated with the four types of triadic relationships considered in this work.***

unbalanced states.The three-body interaction solely differentiates between balanced and unbalanced triads. We wish to further differentiate between strongly and weakly balanced/frustrated triads (Fig.2.1). To this end, we add a one-body and a two-body part to the Hamiltonian. In order to gain a more detailed insight into the underlying dynamics, we propose to add a two-body and a one-body interaction between the edges. The effects of the three-body, two-body, one-body terms are regulated by strength parameters α, γ, and ω respectively. The three contributions to the interactions between the nodes are designed such that the values of those parameters (α, γ, ω) provides detailed information about the node's incentives with regard to establishing relationships.

Let $s_{ij} \in \{-1, +1\}$ encode the relationship (enemy or friend) between nodes i and j. Then the proposed Hamiltonian acting in the space of a given number M of complete triads is

$$\begin{aligned}\mathcal{H} = \sum_{i<j<k} [&\underbrace{-\alpha s_{ij}s_{jk}s_{ki}}_{\text{three-body}} \\ &\underbrace{-\gamma(s_{ij}s_{jk} + s_{ij}s_{ki} + s_{jk}s_{ki})}_{\text{two-body}} \\ &\underbrace{+\omega(s_{ij} + s_{jk} + s_{ki})}_{\text{one-body}}] \,.\end{aligned} \tag{2.6}$$

Here, the sum $\sum_{i<j<k}$ extends over all unique 3-node cycles $\{1, 2, \ldots, M\}$ in the network. The indices i, j, k run over nodes that are part of complete triads. In this way the Hamiltonian for each triad type is derived directly from the valences of the edges that make them up and shown in Table 2.1. Fig. 2.3 shows a schematic diagram of the three terms in the proposed Hamiltonian.

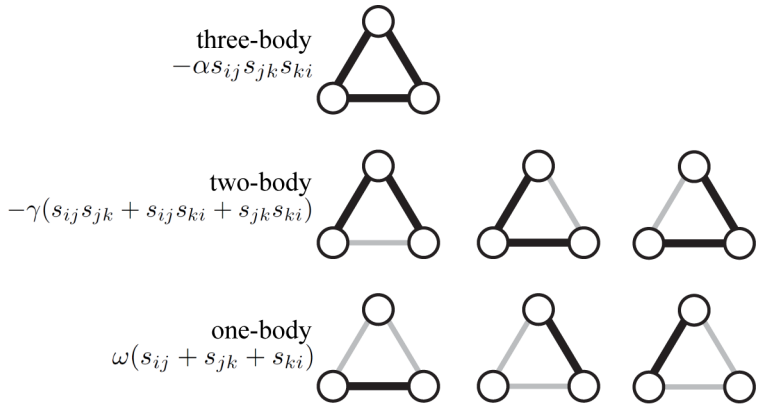


Figure 2.3: ***Schematic diagram illustrating the three terms in the proposed Hamiltonian of Eq 2.6.***

The two-body term in the Hamiltonian with strength γ can be interpreted as a force term that attempts to "homogenize" the relations in the triad. It introduces a fine-splitting of magnitude 4γ in the energy spectrum depending on whether the balanced or unbalanced triad is symmetric or not (Table 2.1). The parameter γ makes clustering energetically more favorable, even if it concerns a triad with solely enemies. The energy of the balanced and unbalanced triads are lowered by an amount 3γ for the symmetric $[+++]$ and $[---]$ triads and increased by an amount γ for the asymmetric triads $[+--]$ and $[++-]$. The strength ω of the one-body term in Eq. 2.6 encodes the "reward" that corresponds with the creation of "$-$" links.

The values of the strength parameters can be determined by analyzing data of triadic relations in empirical networks, and doing so tells us about the underlying dynamics of such systems. To see how this works, we now demonstrate this statistical approach to structural balance theory with a few example applications.

2.3.3 Virtual World Data: Political Network of Alliances in EVE Online

2.3.3.1 Data

The extended dataset that we first analyze is extracted from recorded and time-stamped data from EVE Online, which is an MMOG developed by CCP Games [62]. In this virtual world more than 500,000 players trade, collaborate and fight in a futuristic galaxy. The players get together in social

$s_{1\to2}$	$s_{2\to1}$	s_{12}	Multiplicity
Not Set	Not Set	$\times$	1
Not Set	Friend	$+$	2
Not Set	Neutral	$\times$	2
Not Set	Enemy	$-$	2
Friend	Friend	$+$	1
Friend	Neutral	$+$	2
Friend	Enemy	$-$	2
Neutral	Neutral	$\times$	1
Neutral	Enemy	$-$	2
Enemy	Enemy	$-$	1

Table 2.2: ***Conversion table for the network of alliance relations in EVE.*** *The rules used to transform the directed network of alliance relations $(s_{1\to2}, s_{2\to1})$ in EVE Online into an undirected network of alliance relations s_{12}. The multiplicity is 1 for situations whereby $s_{1\to2} = s_{2\to1}$ and is 2 otherwise.*

structures called alliances with sizes between one and about 25,000 players. The alliances can conquer territory, where they can impose their own taxes, exploit mineral resources, and so on. The relations between the alliances represent an important social aspect of the game including the fact that the leader of the alliance has four choices with regard to setting the relationship to the other alliances in the game. Indeed, the relationship to any other alliance can be set as friendly, hostile, neutral or undetermined. This is important because it will change how the players of one alliance deal will the players in the other alliances, facilitating the process of discriminating between friends, enemies, and others.

Although one might expect a symmetric relationship for the standings between the alliances 1 and 2 $(s_{1\to2}, s_{2\to1})$, the data reveal a slight degree of asymmetry in the directed network. The degree of reciprocity (having the same link valence in both directions) is in the range 0.9-0.98 across our time series, so the assumption of symmetry is still largely justified. Because triadic balance theory works best on undirected edges, we transform the directed edge standings data into symmetric relationships via the conversion rules summarized in Table 2.2. These conversion rules are inspired by the rules of the game so that the status of the relationship between the alliances really reflects the dynamics of EVE Online.

The dataset under analysis consists of a time series of the relations between the alliances between February 15, 2015 and April 16, 2016. We consider "politically active" alliances, defined by the criterion that they set their standings with at least one other alliance. We study two distinct classes of alliances. First, the class of alliances with more than 200 mem-

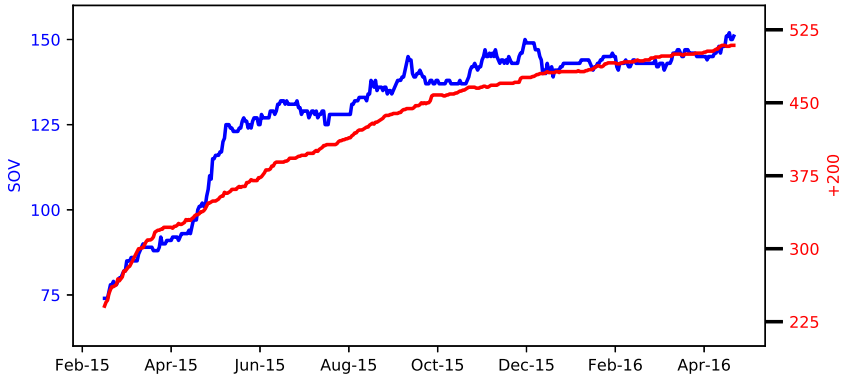


Figure 2.4: ***Evolution of the number of nodes in the network of alliance relations.*** *Daily evolution of the number of alliances in EVE Online from February 2015 through April 2016. We discriminate between alliances with sovereignty (SOV) and alliances with more than 200 members (+200)*

bers (so called "+200" alliances), and second, the class of alliances that hold sovereignty over at least one solar system ("SOV" alliances for short). These two classes of alliances are key to the political dynamics of the game. We proceed with providing information about the network structure of the two classes of alliances.

2.3.3.2 Network Properties of the Alliances in EVE Online

Political networks are not static. In EVE Online, new alliances come into the system and others disappear while the status of the relations between alliances are volatile as a result of all ongoing activities. In an effort to provide some feeling about the size of the data, we display in Fig. 2.4 the time evolution of the number of alliances in the studied period. The number of SOV (+200) alliances grows from 70 to 150 (250 to 500) and is subject to temporal fluctuations on top of the growing trend.

We now study the time evolution of some relevant network metrics of the SOV and +200 alliances. We focus on the evolution of the density of the triads, of the density of the edges and of the clustering coefficients. The results are summarized in Fig. 2.5. For the density of the triads and the edges there is a markedly different behavior of the SOV and the +200 alliances. The network of SOV alliances is marked by a higher density of edges and triads. We also observe strong temporal fluctuations in the densities of edges and triads for the SOV alliances. The density of complete

triads fluctuates between 4% and 14%. The big alliances have a smaller and rather stable density of triads (order 1%) and edges (5-6%). Our analysis based on uncorrelated triads is a reasonable approximation for networks with such an observed low triad density.

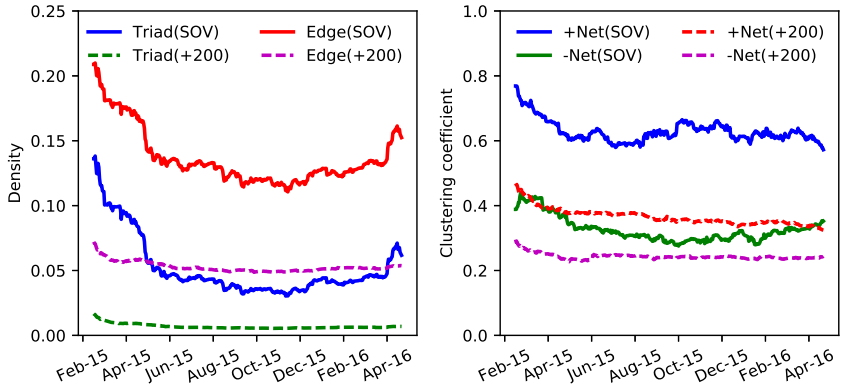


Figure 2.5: ***Properties of the network of EVE's alliances.*** **Left:** *Daily values of the density of complete triads and the density of edges for two classes of alliances in EVE Online.* **Right:** *Daily values of the clustering coefficients of the network of positive ("+Net") and negative relationships ("−Net") for the SOV and +200 alliances in EVE Online.*

For the clustering coefficients, we discriminate between the subnetworks of the hostile ("+") and friendly ("−") relationships. Clearly, the SOV and +200 alliances display a markedly different behavior in their clustering coefficients. For the large alliances, the clustering coefficient in the positive and negative networks is comparable and rather stable across the studied time period. In the SOV alliances, on the other hand, the clustering coefficients are larger and more volatile. For the SOV alliances, the positive and negative networks have a sizable different clustering coefficient. Comparing our EVE Online results for the clustering coefficients with those for PARDUS [58], we also find a higher clustering coefficient in the positive networks than in the negative ones. The reported cluster coefficients in the PARDUS negative subnetwork are of the order of 0.01-0.06, which is significantly lower than the observed values in EVE Online. We stress that the analysis of the PARDUS network was done on a social network of players, whereas our focus is on a political network of alliances.

From Fig. 2.5, we conclude that the SOV alliances maintain more triadic relations than the +200 ones, that their network is more complete and that they have a stronger tendency to create positive clusters. In the re-

sults section we will study how far the +200 and SOV alliances display a more consistent behavior in the way they create the four types of triadic relationships of Fig. 2.1.

2.3.3.3 Real World Data: Syrian Civil War and International Relations during the Cold War Era

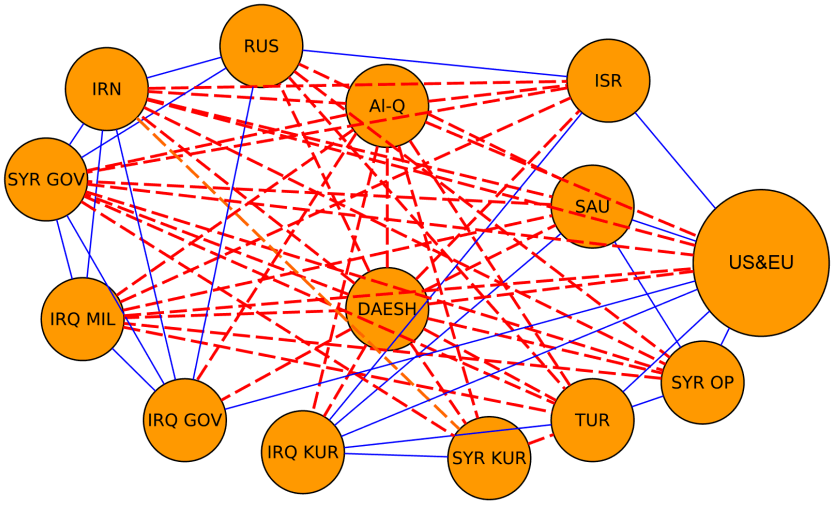


Figure 2.6: ***Network of relationships in the Middle East.*** *Graph with the relationships between 21 different "agents" in the Middle East as reported by The Economist in December 2015 [92]. We discriminate between enemies (red), allies (blue) and neutrals (no link). We use the following abbreviations to specify the nodes: Moderate Sunni Arab opposition (SYR OP); Syrian government (SYR GOV); Syrian Kurds (SYR KUR); Iraqi Kurds (IRQ KUR); Iraqi Shia militias (IRQ MIL); Iraqi government (IRQ GOV); Turkey (TUR); Israel (ISR); Russia (RUS); USA and E.U (US&EU); Saudi Arabia & Arab League (SAU); Iran (IRN); Al-Qaeda/Jabhat al-Nusra (AL-Q); DAESH or ISIS (DAESH). The node "US&EU" represents 8 agents (US, Australia, Belgium, Canada, France, Germany, Netherlands and United Kingdom).*

Here we describe our application of the proposed methodology to the triadic relations of two real-world systems. First, we analyze the triadic relations between 21 "agents" in the Middle East in the military intervention against ISIS (or DAESH) as reported by The Economist [92] on December 23, 2015. The corresponding graph is displayed in Fig. 2.6 and is complete in the sense that each possible edge exists with a valence of either "friendly", or "unfriendly/enemy" or "neutral/mistrust". The data provided in [92]

aggregates all the information about the status of the relationships of the European countries, the US and Australia in one single "US & EU" node. In order to reach a higher level of detail for the occupation probabilities of the different types of triads, we duplicate the "US & EU" node and introduce nodes (extra agents) for the US, Australia, Belgium, Canada, France, Germany, Netherlands and U.K, as countries that are involved in the conflict [93].

As a second system, we study the time series of signed edges of international relations during the Cold War era (1949-1989) from the Correlates of War project [94–96]. The edge between two countries is considered "+" whenever there is an active military alliance or defense treaty between them, and we retrieved this data from the dataset for Formal Alliances (version 4.1) [95].

Correlates of War maintains a database of Militarized Interstate Disputes (MIDs) [96], and we assign a "−" value to the edges between countries whenever they were involved in an unresolved MID in a time window of 50 years. Countries can resolve their disputes and forge an active military alliance and/or a defense treaty. We mark this as a transition from a "−" to a "+" edge between countries. This procedure gives rise to a degeneracy of two for the positive edges and of one for the negative ones. As an illustrative example of how we construct the status of international relations we present the case of USA and Germany. In 1943, both countries are involved in a MID, implying the start of a period of a "−" relationship. In 1950, an alliance between West Germany and the USA comes into being and their relationship changes from "−" to "+". This does not happen with East Germany which implies that in 1950 its relationship with the USA remains "−".

2.4 Results

2.4.1 Triadic Relations between Alliances in EVE Online

As pointed out in the previous section and summarized in Table 2.1, the four types of triadic relationships can be characterized by a degeneracy and an energy. The lower the energy the more stable the triadic relationship is perceived. In the undirected network formulation of SBT, the degeneracy is purely geometric (Table 2.1). In EVE Online, we have an extra multiplicity, associated with the conversion rules (Table 2.2) used to convert the directed into undirected edges. The total degeneracy g is the product of the geometrical degeneracy g_G and the multiplicity (see Table 2.2) associated with each

of the 3 edges in a triad. As can be inferred from Table 2.2 the aggregated weight of the "+" relationships is 5 and for "−" it is 7. For example, for the $[+--]$ configuration, the total degeneracy is $g(E_D) = 3\times(5\times 7\times 7) = 735$.

In Fig. 2.7, the time series of the occupation probabilities defined in Eq. (2.1) is shown. We compare the data extracted from the alliances in EVE Online with those from a randomization whereby the following procedure has been followed. At each time step in the data we fix the topology of the network as it appears in the data. In a next step, we randomly assign the values $s_{i\to j} \in \{+,-\}$ and $s_{j\to i} \in \{+,-\}$ to the edges. Next, we use the rules of Table 2.2 to obtain the status of the edges in the associated undirected network and determine the occupation probabilities $p_i(t)$ for the realization of the network at a particular time step. The procedure of randomly assigning a status to the directed edges is repeated 500 times at each time step in the data. This produces the bands for the $p_i(t)$ in Fig. 2.7. The comparison between the occupation probabilities of the simulations and the data from the alliances in EVE Online, displays clear patterns that are persistent over time. The stochastic networks produce a higher occupation of the classical unbalanced $[++-]$ and $[---]$ triads than the data. The rate of incidence of the other types of triads is systematically higher in the data than in the simulations. This clearly illustrates that there is dynamics at play and that randomness cannot give rise to the observed patterns in the occupation probabilities. Remarkably, we observe the same ordering of the configurations as seen in the PARDUS data [58]. In the EVE Online virtual world, Fig. 2.7 shows that the balanced $[+++]$ triad are on average less populated than the unbalanced $[---]$ one. This illustrates that the total degeneracies g are a key element in the outcome of the relative occupation probabilities of the different triads.

The actual values of the energies E_i are more informative with regard to the underlying dynamics. Eq. 2.5 connects the occupation probability p_i for each triadic state to its energy E_i and the "temperature". Fig. 2.8 shows the time series of the extracted energies (as $\beta E_i + \ln Z$). For both classes of alliances the same ordering of the energy levels emerges. This is a remarkable result given the substantial variations and fluctuations in the number of nodes and clustering properties of the political network of alliances as time progresses (Figs 2.4 and 2.5).

From the time series of Fig. 2.8 one can infer information about the relative triadic energies. The extracted values for the $\beta E_i + \ln Z$ are summarized in Table 2.3. Both the hierarchy and the values of the triadic energies $\beta E_i + \ln Z$ are comparable for the +200 and SOV alliances. The highly balanced triad $[+++]$ has the lowest energy. The energy gap to the second balanced state $[++-]$ is somewhat larger for the SOV alliances. The

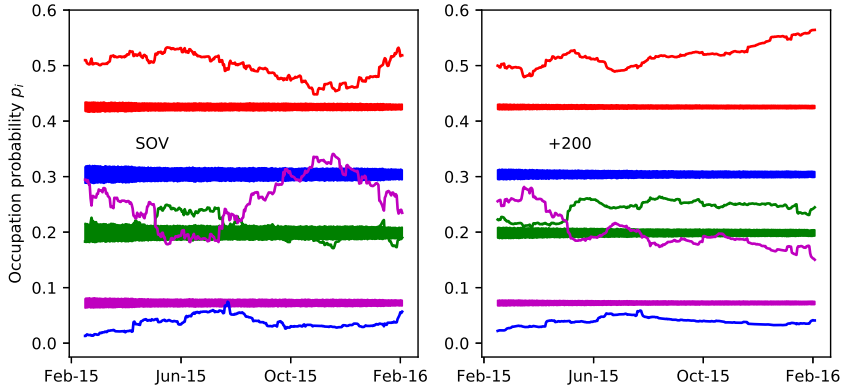


Figure 2.7: ***Comparison between a randomly signed network and EVE's real network of alliances.*** *Daily changes of the occupation probabilities for the four types of triads for the relationships between the alliances in EVE Online. We compare the data with a randomly signed directed network. Blue:* $[++-]$*; green:* $[---]$*; red:* $[+--]$*; magenta:* $[+++]$*. The left (right) panel is for alliances with sovereignty (alliances with more than 200 members).*

highly frustrated $[++-]$ triad has the high energy, reflecting its perceived outspoken instability. The balanced $[+--]$ and unbalanced $[---]$ triads have a comparable energy.

Type of triad	Total degeneracy		$-\ln \frac{p_i}{g(E_i)} = \beta E_i + \ln Z$			
	EVE	Cold War	EVE (SOV)	EVE (+200)	Middle-East	Cold War
A: $[++-]$	$g(E_A) = 525$	$g(E_A) = 12$	9.62 ± 0.33	9.52 ± 0.18	4.47	6.38 ± 0.49
B: $[---]$	$g(E_B) = 343$	$g(E_B) = 1$	7.41 ± 0.09	7.26 ± 0.07	2.11	3.24 ± 0.39
D: $[+--]$	$g(E_D) = 735$	$g(E_D) = 6$	7.30 ± 0.04	7.26 ± 0.03	1.80	4.25 ± 0.40
C: $[+++]$	$g(E_C) = 125$	$g(E_C) = 8$	6.17 ± 0.18	6.45 ± 0.14	1.06	2.24 ± 0.06

Table 2.3: ***Extracted triadic energies****. The triadic energies and corresponding error bars are obtained by time averaging the quantity* $-\ln p_i/g(E_i) = \beta E_i + \ln\ Z$ *for the alliances (SOV and +200) in EVE Online (left and middle panel of Fig 2.8) and for the international relations during the Cold War era (right panel of Fig 2.8).*

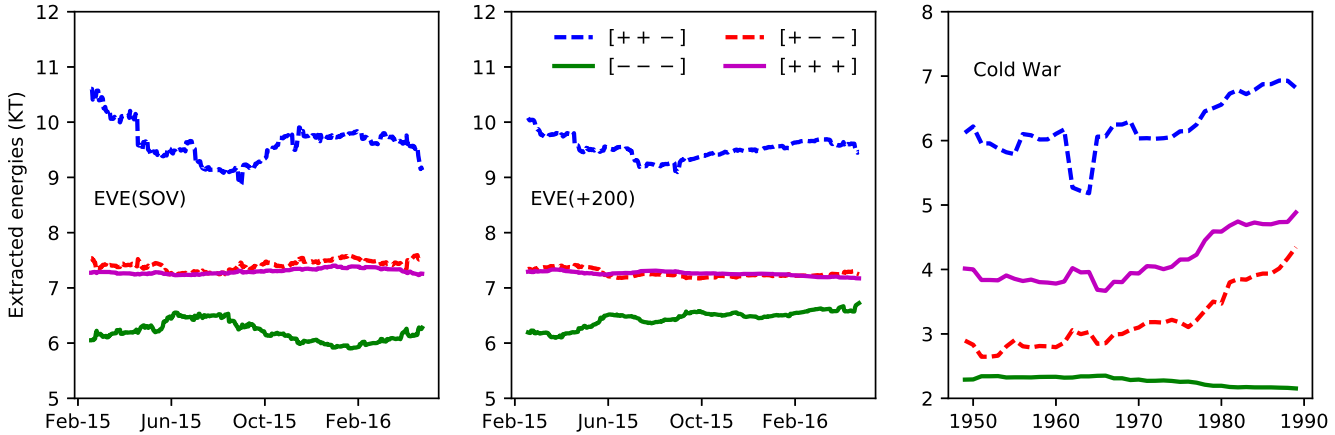


Figure 2.8: ***Time series of the extracted triadic energies.*** *We display the daily values of the log occupation probabilities* $-\ln p_i/g(E_i)$ *for the four types of triadic relationships between the +200 (left) and SOV (middle) alliances in EVE Online. These results are based on the data of Fig. 2.7. The right panel displays the yearly values of the* $-\ln p_i/g(E_i)$ *for the four types of triadic relationships in the international relationships during the Cold War era.*

Fig. 2.9 shows the time series of the systemic entropy values S_T defined in Eq. 2.2. In the networks with randomly assigned edges (see Fig. 2.7) the time-averaged systemic entropies are $S_T = 8.152 \pm 0.011K$ (SOV) and $S_T = 8.152 \pm 0.066K$ (+200). Obviously, the systemic entropy corresponding with the occupation probabilities for the alliances in EVE Online is substantially smaller. This indicates that even in the worst periods with a lot of turmoil and tensions in EVE Online, the system is still quite a bit more organized than random.

The time series of the entropy S_T could potentially teach us about the change-points in the time series of triadic relations in the political network. Strong variations in the time series of S_T for the SOV alliances occur in March 2015 and in April 2016. From the documented history of EVE Online, it is known that in 19-24 March 2015, a large coalition of collaborating alliances known as the "N3" tried to invade the space controlled by the "The Imperium". After this invasion failed, the Imperium counterattacked and N3 fell apart during the following months. The gradual disintegration of N3 is reflected in a continuous increase of entropy (more randomness in the system). This rise of the entropy comes to an end during August-September 2015 by which time the N3 alliances had formed new coalitions. Another distinctive feature in the time series of the entropy is the spike in April 2016. This marks the next great war, known as "World War Bee" or "The Casino War", that struck EVE's virtual world. As a result of this war, The Imperium coalition fell apart in mid April 2016. This is clearly visible as a strong rise in the S_T for both types of alliances.

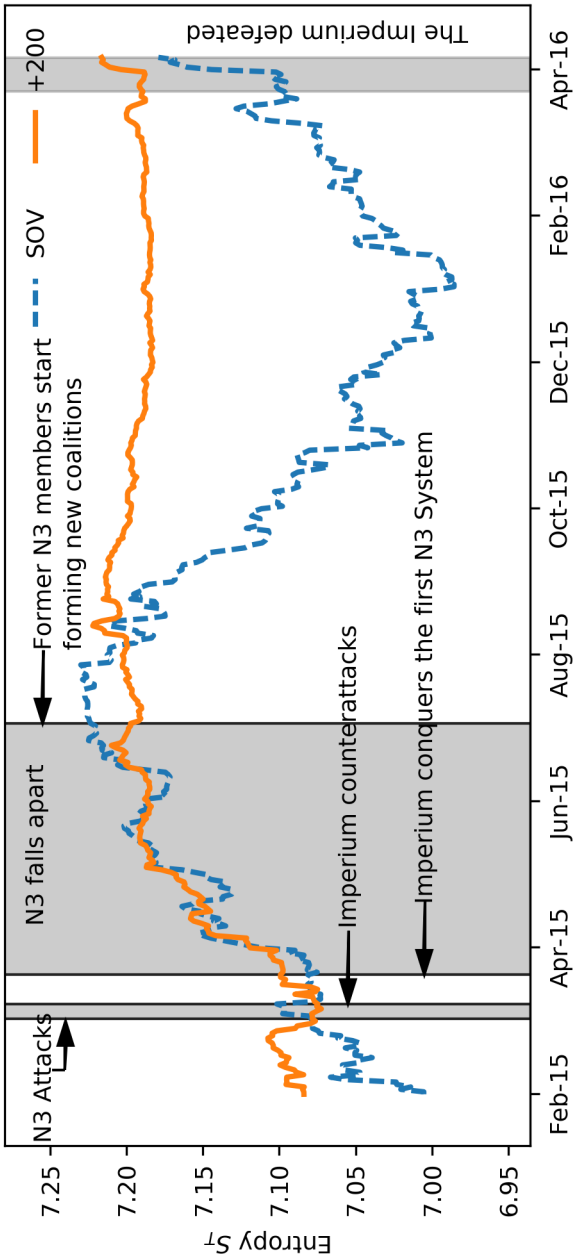


Figure 2.9: ***Time series of the entropy for the triadic relationships in EVE Online.*** *Daily values of the entropy S_T for the occupation probability of the four types of triads in the relationships between the SOV (dashed line) and the +200 (solid line) alliances in EVE Online. The entropies associated with the networks of randomly assigned edges are $S_T = 8.152 \pm 0.011K$ (SOV) and $S_T = 8.152 \pm 0.066K$ (+200).*

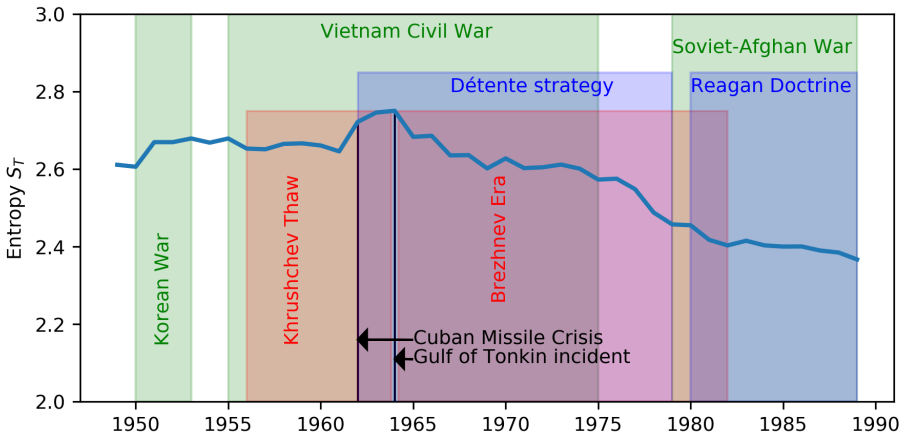


Figure 2.10: ***Time series of the entropy for the triadic relationships among countries during the Cold War era.***

2.4.2 Real World data: Syrian Civil War and International relations during the Cold War

The information about the relationships of 21 agents involved in the Syrian Civil of Fig. 2.6 can be converted into the occupation probabilities p_i for the four types of triads. Using the geometrical degeneracies $g_C(E_i)$ of Table 2.1, we obtain the $\beta E_i + \beta \ln Z$ values as contained in Table 2.3. Using data from the Correlates of War project [94] we extracted the times series for the triadic energies (Fig. 2.8) and entropies (Fig 2.10). The highest entropy value occurred during the 1962-1964 period that marks the Cuban Missile Crisis (October 16-28, 1962) and the Gulf of Tonkin incident (1964) that triggered the USA intervention in Vietnam. The impact of those incidents is also clearly visible in the extracted energies (right panel of Fig. 2.8) which show a decreasing trend from 1964 onwards. This trend can be attributed to a systematically increasing occupation probability for the $[+++]$ triads at the cost of the occupation probabilities for triads including a "$-$" edge. This is likely the result of the easing of the strained relations between the blocks of countries (often referred to as the "détente"). Thus one can identify major change points in each political system from the changes in the entropy over time, which the direction of the changes intuitively translating into increased of decreased political stability.

2.4.3 Model Hamiltonian for SBT

The extracted triadic energies $\overline{\beta E_i + \ln Z} \pm \sigma_{\beta E_i}$ shown in Table 2.3 can be used to determine the strength parameters for the Hamiltonian of Eq. 2.6, as well as the unknown zero level Z_0. From this we can learn about the underlying dynamics in the formation of the triadic relationships. We use the maximum likelihood method to determine the $(\alpha, \gamma, \omega, Z_0)$ that has the highest probability of generating the extracted $\beta E_i + \ln Z$ sets. The likelihood is computed from

$$\mathcal{L}(\beta E_i + \ln Z | \alpha, \gamma, \omega, Z_0) = \prod_{i \in (A,B,D,C)} \exp\left(\frac{-\left(\overline{\beta E_i + \ln Z} - (\beta H_i(\alpha, \gamma, \omega) + Z_0) \right)^2}{\sigma^2_{\beta E_i}} \right) \tag{2.7}$$

The values for the (α, γ, ω) are listed in Table 2.4. Remarkably, similar features and values emerge for the four datasets. The three-body and two-body interactions are clearly the driving forces in the creation of the triadic relationships. The fact that the ω adopts a positive value indicates that we are dealing with political systems with incentives for enmity. The extracted value of γ for the Cold War dataset is significantly higher than in the other three systems. This alludes to a stronger homogenization tendency. A possible explanation is the peculiar political situation during the Cold War era, with 3 blocks of countries (Capitalist, Communist and Non-Aligned Movement). We find that the first two groups maintain a high degree of internal homogeneity in their inter-country relationships and their attitudes toward confrontation between the blocks. The origin of this clustering effect can be attributed to the Cold War military strategy of Mutually Assured Destruction. As a consequence, the countries tend to align themselves with a country possessing nuclear weapons.

2.5 Conclusion

We have proposed a methodology to qualitatively study balance theory for triadic relations in political networks. The crux of our framework is the analysis of the occupation probabilities for the four different types of triads. The model uses elements from Boltzmann-Gibbs statistical physics to assign an energy and a degeneracy to each type of triad and introduce an overall systemic temperature to account for the disorder generating effects. The

Dataset	Strength parameters in units KT			
	α	γ	ω	Z_0
EVE (SOV)	0.95 ± 0.03	0.38 ± 0.02	0.18 ± 0.02	8.00 ± 0.03
EVE (+200)	1.02 ± 0.04	0.41 ± 0.02	0.14 ± 0.02	8.04 ± 0.04
Middle-East	1.09	0.38	0.22	2.70
Cold War	0.89 ± 0.07	0.61 ± 0.07	0.14 ± 0.07	4.67 ± 0.07

Table 2.4: ***Hamiltonian parameters from a theory-data comparison.*** *The extracted values for the strength parameters for the +200 and SOV alliances in EVE Online, for the status of the relationships in the Middle East in December 2015 [92], and for the international relations during the Cold War era.*

Dataset	H_i in units KT from Hamiltonian of Eq (2.6)			
	$H_A : [++-]$	$H_B : [---]$	$H_D : [+--]$	$H_C : [+++]$
EVE (SOV)	$+1.57 \pm 0.33$	-0.63 ± 0.10	-0.75 ± 0.05	-1.85 ± 0.18
EVE (+200)	$+1.51 \pm 0.33$	-0.75 ± 0.06	-0.75 ± 0.04	-1.56 ± 0.14
Middle-East	+1.69	-0.71	-0.93	-1.55
Cold War	$+1.64 \pm 0.44$	-1.36 ± 0.46	-0.42 ± 0.38	-2.29 ± 0.07

Table 2.5: ***Hamiltonian parameters from a theory-data comparison.*** *The corresponding energies H_i (see Table 2.1) of the Hamiltonian of Eq (2.6) for the +200 and SOV alliances in EVE Online, for the status of the relationships in the Middle East in December 2015 [92], and for the international relations during the Cold War era.*

energies can be investigated using a generic Hamiltonian with three-body, two-body and one-body interactions between the edges in 3-node cycles. The interactions have their own characteristic strength parameters that can be extracted from a model/data comparison.

We have tested the underlying assumptions of our model with two datasets from the virtual world EVE Online and two datasets from the real world. We have demonstrated that the proposed model allows one to quantitatively study social balance and gain insight into the mechanisms driving triadic relationships in political networks. For example, we can separate the dynamical mechanisms (regulated by the values of the energies) and stochastic aspects (degeneracy of the different energy levels). The model/data comparison for our four political networks lead to comparable energy and strength parameters. We furthermore find a persistent hierarchy among the four types of triads. The $[+++]$ triad is consistently the most balanced and the $[++-]$ the most unbalanced triadic relation whereas the $[+--]$ and $[---]$ triads are comparably stable. The time series of the Shannon entropy corresponding to the occupation probabilities of the different triads allows one to study the activity of the system and to detect

the change points in the time series. For EVE Online, the change points are connected with wars or collapses of clusters of alliances.

The data that we analyzed clearly indicate that there is no clear tendency toward an increased occupation of the balanced triads as time progresses. This is in line with observations for the triadic relations between countries [44] and can be efficiently captured by the introduction of a finite systemic temperature. The determined values for the strength parameters confirm the importance of SBT, namely that three-body forces play a key role in the formation of the network of relationships. We also find, however, strong corrections via the two-body force. This is indicative for a strong inclination to homogenize the status of the relationships in the triads.

2.6 Acknowledgments

The authors are greatly indebted to Eðvald Gíslason and Andie Nordgren at CCP Games for their help with extracting the data for EVE Online. We are also grateful for many fruitful discussions with Benjamin Vandermarliere.

3

Social Stability and Extended Social Balance - Quantifying the Role of Inactive Links in Social Networks

The content of this chapter reflects the work published in Physica A: Statistical Mechanics and their application on March 15 of 2019 [69]. The format of the document has been adapted to match the rest of the exposition. In addition, the references are unified with the remaining chapters. y role in this work cover Conceptualization, Data curation, Formal analysis, Software, Visualization, Writing – original draft.

3.1 Abstract

Structural balance in social network theory starts from signed networks with active relationships (friendly or hostile) to establish a hierarchy between four different types of triadic relationships. The lack of an active link also provides information about the network. To exploit the information that remains uncovered by structural balance, we introduce the inactive relationship that accounts for both neutral and nonexistent ties between two agents. This addition results in ten types of triads, with the advantage that the network analysis can be done with complete networks. To each type of

triadic relationship, we assign an energy that is a measure for its average occupation probability. Finite temperatures account for a persistent form of disorder in the formation of the triadic relationships. We propose a Hamiltonian with three interaction terms and a chemical potential (capturing the cost of edge activation) as an underlying model for the triadic energy levels. Our model is suitable for empirical analysis of political networks and allows to uncover generative mechanisms. It is tested on an extended data set for the standings between two classes of alliances in a massively multiplayer on-line game (MMOG) and on real-world data for the relationships between countries during the Cold War era. We find emergent properties in the triadic relationships between the nodes in a political network. For example, we observe a persistent hierarchy between the ten triadic energy levels across time and networks. In addition, the analysis reveals consistency in the extracted model parameters and a universal data collapse of a derived combination of global properties of the networks. We illustrate that the model has predictive power for the transition probabilities between the different triadic states.

3.2 Introduction

Balance theory [47, 48] captures a lot of the emergent phenomena in the relationships observed in political networks. Principles like "a friend of a friend is also a friend" and "an enemy of my enemy is my friend" are key driving forces in those networks. The original formulation of balance theory by Heider [47] and its extension to graphs by Cartwright [48] is based on active relationships that can be friendly ("+") or unfriendly ("−"). The four types of emerging triadic relationships are categorized in two stable (= balanced) and two unstable (= unbalanced) ones, whereby one anticipates an overall tendency to create more balanced triads. Balanced triads $[+++]$ and $[+--]$ have an even number of "−" edges. Unbalanced $[++-]$ and $[---]$ triads, however, are a key ingredient in real-life political networks. Balance theory has found applications in many branches of sciences including psychology [47], studies of international networks [43–45, 50, 97], sociology [49, 83, 86, 98, 99] and ecology [84].

As balance theory introduces correlations between the edge attributes in triads, in a physics framework [51,53,68,100] it maps onto a system with predominant three-edge interactions. Associating the existence of unbalanced triads to the occurrence of a non-vanishing excitation energy, the principles of social balance can be mapped onto a model with variations in an energy landscape. Marvel et al. [51] investigated the energy landscape of a social-balance inspired system and stressed the occurrence of so-called jammed

states: network configurations with a finite number of unbalanced triads that cannot evolve to a fully balanced situation. A thermodynamical formulation of balance theory inspired by the Ising model is introduced in [53] by Lee and collaborators. With the aid of Markov chain Monte Carlo sampling, one determines probabilities to find a certain number of unbalanced triads. Connections between the simulations and real-world networks are established through the introduction of a social temperature that controls the fraction of unbalanced triads.

Conflict and the role of coalitions in networks have been addressed from the physics perspective. Galam S. in [43] proposed a model for international relationships inspired by the Ising model. The countries can belong to two coalitions and this information is encoded in their spin value $s_i \in \{-1, +1\}$. The total energy to minimize is $-\sum_{i<j} J_{ij} s_i s_j$ whereby the signs of the J_{ij} are inferred from the history of the inter-country relationships. The system works as a fully connected, weighted and signed network, as the J_{ij} can adopt a real value. The system reaches the ground state if friendly countries ($J_{ij} > 0$) are both members of the same alliance ($s_i = s_j$) and hostile countries ($J_{ij} < 0$) are in different coalitions. This system displays a certain degree of frustration for triads with hostile ties between the nodes. Vinogradova and Galam in [46] extended this model with multiple layers and used it to study the internal dynamics of the European Union. Diep, et al. in [100, 101] propose a model that describes the interaction between groups. In this model, the preference or attitude is encoded as a continuous spin, and the energy is defined as the interaction between these attitudes inside and between groups.

Next to a statistical physics framework, other approaches have been proposed to account for the occurrence of unbalanced triads. For example, Deng et al. [52] and Du et al. [102] suggested that the counteracting forces of homophily and heterophobia inhibit the network to reach a fully balanced state. Various alternative measures of balance have been proposed (for example, see [50]) and a comparative study for different networks can be found in Kirkley et al. [103].

Obviously not all relationships can be categorized as positive or negative. In international relationship studies, for example, the Swiss Confederation is an example of an agent that maintains neutrality. An extension of balance theory to include neutrals is described in [89]. This work introduces continuous edge values in the range $[-1, +1]$ and reformulates balance theory as a set of dynamical equations. Thereby, 0-valued links can be interpreted as the representation of a neutral inter-nodal relation. Under very specific circumstances, the fully connected network is found to evolve to Heider's balanced situation (or "utopia").

Lerner [82] has investigated the correlation between the geographical parameters (attributes of the countries) and the status of the inter-country relationships between 1885 to 2001. A conclusion was that social balance does not predict well the value of a possible link, but is only able to predict the value of a link conditional on its activation. This means that given that there is an active ("+" or "−") relationship, balance theory can predict its sign.

In a previous publication [68] we have shown that structural balance can be mapped onto a Boltzmann-Gibbs type of Hamiltonian model. The model assigns a specific energy to the four types of triads from structural balance and allows one to quantify their relative occupation probabilities. The proposed Hamiltonian has three-edge, two-edge and one-edge interactions that induce correlations between the edge values. In line with social balance, the three-edge term has the greatest coupling strength. Here we build on the work of [68] and take also into account the effect of inactive (neutral or nonexistent) edges in political networks. In this way, we add a dynamical aspect to the model and account for the temporal evolution of the network's topology by means of the activation or deactivation of the links. This implies that we go beyond the restrictions of signed networks and introduce inactive ("0") as a possible status of relationship between two agents in the system. This extension makes the network complete which facilitates analytical derivations in the mean-field approximation for example. In addition, triads with only active edges usually represent but a few percent of all possible triads in the political network. We anticipate that the lack of an active link is a source of information about the structure of the network and of the political relationships. Our proposed methodology can unravel that information. We test the proposed extension of balance theory against two different datasets. The first dataset is the international relationship network for the Cold War era (1949-1993) [94]. As outlined in [68], we constructed that network and its time evolution by combining data of military alliances and data of inter-state disputes [95, 96]. The second dataset stems from a virtual world called EVE Online [62]. These detailed data provide the daily evolution of the standings between the alliances of players in EVE and cover a period of over a year.

In the forthcoming section 3.3 we provide a description of the datasets used in this work. We then proceed (section 3.4) with introducing the Hamiltonian that corresponds with the proposed extended version of balance theory. We introduce some global properties that can be computed from the model (section 3.4.1). We also provide a mean-field approximation for the proposed Hamiltonian (section 3.4.2) that can be used to compute those global properties. We continue with providing detailed theory-data compar-

isons for both static (section 3.5.1) and dynamic (section 3.5.2) properties of the studied political networks. Section 3.5.3 is devoted to a comparison of the recorded network properties and the predictions from the mean-field approximation to the proposed Hamiltonian for the triadic states.

3.3 Datasets for relationships in political networks

Throughout the paper we use two different datasets on political relationships to challenge our proposed extension of balance theory.

The first dataset reports on international relationships during the Cold War era (1949-1993) and was collected by the Correlates of War initiative [94]. This dataset has records on international alliances [95] as well as interstate disputes [96] which can range from commercial or military blocks, to threats and open war. For every year we construct a signed network from the bilateral records. Thereby, two allies share a positive link whereas two countries in a dispute share a negative link. In those situations where both are present, we consider the most recent one. A pair of countries is considered to have a neutral link when no alliances nor disputes between them are registered. As an illustration, the international alliance network and its emerging structure can be seen in figure 3.1 for the year 1960. One can e.g. distinguish a clear group of positively linked Latin-American countries with also a positive link to the USA. For reasons of clarity, neutral links were not drawn. Further details on the dataset and the construction of the network are outlined in [68] together with an analysis of the resulting yearly networks.

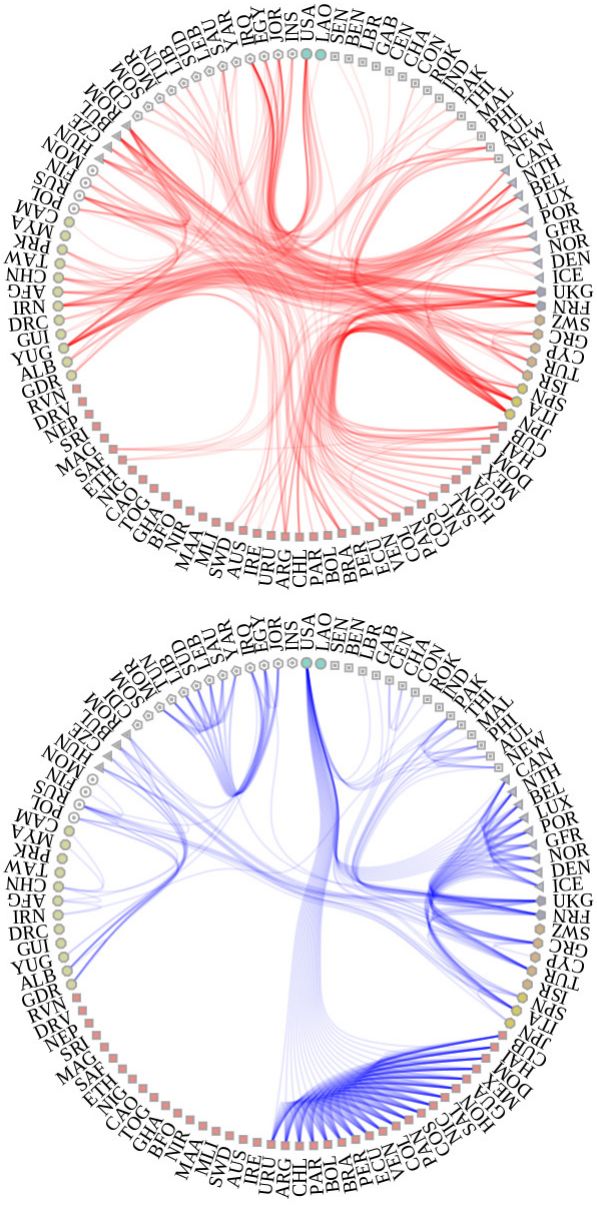


Figure 3.1: Prototypical example (1960) of the yearly Cold War alliance network. Each block (color-coded) is a group of alliances that have a similar position/play a similar role in the alliance network. For clarity, we disentangled the positive (left panel) and negative links (right panel). The inactive links are not drawn and some newly established countries (mostly former colonies gaining independence) do not have any signed edges yet. The displayed hierarchical block structure was detected with the methodology described in [104].

The second dataset was extracted from a virtual world called EVE Online [62]. EVE Online is a sandbox massive multi-player on-line game (MMOG) developed by CCP games. In this virtual world, more than 500,000 players trade, collaborate and fight in a futuristic galaxy. The players organize themselves in social structures called alliances with sizes between one and about 25,000 players. The alliances can conquer territory, where they can impose their own taxes, exploit mineral resources, and so on. Because the data comes from a virtual world, complete and accurate records of all alliance standings across time are available and advanced statistical analyses become feasible. The dataset tracks the daily evolution of the standings between the alliances of players in the MMOG and covers a period of over a year. The relations between the alliances represent an important aspect of the game as they impact a wide range of game-play experiences. The leadership of an alliance can explicitly and publicly set these relationships to friendly, hostile, neutral or undetermined. This is important because these standings affect how players from one alliance react to players from another alliance by facilitating the process of discriminating between friends, enemies, and others. Most alliances in EVE follow a "Not Blue, Shoot It" policy: friends are to be left in peace, while any other relationship means to shoot on sight. However, military, planed and coordinate actions between players have mainly enemy alliances as objectives.

For the entire data period ranging from March 2015 to April 2016, we use the standings data to construct two versions of the daily signed alliance relationship network. One version consists of the alliances that have more than 200 members (so called '+200' alliances), the other of the alliances that hold sovereignty over at least one solar system ('SOV' alliances for short). These two classes of alliances are key to the political dynamics of the game.

We note that before one can construct the undirected signed network which is required by our framework, it is necessary to make the sporadically asymmetrical standings symmetrical. The conversion rules for this symmetrization are the following: if at least one of the alliances considers the other an enemy, the link is negative. If this is not the case, and at least one of them thinks the relation is amicable, the link is positive.

The remaining combinations give rise to inactive links: neutral/neutral and neutral/not set. This symmetrization process impacts the degeneracy, i.e. the total number of possible ways to create a specific type of triad. The degeneracies are listed in table 3.1. The conversion rules are inspired by the rules of the game so that the status of the relationship between the alliances reflects the dynamics of EVE Online. For more details on the data and network construction, we refer the reader to the discussion in [68].

3.4 A Hamiltonian approach to extended social balance

3.4.1 Formalism

In a previous work [68], we proposed a Hamiltonian model with four emerging types of triads that can capture the general features of social balance, including the preference for balanced states and the importance of the three-edge forces in political graphs. The simplest version of social balance only discriminates between balanced and unbalanced triads in the graph. In our approach, social balance is the result of ordering and disordering mechanisms. The ordering is induced by the fact that the balanced triads possess a lower energy than the unbalanced ones. Exogenous effects, like resource interests, geographical and socio-economic parameters, have a disordering impact and introduce unbalanced triads. We introduce "social temperature" as a global measure for the disordering effect.

We suggest that the parameter s_{ij} that quantifies the relation between agents i and j can adopt three discrete values: $s_{ij} = +1$ for friendship, $s_{ij} = -1$ for enmity, and $s_{ij} = 0$ for a non-existent or a neutral relation. We propose to extend the Hamiltonian introduced in [68] with the possibility to activate and deactivate ties in the network. This can be achieved through the introduction of a chemical potential μ, which is the energy cost of activating a link ($s_{ij} = 0 \rightarrow s_{ij} = \pm 1$). μ is also the energy gain upon deactivating ($s_{ij} = \pm 1 \rightarrow s_{ij} = 0$) a link. We refer systematically to $s_{ij} = 0$ links as "inactive" ones.

The state of the political network consisting of N agents is uniquely defined by the numbers

$$\{s_{ij}\} \equiv s_{12}, s_{13}, \ldots, s_{1N}, s_{21}, s_{23}, \ldots s_{2N}, \ldots, s_{N1}, s_{N2}, \ldots, s_{NN-1} \; . \quad (3.1)$$

We consider undirected networks implying that $s_{ij} = s_{ji}$. The proposed extended Hamiltonian in the space of edge values reads

$$\begin{aligned} \mathcal{H}(\{s_{ij}\}) \;=\; & \frac{1}{6} \sum_{i \neq j \neq k=1}^{N} \Big[\underbrace{-\alpha \; s_{ij} s_{ik} s_{jk}}_{\text{three-edge interaction}} \; \underbrace{-\gamma \; (s_{ij} s_{ik} + s_{ij} s_{jk} + s_{jk} s_{ik})}_{\text{two-edge interaction}} \Big] \\ & + \frac{1}{2} \sum_{i \neq j=1}^{N} \Big[\underbrace{+\omega \; s_{ij}}_{\text{one-edge interaction}} \; \underbrace{+\mu \; s_{ij}^2}_{\text{chemical potential}} \Big] , \end{aligned} \quad (3.2)$$

with $s_{ij} \in \{-1, 0, +1\}$ and N the total number of agents in the network. In practice, the sum $\sum_{i \neq j \neq k=1}^{N}$ runs over all possible triads in the network. The factor $\frac{1}{6}$ accounts for the fact that the triad T_{ijk} defined by the agents

$\{i, j, k\}$ appears six times in the summation. Similarly, the factor $\frac{1}{2}$ accounts for the fact that each edge occurs twice in the summation $\sum_{i \neq j}^{N}$.

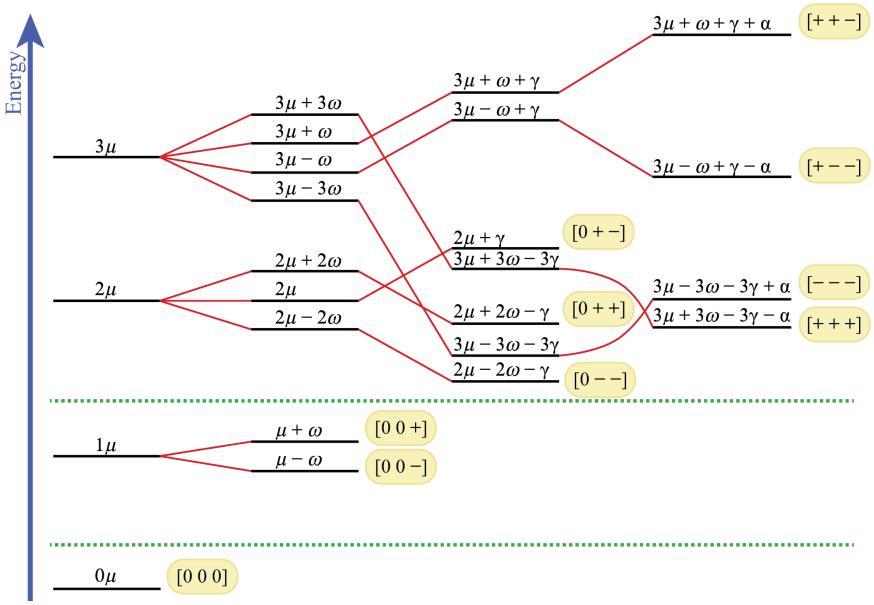


Figure 3.2: The different triadic energy levels that correspond with the Hamiltonian of equation (3.2). The order of occurrence of the different energy levels depends on the parameters. The displayed triadic energies are for $\{\alpha = 0.40, \gamma = 0.36, \omega = 0.10, \mu = 1.0\}$*, numbers that are in the ballpark of what we extract from data (see table 3.2).*

Type of triad	Symbol σ	Associated energy from Hamiltonian (3.2)	Degeneracy (Cold War)	Degeneracy (EVE)
Highly frustrated	A: $[+\,+\,-]$	$E_A = +\alpha + \gamma + \omega + 3\mu$	$g_G(E_A) = 24$	$g_T(E_A) = 525$
Lowly frustrated	B: $[-\,-\,-]$	$E_B = +\alpha - 3\gamma - 3\omega + 3\mu$	$g_G(E_B) = 8$	$g_T(E_B) = 343$
Lowly balanced	C: $[+\,-\,-]$	$E_C = -\alpha + \gamma - \omega + 3\mu$	$g_G(E_C) = 24$	$g_T(E_C) = 735$
Highly balanced	D: $[+\,+\,+]$	$E_D = -\alpha - 3\gamma + 3\omega + 3\mu$	$g_G(E_D) = 8$	$g_T(E_D) = 125$
Singly inactive v1	E: $[0\;+\,-]$	$E_E = +\gamma + 2\mu$	$g_G(E_E) = 24$	$g_T(E_E) = 840$
Singly inactive v2	F: $[0\;+\,+]$	$E_F = -\gamma + 2\omega + 2\mu$	$g_G(E_F) = 12$	$g_T(E_F) = 300$
Singly inactive v3	G: $[0\;-\,-]$	$E_G = -\gamma - 2\omega + 2\mu$	$g_G(E_G) = 12$	$g_T(E_G) = 588$
Doubly inactive v1	H: $[0\;\;0\,+]$	$E_H = +\omega + \mu$	$g_G(E_H) = 6$	$g_T(E_H) = 240$
Doubly inactive v2	I: $[0\;\;0\,-]$	$E_I = -\omega + \mu$	$g_G(E_I) = 6$	$g_T(E_I) = 336$
Fully inactive	J: $[0\;\;0\;\;0]$	$E_J = 0$	$g_G(E_J) = 1$	$g_T(E_J) = 64$

Table 3.1: The energies and degeneracies corresponding with the ten types of triadic relationships in a complete political network with edges that can adopt the values "+", "−" and "0". We refer to "+" and "−" edges as active ones, and to "0" edges as inactive ones. The triads can be separated in groups with 0, 1, 2 and 3 active edges.

The triads can be in 10 different states $\sigma \in \{A, B, C, D, E, F, G, H, I, J\}$ and the Hamiltonian (3.2) determines the corresponding energies E_σ (see table 3.1 and figure 3.2). The Hamiltonian (3.2) bears resemblance with the one from the Blume-Capel model [66,67,105,106], with the addition of a three-edge interaction related to standard social balance. The corresponding strength parameter α is anticipated to be positive. The two-edge interaction term reflects the tendency to "homogenize" the active relations in the triad. It introduces a fine-splitting of magnitude 4γ in the energy spectrum of the triads with three active edges depending on whether the balanced or unbalanced triad is symmetric or not (see table 3.1 and figure 3.2). The creation of a "+" or a "−" link between two agents in a political network comes with a "cost", and that is why we anticipate positive values of the chemical potential μ. This is represented by the last term in the Hamiltonian of equation (3.2), where s_{ij}^2 has a value of 1 if a link is active, and 0 otherwise. The proposed Hamiltonian model allows one to describe a political system as a fully connected network, where the links must be one of positive, negative or neutral/nonexistent. An alternative and more dynamic picture is that of a political system as an incomplete and irregular network with a varying number of positive or negative edges that are deactivated to and activated from neutral/nonexistent.

With the aid of the Hamiltonian of equation (3.2), the probability of finding the political network in a state $\{s_{ij}\}$ is determined by an expression of the type

$$p\left(\{s_{ij}\}\right) = \frac{e^{-\beta \mathcal{H}(\{s_{ij}\})}}{\sum_{s_{ij}=\pm 1,0} e^{-\beta \mathcal{H}(\{s_{ij}\})}}. \tag{3.3}$$

Hereby, we have introduced the positive-valued inverse temperature β in the standard definition used in physics. Accordingly, the value of the temperature is a measure for the coupling of the political network with its environment. At vanishing temperature, all T_{ijk} reside in the [0 0 0] state.

In figure 3.3 we display the time-averaged occupation probabilities for all 10 types of triads for the three political networks introduced in section 3.3. A similar type of hierarchy for the occupation probabilities is found for the EVE and Cold War data. Roughly speaking, the larger the amount of active edges in a triad, the less probable its occurrence. There is a clear difference in the occupation probabilities for the triads with no activated edge and one activated edge. More than 40% of all triads are in a [0 0 0] state. The two types of triads with one activated edge [0 0 ±] represent each about 20% of the total number of triads. The seven types of triads with three activated edges (structural balance) and two activated edges are characterized by occupation probabilities at the single-digit percent level. This implies that the four types of triads with three activated edges that

define standard structural balance, represent at most a couple of percent in the total population of triads. In section 3.5.1, the full time series of the empirical occupation probabilities of figure 3.3 will be used to learn something about the generative mechanisms in the political network. This will be done by optimizing the values of the strength parameters in the Hamiltonian of equation (3.2) so that they can capture the observations.

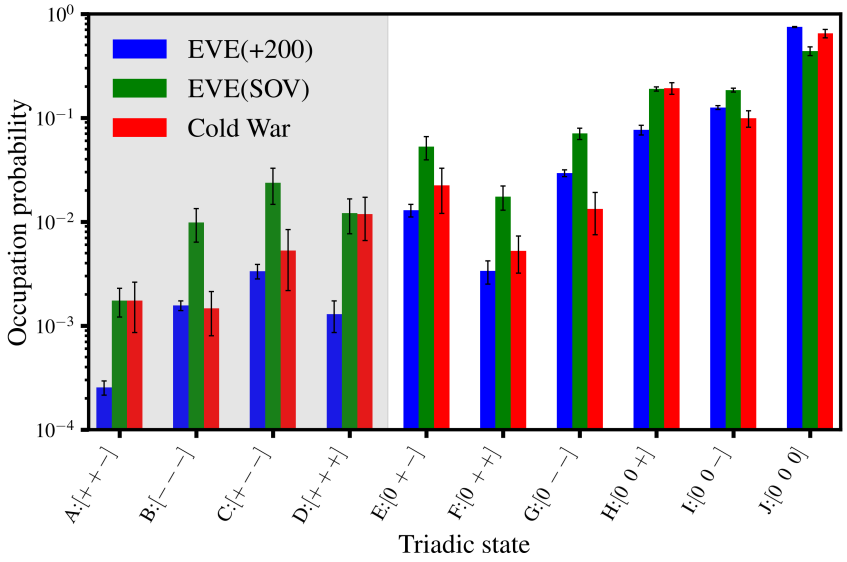


Figure 3.3: Time-averaged occupation probabilities for each of the 10 types of triads and for each of the three datasets. The shaded area corresponds with the four triads of standard structural balance theory.

Given a measure M of the political network, its expectation value is given by

$$\langle M \rangle = \sum_{s_{ij}=\pm 1,0} M\left(\{s_{ij}\}\right) p\left(\{s_{ij}\}\right). \tag{3.4}$$

For a given combination of strength parameters $\{\alpha, \gamma, \omega\}$ and an inverse temperature β, the equation (3.1) determines the probabilities $p\left(\{s_{ij}\}\right)$ for all possible network structures and the expectation values $\langle M \rangle$.

We propose two global measures that are characteristic for the political network. The first measure is related to the expectation value of the edges in the network

$$L \equiv \langle s_{ij} \rangle \qquad (-1 \leq L \leq +1)\ , \tag{3.5}$$

and can be interpreted as the "average magnetization" in the system. The

quantity L determines the average relationship between the agents in the network. The second quantity A is indicative for the fraction of the links that are part of an active ($s_{ij} = \pm 1$) relationship

$$A \equiv \left\langle s_{ij}^2 \right\rangle \qquad (0 \leq A \leq +1) \ . \tag{3.6}$$

We will refer to the A as the "average activation" in the network. In our case the edge values are one of [-1,0,1] so the use of s_{ij}^2 acts like the absolute value, but we follow the BC model [67, 105] here because the squared value has methodological advantages in the derivations below. The value of A is a measure of the tendency to create active ("+" or "−") edges in the network. Obviously, there is a strong correlation between the value of A, the value of the chemical potential μ and the temperature. Without loss of generality, in what follows we set the chemical potential as the unit of energy ($\mu \equiv 1$) and the zero of energy at the energy of the fully inactive triad [0 0 0] ($E_J = 0$).

3.4.2 Mean-field approximation

We now apply a mean-field approximation [107, 108] to the Hamiltonian of equation (3.2). This is done in the standard way, by isolating the fundamental degrees of freedom (for our purposes, the edges s_{ij}) and setting them in their environment that is due to their couplings with the other edges in the political network. To this end, we rewrite the Hamiltonian of equation (3.2) as

$$6\mathcal{H} = \sum_{i \neq j} \Bigg[\sum_{k \neq i, k \neq j} \Big[-\alpha s_{ij} s_{ik} s_{jk} - \gamma s_{ij} \left(s_{ik} + s_{jk} \right) - \gamma s_{ik} s_{jk} \Big] + 3\omega s_{ij} + 3\mu s_{ij}^2 \Bigg] \ . \tag{3.7}$$

Proceeding in the standard way by writing the edges s_{ik} and s_{jk} as the sum over their expectation value $L = \langle s_{ij} \rangle = \langle s_{jk} \rangle$ and the deviation Δ_{ij} through $s_{ij} = L + (s_{ij} - L) = L + \Delta_{ij}$, one obtains

$$\begin{aligned} 6\mathcal{H} \ = \ & \sum_{i \neq j} \Bigg[\sum_{k \neq i, k \neq j} \Big[-\alpha s_{ij} \left(L^2 + L(\Delta_{ik} + \Delta_{jk}) + \Delta_{ik} \Delta_{jk} \right) \\ & - \ \gamma s_{ij} \left(2L + \Delta_{ik} + \Delta_{jk} \right) \\ & - \ \gamma \left(L^2 + L(\Delta_{ik} + \Delta_{jk}) + \Delta_{ik} \Delta_{jk} \right) \Big] + 3\omega s_{ij} + 3\mu s_{ij}^2 \Bigg] \ . \end{aligned} \tag{3.8}$$

Expanding up to zeroth order in the deviations Δ_{ab} from the mean L, we have that

$$\begin{aligned}\mathcal{H} &= \frac{1}{2}\sum_{i\neq j}\mathcal{H}_{MF}^{ij} + \theta(\Delta) \\ &= \frac{1}{2}\Bigg[\sum_{i\neq j} -\frac{1}{3}\alpha s_{ij}L^2(N-2) - \frac{1}{3}\gamma s_{ij}2L(N-2) - (N-2)\gamma L^2 \\ &+ \omega s_{ij} + \mu s_{ij}^2\Bigg] + \theta(\Delta)\ . \end{aligned} \tag{3.9}$$

This means that in the mean-field approximation the Hamiltonian corresponding with an edge s_{ij} adopts the form

$$\begin{aligned}\mathcal{H}_{MF}^{ij} &= \mu s_{ij}^2 + \left(\omega - \frac{2\gamma}{3}(N-2)L - \frac{\alpha}{3}(N-2)L^2\right)s_{ij} + \frac{\gamma}{3}(N-2)L^2 \\ &= c_2(\mu)s_{ij}^2 + c_1(N,L,\alpha,\gamma,\omega)s_{ij} + c_0(N,L,\gamma)\ . \end{aligned} \tag{3.10}$$

Remark that c_2 is uniquely defined by the chemical potential, whereas c_1 depends on the strength parameters of the three interaction types in the Hamiltonian of equation (3.2).

The partition function corresponding with one edge can be readily obtained in the mean-field approximation by performing the summation over all possible s_{ij} in $\mathcal{H}_{MF}^{ij}$. One obtains

$$Z_{MF}^{ij} = (e^{\beta c_2} + 2\cosh\beta c_1)e^{-\beta(c_2+c_0)}\ . \tag{3.11}$$

The values of the measures L and A of Eqs. (3.5) and (3.6) can be directly computed from the partition function

$$L = \langle s_{ij}\rangle = \frac{-\partial \ln Z_{MF}^{ij}}{\partial(\beta c_1)} = \frac{-2\sinh\beta c_1(N,L,\alpha,\gamma,\omega)}{(2\cosh\beta c_1(N,L,\alpha,\gamma,\omega) + e^{\beta c_2(\mu)})}\ , \tag{3.12}$$

$$A = \langle s_{ij}^2\rangle = \frac{-\partial \ln Z_{MF}^{ij}}{\partial(\beta c_2)} = \frac{2\cosh\beta c_1(N,L,\alpha,\gamma,\omega)}{(2\cosh\beta c_1(N,L,\alpha,\gamma,\omega) + e^{\beta c_2(\mu)})}\ . \tag{3.13}$$

The equation for the L is a self-consistent equation. The self-consistent equation for the magnetization of the infinite-range Ising Hamiltonian can be retrieved from equation (3.12) in the limit of $\mu \to -\infty$ and vanishing three-edge interactions.

Whereas the average activation A is rather insensitive to changes in the strength parameters, distinct regimes for the average magnetization L can be identified. At low temperatures ($\beta^{-1} \ll \mu$) there is not enough energy to activate the links and one has $L \to 0$ and $A \to 0$. For an intermediate range of temperatures ($\beta^{-1} \approx \mu$), links start to get activated and a non-zero magnetization dependent on the parameters α, γ and ω emerges. For

combinations of parameter values that result in a positive (negative) c_1, the average magnetization is negative (positive). At high temperatures ($\beta^{-1} \gg \mu$) all triadic states are equiprobable. Accordingly, one has $L \to 0$ and $A \to \frac{2}{3}$, as there are two active ("+" and "−") and one inactive ("0") possibilities for the edge attributes.

As the system measures L and A are rather simple and common measures, the mean-field results restate some well-known features of the dynamics of political networks. Indeed, in the absence of incentives to create active ties, the network will be fully inactive. As soon as some incentives appear, the relations start to form and the links get activated. When these incentives are abundant and omnipresent, the network behaves like a random network. Many systems are expected to fall in the intermediate range where the incentives are present and create some form of order. The derived mean-field expressions (3.12) and (3.13) give rise to some non-trivial and intriguing relation between the system measures L and A and other system properties. This prediction will be the topic of investigation in section 3.5.3.

3.5 Triadic relations in empirical networks

In this section, we put the developed theoretical framework to the test for both static (subsection 3.5.1) and dynamic (subsection 3.5.2) properties of the triads in the three political networks described in section 3.3. Also the predictions of the mean-field approximation to the proposed model are compared to the observations (subsection 3.5.3).

3.5.1 Static probabilities from a quasi-equilibrium approximation

Under the assumption that the states of any two triads are independent, the probability p_σ of finding the triad T_{ijk} in a certain state σ is determined by

$$p_\sigma = \frac{g_{G,T}(E_\sigma)e^{-\beta E_\sigma}}{\sum_{\sigma'} g_{G,T}(E_{\sigma'})e^{-\beta E_{\sigma'}}} = g_{G,T}(E_\sigma)e^{-\beta E_\sigma - \ln Z_0} \,, \tag{3.14}$$

where we have introduced the normalization factor

$$Z_0 \equiv \sum_\sigma g_{G,T}(E_\sigma)e^{-\beta E_\sigma} \tag{3.15}$$

and the degeneracies $g_G(E_\sigma)$ and $g_T(E_\sigma)$ as listed in table 3.1. These degeneracies measure the number of possible combinations that give rise to a certain triad. In the absence of any preference with regard to the triadic state all energies E_σ are identical and the p_σ are proportional to the

degeneracies. By measuring the occupation probability p_σ for each of the ten triadic states and by counting the corresponding degeneracies, one can extract the triadic energies E_σ in any given time window. This allows one to determine the triadic states that are overpopulated (low energies E_σ) and are underpopulated (high energies E_σ) relative to a random network.

In reality, the states of two triads that share an edge, are not completely independent and taking those correlations into account is a notoriously challenging problem. To our knowledge, there is no formalism dealing with structural balance and its extensions that takes these correlations into account. In the forthcoming we first introduce a quantity that allows us to get a handle on the correlations in the studied network. We then proceed with extracting triadic energies E_σ under the independence assumption of equation (3.14) and discuss the results for the three empirical networks described in section 3.3.

As correlations can be anticipated to be strongest in the local neighborhood, we introduce a quantity that measures the correlations between the state σ of a triad T_{ijk} and the state of its neighbors. We define the probability of finding a triad of the type σ given that it shares n nodes with a triad of the type σ' as $G(\sigma|\sigma', n)$. The n-dependence of $G(\sigma|\sigma, n)$ is instructive about the effect of triad clustering. Figure 3.4 displays $G(\sigma|\sigma, n = 0, 1, 2)$ for the ten triadic states in the EVE(SOV) network relative to the same quantity for a network where the same number of triads are randomly assigned. For $n = 0$ one retrieves the occupation probabilities of figure 3.3 for the ten triadic states. We find that $G(\sigma|\sigma, n = 1, 2)$ significantly differs from $G(\sigma|\sigma, n = 0)$. As the number of shared nodes increases, the values of $G(\sigma|\sigma, n)$ for all triads σ tend to converge which alludes to a decreasing degree of correlations as one moves out of the local neighborhood.

From figure 3.4 we thus conclude that triads in the local neighborhood are not fully independent.

Application of the equation (3.14) to network data with few nodes is not fully justified. When zero nodes are shared, however, equation (3.14) becomes more appropriate. With increasing system size, the proportion of local triads decreases and it can be anticipated that the validity of equation (3.14) improves.

From the observed occupation probabilities p_σ, we can extract the energies E_σ for each time window using (3.14). The time series of the extracted E_σ are displayed in figure 3.5 for the EVE data and in figure 3.6 for the Cold War data. From figure 3.5 one can conclude the following: For both versions of the EVE network, one can discern a similar hierarchy in the relative position of the different types of triads. First, unbalanced triads (dashed lines) have a higher energy than balanced triads (solid lines) which

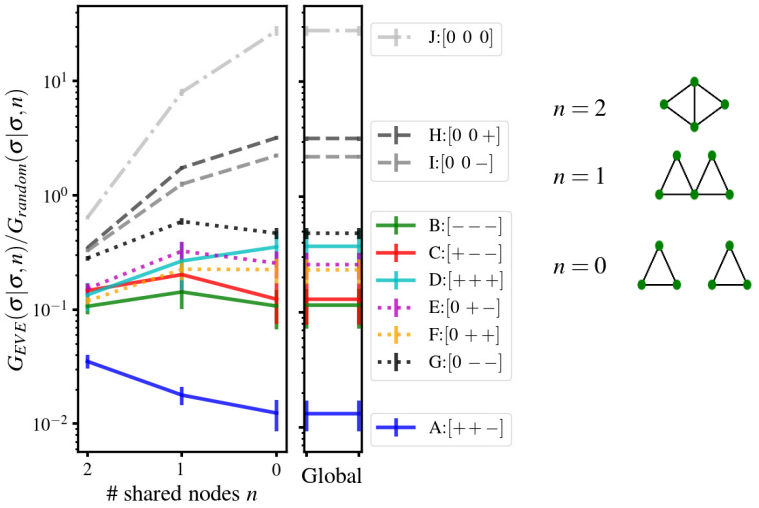


Figure 3.4: The conditional probabilities $G(\sigma|\sigma,n)$ in the EVE(SOV) dataset $(= G_{EVE}(\sigma|\sigma,n))$ relative to a random network $(= G_{random}(\sigma|\sigma,n))$. The $G(\sigma|\sigma,n)$ expresses the probability to find a triad in the state σ sharing n nodes with another triad in the state σ. We use a dot-dashed line for the triads with no activated edges, dashed line for triads with 1 activated edge, dotted line for triads with 2 activated edges, and a solid line for the triads with 3 activated edges. As a consistency check we also show the probability of finding a triad of type σ without taking into account the triad's neighborhood.

confirms the predictions of standard structural balance. Second, complete triads (non-gray colors) consistently have higher energies (less stable) than incomplete triads. Third, incomplete triads with only one inactive link are more energetic (less stable) than triads with two inactive links. This hierarchy inferred from the latter two observations suggests that the chemical potential is an essential term in the dynamics governing the formation of triads. As could already be inferred from figure 3.3, there is clear separation in energy scale between the triads with zero and one active link. There is not a clear separation between the energies of the triads with two and three active edges.

Furthermore, the close relative energy position for triads that are similarly affected by the chemical potential μ (e.g. $[0 \quad 0 \; +]$ and $[0 \quad 0 \; -]$), hints that ω is only a small perturbation. Finally, we note that the stability of the hierarchy of the energy levels over time suggests a roughly constant temperature.

As can be seen from figures 3.5 and 3.6, the ordering of the extracted triadic energies is similar for the Cold War and EVE political networks.

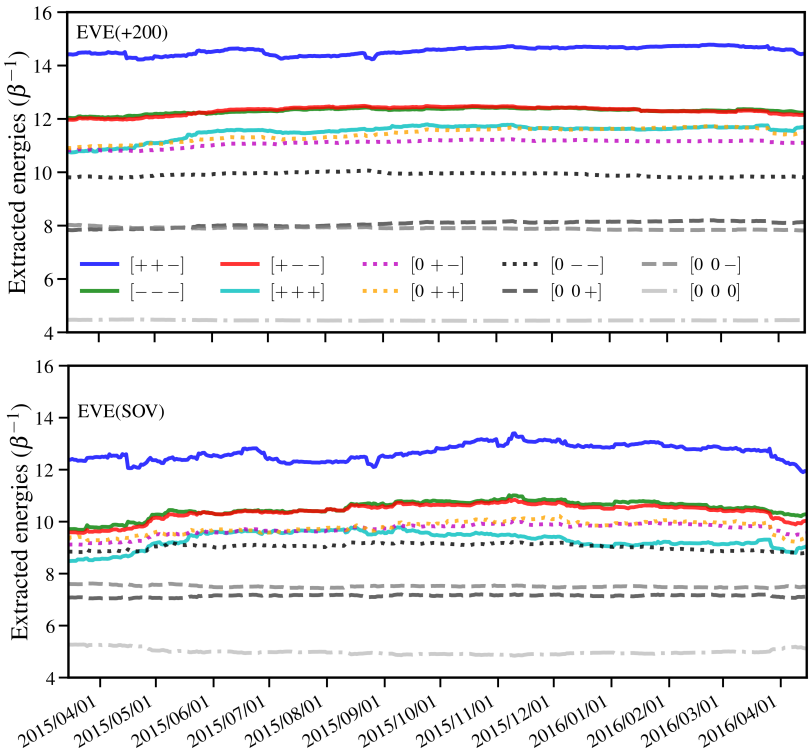


Figure 3.5: The logarithm of the daily occupation probabilities $-\ln p_\sigma / g_T(E_\sigma)$ *for the ten triadic states* σ *for the (top panel) EVE(+200) and (bottom panel) EVE(SOV) of the alliances' network.*

However, there are a few differences for the complete triads. The energy gap for the strongly balanced $[+\,+\,+]$ and the other three triads is higher for the real world than for EVE. Also, the energy separation between the other complete triads is smaller. Even so, these differences are rather small, so we expect similar values for the parameters. One also observes, as time progresses, a slight divergence between the higher and lower energy levels which hints at a cooling process in the political network of countries.

From the extracted energy levels, one can infer the force strength parameters of the Hamiltonian $\{\alpha, \gamma, \omega\}$ and the value of the chemical potential μ using the relations contained in table 3.1. To this end, we use the time-averaged values $\overline{E_\sigma}$ with the standard deviation as a measure for the error. Through a χ^2 fit we find the parameters as listed in table 3.2. The results show a consistent hierarchy between the values of $\{\alpha, \gamma, \omega, \mu\}$ with the chemical potential as dominant term. In line with the underlying ideas of

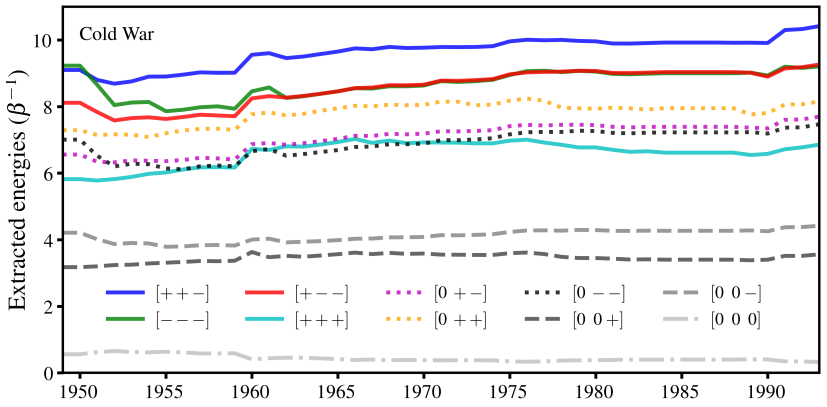


Figure 3.6: The logarithm of the yearly occupation probabilities $-\ln p_\sigma / g_G(E_\sigma)$ for the ten triadic states σ of the inter-country relationships network during the Cold War era (1949-1993).

	α (β^{-1})	γ (β^{-1})	ω (β^{-1})	μ (β^{-1})	$\ln(Z_0)$
EVE (SOV)	0.76 ± 0.20	0.43 ± 0.08	-0.10 ± 0.06	2.08 ± 0.06	5.22 ± 0.08
EVE (+200)	0.73 ± 0.14	0.54 ± 0.06	$+0.06 \pm 0.05$	3.06 ± 0.05	4.48 ± 0.06
Cold War	0.68 ± 0.23	0.42 ± 0.09	-0.09 ± 0.07	3.00 ± 0.07	0.57 ± 0.11

Table 3.2: The extracted values for the strength parameters in the Hamiltonian (3.2) for the +200 and SOV alliances in EVE Online, and for the international relations during the Cold War era. In a random network $\alpha, \gamma, \omega, \mu = 0$

structural balance and with the results obtained in [68], the strength parameter α of the three-edge force is larger than the strength parameter γ of the two-edge force. The positive value of μ reveals that there is a substantial cost to establishing a relationship between two agents. The sizable positive value of the two-edge term γ indicates that there are strong tendencies for homogenization in the three-node cycles. For the EVE data, the extracted values for ω are small. This alludes to the fact that in EVE there are comparable incentives for "+" and "−" relationships between the alliances of players. It is remarkable, however, that the extracted parameters $\{\alpha, \gamma, \omega, \mu\}$ for all three studied networks are of the same order of magnitude. This gives rise to a rather universal hierarchy of the triadic energy states as it is shown in figure 3.2 and could also be inferred from the occupation probabilities of figure 3.3.

3.5.2 Dynamic transition probabilities

Starting from equation (3.14) one can predict the transition probabilities between the 10 triadic states listed in table 3.1 and displayed in figure 3.2. In equilibrium, the relative occupation probabilities between the different triadic states are maintained. The normalized single-edge transition probabilities for a triad to change from state σ to state σ' define the 10×10 non-symmetric probability matrix $\mathcal{M}$[1]

$$\mathcal{M}_{\sigma\sigma'} = \frac{g_{\sigma\to\sigma'} \exp -\beta \left(E_{\sigma'} - E_{\sigma}\right)}{\sum_{\sigma''} g_{\sigma\to\sigma''} \exp -\beta \left(E_{\sigma''} - E_{\sigma}\right)} \,. \tag{3.16}$$

Here, $g_{\sigma\to\sigma'}$ denotes the total number of micro-states identified as a σ' triadic state that are accessible from σ through changing exactly one of the attributes of the micro-state. Examples of attribute changes can be a sign change in one of the directed links between EVE alliances, or a change in an alliance treaty that corresponds with war declaration in the real-world data of inter-country relationships. The $g_{\sigma\to\sigma'}$ can be interpreted as the dynamical equivalent of the degeneracies g_G and g_T listed in table 3.1.

For a considerable amount of transitions in the undirected network of inter-country relationships during the Cold War era, one has $\mathcal{M}_{\sigma\sigma'} = 0$ as the corresponding $g_{\sigma\to\sigma'} = 0$. This captures the fact that some triadic transitions are forbidden by means of changing the value of a single edge (e.g. the transition from $[+\,+\,+]$ to $[-\,-\,-]$). For the network of alliances in EVE the situation is different. Specifically, in determining the $g_{\sigma\to\sigma'}$ we start from the raw data for the edge attributes with occasional asymmetrically directed edges between two alliances in EVE.

In figure 3.7 we compare the predicted probability matrix $\mathcal{M}$ for the network of EVE(+200) alliances with the recorded transitions on April 9-10, 2015. This choice is inspired by the fact that on that particular day all possible single-edge transitions are recorded. We stress that some of those intra-day transitions are very infrequent. To compare the data with the model, we use (3.16) in combination with the extracted values for the energy values E_σ (see figure 3.5). The fact that the $g_{\sigma\to\sigma'}$ are computed with the aid of directed networks, explains why the diagonal $g_{\sigma\to\sigma}$ are non-vanishing, resulting in large diagonal $p_{\sigma\to\sigma}$. For example, the combinations $(s_{1\to2} = +1, s_{2\to1} = 0)$ and $(s_{1\to2} = +1, s_{2\to1} = +1)$ both give rise to $s_{12} = +1$. The event of agent 2 changing its attitude towards agent 1 (denoted as $s_{2\to1}$) from "0" to "+" is recorded as a transition. Nevertheless, the initial and final edge value $s_{12} = +1$ in the undirected network did not change.

[1] This is not the only matrix that can maintain the occupation probabilities under a Markov chain Monte Carlo, however this is the consistent with classical statistical physics.

We find that our model for the transition probabilities has predictive power in that the transitions above the diagonal (that correspond to transitions to states with a higher energy) are less common than those below the diagonal. The recorded and computed values of $\mathcal{M}_{\sigma\sigma'}$ extend over several orders of magnitude. In the recorded daily transition data we find ample transitions with a finite transition probability whereas the corresponding single-step transition probability $\mathcal{M}_{\sigma\sigma'} = 0$. This discrepancy is a clear indication that multiple changes of a triad's state can occur during the measurement interval of a single day. Hence, we face a situation where the time scale of the model and of the data recording is different. As a matter of fact, an extension of the single-edge triad transitions of equation (3.16) to multiple-edge triad transitions is in order.

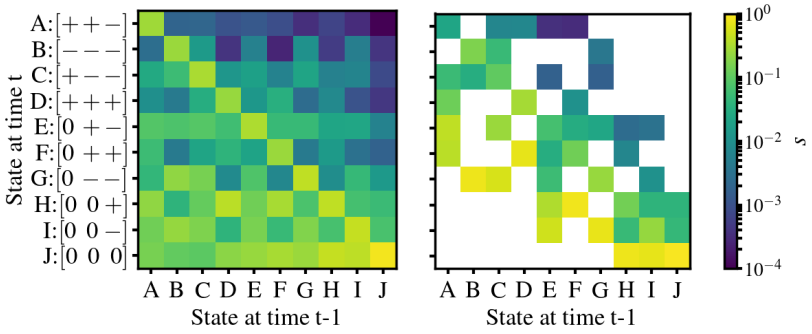


Figure 3.7: The matrix $\mathcal{M}_{\sigma'\sigma}$ for the intra-day single-edge transition probabilities between the ten triadic states in the EVE(+200) network. Left panel: the recorded transitions from April 9 to April 10, 2015. Right panel: the predicted single-edge transitions of equation (3.16). White sites correspond with $\mathcal{M}_{\sigma\sigma'} = 0$.

Let us define $W(n;\Delta t)$ as the probability that an edge is subjected to n updates in the time-measurement interval Δt. This allows us to write the multiple-edge transition probability $P_{\sigma\to\sigma'}(\Delta t)$ to move from triadic state σ to triadic state σ' in an arbitrary number of edge changes during the time interval Δt as:

$$\begin{aligned} P_{\sigma\to\sigma'}(\Delta t) &= W(0;\Delta t)\delta_{\sigma,\sigma'} + W(1;\Delta t)\mathcal{M}_{\sigma\sigma'} \\ &+ W(2;\Delta t)\sum_i \mathcal{M}_{\sigma i}\,\mathcal{M}_{i\sigma'} \\ &+ W(3;\Delta t)\sum_{i,j} \mathcal{M}_{\sigma i}\,\mathcal{M}_{ij}\,\mathcal{M}_{j\sigma'} + \cdots . \end{aligned} \tag{3.17}$$

For brevity of notation, we drop the explicit dependence of $W(n;\Delta t)$ and $P_{\sigma\to\sigma'}(\Delta t)$ on Δt. In order to normalize the multiple-edge transition probabilities $P_{\sigma\to\sigma'}$ of equation (3.17), we make use of the fact that

the single-edge transition probabilities of equation (3.16) are normalized: $\sum_{\sigma'} \mathcal{M}_{k\sigma'} = 1, \forall k$. This implies that

$$\begin{aligned}\sum_{\sigma'} P_{\sigma\to\sigma'} &= \sum_{\sigma'}\Big[W(0)\delta_{\sigma,\sigma'} + W(1)\mathcal{M}_{\sigma\sigma'} + W(2)\sum_i \mathcal{M}_{\sigma i}\,\mathcal{M}_{i\sigma'} \\ &\quad + W(3)\sum_{i,j}\mathcal{M}_{\sigma i}\,\mathcal{M}_{ij}\,\mathcal{M}_{j\sigma'} + ...\Big] \\ &= W(0) + W(1) + W(2) + W(3) + \dots \ . \end{aligned} \tag{3.18}$$

This means that the multiple-edge transition probabilities $P_{\sigma\to\sigma'}$ can be normalized under the condition that

$$\sum_{n\geq 0} W(n) = 1 \ . \tag{3.19}$$

Assuming that $W(n \geq 1) = [W(1)]^n$ one obtains an expression that connects $W(0)$ to $W(1)$

$$1 = W(0) + \sum_{n\geq 1}[W(1)]^n = W(0) + \frac{W(1)}{1 - W(1)} \ . \tag{3.20}$$

As the $W(n)$ are probabilities, the required convergence criteria are met. As a matter of fact, as $W(0)$ is a probability one has $0 \leq W(n) < \frac{1}{2}$.

In view of our above derivations, our final expression for the multiple-edge transition probability between the triadic states σ to σ' in the time interval Δt reads

$$P_{\sigma\to\sigma'}(\Delta t) = \frac{1 - 2W(1;\Delta t)}{1 - W(1;\Delta t)}\delta_{\sigma,\sigma'} \ + \ \left[\sum_{n\geq 1} W^n(1;\Delta t)\mathcal{M}^n\right]_{\sigma\sigma'} \ . \tag{3.21}$$

To determine the functional form of the probability $W(1;\Delta t)$ we note there should be a dependence on the temperature and the constraint $0 < W(1;\Delta t) < 0.5$. We propose the expression $W(1;\Delta t) = 0.5\exp(-E_p/T)$ inspired by equation (3.14) with a positive-valued E_p. In line with our intuition, this expression gives rise to a higher number of transitions for higher temperatures.

To test the accuracy of the proposed model, we compare the theoretical multiple-edge transition probabilities of equation (3.21) with the observed ones. Given a set of parameters $\left\{\frac{\alpha}{\mu}, \frac{\gamma}{\mu}, \frac{\omega}{\mu}\right\}$, an energy-activation-cost (E_p) and a temperature (β^{-1}), one can compute the transition probabilities of equation (3.21) for each time window (intra-day for the EVE data and intra-year for the Cold War data). We stress that diplomatic relationships

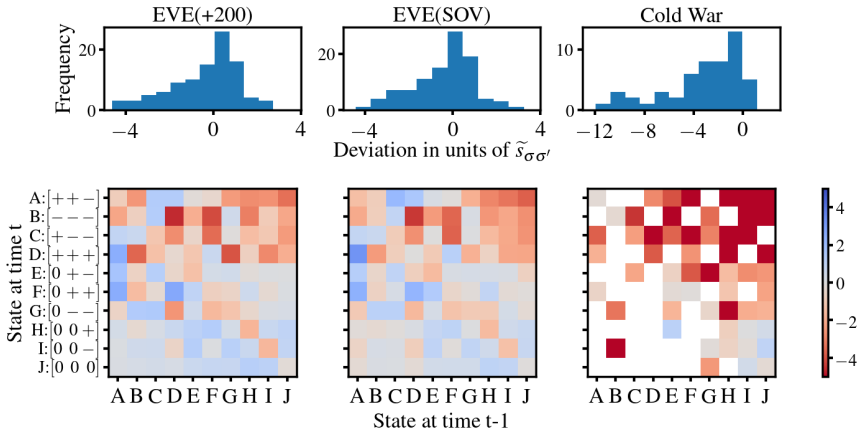


Figure 3.8: The quality of agreement (quantified according to the equation (3.22)) between the predicted and the observed multiple-edge transition probabilities $P_{\sigma\to\sigma'}(\Delta t)$ between the ten triadic states. The difference is positive (negative) when the model overestimates (underestimates) the mean of the observed transition probabilities. Left panels: EVE(+200); middle panels: EVE(SOV); right panels: Cold War era. In the upper row we show the histograms of the $\chi_{\sigma\sigma'}$ values. In the lower row we show the matrix with $\chi_{\sigma\sigma'}$ values whereby light colors correspond to transition probabilities that are well predicted by the model.

between alliances in EVE are highly volatile, in the sense that multiple alliances can change their attitudes within a time frame of a few days. In contrast, the network of international relationships during the Cold War era is rather static and an analysis of the transition probabilities is really challenging. From the data described in section 3.3 we have collected the recorded transition probabilities for each of the 100 $\sigma \to \sigma'$ combinations. From those samples we determined the expectation value $\widetilde{\mu}_{\sigma\sigma'}$ and the standard deviation $\widetilde{s}_{\sigma\sigma'}$ of the logarithm of the transition probabilities. For every transition $\sigma \to \sigma'$, the quality of agreement between the model of equation (3.21) and the data is quantified by means of the difference between the predicted and measured transition probability relative to the variance of the distribution of the recorded $\sigma \to \sigma'$ transitions

$$\chi_{\sigma\sigma'} = \frac{\log\left(P_{\sigma\to\sigma'}(\Delta t)\right) - \widetilde{\mu}_{\sigma\sigma'}}{\widetilde{s}_{\sigma\sigma'}} \,. \tag{3.22}$$

Figure 3.8 shows the quality of agreement for the transition probabilities between the model and the data for the three networks considered in this work. For both versions of the EVE alliance network, the majority of the values for the difference between predicted and observed transition probabilities is smaller than $2\widetilde{s}_{\sigma\sigma'}$. We also learn that the model underestimates

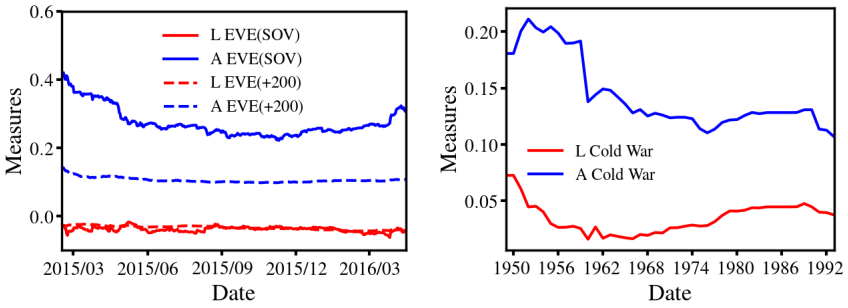


Figure 3.9: Time series of the global measures (A, L) for the networks of inter-alliance relationships in EVE (left) and of the inter-country relationships during the Cold War (right).

the perseverance of the ground state, predicting a lower probability to stay in [0 0 0] than observed in the data. The model underestimates some transitions from balanced to unbalanced triangles (e.g. $[+\,+\,+]$ to $[-\,-\,-]$). The edge attributes of the network of inter-country relationships during the Cold War era are far more static than for the EVE networks. Accordingly, the number of recorded transitions is relatively small and about 40% of the possible triadic transitions are not even recorded during the half-century covered by our data. As for the EVE data, the model severely underestimates the transition from a triadic state with at most one active edge, to one with three active edges (upper right corner of the transition matrix). Remarkably, the reversed transitions (lower left corner of the transition matrix) are reasonably well predicted. This seems to indicate that there is an incentive to generate triads with active edges that is not captured by our Hamiltonian approach. This could be related to another common idea in political science: "You're either with us, or against us". In other words, once one has revealed a few standings it often implies one is forced to announce others. In the model proposed here, this principle could be translated in an extra cost to maintain unclosed triangles.

3.5.3 Global measures

We now provide an analysis of the two global measures L (average magnetization) and A (average activation) and compare them with the mean-field predictions of section 3.4.2. Figure 3.9 displays the time series of the extracted L and A for the three political networks considered in this work. In the virtual world of EVE, L is small and negative for both types of alliances. This indicates a small preference for a negative interaction between

alliances in the virtual EVE world. The activation A is markedly different for the two versions of EVE's alliance networks. For the network of large alliances EVE(+200), about 10% of the links are active. For the alliances with sovereignty EVE(SOV), this is about 30%. This substantial difference can be attributed to a size effect. We observe that the number of activated edges does not increase quadratically with the number of nodes. As a result the fraction of active links decreases with the number of nodes in the network. Even with an infinite number of nodes, each node interacts with a finite number of other nodes. For the Cold War network, the activation is subject to substantial time variations. Starting at around 20% in the fifties, A drops to 12-15% in the sixties. In those days many countries gained independence and it took them some time to create active ties. For a similar reason the collapse of the USSR (1990-1991) leads to a sudden drop in activation. In contrast to the EVE network, we observe a positive magnetization L.

The Eqs. (3.12) and (3.13) express $A = \langle s_{ij}^2 \rangle$ and $L = \langle s_{ij} \rangle$ as a function of the temperature and two functions $c_2 = \mu$ and $c_1(N, L, \alpha, \gamma, \omega)$. From Eqs. (3.12) and (3.13) we can extract βc_1 and βc_2 from given samples of the combination (L, A)

$$\beta c_1 = \operatorname{arctanh}\left(\frac{-L}{A}\right) = \beta\left(\omega - \frac{2\gamma}{3}(N-2)L - \frac{\alpha}{3}(N-2)L^2\right) \quad (3.23)$$

$$\beta c_2 = \ln\left(\left[\frac{1}{A} - 1\right] 2\cosh\left[\operatorname{arctanh}\left(\frac{-L}{A}\right)\right]\right) = \beta\mu\,. \quad (3.24)$$

Upon close inspection of figure 3.9 it is clear that L is of the order of 0.1. As N of the order of a few hundred, one has $(N-2)\cdot L \approx 1$. This implies that the ratio of Eqs. (3.23) and (3.24) can be approximately written as

$$\begin{aligned} \mathcal{G}_{MF}(L, A) &\equiv \frac{\operatorname{arctanh}\left(\frac{-L}{A}\right)}{\ln\left(\left[\frac{1}{A} - 1\right] 2\cosh\left[\operatorname{arctanh}\left(\frac{-L}{A}\right)\right]\right)} \\ &= \frac{\operatorname{arctanh}\left(\frac{-L}{A}\right)}{\beta\mu} \approx \frac{\omega}{\mu} - \frac{2\gamma}{3\mu}(N-2)L\,. \end{aligned} \quad (3.25)$$

As a matter of fact, this means that under conditions of constant parameters $\{\alpha, \gamma, \omega, \mu\}$, the mean-field model developed in section 3.4.2 predicts a linear relationship between the well-defined scaling function $\mathcal{G}_{MF}(L, A)$ and $L(N-2)$ for small values of L. The scaling function can be rewritten in a more convenient form

$$\mathcal{G}_{MF}(L, A) \equiv \frac{\ln(1 - \frac{L}{A}) - \ln(1 + \frac{L}{A})}{2\ln(2) + 2\ln(\frac{1}{A} - 1) - \ln(1 - \frac{L^2}{A^2})}\,. \quad (3.26)$$

We evaluate this predicted non-trivial data collapse in figure 3.10 by plotting the l.h.s. of equation (3.25) versus $L(N-2)$ for the data samples of (L, A) collected from the time series of the three studied political networks. The results are encouraging. The values of $\mathcal{G}_{MF}(L, A)$ for the EVE(+200) network nicely cluster along a line. For the EVE(SOV) network two distinct lines are discerned. One of the lines of the EVE(SOV) data concerns the first 100 days in the time series, whereas the second one refers to the rest of the data. Also the Cold War data cluster in two distinct lines separating the first decade of the data from the remainder. Apparently, at the end of the 1950s-start of the 1960s, a series of events has induced changes in the inner workings of the network of inter-country relationships. This trend change coincides with the period between the Hungarian Revolution of 1956 and the construction of the Berlin wall (1961). The Hungarian Uprising of 1956 started the outflow of people out the East block that culminated with the surprise erection of the wall in 1961. Before this period, there was still some communist movement in Europe, and there were open lines and co-operation between the blocks. It could have still gone in the cooperative direction, but the violent response to the Budapest revolution and the construction of the Berlin wall destroyed this idea, creating a schism between the communist world and the western powers. After this, the world enters into a situation of clear blocks and every country was almost forced to affiliate itself with one or the other power. This new rule of "if you are not a friend, you are an enemy" increases the pressure of the triadic closure. This change in the dynamics gives rise to two different slopes in the scaling analysis of the data. The mean-field scaling function of equation (3.25) allows one to detect changes in the way the international network rules itself. Indeed, it goes without saying that parameters ruling complex dynamic systems may fluctuate over time [109]. We mention that for both the EVE(SOV) and the Cold War data there are indications for the change points in the time series displayed in figure 3.9 but that the analysis of figure 3.10 provides a clearer way to separate the different dynamical regimes. The scatter plot of A versus L in figure 3.10 illustrates that the two global measures of the networks are not linearly correlated. In those scatter plots, however, the data for the three political networks tend to cluster in regions. For the Cold War and EVE(SOV) data two distinct clusters can be discerned.

3.6 Conclusion

Standard structural balance builds on active (i.e. "+" and "−") edges and often succeeds in capturing the major features of the formation and evolution of alliances in political networks. Standard social balance, however,

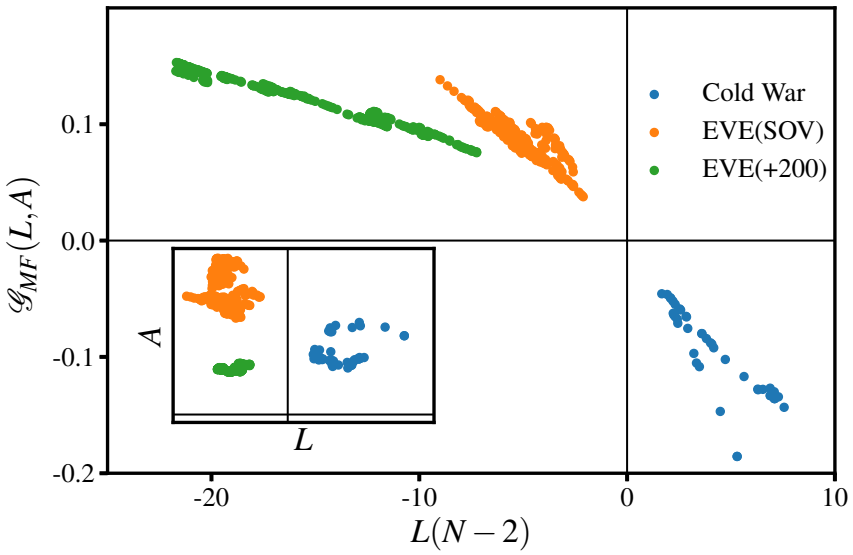


Figure 3.10: Results of the empirical analysis of the predicted mean-field result (3.25) for the (L, A) data of the EVE(SOV), EVE(+200) and Cold War networks. The inset shows the (L, A) scatter plots.

corresponds with a very small fraction (order of a single percent) of the triads in political networks which raises the question whether all information is being used. We have added the inactive edges and this has some clear advantages. Indeed, it introduces an additional layer of dynamics as active edges can change to inactive ones and vice versa. In addition, adding the inactive edge attribute allows one to analyze the political networks as complete networks. This facilitates a mean-field approach and the introduction of some representative global quantities of the network like the activation (fraction of active edges in the network) and the magnetization (average over edges states).

We have proposed a Hamiltonian that deals with positive, negative and neutral/nonexistent relationships as a representation of the generative dynamics in political networks. The Hamiltonian contains single-edge, double-edge and triple-edge interactions inducing correlations in the edge attributes ("+", "−" or "0"). The Hamiltonian adds a three-edge interaction (that accounts for the major effect from structural balance) to the one from the Blume-Capel model. In a Boltzmann-Gibbs framework the model is suitable for theory-data comparisons and allows one to uncover and quantify the major dynamical mechanisms (including change points) at play in the

political network. We have put the Hamiltonian model as well as its mean-field approximation to the test with three extended sets of time-series data for the relationships between agents in political networks. Two sets of data are from the virtual world EVE Online and one set from the data about the inter-country relationships during the Cold War era. Topics of investigation across the three networks considered have been the occupation probabilities of the ten triadic states as well as the transition probabilities among those states. The occupation probabilities can be connected to the force terms in the Hamiltonian and some features are universal across the three network. We found that the cost of creating an active (hostile or friendly) connection is higher than any other interaction component between the nodes. This locates the ground state of the system at a set of nodes with exclusively inactive connections in contrast to previous models which took universal utopia as the minimal energy system state. An empirical analysis of the three political networks revealed that about half of the possible triads reside in the [0 0 0] state. There is a plethora of incentives (captured by the concept of temperature in our model) that result in agents interacting in hostile or friendly ways and causes the political system to reside in an excited state.

With regard to the transition probabilities a fair model-data agreement is obtained for the EVE data where changes in the edge attributes are abundant. In the Cold War data the edge attributes are rather static and the model does not describe the data very well. Across the three networks the model systematically underestimates a peculiar type of transition that is referred to as triadic closure. This corresponds with the transition from a triadic state with one active and two inactive edges, to a triadic state with exclusively active edges. This implies that upon activating one edge in the triad there are strong incentives to close the triad that are not captured by our current model.

The mean-field approximation for our model of political networks can be developed along the lines of the Blume-Capel model and results in a set of two self-consistent equations for the average magnetization L and the average activation A of the edges. Those two global variables allow one to sketch the phase boundaries between ordered and disordered realizations of the political networks. We find that the three studied networks fall in an intermediate partially ordered state. Thanks to the mean-field theory we could uncover a linear scaling between a non-trivial function of (L, A) and L. An analysis of the empirical (L, A) data provided evidence for the predicted linear scaling. The mean-field prediction also turned out to be an efficient method for change-point detection and identify the different modes of operation as the political networks evolve over time.

Acknowledgments

This work was supported by Fund of Scientific Research - Flanders (FWO) (JR, KS, MvdH,BV) and Research Fund of Ghent University (AMB, KH). The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

4

Social Balance Theory, Supplementary Material

This chapter contains supplemental material with a detailed study of a network of interacting nodes determined by the three-body Hamiltonian presented in [69] that is a translation of the principles of social balance and captures well the structure of political networks in EVE online and in the real world (network of countries during the cold-war era). In what follows we develop a mean-field approximation for the Hamiltonian and present results of Monte-Carlo simulations based on it. With the presented mean-field results and Monte-Carlo simulations we aim at investigating the phase diagram that corresponds to the Hamiltonian defined to capture the principles of social balance.

4.1 Mean-field approximation

Building on the work of [110], we develop a mean-field approximation to obtain mean-field estimates of $\langle s_{ij} \rangle$ (the expectation value for a link between two nodes) and of $\langle s_{ik} s_{kj} \rangle$ (the average correlation between two links that share a node). To this end we start from the Hamiltonian as it was

introduced in [69] and discussed in great detail in Chapter 3 of this work

$$\begin{aligned}\mathcal{H}\left(\{s_{mn}\}\right) &= \frac{1}{6}\sum_{i\neq j\neq k=1}^{N}\Big[-\alpha\, s_{ij}s_{ik}s_{jk} - \gamma\,\left(s_{ij}s_{ik}+s_{ij}s_{jk}+s_{jk}s_{ik}\right)\Big] \\ &\quad +\frac{1}{2}\sum_{i\neq j=1}^{N}\Big[\,+\omega\, s_{ij} \;+\; \mu\, s_{ij}^{2}\Big]\,. \end{aligned} \tag{4.1}$$

The Hamiltonian has four "strength parameters" and is defined in the space of configurations of all possible edge combinations $\{s_{mn}\}$ whereby each value $s_{mn} \in \{-1, 0, +1\}$. Efficient counting over the possible nodes, dyads and triads results in

$$\begin{aligned}\mathcal{H}\left(\{s_{mn}\}\right) &= \sum_{i<j<k=1}^{N}\Big[-\alpha\, s_{ij}s_{ik}s_{jk} - \gamma\,\left(s_{ij}s_{ik}+s_{ij}s_{jk}+s_{jk}s_{ik}\right)\Big] \\ &\quad +\sum_{i<j=1}^{N}\Big[\,+\omega\, s_{ij} \;+\; \mu\, s_{ij}^{2}\Big]\,. \end{aligned} \tag{4.2}$$

We can write the above Hamiltonian $\mathcal{H}\left(\{s_{mn}\}\right)$ as the sum of two parts, whereby the first part contains all the terms that depend explicitly on a tagged link s_{ij} and the second part contains all terms that do not depend on the tagged link s_{ij}

$$\mathcal{H}\left(\{s_{mn}\}\right) = \mathcal{H}_{I}\left(s_{ij}\right) + \mathcal{H}'\left(\{s_{mn} \neq s_{ij}\}\right)\,, \tag{4.3}$$

whereby one has that

$$\mathcal{H}_{I}\left(s_{ij}\right) = s_{ij}\left(\omega - \sum_{k\neq j,i}\left[\alpha s_{ik}s_{kj} + \gamma(s_{ik}+s_{kj})\right]\right) + s_{ij}^{2}\mu\,. \tag{4.4}$$

We now compute the expectation value of a tagged link s_{ij} in the standard fashion

$$\begin{aligned}\langle s_{ij}\rangle &= \sum_{s_{ij}=\pm 1,0} s_{ij}\mathcal{P}(s_{ij}), \\ &= \frac{\sum_{s_{ij}=\pm 1,0}\sum_{s_{mn}\neq s_{ij}} s_{ij}e^{-\beta\left(\mathcal{H}_{I}(s_{ij})+\mathcal{H}'(\{s_{mn}\neq s_{ij}\})\right)}}{\sum_{s_{ij}=\pm 1,0}\sum_{s_{mn}\neq s_{ij}} e^{-\beta(\mathcal{H}_{I}(s_{ij})+\mathcal{H}'(\{s_{mn}\neq s_{ij}\}))}}\,. \end{aligned} \tag{4.5}$$

In the mean-field approximation one can approximate this expectation value for the value of an edge $-1 \leq \langle s_{ij}\rangle \leq +1$ as

$$\langle s_{ij}\rangle \approx \frac{\sum_{s_{ij}=\pm 1,0} s_{ij}e^{-\beta\mathcal{H}_{I}(s_{ij})}}{\sum_{s_{ij}=\pm 1,0} e^{-\beta\mathcal{H}_{I}(s_{ij})}}\,. \tag{4.6}$$

In this expression the summation over the three values of s_{ij} can be done and one arrives at

$$\langle s_{ij}\rangle \approx \frac{-2\sinh\beta\left(\omega-\sum_{k\neq j,i}\left[\alpha s_{ik}s_{kj}+\gamma(s_{ik}+s_{kj})\right]\right)}{2\cosh\beta\left(\omega-\sum_{k\neq j,i}\left[\alpha s_{ik}s_{kj}+\gamma(s_{ik}+s_{kj})\right]\right)+e^{\beta\mu}}\,. \tag{4.7}$$

In the spirit of the mean-field approximation the remaining sums $\sum_{k\neq j,i}$ can be replaced by averages using $\langle x\rangle=\frac{1}{N}\sum_{i=1}^{N}x$ and one obtains

$$\langle s_{ij}\rangle \approx \frac{-2\sinh\left(\beta\left(\omega-\alpha(N-2)\langle s_{ik}s_{kj}\rangle-\gamma(N-2)2\langle s_{ij}\rangle\right)\right)}{2\cosh\left(\beta\left(\omega-\alpha(N-2)\langle s_{ik}s_{kj}\rangle-\gamma(N-2)2\langle s_{ij}\rangle\right)+e^{\beta\mu}\right.}\,. \tag{4.8}$$

For each combination of the parameters $(\alpha,\gamma,\omega,\mu)$ and the temperature β, this expression connects the expectation value for a link to the expectation value of a product of two links sharing a node $\langle s_{ik}s_{kj}\rangle$.

We compute the expectation value $\langle s_{ik}s_{kj}\rangle$ that occurs in Eq. (4.8) in the mean-field approximation using a methodology that is similar to the one used to compute $\langle s_{ij}\rangle$. We start from writing the Hamiltonian as a sum of two parts, whereby the first part contains all terms that depend on two tagged links sharing a node (s_{ik} and s_{kj}), and the second part contains the remaining terms

$$\mathcal{H}\left(\{s_{mn}\}\right)=\mathcal{H}_{II}\left(s_{ik},s_{kj}\right)+\mathcal{H}''\left(\{s_{mn},s_{mn}\neq s_{ik},s_{kj}\}\right)\,, \tag{4.9}$$

with

$$\begin{aligned}\mathcal{H}_{II}\left(s_{ik},s_{kj}\right) &= s_{ik}\left(\sum_{l\neq j,i,k}^{N}\Big[-\alpha\, s_{il}s_{lk}-\gamma\,\left(s_{il}+s_{lk}\right)\Big]+\omega\right)+\mu s_{ik}^2\\ &+ s_{jk}\left(\sum_{l\neq j,i,k}^{N}\Big[-\alpha\, s_{jl}s_{lk}-\gamma\,\left(s_{jl}+s_{lk}\right)\Big]+\omega\right)+\mu s_{jk}^2\\ &+ s_{jk}s_{ik}\left(-\alpha\, s_{ij}-\gamma\right)\,.\end{aligned} \tag{4.10}$$

As above, we replace the $\sum_{l\neq j,i,k}^{N}s_{il}s_{lk}$ by the average $(N-3)\langle s_{ik}s_{kj}\rangle$ to obtain an approximate "mean-field" expression of $\mathcal{H}_{II}\left(s_{ik},s_{kj}\right)$

$$\begin{aligned}\mathcal{H}_{II}\left(s_{ik},s_{kj}\right) &\approx s_{ik}\Big((N-3)\Big[-\alpha\langle s_{ik}s_{kj}\rangle-\gamma 2\langle s_{ij}\rangle\Big]+\omega\Big)+\mu s_{ik}^2\\ &+ s_{jk}\Big((N-3)\Big[-\alpha\langle s_{ik}s_{kj}\rangle-\gamma 2\langle s_{ij}\rangle\Big]+\omega\Big)+\mu s_{jk}^2\\ &+ s_{jk}s_{ik}\Big(-\alpha\langle s_{ij}\rangle-\gamma\Big)\,.\end{aligned} \tag{4.11}$$

For the sake of notation conciseness, we define

$$C_1(\langle s_{ik}s_{kj}\rangle, \langle s_{ij}\rangle) \equiv \left[-\alpha N\langle s_{ik}s_{kj}\rangle - \gamma N 2\langle s_{ij}\rangle + \omega\right] \quad (4.12)$$

$$C_2(\langle s_{ij}\rangle) \equiv \left[-\alpha\langle s_{ij}\rangle - \gamma\right], \quad (4.13)$$

which allows one to convert the mean-field expression of $\mathcal{H}_{II}\left(s_{ik}, s_{kj}\right)$ into the following form

$$\mathcal{H}_{II}\left(s_{ik}, s_{kj}\right) \approx \left(s_{ik} + s_{jk}\right) C_1 + \mu\left(s_{ik}^2 + s_{jk}^2\right) + s_{jk}s_{ik}C_2\,, \quad (4.14)$$

whereby use was made of the fact that for $N >> 1$ one has $(N-3) \approx N$ and $(N-2) \approx N$. Similar arguments that led to the mean-field result of $\langle s_{ij}\rangle$ in Eq. (4.8) lead to the following expression of $\langle s_{ik}s_{kj}\rangle$

$$\langle s_{ik}s_{kj}\rangle \approx \frac{\sum_{s_{ik}=\pm1,0}\sum_{s_{kj}=\pm1,0} s_{ik}s_{kj}e^{-\beta\mathcal{H}_{II}(s_{ik},s_{kj})}}{\sum_{s_{ij}=\pm1,0}\sum_{s_{kj}\pm1,0} e^{-\beta\mathcal{H}_{II}(s_{ik},s_{kj})}}\,,$$

$$\approx \frac{2e^{-\beta(2\mu+C_2)}\cosh\left(\beta 2C_1\right) - 2e^{-\beta(2\mu-C_2)}}{1 + e^{-\beta\mu}\left[4\cosh\left(\beta C_1\right) + e^{-(\beta\mu)}\left(2\cosh\left(\beta 2C_1\right)e^{-(\beta C_2)} + 2e^{(\beta C_2)}\right)\right]}\,. \quad (4.15)$$

As one can express Eq. (4.8) in terms of C_1 one finally obtains a set of two non-linear equations for $\langle s_{ij}\rangle(C_1)$ and $\langle s_{ik}s_{kj}\rangle(C_1, C_2)$ with $C_1 = C_1(\langle s_{ik}s_{kj}\rangle, \langle s_{ij}\rangle)$ and $C_2 = C_2(\langle s_{ij}\rangle)$. Equations (4.17) and (4.18) in combination with the definitions in Eqs. (4.12) and (4.13) constitute a non-trivial set of self-consistent equations for the expectation values of a link ($\langle s_{ij}\rangle$) and the product of two links ($\langle s_{ik}s_{kj}\rangle$).

In order to find the possible solutions of the self-consistent set of two equations we transform them in one self-consistent equation. To this end, we first insert Eq. (4.17) in Eq. (4.13) to express C_2 as a function of C_1.

$$C_2(C_1) = -\alpha\frac{2\sinh\beta C_1}{2\cosh\beta C_1 + e^{\beta\mu}} - \gamma\,. \quad (4.16)$$

Next, inserting this expression for $C_2(C_1)$ in Eqs. (4.17) and (4.18), we have both average values $\langle s_{ij}\rangle$ and $\langle s_{ik}s_{kj}\rangle$, written as a function of C_1. In the final step, one makes use of Eq. (4.12) to obtain a self-consistent expression in C_1 (Eq. (4.19). The solutions of this self-consistent equation provide values for the C_1 and can be obtained as follows.

The $f(C_1)$ can be plotted versus C_1 for any given set of parameters $(\alpha, \gamma, \omega, \mu)$ and a temperature β. For a specific C_1, we can calculate $\langle s_{ij}\rangle$

using Eq. (4.17). Then, using $\langle s_{ij} \rangle$ and C_1, we can solve Eq. (4.12) to obtain a value for the measure $\langle s_{ik} s_{kj} \rangle$. In order to produce plots of the Eq. (4.19) we use the Python module SymPy [72]. SymPy is a Python module for symbolic mathematics that allows one to evaluate the exponential function with an arbitrary precision thereby avoiding the overflows.

$$\langle s_{ij} \rangle = \frac{-2 \sinh \beta C_1}{2 \cosh \beta C_1 + e^{\beta \mu}} , \tag{4.17}$$

$$\langle s_{ik} s_{kj} \rangle = \frac{2e^{-\beta(2\mu + C_2)} \cosh(\beta 2 C_1) - 2e^{-\beta(2\mu - C_2)}}{1 + e^{-\beta\mu}\left[4 \cosh(\beta C_1) + e^{-(\beta\mu)}\left(2 \cosh(\beta 2 C_1) e^{-\beta C_2} + 2e^{\beta C_2}\right)\right]} . \tag{4.18}$$

$$\begin{aligned} C_1 &= f(C_1) \\ &= -\alpha N \frac{2e^{-\beta C_2(C_1)} \cosh(\beta 2 C_1) - 2e^{\beta C_2(C_1)}}{e^{2\beta\mu} + e^{\beta\mu}\left[4 \cosh(\beta C_1) + e^{-(\beta\mu)}\left(2 \cosh(\beta 2 C_1) e^{-\beta C_2(C_1)} + 2e^{\beta C_2(C_1)}\right)\right]} \\ &\quad + \gamma N 2 \frac{2 \sinh \beta C_1}{2 \cosh \beta C_1 + e^{\beta\mu}} + \omega . \end{aligned} \tag{4.19}$$

4.1.1 Solutions of the self-consistent mean-field equations

In this subsection, we study the solutions of Eq. (4.19) for different combinations of parameter sets $(\alpha, \gamma, \omega, \mu)$ and the temperature β. We will start with the set of extracted values of parameters from theory-data comparisons reported in [69] (Chapter 3 of this thesis). Next, we study the mean-field solutions under variations of the ratio of α to γ, that are crucial to the definitions of C_1 and C_2 (Eqs. (4.12) and (4.13)). We also determine the behavior of the system under a change of the sign of the parameters α and γ and study the effect of the chemical potential μ. As a robustness check for the adopted numerical procedures, we determine the solutions of the self-consistent equation (4.19) for $(\alpha = 0, \omega = 0, \mu \sim -\infty)$ and check whether we can retrieve the well-known mean-field results of the Ising Hamiltonian with $J = 2\gamma$.

In the upper panel of Fig. 4.1 we observe that for positive parameters $(\alpha > 0, \gamma > 0, \omega > 0)$, the number of solutions to the self-consistent equation $C_1 = f(C_1)$ strongly depends on the temperature β. Independently of the value of β, $C_1 = 0$ is always a solution and for high temperatures (small β) it is the sole solution to $f(C_1) = C_1$. With increasing temperatures, the value of Eq. (4.19) approaches a straight line with a small inclination and not alternate solutions to $C_1 = 0$ can be found. With increasing β the $f(C_1)$ is an increasing function for $C_1 > 0$, creating a solution when the maximum touches on the line $f(C_1) = C_1$. The value of the C_1 that obeys the self-consistent condition depends on α. Note that the function $f(C_1)$ is not symmetric with regard to the substitution $C_1 \longrightarrow -C_1$ which implies that the absolute value of the non-vanishing solution to the self-consistent equation for $C_1 < 0$ is not equal to the non-vanishing solution for positive C_1. Within the studied ranges of $\alpha > 0$ and $\gamma > 0$ we find for sufficiently small temperatures a similar amount of non-vanishing solutions C_1.

For $(\alpha < 0, \gamma > 0, \omega > 0)$ (middle panel in Fig. 4.1) the solutions to $f(C_1) = C_1$ have a similar but in some sense "inverted" pattern in comparison to the $(\alpha > 0, \gamma > 0, \omega > 0)$ situation displayed in the upper panel of Fig. 4.1. Indeed, the upper and middle panel of Fig.4.1 can be roughly related by means of the transformations $f(C_1) \longrightarrow -f(C_1), C_1 \longrightarrow -C_1$.

In the lower panel in Fig. 4.1 we study the solutions to $f(C_1) = C_1$ for $(\alpha < 0, \gamma < 0, \omega > 0)$. We find that not all parameter combinations give rise to a non-vanishing solution C_1. For $C_1 < 0$, one has $f(C_1) > 0$ and there is no solution to the self-consistent equation. For small values of the strength parameter α ($\alpha = -0.1$ and $\alpha = -0.3$), there is also no solution to $f(C_1) = C_1$ for positive values of C_1. For larger negative values of α, the $f(C_1)$ increases for $C_1 > 0$ and a pair of non-vanishing and positive

solutions to the self-consistent equation are obtained.

Figure 4.1 shows the solutions to the self-consistent equation $f(C_1) = C_1$ for various $(\alpha, \gamma, \omega, \beta)$ combinations and a fixed value of the chemical potential μ. The impact of the value of the chemical potential on the solutions is illustrated in the upper panel of Fig. 4.2. The impact of the chemical potential on the position of the non-vanishing solutions C_1 is rather modest except for large μ, which corresponds to a situation whereby the edges remain mostly inactive in the range of temperatures studied here. The bottom panel of Fig. 4.2 illustrates that for the Ising-like Hamiltonian we retrieve the well-known tanh shape of the mean-field solution. For $\gamma > 0$ we have a ferromagnetic system and for $\gamma < 0$ we have an anti-ferromagnetic one. Note the qualitative similarity between the mean-field solutions to the social balance Hamiltonian of Eq. (4.1) (upper panel of Fig. 4.2) and the well-known Ising Hamiltonian (lower panel of Fig. 4.2).

The Figs. 4.1 and 4.2 show the function $f(C_1)$ versus C_1 for various combinations of the temperature and the parameters that appear in the social balance Hamiltonian of Eq. (4.1). In Figs. 4.3 and 4.4 we display the temperature dependence of the corresponding mean-field solutions to the physical measures $\langle s_{ij} \rangle$, $\langle s_{ik} s_{kj} \rangle$, and $\langle s_{ij}^2 \rangle$. Thereby, the left, middle and right columns of Fig. 4.3 correspond with the parameter combinations discussed in the upper, middle and lower panels of Fig. 4.1. The left and right columns of Fig. 4.4 correspond with the parameter combinations discussed in the upper and lower panels of Fig. 4.2. In the left column in Fig. 4.3 corresponding with $(\alpha > 0, \gamma > 0, \omega > 0)$ we observe non-vanishing physical measures for $\beta > \beta_c$. The value of β_c depends on α. For low β (high temperatures) all values (i.e., +1, -1, 0) for the edges have a similar probability and we have $\langle s_{ij} \rangle \sim \langle s_{ik} s_{kj} \rangle \sim 0$, and $\langle s_{ij}^2 \rangle \sim 2/3$. For $\beta > \beta_c$ a positive and a negative non-vanishing branch for $\langle s_{ij} \rangle$ appear. At low temperatures the positive branch approaches the limit $\langle s_{ij} \rangle = 1$. The β dependence of the local correlation $\langle s_{ik} s_{kj} \rangle$ has a similar amount of solutions as the magnetization $\langle s_{ij} \rangle$. For $\beta > \beta_c$ one has the zero magnetization solution as well as a positive branch corresponding to a fully aligned system $\langle s_{ij} \rangle \to +1$ and $\langle s_{ik} s_{kj} \rangle \sim 1$.

For the set of parameters $(\alpha < 0, \gamma > 0, \omega > 0)$ (central column in Fig. 4.3) the β dependence of the mean-field results for $\langle s_{ij} \rangle$ are similar to what is observed for $(\alpha > 0, \gamma > 0, \omega > 0)$ (Left column in Fig. 4.3) but the sign of the magnetization is reverted. For the parameter combinations corresponding with $(\alpha < 0, \gamma < 0, \omega > 0)$ (Right column in Fig. 4.3) the mean-field results for $\langle s_{ij} \rangle$ are markedly different. For small values of the strength parameter α ($\alpha = -0.1$ and $\alpha = -0.3$), there is also no solution to the self-consistent equation apart from the $\langle s_{ij} \rangle = 0$ one that corresponds to

a situation whereby all values (i.e., +1, -1, 0) for the edges occur with equal probability. For large negative values of α, a pair of non-vanishing solutions for the magnetization appears for $\beta > \beta_c$. The two solutions correspond with a high level of activation $\langle s_{ij}^2 \rangle$ and a negative magnetization $\langle s_{ij} \rangle$. The difference between the two non-vanishing solutions grows with increasing β.

The left column of Fig. 4.4 provides a feeling of the impact of the value of the chemical potential on the mean-field results for the three measures. Overall, the effect of the value of the chemical potential is rather modest, apart from situations with very values of the chemical potentials. Under that condition, the edges remain mostly inactive in the range of temperatures studied. This corresponds with the situation that the cost of an active edge is so high that none of the other interaction terms in the Hamiltonian is strong enough to compensate for this effect, leaving all edges inactive. The Ising-limit of the studied social balance Hamiltonian (right column in Fig. 4.4) provides the expected β dependence for the measures studied here. For a ferromagnetic system ($\gamma > 0$) we retrieve the two symmetric solutions $\langle s_{ij} \rangle > 0$ and $\langle s_{ij} \rangle < 0$ for $\beta > \beta_c$. For $\beta < \beta_c$ the system is disordered which is reflected in $\langle s_{ij} \rangle = 0$. In the antiferromagnetic situation($\gamma < 0$), we retrieve the perfect balance between the two active states $\langle s_{ij} \rangle \pm 1$ for the full range of β values studied here.

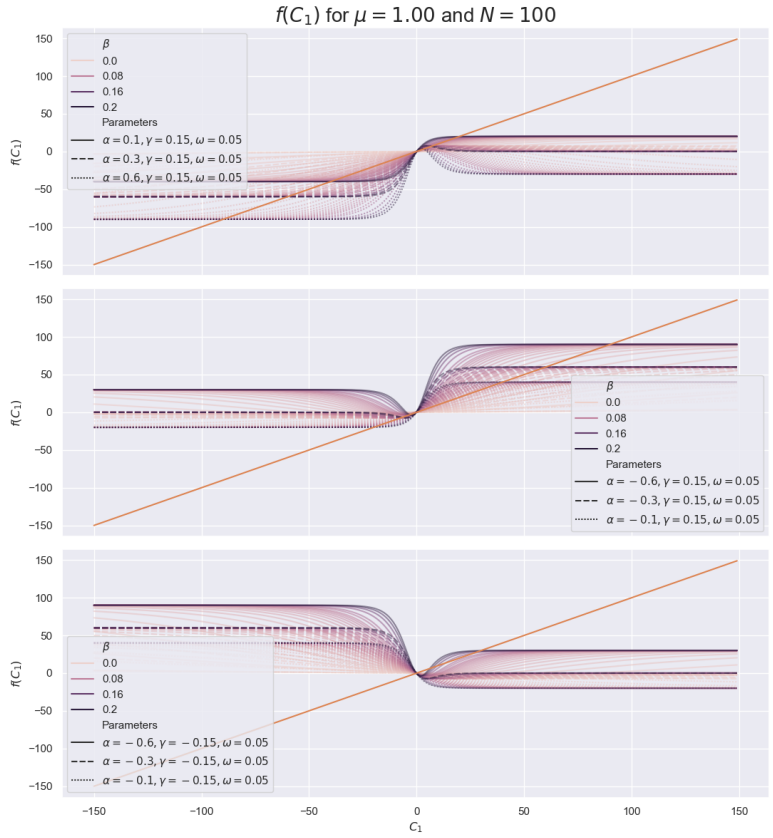


Figure 4.1: The function $f(C_1)$ of Eq. (4.19) for 23 different values of β [0.001, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05, 0.055, 0.06, 0.065, 0.07, 0.075, 0.08, 0.09, 0.1, 0.125, 0.15, 0.175, 0.2] and combinations of parameters (α,γ,$\omega = 0.05$) for $\mu = 1$ and $N = 100$. The mean-field solutions are those values of C_1 where $f(C_1) = C_1$. Upper panel: $f(C_1)$ for $\alpha > 0, \gamma > 0$; Middle panel: $f(C_1)$ for $\alpha < 0, \gamma > 0$; Bottom panel: $f(C_1)$ for $\alpha < 0, \gamma < 0$.

4.1.2 Searching for the critical point

The plots for $f(C_1)$ in Figs. 4.1 and 4.2 that cover a wide range of strength parameters and temperatures identify some emerging trends. For all parameter combinations $C_1 = 0$ is a solution to the self-consistent equation (4.19). Other solutions occur whenever the function $f(C_1)$ meets the line C_1. For the Ising system, this condition can be guaranteed whenever $f(C_1)$ has a slope in the origin larger than 1 (the gradient of $f(C_1) = C_1$). For slopes higher than 1, the function $f(C_1)$ intersects C_1 in two points away from the

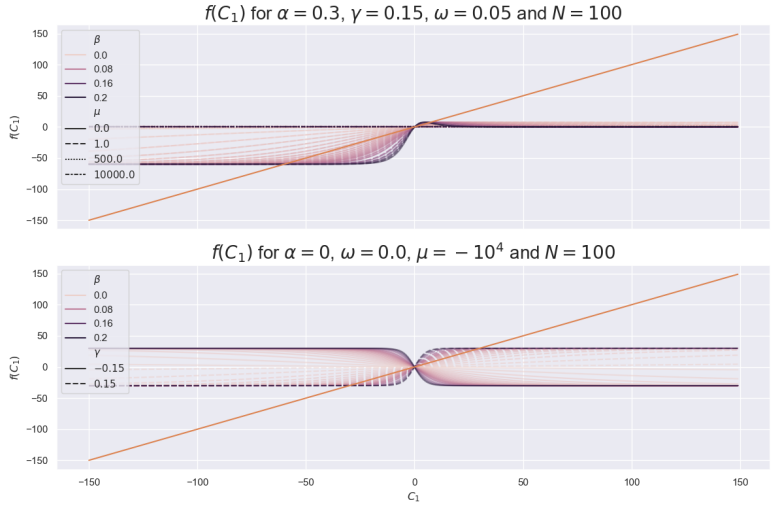


Figure 4.2: The function $f(C_1)$ of Eq. (4.19) for 23 different values of β [0.001, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05, 0.055, 0.06, 0.065, 0.07, 0.075, 0.08, 0.09, 0.1, 0.125, 0.15, 0.175, 0.2] and combinations of parameters (α,γ,ω) for $N = 100$. The mean-field solutions are those values of C_1 where $f(C_1) = C_1$. Upper panel: $f(C_1)$ for $(\alpha = 0.3, \gamma = 0.15, \omega = 0.05)$ and various values for the chemical potential μ; Bottom panel: $f(C_1)$ for an Ising-like system ($\alpha = 0.0$,$\omega = 0.00$,$\mu = -10^4$).

origin. The calculation of the derivative of $f(C_1)$ with respect to C_1 in the origin, allows one to determine whether solutions with $C_1 \neq 0$ occur. The derivative of Eq. (4.19) is:

$$\begin{aligned}\left.\frac{df(C_1)}{dC_1}\right|_{C_1=0} = & -\frac{4N^2\beta^2\alpha^2\left(1-e^{2\beta\gamma}\right)\left(\left(e^{\beta\mu}+4\right)e^{\beta(\gamma+\mu)}+4\right)}{\left(e^{\beta\mu}+2\right)\left(\left(e^{\beta\mu}+4\right)e^{\beta(\gamma+\mu)}+2e^{2\beta\gamma}+2\right)^2} \\ & +\frac{8N^2\beta^2\alpha^2}{\left(e^{\beta\mu}+2\right)\left(\left(e^{\beta\mu}+4\right)e^{\beta(\gamma+\mu)}+2e^{2\beta\gamma}+2\right)} \\ & +\frac{4N\beta\gamma}{\left(e^{\beta\mu}+2\right)} \ . \end{aligned} \tag{4.20}$$

The fully-connected Ising model can be retrieved from the Hamiltonian of Eq. 4.1 by setting $\alpha = 0$, $\omega = 0$ and $\mu = -\infty$. In that particular "Ising"

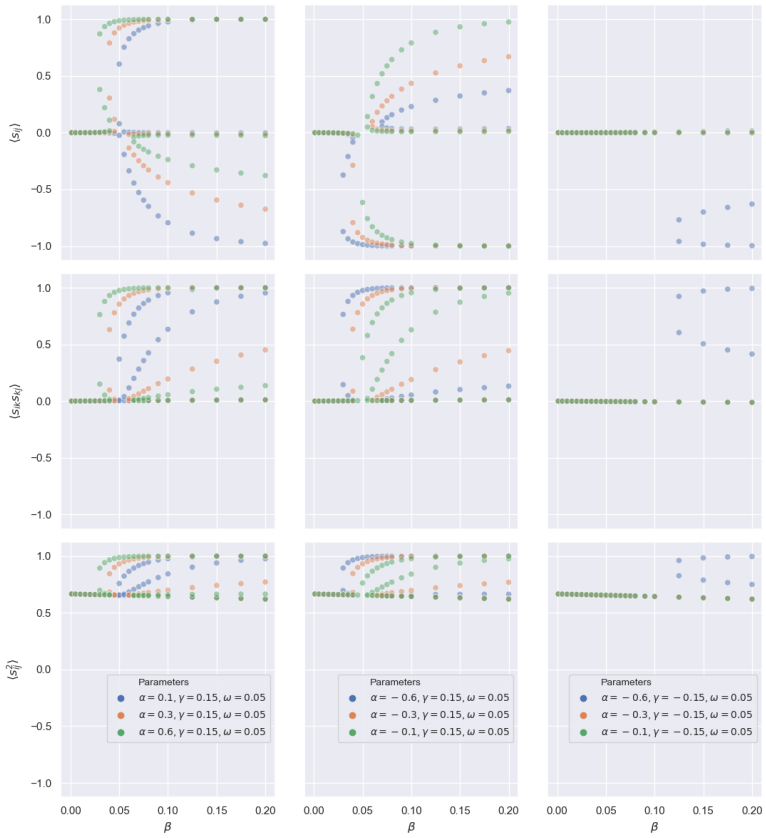


Figure 4.3: The mean-field prediction for the β dependence of the measures $\langle s_{ij} \rangle$, $\langle s_{ik} s_{kj} \rangle$, and $\langle s_{ij}^2 \rangle$ for various sets of parameters (α, γ, ω) and $\mu = 1, N = 100$. Left: $\alpha > 0$ and $\gamma > 0$ (parameter combinations from the upper panel of Fig. 4.1); Center: $\alpha < 0$ and $\gamma > 0$ (parameter combinations from the middle panel of Fig. 4.1); Right: $\alpha < 0$ and $\gamma < 0$ (parameter combinations from the bottom panel of Fig. 4.1).

limit the derivative of Eq. (4.20) in the origin becomes

$$\left.\frac{df(C_1)}{dC_1}\right|_{C_1=0,\alpha=0,\omega=0,\mu=-\infty} = \frac{4N\beta\gamma}{(e^{\beta-\infty}+2)} = 2N\beta\gamma = N\beta J \,. \tag{4.21}$$

Obviously, we retrieve the well-known critical point for the Ising system for an effective number of neighbors N. For most strength combinations (α, γ, ω), the Figs. 4.1 and 4.2 have a non-trivial behavior close to the ori-

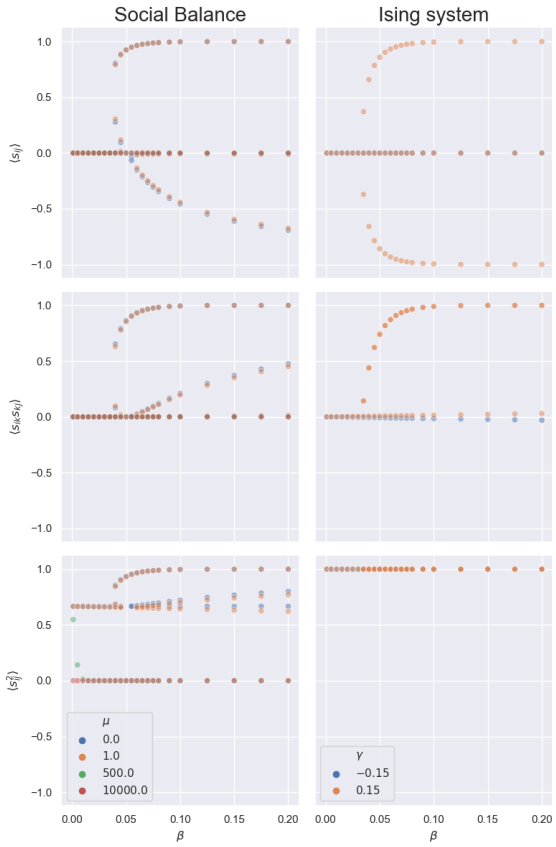


Figure 4.4: The mean-field prediction for the β dependence of $\langle s_{ij} \rangle$, $\langle s_{ik}s_{kj} \rangle$, and $\langle s_{ij}^2 \rangle$ for various chemical potentials μ. Left: $(\alpha = 0.3, \gamma = 0.15, \omega = 0.05)$ and $N = 100$ (parameter combinations from the upper panel of Fig. 4.2). Right: results for an Ising-like system $(\alpha = 0.0, \omega = 0.00, \mu = -10^4)$ and $N = 100$ (parameter combinations from the bottom panel of Fig. 4.2).

gin and the above-mentioned method cannot be used to seek for the critical temperature that delivers non-vanishing values for C_1. For any given precision $\Delta\beta$, however, one can compute the number of times where the two values of the functions $f(C_1)$ and C_1 coincide. Table 4.1 lists the critical points for the transition with the lowest $\beta \in (0, 1.0]$ determined in this way.

α	γ	ω	μ	β_c (Mean field)	β_c (Monte Carlo) *
+0.3	+0.15	0.05	1.0	0.035-0.040	0.040-0.045
+0.6	+0.15	0.05	1.0	0.025-0.030	0.035-0.040
+0.1	+0.15	0.05	1.0	0.045-0.05	0.050-0.060
-0.3	+0.15	0.05	1.0	0.035-0.04	0.040-0.045
-0.6	+0.15	0.05	1.0	0.025-0.030	0.035-0.04
-0.1	+0.15	0.05	1.0	0.050-0.055	0.045-0.05
-0.3	-0.15	0.05	1.0	NF	NF
-0.6	-0.15	0.05	1.0	NF	NF
-0.1	-0.15	0.05	1.0	NF	NF
+0.3	+0.15	0.05	0.0	0.035-0.04	0.04-0.045
+0.3	+0.15	0.05	5e2	NF	NF
+0.3	+0.15	0.05	1e3	NF	NF
+0.0	+0.15	0.00	-1e5	0.030-0.35	0.035-0.04
+0.0	-0.15	0.00	-1e5	NF	NF

Table 4.1: Extracted values of the critical temperature β_c for $\beta \in (0, 1.0]$. With "NF" we indicate those situations where no critical point was found in the selected interval for β. The Monte-Carlo results for β_c are those from simulations with a "random" initial state. We determine the critical temperature through determination of the temperature interval where the absolute average mangetization jumps from $|\langle s_{ij}\rangle| < 0.05$ (" vanishing magnetization") to $|\langle s_{ij}\rangle| > 0.20$ ("finite magnetization").

4.2 Monte-Carlo simulations

In this subsection, we present results of Monte-Carlo simulations based on the Hamiltonian of Eq. (4.1). We wish to measure the expectation value of the three measures studied in the previous sections: $\langle s_{ij}\rangle$, $\langle s_{ik}s_{kj}\rangle$, and $\langle s_{ij}^2\rangle$. The expectation value of a measure A is:

$$\begin{aligned}\langle A\rangle &= \sum_{\{s_{ij}\}} A(\{s_{ij}\})P(\{s_{ij}\}) \\ &= \sum_{\{s_{ij}\}} A(\{s_{ij}\})e^{-\beta\mathcal{H}(\{s_{ij}\})}/Z \ , \end{aligned} \tag{4.22}$$

where $Z=\sum_{\{s_{ij}\}} e^{-\beta H(\{s_{ij}\})}$ is the partition function. The sum $\sum_{\{s_{ij}\}}$ extends over all possible states of the system and can only be executed exactly for networks with a relatively small number of nodes N. For networks with a larger number of nodes N, a Monte-Carlo algorithm provides a means to run over the state space (=the set of all possible configurations of a system) in a weighted run. Thereby, the algorithm is designed so as to spend more time in that part of the state space that provides the largest contributions to the sum in Eq. (4.22) thereby providing good estimates of the $\langle A\rangle$. Here, we use Monte-Carlo simulations to test the major results of the mean-field approach with regard to the existence of a phase transition and the phase diagram of the Hamiltonian under study.

The Metropolis algorithm is implemented as follows:

Step1: Generate a set of symmetric $\{s_{ij}\}$ with $s_{ij}=\pm 1,0$ and $1\leq i,j\neq i\leq N$. The generated set $\{s_{ij}\}$ serves as the starting configuration of the simulations. Four different types of initial conditions are considered: $s_{ij}=+1$ (a network of friends), $s_{ij}=-1$ (a network of enemies), $s_{ij}=0$ (a network of neutrals), $s_{ij}=0,\pm 1$ with equal probability.

Step2: Randomly select a pair l and f and change the status of s_{lf} to $s'_{lf}=\pm 1,0$ with $s'_{lf}\neq s_{lf}$

Step3: Calculate the energy difference $\Delta\mathcal{H}$ between the initial configuration $\{s_{ij}\neq s_{lf}, s_{lf}\}$ and the proposed configuration $\left\{s_{ij}\neq s'_{lf}, s'_{lf}\right\}$ with one modified value for the link s_{lf}. Using the Eq. (4.2) one finds:

$$\begin{aligned}\Delta\mathcal{H} &= \mathcal{H}(\left\{s_{ij}\neq s'_{lf}, s'_{lf}\right\}) - \mathcal{H}(\{s_{ij}\neq s_{lf}, s_{lf}\}) \\ &= +\omega(s'_{lf}-s_{lf}) + \mu(s'^2_{lf}-s^2_{lf}) \\ &\quad -(s'_{lf}-s_{lf})\sum_{k\neq l,f}\left[\alpha s_{lk}s_{kf}+\gamma(s_{lk}+s_{kf})\right] . \end{aligned} \tag{4.23}$$

Step4: Accept the change with a probability $P(s_{lf} \to s'_{lf}) = \min(1, e^{-\beta\Delta\mathcal{H}})$

Step5: After a substantial amount of iterations the system's measures $\langle s_{ij}\rangle$, $\langle s_{ik}s_{kj}\rangle$, and $\langle s_{ij}^2\rangle$ are computed.

Step6: We go back to Step 2 until one reaches thermal equilibrium and convergence.

The above algorithm is executed for a relevant range of inverse temperatures β at any given set of the parameters $(\alpha, \gamma, \omega, \mu)$. For each combination $(\alpha, \gamma, \omega, \mu)$ one can measure the $\langle s_{ij}\rangle$, $\langle s_{ik}s_{kj}\rangle$, and $\langle s_{ij}^2\rangle$ as a function of β. All presented results are obtained with simulations for $N = 100$ nodes and $N(N-1)/2 = 4950$ pairs and realizing 10^7 steps in the above-mentioned iterative procedure. The Monte-Carlo results are displayed in Figs. 4.5 and 4.6 and are compared with the mean-field predictions. The studied interval of inverse temperatures $0 \leq \beta \leq 0.17$ covers the range for which we observed significant changes in $\langle s_{ij}\rangle$ (critical β in Table 4.1).

The results of the "exact" Monte-Carlo simulations and of the mean-field approximation have overall a quantitatively similar behavior. Far from the critical point, the numerical values of the measures obtained with the Monte-Carlo simulations are scattered around one of the solutions from the mean-field equations. Some solutions of the mean-field approach are not retrieved by the Monte-Carlo algorithm and are likely connected to solutions that are related unstable configurations of the network. The mean-field approach typically predicts lower values of the critical β_c when compared to the Monte-Carlo ones.

For $\gamma > 0$, the results agree with $\langle s_{ij}\rangle$ and the positive part of $\langle s_{ik}s_{kj}\rangle$ (Fig. 4.5). For low β (high temperature), the system remains as expected in a low-magnetization state, with weak correlations between the nodes and with two thirds of the links active. All these results are in line with a system for which the values of the edges are randomly assigned. For high β (low temperatures) the system resides in an ordered state which is reflected in increasing values of the absolute value of $\langle s_{ij}\rangle$ and a $\langle s_{ik}s_{kj}\rangle > 0$. The critical beta that marks the transition between the ordered and disordered state is $\beta_c \approx 0.05$. The mean-field and Monte-Carlo simulations predict different values for the position of the critical point, and the corresponding displacement between the two values increases with growing α. The negative (positive) branch of the magnetization para positive (negative) α, appear to be unstable except for low γ. The combination of negative α and negative γ leads to systems for which solely the vanishing magnetization situation is stable. For low-temperature systems subject to strong three-edge interactions (large α) the system presents itself in a fully activated

state ($\langle s_{ij}^2 \rangle \approx 1$) with vanishing magnetization $\langle s_{ij} \rangle \approx 0$ and vanishing local correlations $\langle s_{ik} s_{kj} \rangle \approx 0$. Physically this corresponds to a situations whereby half of the of the edges are positive and the other half of the edges are negative.

For the different small μ (Fig. 4.6), the system behavior as we discussed before, and became inactive for higher μ. Ising system presents a strong match between the Monte-Carlo simulations and the mean-field approximation, matching the phase transition and the curve of measures with excellent results.

4.3 Conclusion

From the perspective of statistical physics, the Hamiltonian of Eq. (4.2) extensively discussed in [69] implements the ideas from social balance theory. It can be considered as a generalization of the basic Hamiltonian for social balance that has only a three-edge interaction and has been discussed in the literature [51, 110]. The extended Hamiltonian that is being proposed in this work can accommodate a few additional features including the differences between strong and weak balance observed in the real-world systems [58, 68]. In addition, the proposed generalization has the advantage that it includes other common models such as the well-known Ising model for magnetic interactions [111]. In this chapter, the focus was on the study of the properties of a generic system that obeys the Hamiltonian of Eq. (4.1) and no reference has been made to specific data for political networks. Using a mean-field approximation and Monte-Carlo simulations, we found that the proposed Hamiltonian gives rise to phase transitions whose existence is controlled by the relation between the model parameters. The mean-field approximation gives rise to different branches of solutions, of which only selected solutions could be retrieved in the Monte-Carlo simulations. The other solutions of the mean-field equations are likely connected to unstable solutions corresponding with local minima/maxima in the complex free-energy surface of the system under study. For the set of parameters and temperatures studied, both methods agree qualitatively on the position of the critical temperature.

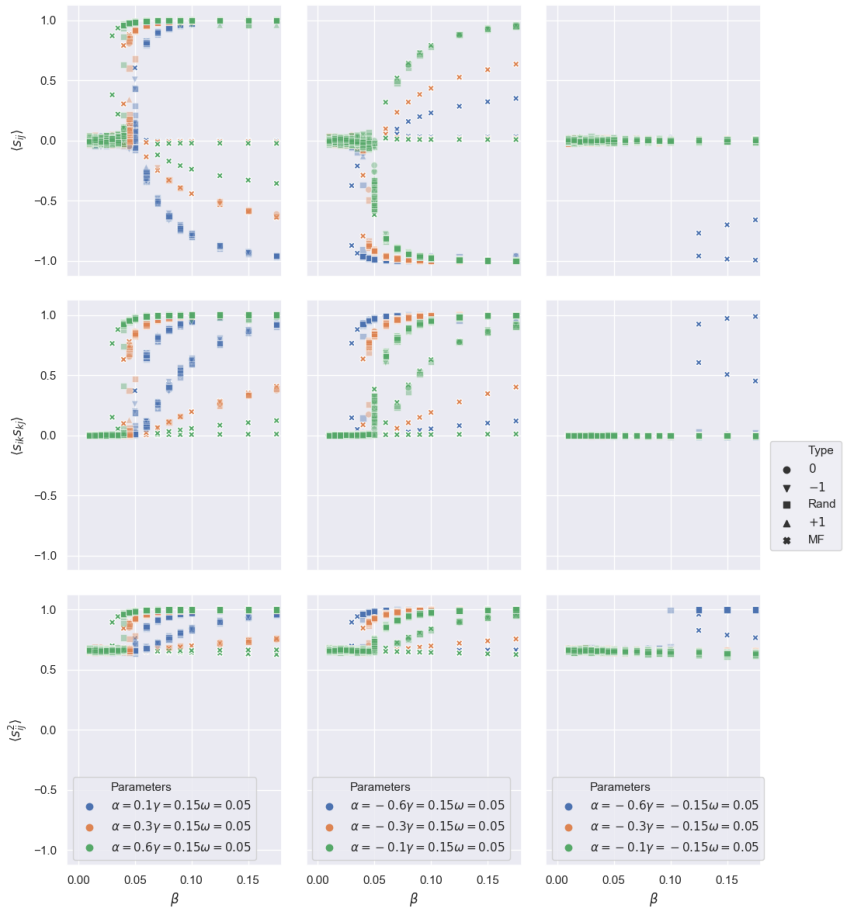


Figure 4.5: The β dependence of the measures $\langle s_{ij} \rangle$, $\langle s_{ik} s_{kj} \rangle$, $\langle s_{ij}^2 \rangle$ for various combinations of (α, γ, ω) and $\mu = 1$. Left: $\alpha > 0, \gamma = 0.15, \omega = 0.05$. Center: $\alpha < 0, \gamma = 0.15, \omega = 0.05$. Right: $\alpha < 0, \gamma = -0.15, \omega = 0.05$. We compare the mean-field results of Fig. 4.3 with those of Monte-Carlo simulations. The MC results are shown for four different starting conditions (randomly assigned edges or edges that are exclusively $+1$,0 or-1). Also the mean-field results are shown.

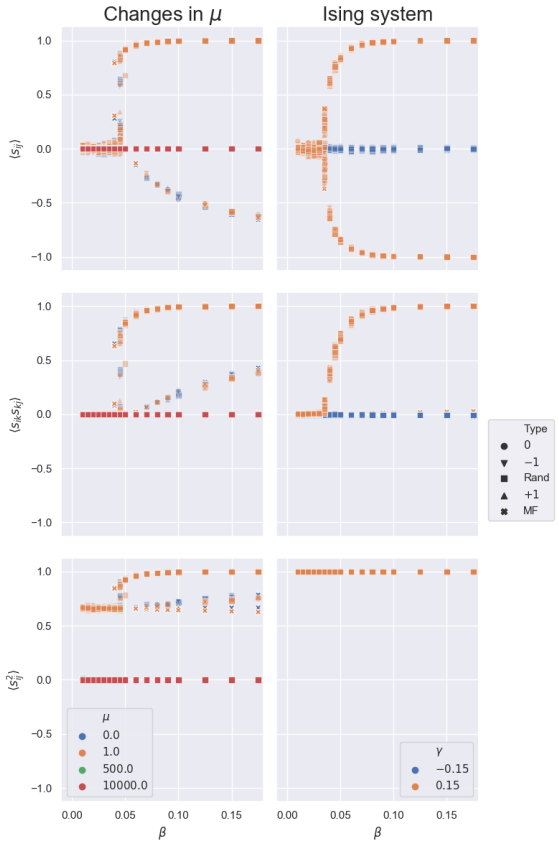


Figure 4.6: The β dependence of the measures $\langle s_{ij}\rangle$, $\langle s_{ik}s_{kj}\rangle$, $\langle s_{ij}^2\rangle$. We show the solutions of Fig. 4.4 as well as the Monte-Carlo results. The MC results are shown for four different starting conditions (randomly assigned edges or edges that are exclusively $+1$,0 or-1). Left: ($\alpha = 0.3$,$\gamma = 0.15$,$\omega = 0.05$) and $N = 100$ and various chemical potentials μ; Right: results for an Ising-like system ($\alpha = 0.0$, $\omega = 0.00$, $\mu = -10^4$) and $N = 100$.

Part II

Economy in a virtual world

5

Connections between real and virtual world in human behavior

5.1 Abstract

Using data on the virtual world "EVE Online", we investigate whether the in-game behavior of players gets influenced by their real-world nation-level context. We test whether levels of in-game aggressiveness can be related to the real-world nation-wide aggression levels as measured by the Global Peace Index and the Global Terrorist Index. We observe that both higher levels of in-game aggression and whether a player is perceived as 'friendly', have a negative relation with the Global Peace Index. We also test whether in-game trading behavior can be linked to a player's real-world socioeconomic environment. We observe a positive relation between the level of in-game trading activity of a country and its unemployment rate and local dollar exchange rate. We conclude that the real world can permeate to a virtual world, as the players in the virtual world also get affected by the real world. With our analysis, we present evidence that data extracted from the virtual world can be used to infer the real-world context.

5.2 Introduction

The last two decades has seen the rise of studies of virtual worlds and of their connection to the real world [56–58,112,113]. The use of data extracted from virtual worlds allows researchers to test theories on human behavior in highly controlled environments. For example, virtual worlds have provided researchers with data that enables them to test theories of sociological behavior [58,114], of human mobility [60], of international relations [68,69] and of economic models [59, 113, 115, 116]. A major issue, however, is whether the data extracted from virtual worlds can be considered representative to be used in studies of real-world socioeconomic phenomena.

In a comparative study of virtual and real worlds, Martoncik [117] observed that the players of the virtual world World of Warcraft (WoW) are on average more social in the game than in the real world. WoW is a virtual world in a fantasy setting where players can either fight each other or collaborate in order to meet the game's preset challenges. Upon interacting in the virtual world the players form different feelings and experience less loneliness and social anxiety compared to the real world. This difference between the real and virtual world is more prominent for players that prefer to operate in in-game groups. Zhang and Kaufman [118] have also studied the connection between online social interactions in WoW and the socio-emotional well-being of adults. Grown-ups that are part of an online club inside WoW can develop meaningful online relationships with game friends. The online world Destiny was used as a medium by Perry et al. [119] to compare how relationships (in-game and out-game friendships and strangers) impact the behavior of players in an online game. Destiny is also a virtual world in a futuristic setting, where players can play either on an individual basis, or in groups (with friends or strangers) to meet a number of the challenges as they have been designed by the game's developers.

Another virtual world that has been extensively studied [58,60,99,112, 120] is the online game Pardus [61] that can be accessed via an internet browser. Pardus represents a futuristic universe where players compete for the limited space whilst interacting in many ways (including fighting) with the other players. Szell et al. [99] have made a study of the population of Pardus and of the different interaction channels of its players. The resulting in-game communication, friendship, and animosity interaction networks have non-trivial structural properties. Using both the friendship and enmity networks, for example, they found strong evidence for a generative mechanisms known as social balance theory [47,48]. This mechanism is also found in real-world networks [44, 49, 84]. Furthermore, the typical number of interactions per player in Pardus corresponds with the standard group

size for human interaction throughout history known as Dunbar's number [40]. Dundar's number has been proposed to lie in the range 100-250 and has emerged from the analysis of social networks [41]. It is a suggested cognitive limit to the number of people with whom one can maintain stable social relationships. Thurner et al. in [120] have also studied sequences of in-game actions in Pardus. They observe a higher probability of "positive" ("negative") action if the player has received a "positive" ("negative") action in the preceding time step.

In this paper we work with data extracted from the behavior of players in Eve Online [62]. EVE Online is a sandbox massive multi-player on-line game (MMOG) developed by CCP games [62]. In this EVE Online virtual world, over half a million players fight, trade, collaborate, explore, ... in a futuristic galaxy. As the players operate in a virtual world, their actions are recorded with a high level of precision and detail which sparked off the investigations reported in [68, 69, 116, 121, 122]. Feng et al. in [122] have studied the temporal evolution of EVE's in-game activity and EVE's in-game population for three consecutive years. They found daily cycles and a weekly periodicity in the number of player connections. EVE Online supports a player-driven complex economy as all items need to be produced from raw materials and traded by players with minimal intervention of the game developers. Hoefman et al. [116] have studied the price of the items in the EVE Online free market and found convincing evidence in support of the theory of hedonic pricing. Hedonic pricing theory [123] posits that, next to functional value, people or willing to pay for the social value of a product. Using data from EVE Online, Carter has conducted a study of the use of propaganda during war times, as an essential factor for maintaining the morale [121]. Connected with war times, in previous work [68, 69], we have developed a framework inspired by the methodology of statistical physics and structural balance to model international relationships. The model has been tested against the network of relationships between alliances in EVE Online, and has been compared with the data for international relationships during the Cold War era. We found a convincing example of how the virtual world can be used as a cleaner and a more extended test-bed for evaluating generative mechanisms also present in the real world. Indeed, the networks of alliances in EVE Online and of countries during the Cold War era seem to have a lot in common.

In this work, we use data on the activities developed by the players in EVE Online to test whether a player's real-world socioeconomic environment permeates into its in-game behavior. Using country-averaged activity profiles, we compare the in-game behavior with real-world measures at the level of countries. This work contributes to the question to what extent the

accurate and flexible data of virtual worlds can be used as a reliable proxy for real-world data.

5.3 Data for the socioeconomic profiles of countries in EVE Online and in the real world

5.3.1 Levels of aggressiveness, of social interactions, and of production activities

In the EVE Online virtual world, players can choose to engage in a host of different activities, ranging from mining asteroids to waging war. The player's experience in EVE Online is shaped by the activities of the other players. For example, the game developers set the distribution of raw materials in the game, but the extraction, processing and manufacturing of items is in the hands of the players, as is the transport and trading of the goods. This gives rise to an emergent economy with specializations among players and intricate economic interactions. These inter-player interactions also play a vital role on another level. Indeed, they create complex social hierarchies targeted to control territory and to exploit its resources. With these structures and the abundant inter-player interactions, EVE Online represents a complex world with for example espionage and common game-wide wars as emergent phenomena.

From this broad set of possible activities in EVE Online, we select those that we deem to be the most relevant ones pertaining to the levels of social behavior, of aggressiveness, and of production activities of the individual players. A description of the in-game activities included in our analysis is included in table 5.1. To capture the level of in-game social behavior of a player, we use two variables. The variables "*Added As Negative*" and "*Added As Positive*" encode whether a particular player gets tagged as friend or enemy during inter-player contacts. To capture the level of in-game aggressiveness of a player, we use the number of times a player gets involved in combats with other players both as the aggressor and the defender. As a measure of the level of in-game aggressiveness we also include the number of events that involve the killing of an NPC (Non-Player Character), which is a ship controlled by the game. To measure the level of in-game production activities of the individual players we use three variables: the actual production of items ("*Production*"), the extraction of raw materials from Asteroids ("*Mining*") and extraction of material from destroyed ships ("*Salvaging*").

Name	Category	Description
Added As Negative	Level of social interactions	Interacting players can mark the counter-party as friend,
Added As Positive	Level of social interactions	enemy or neutral. This mark is visible whenever players meet.
Reputation	Level of social interactions	*Added As Positive* minus *Added As Negative*.
Engaged Attack	Level of aggressiveness	Attack another player.
Engaged Defense	Level of aggressiveness	Being attacked by another player.
Aggression	Level of aggressiveness	Conflict against others: *Engaged Attack*-*Engaged Defense*.
Kill NPC	Level of aggressiveness	Destroy ships controlled by the game, usually "pirates".
Production	Level of production activities	Produce an item, using other items as input.
Mining	Level of production activities	Extract materials from asteroids.
Industrial	Level of production activities	*Production* and *Mining*
Salvaging	Level of production activities	Recuperate some materials from destroyed ships.

Table 5.1: In-game activities in EVE Online considered as representative to capture the levels of aggressiveness, of social interactions, and of production activities of the individual players.

Our goal is to investigate whether there is a connection between in-game behavior and corresponding real-world quantitative measures. As these real-world measures are commonly available at the level of countries, we construct an activity-based in-game profile per country starting from the selection of activities listed in Table 5.1. A country's profile is based on the average profile of the players that come from that country. We record the number of times that the in-game events listed in Table 5.1 is realized by the players, except for *Mining*, *Salvaging*, and *Production* for which we use the provided market prices of the outcome. We aggregate on a weekly basis, creating a profile per player per week. Special care needs to be taken when combining the temporal aspect of some activities. Indeed, there is a great variation in the time scales of the different in-game events: marking someone as a friend is an instantaneous event, but the process of mining an asteroid can take half a minute. The variables "*Production*", "*Mining*" and "*Salvaging*" are measured through monetary value instead of through counting the number of times that the activity has been completed. In order to deal with these different scales and metrics, we define an adimensional "activity coefficient" that combines the in-game events and is representative for the activity of a player during a particular week. The idea behind this is to re-scale the measurements for all activities to a common metric. In order to achieve this, for each activity and week the average across players is computed. In the forthcoming analysis, for each player and for all activities listed in Table 5.1 we use a measure relative to this average. In this process the zeros are not counted. The ratio of the action and the normal activity determines the adimensional coefficient. For example, if all players killing NPCs kill on average 200 times per week, and a specific player kills 500 times in a week, his activity will be recorded as "*Kill NPC* $500/200 = 2.5$" for that particular week. Finally, we normalize each player's profile by the sum of the adimensional measures of the in-game activities listed in Table 5.1. As a matter of fact, the resulting profile of a particular player provides information about the way it spends his time in the game. For the country profile, we use the average of the profiles of the players that belong to that country, and we normalize to the standard score between all countries with at least 15 players in EVE online. We have checked that we obtain noisy profiles when including countries with only a few players. The required minimum amount of players in a country is set to a level so that we keep an extended sample of countries for further analysis. Finally, the country-based measures for the in-game activities of Table 5.1 are re-scaled to zero mean and unit variance. For illustration aims, some countries profiles are displayed in Table 5.2.

As a sanity check for the outlined procedure we have looked at the

similarity between pairs of country-level profiles. This has been done for the subset of countries with more than 100 players in EVE online and the results are displayed in Figure 5.1. We calculate similarity using the cosine similarity for the hyper-dimensional vectors that correspond to the country profiles based on the nine in-game activities of Table 5.1. The different countries are aggregated with the aid of the function *hierarchy.linkage* in the Python scipy.cluster module. The algorithm uses an iterative procedure whereby at each step the two closest clusters are aggregated (See [72,124] for more details of the workings of the algorithm). After running the algorithm, the countries are ordered in clusters with similar profiles. We discern two clusters in the 62 countries with more than 100 players in EVE Online. Figure 5.1 shows that there is a cluster of countries (upper-left part of the matrix) with highly similar profiles. We also discern strong similarities in the profiles of a group of countries in Eastern Europe (Ukraine, Russia, Belarus, Moldova, Bulgaria, Lithuania, ...). Furthermore, France and the United Kingdom have a similar profile, as well as Germany and Austria. Canada and Ukraine are the countries with high differences (the cosine similarity is -0.98). The observed inter-country similarities in Figure 5.1 lend strong support for the adopted procedure. For the sake of illustration, the profiles of some countries are listed in Table 5.2.

As measures for the levels of aggressiveness in the real world, we use both the Global Terrorist Index (GTI) and the Global Peace Index (GPI) from the year 2016. The GTI [125] is the product of a study that analyzes the impact of terrorism in 196 different countries and is extracted from the Global Terrorist Database by the Institute for Economics & Peace. The GPI [126], published by the same institute, combines data on terrorist attacks with other indicators including criminality rates and military expenses for example. Note that a low value of the GPI reflects higher levels of peace. As measures for the real-world level of trade and economic activity per country we use economic data for the year 2016.

Country	Russia	Belarus	Ukraine	Canada	France	UK	Germany	Austria
Added As Negative	0.177	0.357	0.325	-0.024	0.133	0.209	0.204	0.198
Added As Positive	0.291	0.314	0.315	0.068	0.283	0.232	0.110	0.136
Kill NPC	0.849	1.207	1.556	-0.230	-0.095	-0.169	0.055	-0.068
Engaged Attack	0.384	0.444	0.439	0.355	0.389	0.413	0.360	0.384
Engaged Defense	0.560	0.684	0.670	0.512	0.542	0.570	0.457	0.476
Production	0.559	1.344	1.121	0.015	0.010	0.054	0.275	-0.033
Salvaging	0.615	0.691	0.597	0.255	0.339	0.275	0.389	0.310
Mining	0.610	0.856	0.702	0.335	0.408	0.322	0.617	0.590
Cosine Similarity	0.950		-0.982		0.792		0.823	

Table 5.2: Examples of country profiles as determined with the nine in-game activities listed in Table 1. For four pairs of countries the cosine similarity of the profiles are shown.

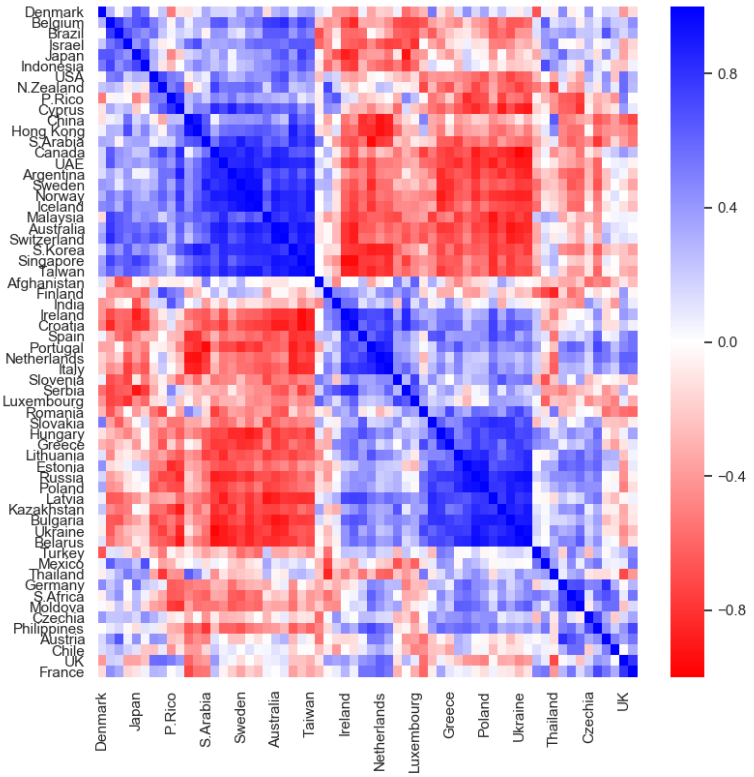


Figure 5.1: Cosine similarity between profiles for countries with more than 100 players in EVE Online. The profiles are represented as hyper-dimensional vectors and are obtained with the variables of Table 5.1 after averaging over all players from a particular country. The precise procedure is described in the text.

5.3.2 Levels of trade activity

To capture the in-game trade activity, we define a number of measures that are reminiscent for the in-game trading behavior of players from a particular country. The EVE market operates according to a double auction system. Players can set a sell order or a buy order for a specific item at a certain price. They can also see the full list of active orders and opt to fulfill them. For each country c and item i, we determine the monthly sell price $P^c_{S,i}$, the monthly buy price $P^c_{B,i}$, and the monthly volumes for sell and buy orders

$V^c_{S,i}$ and $V^c_{B,i}$. We obtain those numbers for the 106 different countries c with more than 15 players and the 6882 different items i. The in-game measures for trade activity of players from a particular country are defined in Table 5.3. All measures include an aggregation over the different items.

From this transaction-level data we create monthly measures spanning the period Dec 2011 up to and including Dec 2016. We normalize the data using the average and the standard deviation. In addition we winsorize the data by removing the 1.5% outliers (Points farther than 3 standard deviations). Summary statistics are reported in table 5.4.

EVE online is an economic game as well as a social. One of the main activities in the game is trade. Here, we want to test if the in-game trade is connected with the real-world economic context. As a proxy for real-world socioeconomic context we use three variables that are monthly reported and that are expected to affect players directly: the inflation in a player's country, the ratio of the local currency vs. the US Dollar, and the unemployment rate.

The Consumer Price Index (CPI) is a measure of inflation which enables to compare across countries the price of a fixed set of consumer goods and services like food and medical care. The Real Effective Exchange Rate (REER) is an index that describes the strength of the local currency respect the rest. It is calculated as a weighted average of all exchange rates with other trade countries' currency, using as weight the trade volume. The last is the unemployment rate (UNEMP), that is the fraction of the job force that is jobless.

To play EVE Online, a player needs to pay a subscription fee of 15 dollars or 15 euros per month. Another option, is to buy for the same price an item (called PLEX) inside the game that, that once activated and consumed, increases the subscription for a month. So, a player can pay the subscription directly with real-world money or by buying in-game PLEX from another player that expend cash from the real world. As a consequence, a player that earns enough in-game currency can play paying the subscription without expending real-world currency. Through this subscription mechanism; the chances of being unemployed, how expensive a subscription is in local real-world currency or how prices change in the real world, can be transmitted to the economic behavior in the game. One could expect that being unemployed will make the player more careful in their in-game transaction, as the chances that they were unemployed and be cautious with the out-game purchase. Exchange rates would be different, as they affect the ratio between in-game currency and real-world currency. The player can react, asking for higher prices as they sell, and be more eager to earn in-game money to pay for their subscription. Finally, inflation represents the con-

nection between real-world currency and the real-world goods, so we will expect that increase the prices (in both, buy and sell), bringing inflation to the game — this last connection relay in imitation of the real-world behavior.

Name	Expression	Description
Buy Price	$\sum_i \frac{P^c_{Bi} - \langle P_{Bi} \rangle_c}{\langle P_{Bi} \rangle_c}$	Buy prices in country c relative to the average across countries.
Bid Ask	$\sum_i \frac{P^c_{Si} - P^c_{Bi}}{P^c_{Si} + P^c_{Bi}}$	Bid-ask spread in prices for country c.
Buy Volume	$\sum_i \frac{V^c_{Bi} - \langle V_{Bi} \rangle_c}{\langle V_{Bi} \rangle_c}$	Volume per transaction of buying orders in country c relative to the average across countries.
Bid Ask Volume	$\sum_i \frac{V^c_{Si} - V^c_{Bi}}{V_{Si} + V_{Bi}}$	Bid-ask spread for the traded volume per transaction in country c.
Bid Ask Transactions	$\sum_i \frac{N^c_{Si} - N^c_{Bi}}{N^c_{Si} + N^c_{Bi}}$	Bid-ask spread in the number of transactions in country c.

Table 5.3: Measures used to capture the in-game trade activity of the individual players in EVE online grouped in the different countries. The symbols $P^c_{Bi}, V^c_{Bi}, N^c_{Bi}$ stand for the price, the volume and the number of transactions of buy orders for the item i in a country c. The symbols $P^c_{Si}, V^c_{Si}, N^c_{Si}$ are the equivalent for the sell orders. The mean values refer to the average for the item i across the M different countries: $\langle P_{Bi} \rangle_c = \sum_{c=1}^{c=M} \frac{P^c_{Bi}}{M}$ and similar definitions for the other quantities.

	count	mean	std	min	25%	50%	75%	max
Buy Price	2113	0.159	0.331	-0.823	-0.005	0.099	0.235	2.946
Bid Ask	1614	-0.088	0.183	-0.991	-0.158	-0.079	-0.006	0.895
Buy Volume	2113	-0.052	0.426	-1.000	-0.283	-0.126	0.065	2.981
Bid Ask Volume	1614	-0.067	0.229	-1.000	-0.155	-0.041	0.029	1.000
Bid Ask Transactions	1614	-0.025	0.209	-0.967	-0.132	-0.027	0.056	0.976

Table 5.4: Summary statistics for the variables capturing levels of in-game trade activities defined in Table 5.3. We include data for the 106 countries with more than 15 players in EVE online.

5.4 Analysis and results

As explained in the introduction our aim is to investigate whether a player's real-world socioeconomic environment permeates into its in-game behavior. We test this connection at the level of countries. In the previous section we have outlined how we designed country profiles from recorded in-game events from its players. Thereby, the focus was on measures for the in-game levels of social interactions, of aggressiveness, of production activities and of trading activities. In what follows we seek to investigate the correlation between (i) the in-game levels of social interactions, of aggressiveness, of production activities and the real-world socioeconomic measures GTI, GPI, and (ii) the in-game trading activities and the real-world measures CPI, REER and UNEMP. For these analyses, we include all countries with more than 15 players in EVE online.

5.4.1 In-game player behavior and the real-world socioeconomic environment

Several studies have established that the real-world levels of aggressiveness in a country as defined by the Global Terrorist Index correlate with a variety of social, economic, and geographical measures, like the linguistic fractionalization, the GDP, and the Human Development Index (HDI) [127, 128]. As players develop activities in EVE online, their in-game activities and attitude can correlate with the different real-world measures. First, we focus on the possible correlation between real-world and in-game aggressiveness. Second, we wish to elucidate whether the levels of real-world aggresiveness activities are correlated with the levels of in-game economic activities and socialo interactions. The research questions are as follows:

- Are the in-game levels of aggressiveness correlated with the levels of aggressiveness in the real world?
- Are the in-game level of social interactions and in-game level of economic activity correlated with the aggressiveness in the real world?

For this Analysis, we use the country profiles based in activities that we calculate in the previous section. In order to test these hypotheses, we develop three models for each real-world measures. We use two real-world measures: GTI and GPI are measures for levels of aggressiveness. As similar classification can be made for the in-game activities of Table 5.1: (*Aggression*, *Kill NPC*) for aggressiveness, (*Industrial*, *Salvaging*) for economic measures and (*Reputation*) for social interactions. *Aggression* is a measure of how likely the players are the aggressors in a fight against other

players and it is measure as: *Engaged Attack*minus *Engaged Defense*. If *Aggression* is positive, the player tend to be the first shooter. *Industrial* is a measure of *Production* plus *Mining*. We also include the in-game social measures (*Reputation*) that we consider as relevant measures for both classes. *Reputation* is measured as the opposite of *Aggression*: *Added As Positive* minus *Added As Negative*. If *Reputation* is positive, other players mark him as a friend more often than as enemy. We propose the following models:

- Model I: This is the minimal model that compares in-game and real-world levels of aggressiveness.We refer to those models as GTI I, GPI I,
- Model II: An extension of the Model I variants that also include the in-game variables that refer to the level of social interactions. These models are referred to as GTI II, GPI II
- Model III: A further extension of the Model II variants that compare the real-world measures with all in-game variables defined in Table 5.1. These models are referred to as GTI III, GPI III

The results of the OLS regressions with all mentioned types of models are contained in Table 5.5. For countries with more than 15 players, we find some correlations between the real-world levels of aggressiveness (GTI and GPI) and in-game behavior of the players.

	GTI I	GTI II	GTI III	GPI I	GPI II	GPI III
Intercept	2.94***	2.95***	3.32***	1.97***	1.97***	2.02***
	(0.29)	(0.29)	(0.64)	(0.05)	(0.05)	(0.11)
Aggression	-0.71**	-0.70**	-0.98***	-0.15***	-0.15***	-0.22***
	(0.28)	(0.28)	(0.29)	(0.05)	(0.05)	(0.06)
Kill NPC	0.42	0.42	1.08**	0.07	0.08	0.19***
	(0.33)	(0.34)	(0.41)	(0.06)	(0.06)	(0.07)
Reputation		-0.20	0.39		0.11**	0.21***
		(0.30)	(0.36)		(0.05)	(0.06)
Industrial			-0.31			-0.04
			(0.56)			(0.10)
Salvaging			-0.96**			-0.19**
			(0.43)			(0.08)
R^2	0.11	0.12	0.21	0.11	0.15	0.23
Adjusted R^2	0.09	0.08	0.14	0.09	0.12	0.18
No. observations	71	71	71	88	88	88
Min Eigenval	5.33e+01	5.31e+01	7.72e+00	7.01e+01	6.98e+01	1.00e+01
Condition number	1.22e+00	1.23e+00	5.04e+00	1.12e+00	1.13e+00	4.84e+00
AIC	3.32e+02	3.34e+02	3.30e+02	1.27e+02	1.25e+02	1.21e+02

Standard errors in parentheses. * p<.1, ** p<.05, ***p<.01

Table 5.5: The level of connection between in-game levels of aggressiveness, of social interactions and of production activities and the real-world Global Terrorist Index (GTI) and Global Peace Index (GPI) per country. Ordinary Least Squared (OLS) regressions are used. The GTI and GPI data are for 2016 and are obtained from [125, 126]

The linear regression for the Global Terrorist Index (GTI) has a lower R^2 than for the Global Peace Index (GPI) (Table 5.5), that takes into account more variables than GTI. Indeed, the GTI only includes intentional acts of violence or threats of violence by a non-state actor' (See [125] for more info). The GPI includes more indicators like the jail population per capita, the size of the armed forces per capita, the number of homicides per capita, as well as the Terrorist index (See [126] for more information about the adopted methodology to construct the indices).

Based on the AIC (Akaike information criterion) we cannot discard any of the models. For all models, the coefficient for the aggressiveness against the A.I. is positive but only significant for the model III in GPI. Considering the aggressiveness against fellow players, *Aggression* is negatively correlated with the GPI whereas NPC is positively correlated. Having a good *Reputation* (being considered more as a friend and less as enemy) is positively correlated with GPI. Countries with higher GPI, can be associated with more peaceful players, as proxied by having more friends and fewer enemies.

We have conducted several robustness checks on the above results. First, we have divided our data into two groups containing 80% and 20% of the data . The linear regression model trained with the 80% group performs comparably to the 20%-group. The signals from the full dataset and the 80%-group are also consistent. Thereby, some signals become less relevant due the increase of the noise. We have also tested the impact of the imposed minimum number of players required — we remind that all results in the tables are for countries with more than 15 players in EVE online. An analysis for a more extended set of countries that results from imposing a minimum of 10 players produces a more noisy signal but most observed tendencies are maintained. We have also tested that imposing a minimum of 25 players per country produces results that are completely in line with those of Tables 5.5 that are obtained with a minimum of 15 players.

5.4.2 In-game trading activity versus the real-world situation

All three regressions in Table 5.6 have low explanatory power (low R^2). Players from countries with higher consumer prices inflation in real-world items (CPI) tend to relatively buy more than they sell in the game (higher *Bid Ask Volume*), which is rational behavior if these players expect that in-game prices will rise too. So it seems that real-world inflation experiences are transmitted into in-game inflation expectations.Players from countries with higher consumer prices inflation in real-world items (CPI) tend to

	CPI	REER	UNEMP
Intercept	108.90***	102.27***	8.17***
	(2.09)	(0.56)	(0.26)
Buy Price	0.37	-2.51**	-1.39***
	(3.74)	(1.03)	(0.47)
Bid Ask	2.48	3.70**	-2.64***
	(5.37)	(1.51)	(0.68)
Buy Volume	-1.18	-1.78***	-0.68**
	(2.46)	(0.66)	(0.32)
Bid Ask Volume	-15.77***	0.39	-0.37
	(4.18)	(1.16)	(0.53)
Bid Ask Transactions	3.20	1.93	1.17**
	(4.56)	(1.23)	(0.58)
R^2	0.02	0.04	0.02
Adj. R^2	0.02	0.03	0.02
No. obser.	1500	1499	1554
Min Eigenval	4.34e+01	3.87e+01	4.38e+01
Cond. number	2.29e+02	2.43e+02	2.31e+02

Standard errors in parentheses. * p<.1, ** p<.05, ***p<.01

Table 5.6: The level of connection between in-game levels of trading activities and the real-world monlty measures Consumer Price Index (CPI), Real Effective Exchange Rate (REER) and Unemployment Rate (UNEMP). Ordinary Least Squared (OLS) regressions are used. The real-world data are for Dec 2011 *up to and including Dec* 2016 *and are from [129].*

relatively buy more than they sell in the game (higher *Bid Ask Volume*), which is rational behavior if these players expect that in-game prices will rise too. So it seems that real-world inflation experiences are transmitted into in-game inflation expectations. The real effective exchange rate is negatively correlated with the in-game *Buy Price* and *Buy Volume*, but positively with the bid-ask spread (*Bid Ask*): if a country's real currency depreciates, in-game money becomes relatively more expensive for the players of this country, urging them to buy smaller quantities at lower prices and to try and trade more efficiently. Finally, the level of unemployment is strongly related to most of our in-game trading behaviors: Players from countries with higher levels of unemployment not only buy less (*Buy Volume*) at cheaper prices (*Buy Price*) and more efficiently (*Bid Ask*) but they have also more sell transactions, relative to buy transactions (*Bid Ask*). These findings are consistent with the view that players of high unemployment countries attach a relatively higher opportunity cost to real-world money than to real-world time, making them more willing to work in the game for earning the in-game money needed to finance their subscription. As a robustness check, we again randomly divided our data in a proportion of 80%/20%. The estimations with the sample of 80% of our data yields results that are consistent with the full sample results. Our model estimated on the remaining 20% of the data is also remarkably accurate. We find that the root-mean-square error (RMSE) of the estimations on these two data samples differ less than 10%.

5.5 Conclusion

Using data on human behavior in a virtual world, we connect in-game player behavior and profiles to the real-world socioeconomic conditions in the players' country of residence. We find evidence of correlations that suggest that conditions in the real-world permeate behavior in the virtual world.

We find a positive connection for in-game aggressiveness against non-player characters and real-world aggressiveness at the country level, as measured by the Global Terrorism Index and Global Peace Index. In contrast, we find a negative connection between real-world aggressiveness and in-game aggressiveness against other players. This suggests that players that do not deal with violent situations and their consequences in their daily experiences are more inclined to be hostile in the game and vice versa. On the other hand countries whose players accumulate more friends than enemies in the game tend to be more peaceful in the real-world. So the ability to make virtual friends in an aggressive virtual environment is related to real-world levels of peace.

We also find small, though significant, and intuitive correlations between in-game trading behavior and variables that capture the macroeconomic conditions of the involved players' countries. These results support the hypothesis that players' interactions on the in-game market are influenced by the macro-economic conditions in their country of residence. We find for example that, strikingly, larger real-world unemployment rates and weaker currencies are associated with more economical and efficient in-game trading behavior and an increased eagerness to earn in-game money, possibly to pay the game subscription with in-game, instead of real-world, money.

The models we study here are obviously incomplete as they only explain a small part of the variation of the real-world measures by in-game behavior. Still they provide stark evidence of a connection between the activities in the virtual world and the socioeconomic conditions of the real-world. This evidence reinforces the idea that real-world conditions permeate player behavior in the virtual world and therefore that observations from the virtual world can, on the margin, be used to make inferences about the real-world.

6

Trade and production in an Online World: Case of EVE-Online

6.1 Abstract

In this work, we used data from the virtual world "EVE Online" to research pricing, which is the process of setting a price for products or services. More specifically, we test which market and production circumstances affect the decisions of players in EVE to sell and buy the manufactured items at specific prices. We found strong correlations with many other market measures, such as material cost and the average size of the transaction, but also with properties set by the developers, such as production time and properties of the production network. We found evidence that pricing is a multivariable process and that other factors apart from the cost are dynamically taken into account. Additionally, we test how the market responds to changes in the requirements of the developers. In the time window considered, EVE Online had monthly patches that can have an impact on production, changing, among others, the production time or the materials needed. We measured the predictability of our model around each one of these patches and found a heterogeneous market response.

6.2 Introduction

Pricing is the process of setting the price of a product or service [130–134]. The limitation of resources and the desire to utilize them have led to the study of economics. As these resources are not distributed, their interchange and trade has been an essential key in human development. These resources can have diverse uses and values, so the setting of a price for different assets became an essential key in human society. Since pricing is a crucial key in the market, many models have tried to illuminate the different methods that humans have used to arrange it, ranging from game theory [135–137] to entropy [138].

As an important key in trade, many models for pricing have been proposed [136, 139, 140]. The key differences between them are the values that they take into account and how they measure them [130, 135, 140]. An element that can be taken into account for complex products is the supply chain [140], which consists of the different processes and steps between the raw material and the final complex products. In [140], Zhang and Huang developed a model of pricing that tries to capture supply chain networks and taxation. As Mizuno et al. [141] showed in their work with Japanese data from 2008-2012, the supply chain network plays a vital role in the growth of companies. In [139], Hinterhuber elaborated on a framework for pricing by trying to consider all relevant dimensions, including cost-volume, customer analysis, company analysis, and competitors. This model set the prices as an iterative process with a feedback loop. Based on game theory, Yaiche and Mazumar [136] developed another theoretical framework focused on the allocation and pricing of telecommunications networks.

Focusing on manufactured products, Marston [142] investigated the export and national markets of Japan in the 1980s and detected evidence that Japanese companies adapt their prices in response to exchange rates. Similarly, Yang and Hwang [143] examined Korean market prices in the export and domestic markets for the period 1976-1990, focusing on a group of commodities. They also found a price response to exchange rates and changes in prices outside Korea. Additionally, Lee [144] found similar results in a later work. Corbo and McNelis [145] used data from periods of liberalization between the mid-1970s to 1985 from three countries: Chile, Israel, and Korea. They found changes in the price dependence on external prices and labor costs before, during, and after liberalization. With more recent data, Lehnert [146] researched differences in the pricing of the stock market between countries. Lehnert used governance indicators from the World Bank and found robust results of higher volatility and negative returns for poorly governed countries.

In recent decades, studies of virtual worlds have been used to study human behavior, including sociology [58, 114], political networks [68, 69], human mobility [60], and economics [59, 115, 116]. The use of a virtual world as a socioeconomic laboratory relies on the similarities between the virtual and real worlds. As in the real world, a person might have no use for a diamond but could still be willing to pay a considerable sum for it. The community determines the value: if some person would pay for the diamond, it has a value. The same is the case for a virtual object in a virtual world [59,115]. The estimation of economic value is set through the willingness to pay with time or currency. The mechanisms that give value in virtual worlds differ little from, for example, the stock market.

One of these virtual worlds is EVE Online [62], which has been used to study international relationships [68,69] and economics [116]. EVE Online is an massively-multiplayer online role-playing game (MMORPG) sandbox where players are free to trade, create complex social structures, go to war against other players, own virtual territories, and exploit resources, among other complex interactions. Compared to other virtual worlds, like World of Warcraft, Destiny, or Second Life, EVE Online has an elaborate production system and marketplace. In EVE's virtual world, there are raw materials and intermediates as well as final goods. Just as in the real world, raw materials need to be mined (or extracted) from the environment, which takes in-game time. Players can improve the rate of extraction by expending currency on in-game equipment and expending time to improve their skills. The players themselves must manufacture almost all in-game items and equipment, and it takes time, blueprints, and input of other materials to create these new manufactured goods. Also, there is a market where the players trade all goods. The creators of the world do not set prices: players decide which price to ask for, and if there is an agreement, the trade takes place. As the items also have a location in the space, they must be carried between extraction places, manufacturing facilities, and trade centers. Players can attack supply lines as well as the mining places to steal their valuable products.

In this work, we use data on the market of EVE online for a time window of two years to study what guides players to set the prices of the manufactured items in-game. Players determined by themselves the prices of the items, but many factors could affect these prices, such as material costs, production time, etc. We study the question of which factors interfere with the pricing of manufactured goods. In addition, we test how the virtual market reacts in the proximity of possible changes in production rules.

6.3 Data on the Virtual World

To produce a new unit of a virtual item in EVE Online, a player must meet the requirements. EVE developers establish these requirements and they are publicly available. In-game, there are different processes of manufacturing, the most common of which is industrial production. We will focus on this type of process. In industrial production, the character that will realize the process must meet some skill requirements. Then, they need to be situated in one of the space stations that can produce the item. In addition, they must have a copy of the blueprint of the virtual item to be produced. Finally, there must also be in the same station a set of items in a certain amount that will be consumed by the production. If the character fulfills all of these requirements, they can start the production, which will take some time (from minutes to days to weeks). After this production time, the final item is added to the inventory of the character. To condense this process, we create the "Production Network." Each node is an item involved in industrial production. We assign a directed link from item A to item B, where item A is a component in the production of B. Each link is weighted with the number of items of A required to create a unit of B and each item has an associated production time. As an example, we present a sub-network of the production network representing the components needed to create a "Raptor" spaceship (Fig 6.1).

The production sub-network for the "Raptor" spaceship has three levels (Fig 6.1), which also is the maximum number of levels for any of the production sub-networks in the complete production network in EVE online. These three levels translate to a maximum distance of two and three item categories: **raw material**, **intermediate item**, and **final item**. We will designate as **raw materials** items that cannot be produced and need to be extracted or found, for example, "Nonlinear Metamaterials" or "Zydrine." A **final item** is an item that cannot be used in the production of any other item, like the "Raptor." Finally, an **intermediate item** is a product that needs to be produced but also is used for the production of final items, like the "Condor" or the "Quantum Microprocessor." This definition is determined only by production, not use. All final items are used for something other than production, but some intermediate items and raw materials also have a use outside of production. In Figure 6.1, we use a color code to differentiate the distance to the final item. To produce a "Raptor" spaceship, all items in the green color are necessary. Some of those items are intermediate materials made with the corresponding blue items. Products that are not raw materials, meaning they need to be produced, are "manufactured items." For example, "Condor" is an intermediate item but it is also a fully

Figure 6.1: Example of a sub-network of the production network: the production tree of the Raptor ship in the Crius patch.

functional ship.

The developers of the game, CCP Games [62], release patches that can change the in-game rules. These patches can modify the methods of conquering areas and the distribution of resources, release new items, change the properties of old items, and make other changes. As a consequence, the production networks after different patches can differ. The different patches and some network properties are shown in Table 6.1. Additionally, we calculate the differences compared to the previous patch. The differences we calculate are the fraction of new nodes and edges, the fraction of nodes and edges that disappeared, and the centralities of the new and missing edges. The changes in the network between patches are small for networks of similar sizes (around 3,100 nodes and 14,000 connections).

	patch Name	Starting Date	Nodes	Edges	Missing Nodes	Missing Edges	Missing Sum Centrality	New Nodes	New Edges	New Sum Centrality
0	Crius	2014-07-22	3171	13861	-	-	-	-	-	-
1	Hyperion	2014-08-26	3174	13879	0.000	0.000	0.000	0.003	0.001	0.003
2	Oceanus	2014-09-30	3056	13795	0.060	0.011	0.025	0.014	0.004	0.011
3	Phoebe	2014-11-04	3083	13909	0.042	0.010	0.023	0.055	0.019	0.045
4	Rhea	2014-12-09	3083	13898	0.033	0.009	0.012	0.026	0.008	0.050
5	Proteus	2015-01-13	3098	13945	0.000	0.000	0.000	0.010	0.003	0.009
6	Tiamat	2015-02-17	3100	13948	0.002	0.001	0.003	0.005	0.001	0.022
7	Scylla	2015-03-24	3105	13957	0.003	0.001	0.003	0.005	0.002	0.006
8	Mosaic	2015-04-28	3055	14148	0.285	0.089	0.233	0.262	0.102	0.206
9	Carnyx	2015-06-02	3051	14220	0.011	0.003	0.006	0.012	0.008	0.023
10	Aegis	2015-07-07	3054	14231	0.000	0.000	0.000	0.004	0.001	0.002
11	Galatea	2015-08-25	3060	14163	0.015	0.008	0.007	0.011	0.004	0.021
12	Vanguard	2015-09-29	3060	14187	0.003	0.001	0.001	0.007	0.003	0.010
13	Parallax	2015-11-03	3060	14187	0.000	0.000	0.000	0.000	0.000	0.000
14	Frostline	2015-12-08	3080	14385	0.014	0.010	0.016	0.035	0.024	0.036

Table 6.1: List of patches in EVE Online and production network properties. Each patch has its own production network. For each new patch, we calculate the fraction of missing and new nodes and edges, as well as the sum of the centrality of the edges..

6.4 Price formation

Our initial hypothesis is that material costs are the driving force for setting prices [131, 140, 147]. We suppose that the players that expend their time and effort in the market would rather see some profit than losses. To attain this profit, they must set a price that at least covers costs. However, the same cost does not necessarily translate to the same price, as some products need exclusive technologies or have more extended times of production. For example, a farmer that only harvests one time per year would try to sell their product at a higher price than similar products that are produced weekly. Besides the intrinsic characteristics of the goods, the market itself can affect the prices. Supply and demand is a common ground for many price determination theories. The existence of monopolies or oligopolies in the market can also be used for price determination [143,148]. Some theories include non-rational price formation [149]. For example, social values can affect prices [123,150]. Here, we focus only on quantifiable values for pricing formation.

For the EVE Online market, we can measure many of these elements and study their connections. In EVE Online, the currency is called InterStellar Kredits or ISK. We will focus on non-raw materials that they have a material cost over 10 ISK associated with them. Trying to cover a broad spectrum of possible factors, we will use the following measures:

- Price (Pr): The average price (weighted by volume) of the item each day. It is set by players using the information they have available.
- Average Volume (V): The average size of the transaction of the item in the patch. Some items (like ships) are sold in low volume (usually one or two in the same transaction), but others are sold in thousands of units per sale (like ores or ammo). Players set the number of items they want to trade in each
- Product Type (Pt): Initial, Intermediate, or Final good.
- Cost (C). The material cost of producing a unit of the item the same day. We calculate using the price and the number of components needed for each item. For Init products, this is zero. This measure depends on the production tree and the market by the prices.
- Product Complexity Index (PCI): A measure of the complexity of a product. We use the method of eigenvalues and extract the second biggest eigenvalue. For more on the methodology, see [151, 152].This measure is the counterpart to the economic complexity index (ECI). Hidalgo et al. build the market matrix M_{ij} where if country i is a

significant producer of the item (o group of items j, $M_{ij} = 1$, and 0 otherwise. Calculating the eigenvector of the product of this matrix with their transpose, they reach to the PCI or ECI. In a market composed of different agents with different items, ECI measures the complexity of the agents while PCI is for items. Intuitively, an item is considered complex if only a few players can sell it in a substantial amount. Note that this measure is only trade-based and does not consider technological complexity. To avoid noise, we consider players who traded a total of at least 6 million ISK in the patch and items traded of at least 6 million ISK.

- Degree Centrality (Dc) and Closeness Centrality (Cc) of the item in the production tree (as an undirected graph) in that patch. These are fixed for each production tree set by the EVE developers. These measures try to capture information about the position of the item in the network.

- Days (D): A time index to measure general trends.

The activity in EVE Online is highly affected by real-world cycles [122]. Weekends have high activity and players from different time zones have different peaks in activity. For this analysis, we will use 7-day moving windows to generate data for the regressions and we exclude the windows that mix two patches. The variables considered are shown in Table 6.2. The Centralities and the Day index are normalized to the range 0-1.

As a first model (Model I), we use a simple economic model, where the price of the item is controlled by just its cost.

$$log(Pr) = \beta_0 + \beta_C log(C) \tag{6.1}$$

As a second model (Model II), we add time dependence.

$$log(Pr) = \text{model I} + \beta_D D \tag{6.2}$$

In the next model (Model III), we add the production information from the production tree.

$$log(Pr) = \text{model II} + \beta_{Pt} Pt + \beta_{Dc} Dc + \beta_{Cc} Cc \tag{6.3}$$

In the next model (Model IV), we add information about how the item is traded to model II.

$$log(Pr) = \text{model II} + \beta_{PCI} PCI + \beta_V log(V) \tag{6.4}$$

Then, we use all of the information in Model V.

$$\begin{aligned} log(Pr) &= \text{model II} + \beta_{Pt}Pt + \beta_{Dc}Dc \\ &\quad +\beta_{Cc}Cc + \beta_{PCI}PCI + \beta_{V}log(V) \end{aligned} \tag{6.5}$$

	count	mean	std	min	25%	50%	75%	max
log(Price)	4626720	14.26	3.93	1.23	12.38	14.44	17.34	21.86
log(Cost)	4626720	14.23	3.12	5.70	12.11	13.94	16.72	22.28
log(Average Volume)	4626720	1.57	2.32	-0.00	0.26	0.78	1.66	11.29
Degree Centrality	4626720	0.04	0.05	0.00	0.02	0.03	0.04	1.00
Closeness Centrality	4626720	0.74	0.20	0.00	0.70	0.77	0.88	1.00
PCI	4626720	-0.03	0.93	-3.02	-0.69	-0.11	0.57	2.95
Days	4626720	0.51	0.29	0.00	0.26	0.51	0.76	1.00

Table 6.2: List of numerical variables considered and their descriptions. For this analysis, we consider items with a cost difference of zero (to consider only manufactured materials) that are traded at least an average of 10 times per day to avoid outliers. The number of data points is 4,626,720. Centralities and Days are normalized to the range 0-1

Finally, we add information for the different patches (Pa) (categorical variable, with respect to patch 0) for Model VI but without the time dependence.

$$log(Pr) = \text{model V} + \beta_{Pa} Pa - \beta_D D \tag{6.6}$$

The results of these regressions are summarized in Tables 6.3, 6.4 and 6.4. We perform the analysis for items with a cost difference of zero (to consider only manufactured materials), that are traded at least an average of 10 times per day, to avoid outliers. The minimum eigenvalue of the matrix and condition number are used as signals for multicollinearity. AIC (Akaike Information Criterion [153]) is used as a measure of model selection.

We observe a qualitative difference between Models I, II, and III compared to Models IV, V, and VI in the R2. The first models have a similar R2, jumping from 0.7 to around 0.92 for Models III, IV, V. With only the information of cost, we can explain almost 70% of the variance in the price. As we expect from the hypothesis, price and cost correlate positively in all models. We do not find significant differences between Model I and Model II or Model III, but the Production tree information returns some unexpected tendencies. Model III's reactions are the opposite, as we would expect, as faster production time, lower centrality, and the intermediate product are associated with higher prices. These tendencies are inverted if we introduce market information. The jump between Models IV, V, and VI and the others is the introduction of market information. This change varies the slopes slightly for the Basic values and completely inverts the production tree information, returning it to a more logical output. Models V and VI are similar in their performance, and the introduction of patches (Model VI) does not have a substantial impact. Model V and Model VI are the best models, according to AIC (Akaike Information Criterion [153]). In all models, all measures have a high significance.

Among the models, we found a positive connection between price and cost, as we would expect under a basic economic premise. Note that the link for Models IV, V, and VI is lower than 1. This indicates that a percentage increment in cost translates to a smaller percentage increase in price. However, if the price is higher than the cost, a proportional change in cost could create a proportionally lower price.

In the complete models (V and VI), Centralities correlate positively with the prices, as do longer production times. However, the size of the transactions (Volume) relates oppositely: products that are sold in high quantities at the same time have lower associated prices. PCI also presents a negative correlation. As PCI measures the knowledge complexity of a product by considering the sales of the players that sell them, this is hard

	Model I	Model II	Model III
Intercept	-0.58***	-0.54***	-0.25***
	(0.00)	(0.01)	(0.01)
log(Cost)	1.04***	1.04***	1.05***
	(0.00)	(0.00)	(0.00)
ProdType Intermediate			0.29***
			(0.00)
Degree Centrality			-3.51***
			(0.02)
Closeness Centrality			-0.49***
			(0.01)
PCI			
log(Average Volume)			
Days		-0.09***	
		(0.00)	
Patches			
R^2	0.68	0.68	0.69
No. observations	4626720	4626720	4626720
Min Eigenval	2.11e+05	1.72e+05	1.27e+04
Condition number	6.84e+01	7.58e+01	2.79e+02
AIC	2.0476e+07	2.0475e+07	2.0431e+07

Standard errors in parentheses. * p<.1, ** p<.05, ***p<.01

Table 6.3: Table - OLS Regressions

	Model IV	Model V	Model VI
Intercept	5.89***	5.95***	6.05***
	(0.00)	(0.00)	(0.00)
log(Cost)	0.70***	0.67***	0.67***
	(0.00)	(0.00)	(0.00)
ProdType Intermediate		-0.49***	-0.49***
		(0.00)	(0.00)
Degree Centrality		4.77***	4.77***
		(0.01)	(0.01)
Closeness Centrality		0.46***	0.47***
		(0.00)	(0.00)
PCI	-0.14***	-0.14***	-0.14***
	(0.00)	(0.00)	(0.00)
log(Average Volume)	-0.93***	-0.96***	-0.96***
	(0.00)	(0.00)	(0.00)
Days	-0.10***	-0.12***	
	(0.00)	(0.00)	
Patches			See Table 6.5
R^2	0.91	0.92	0.92
No. observations	4626720	4626720	4626720
Min Eigenval	1.22e+05	1.20e+04	1.20e+04
Condition number	9.05e+01	2.88e+02	2.88e+02
AIC	1.4458e+07	1.4177e+07	1.4172e+07

Standard errors in parentheses. * p<.1, ** p<.05, ***p<.01

Table 6.4: Table - OLS Regressions

	Model VI
Patch 1	-0.10***
	(0.00)
Patch 2	-0.12***
	(0.00)
Patch 3	-0.17***
	(0.00)
Patch 4	-0.14***
	(0.00)
Patch 5	-0.16***
	(0.00)
Patch 6	-0.18***
	(0.00)
Patch 7	-0.16***
	(0.00)
Patch 8	-0.17***
	(0.00)
Patch 9	-0.22***
	(0.00)
Patch 10	-0.20***
	(0.00)
Patch 11	-0.15***
	(0.00)
Patch 12	-0.18***
	(0.00)
Patch 13	-0.16***
	(0.00)

Standard errors in parentheses. * p<.1, ** p<.05, ***p<.01

Table 6.5: Table - OLS Regressions

to interpret.

6.5 Preparation for the patches

In general, the content of each new patch is not known by the players in advance. The changes in each patch are partially known a few days before the application and rumors spread for weeks before. A patch can change the rules of the game, like the distribution of resources and methods for conquering areas. Patches can also change the properties of items, create new items, or change the way the player produces them. This uncertainty in the proximity of a next patch could affect the behavior of the players and the market. As a possible example, if there are rumors that the production of an item will become harder, the players that trade with it could decide to speculate with it to maximize their profit.

For these possible effects, we will suppose that the market will become harder to predict around the application of a patch. Before the patch, speculation or uncertainty may occur, and after the patch, the player may need some time to adapt their market behavior to the new rules.

We can study how well the regressions perform by conducting regressions using Model IV with the data 7-day window ending that day. As we want to measure the model performance, we consider the determination coefficient (R^2) for each of these regressions around the patch date. The performance of the regression for each day is displayed in Figure 6.2 and the results normalized around each patch are in Fig 6.3. For easy comparison, we compare R^2 with the average R_0^2 of the first 5 days of each window.

The performance among all considered individual window days (Fig 6.2) is similar to the regression performance for all windows (Table 6.3, Model IV). A visual inspection reveals that some patches are more homogeneous, like patch 0; however, the fluctuations inside each patch are similar to the ones between the patches.

The uncertainty of the regression around a new patch does not have a homogeneous response (Fig 6.3). Some patches (like *Aegis*) present a decrease in the determination before the application of the patch, while others do not present a significant change. Some patches (like *Phoebe*) present a higher uncertainty after the patch than before. The spread of different rumors or a diverse release of information before each patch could explain the diversity of effects.

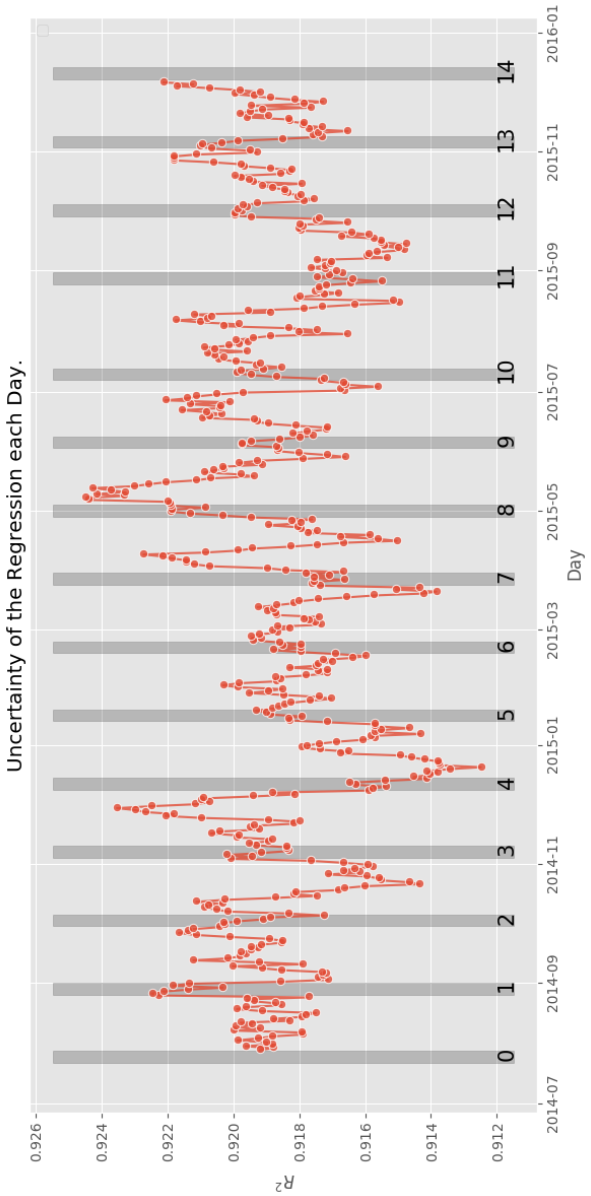


Figure 6.2: Coefficient of determination of the regression near a patch. The gray area represents regressions that cover pre- and post-patch days. The R^2 is compared to the mean R_0^2 of the first 5 days of the graphic.

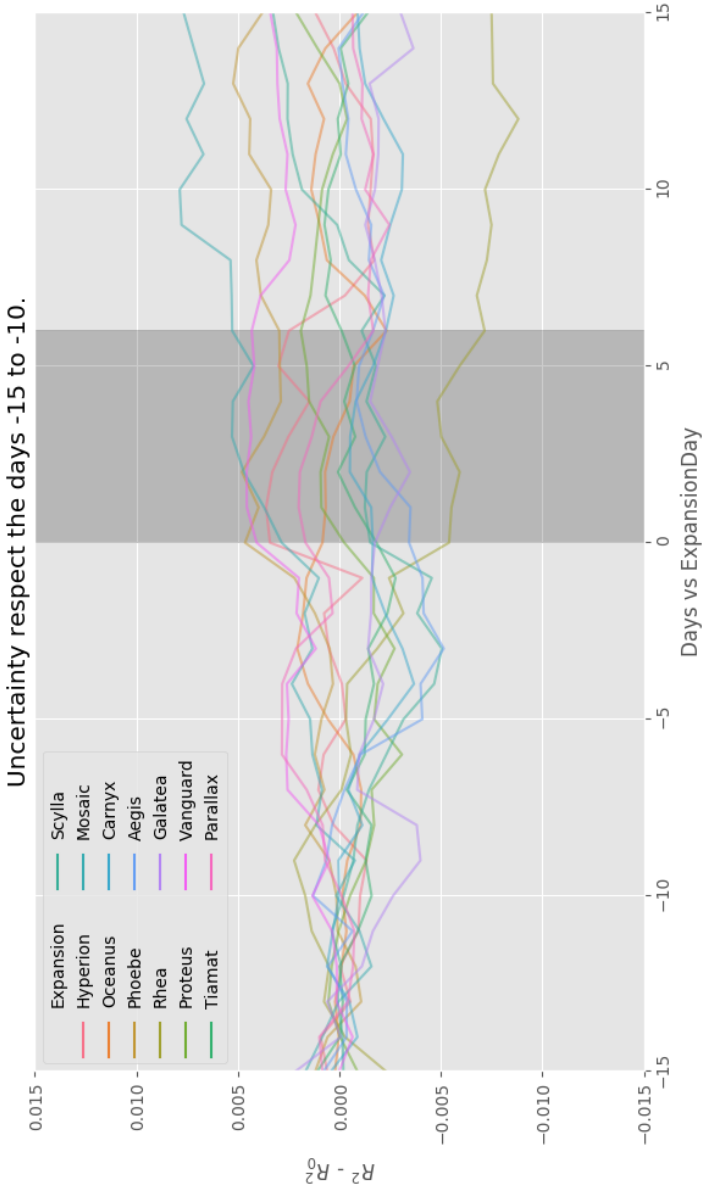


Figure 6.3: Coefficient of determination of the regression near a patch. The gray area represents regressions that cover pre- and post-patch days. The R^2 is compared to the mean R_0^2 of the first 5 days of the graphic.

6.6 Conclusions

In this work, we have used virtual world data to answer real-world questions. More specifically, we study how elements interfere with the pricing of manufactured items in a virtual world. Using different hypotheses, represented as separate models, we performed a series of linear regressions and tested which models best explain the prices. These prices are dynamic and emerged from the market that was conducted by the players.

From the regressions, we found a secure connection between the prices of a manufactured item and the cost of the components (Model I), which can explain almost 70% of the variability. This connection is favorable: higher material costs result in higher market prices. The introduction of other market variables (Model IV) increases the performance sharply (R2 around 0.91). If we also include information about the item fixed by the developers (Model V), we get the best model, according to the Akaike Information Criterion.

Using the complete model, Model V, we found that all variables are firmly significant in explaining the pricing. Both the information set dynamically by the players and the information placed by the developers can be used to describe price formation. Items with lower costs and those sold in high quantities per transaction were associated with reduced prices. Meanwhile, having higher centrality in the production network and being a final product are associated with higher prices. The product complexity index, calculated using the market, correlates negatively with costs but plays a minor role in price formation.

The daily regression shows volatility close to the implementation of some patches. However, this effect does not appear for all patches. The coefficient of determination drops before some patches, while for others it increases. Proximity to changes in the rules could affect the market; however, this is not clear for the peculiarities around each patch. The spread of specific information and rumors would be the transmission belt between the patches and the market, and these are hard to track. The difference in what information is available pre-patch and the information quality would be the factors that would affect the market. The results of this last analysis are inconclusive.

Overall, this analysis of the prices in an online market extracted the quantities that can be used to predict the response of the manufactured items' prices. Our study could help to enlighten real-world questions of supply networks, price formation, and test price theories. However, this work should be expanded to be able to answer these questions.

7

Virtual Economy, Supplementary Materials

This chapter contains some extra material concerning exploratory research on the activity and player categorization in EVE Online. In the first part, we test the effect of weekly cycles on the activity of the players in-game using the market and motion data, exploring whether different weekdays have different activity levels. In the second part, we try to categorize players according to the kind of activities they perform. We create player profiles similar to the country profiles we used in chapter 5 and try to find groups of players that behave similarly.

7.1 Activity

EVE players are subject to real-world time schedules. Work, sleep, and many other real-world activities can impact their activity in the virtual world. Feng et al. [122] have studied the temporal effects in EVE's in-game activity and EVE's in-game population for three consecutive years. In the study, they found a time dependence on the number of players connected with a daily and weekly periodicity. The weekly periodicity would mean that each day of the week presents similar activity with respect to the other days of the week. Here, we will focus on the question of whether we can distinguish the different weekdays using the activity in-game. For activity,

we will use two distinct variables: the EVE Market and the movement of the players in the space.

For market activity, we will use the number of transactions each day. We use the period from 22/07/2014 to 07/12/2015. The data covers 504 days, with $4.38 \pm 0.73 \cdot 10^5$ transactions per day. For movement, we use the number of jumps on each day. EVE Online presents a network of solar systems where the players can move between solar systems (nodes) using gates that connect two solar systems (edges). Each of these events is called a "Gate Jump." The data for jumps covers 01/07/2015 to 1/7/2017, with one month (10/2015) missing. We use 702 days, with $2.41 \pm 0.59 \cdot 10^6$ jumps per day. We consider server time, so all events are recorded in Iceland time (UTC+0). For example, a German player in Germany (UTC+1) performing a jump on Saturday 00:30 would be recorded as Friday 23:30 in Iceland. Aggregating all the data for each weekday, we can see in Figure 7.1 a comparison of the distribution of the counts and in Figure 7.2 the Kernel density estimation for each weekday and the total distribution.

In Figures 7.1 and 7.2, we can perceive differences between days. For market activity, there is a clear difference between the distribution on a workday (Monday to Friday) and the weekend days (Saturday and Sunday). Weekend days appear to be more active than workdays. For jumps, outliers distort the figure and the differences between days are less clear. However, the picture also suggests a more productive weekend than workdays. For testing whether these distributions are different, we set a Kolmogorov–Smirnov (KS) test for two samples [154, 155]. A KS test is a statistical test that compares the empirical distribution function with the cumulative distribution function of a reference. The test could also be used to test if two empirical distributions differ. In Figure 7.3, we present the p-values of the KS-Test. With these p-values, we measure the probability that the distribution for each day is the same underlying distribution (null hypothesis).

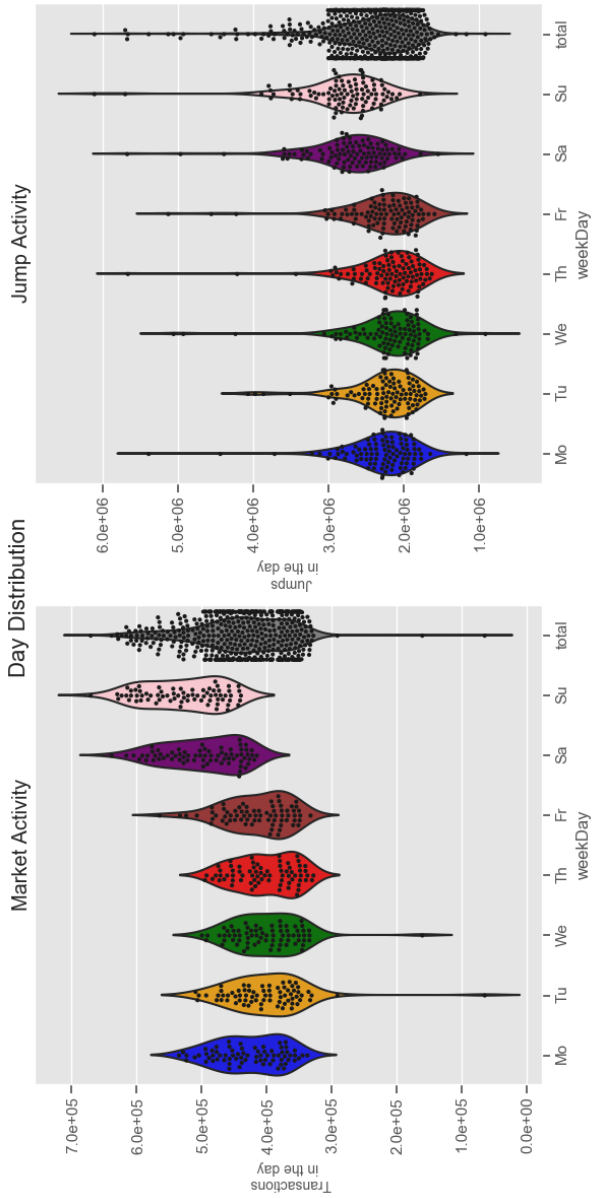


Figure 7.1: Frequency of activities per day.

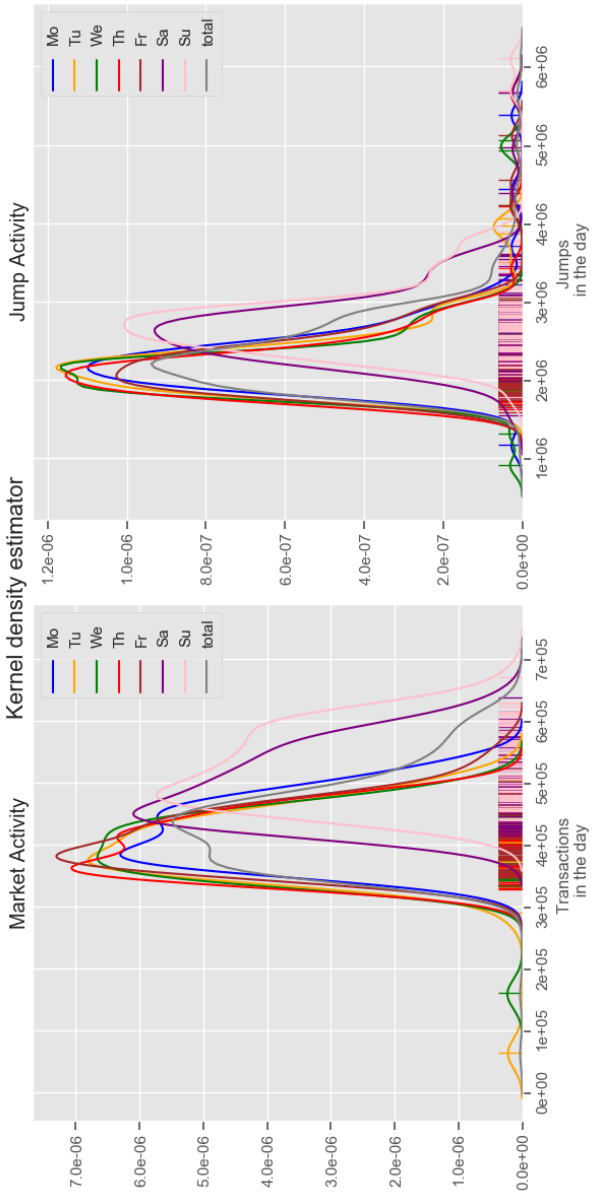


Figure 7.2: Kernel density estimation for activities per day.

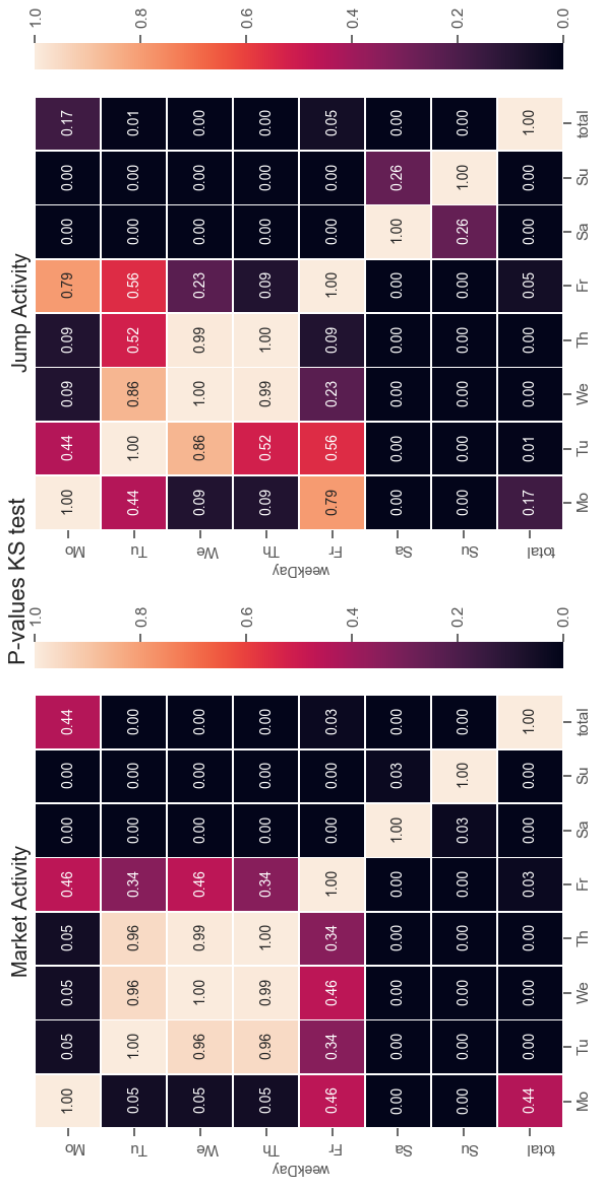


Figure 7.3: P-values of the Kolmogorov–Smirnov test for the distribution of each pair of days and the total distribution.

The results of the KS test (Fig 7.3) showed a clear difference between the weekend days and the other days and gave us a probability that they are the same. For market activity, we can discard the null hypothesis (that both days have the same underlying distribution) with great certainty. Also, for the market activity, we found that Tuesdays, Wednesdays, and Thursdays present similar distributions that are indistinguishable, but it is not clear for Mondays and Fridays. In comparison, weekend days are unique according to market activity. Finally, all days are unique when compared with the global distribution, except for Mondays. For jumps, we obtained similar results: a clear difference between workdays vs. weekends. However, there is not a clear difference between the two weekend days, as there is for market activity.

These results indicate that the level of market activity and mobility are dependent on the day of the week. Players are more active on the weekend than on workdays, both in market and jump activities. The difference is statistically significant and is compatible with the weekly cycles already reported in [122]. This difference could affect any time-series analysis in EVE Online data. A possible method to reduce this effect is to use a time-window average that covers the same number of instances of each weekday.

7.2 Player Profiling

In this section, we will try to categorize players using player profiles, similar to the country profiles we developed in chapter 5. In chapter 5, we create a country profile by adding the activity of the players that belong to that country. With the same process, we can calculate profiles for individual characters and try to categorize them based on their activity. Thawonmas et al. in [156] and in [157] perform character classification and identification based on their actions in massively-multiplayer online role-playing game (MMORPG) designed for research [158]. Using the relative frequency of the player actions, they categorize the player into a certain pre-set group. In this work, we will try to cluster the players into groups without creating group characteristics.

Each player can have several characters, which tend to be specialized. To calculate the character profiles, we perform the same process as is shown in chapter 5. We consider the number of times each character performs a specific action (Table 7.1). Then, we calculate the average of each action for each character that performs this specific action (removing zeros) in a week to use as a unit of reference. Finally, we normalize the actions done per character by dividing each activity by the total activities performed per character. The process assigns to each player a point in hyper-sphere where

each dimension is each kind of action. Players close to one of the axis are performing only the action (Killing NPC, etc.) of that coordinate.

Next, we try to aggregate players by the kind of actions they realize. There are many methods to cluster points in the hyperspace, such as K-means or spectral clustering [159, 160], but they usually need the input of the number of clusters. Here, we will use hierarchical agglomerative clustering (HAC). HAC starts by considering all data points as their own cluster, and each step merges two clusters. It continues until all data points belong to the same cluster. The criteria under which they are selected to merge are the parameters of this algorithm. For this analysis, we will use the libraries SciPy [72] and SciKit-learn [74] from Python. We use as criteria the Ward variance minimization algorithm [161], which merges two clusters if the merger would create a new cluster with less variance.

In our analysis, we use players with high activity for whom the sum of their activity coefficients is more than 100 reference units. The results are that 38,349 different characters are selected to be categorized. We can see the last 12 merges and their variances in Figure 7.4. We mark the three most significant changes in the internal variances in a merge using black horizontal lines. These correspond to the partitions with two, three, and six clusters.

After choosing the partition, we can categorize the players using the clusters to which they belong. In Table 7.2, the different partitions, the size of the cluster measured as a fraction of the total population, and the center of each cluster are displayed. In Figures 7.5, 7.6 and 7.7, 100 random characters are displayed, as well as the centroids of the clusters. When we divide the data into two groups (Table 7.2 and Fig 7.5), the algorithm separates the data into two highly asymmetric groups. The bigger group (Cluster A0) has 83% of the data and presents a generic profile, while the smaller (Cluster A1) has the remaining 17% with a profile more focused on *Mining* (*Min*). With the division of 3 groups (Table 7.2 and Fig 7.6), the bigger cluster was split between a generalist (67% of the total sample, Cluster B0) and a smaller fraction characterized by higher kills of non-player characters (*Kill NPC* or *KNP*), Cluster B1. Cluster B2 is equivalent to Cluster A1. When splitting into 6 Clusters (Table 7.2 and Fig 7.7), we got similar clusters as before, consisting of a big generalist C0, a mining C2 that was the equivalent of B2 and A1, and a cluster C3 that was the equivalent of B1. The new clusters are a combat-focused C1, which focuses on conflicts (*Atk* and *Def*), an industrialist C4, and C5, which was centered in Ratting (*Kill NPC* and *Salvaging*).

Name	Description
Added As Negative *Added As Positive*	Interacting players can mark the counter-party as friend, enemy, or neutral. This mark is visible whenever the players meet.
Fleet	A group of players playing together can form a "fleet." Members of a fleet can track each other's status and location at all times.
Engaged Attack *Engaged Defense*	Attack another player. Be attacked by another player.
Kill NPC	Destroy ships controlled by the game, usually "pirates."
Production	Produce an item, using other items as input.
Mining	Extract materials from asteroids.
Salvaging	Recover some materials from destroyed ships.

Table 7.1: In-game activities in EVE Online considered for character profiling.

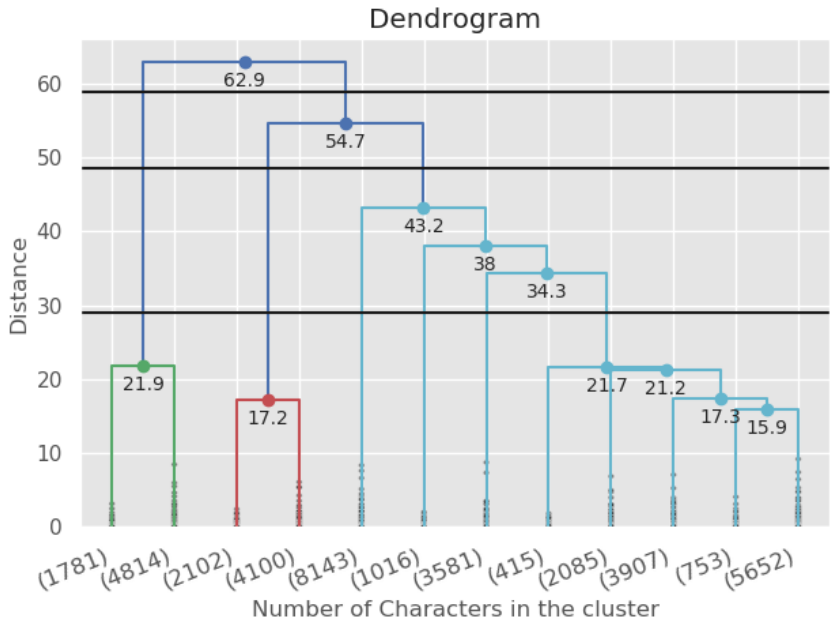


Figure 7.4: Dendrogram for the 38,349 players. Each vertical line represents a cluster and each colored horizontal line represents a merge of two clusters. In the vertical axis, the variance considered to merge two clusters is represented. Points represent the merging points and variance of the new cluster. For simplicity, only the 12 last merges are displayed. The three black lines represent the most significant jumps in the variances, which are the merging points used for clustering.

Partition	Cluster	Size	Profile Center								
			Pro	*Sal*	*Min*	*KNP*	*Atk*	*Def*	*Fle*	*Fri*	*Ene*
A	A0	0.83	0.07	0.09	0.04	0.23	0.13	0.15	0.24	0.04	0.03
	A1	0.17	0.02	0.03	0.58	0.05	0.01	0.02	0.25	0.02	0.01
B	B0	0.67	0.08	0.09	0.04	0.13	0.16	0.17	0.26	0.04	0.03
	B1	0.16	0.03	0.06	0.02	0.63	0.03	0.05	0.16	0.02	0.01
	B2	0.17	0.02	0.03	0.58	0.05	0.01	0.02	0.25	0.02	0.01
C	C0	0.33	0.07	0.06	0.07	0.15	0.10	0.12	0.32	0.06	0.05
	C1	0.21	0.01	0.01	0.01	0.04	0.33	0.34	0.24	0.02	0.01
	C2	0.17	0.02	0.03	0.58	0.05	0.01	0.02	0.25	0.02	0.01
	C3	0.16	0.03	0.06	0.02	0.63	0.03	0.05	0.16	0.02	0.01
	C4	0.03	0.86	0.01	0.02	0.02	0.00	0.01	0.05	0.01	0.00
	C5	0.09	0.02	0.44	0.06	0.25	0.01	0.03	0.14	0.03	0.01

Table 7.2: Clustering and center of the clusters. For each kind of partition, the algorithm divides the players into clusters. It split the players into 2 (A), 3 (B), or 6 (C) clusters of different sizes. The size is calculated as the fraction of total players that belong to a specific group. The number of total players is 38,349. We display the center of each cluster as a representative of the cluster. The coordinates are Production (Pro), Salvaging ("Sal"), Mining (Min), Kill NPC (KNP), Engaged Attack (Atk), Engaged Defense (Def), Fleet (Fle), Added As Positive (Fri), and Added As Negative (Ene).

Figure 7.5: Results for a two-cluster grouping of player profiles: (left) sample of 100 random players and (right) center of the clusters. The colors correspond to the different clusters. The coordinates correspond to: Production (Pro), Salvaging (Sal), Mining (Min), Kill NPC (KNP), Engaged Attack (Atk), Engaged Defense (Def), Fleet (Fle),Added As Positive (Fri), and Added As Negative (Ene).

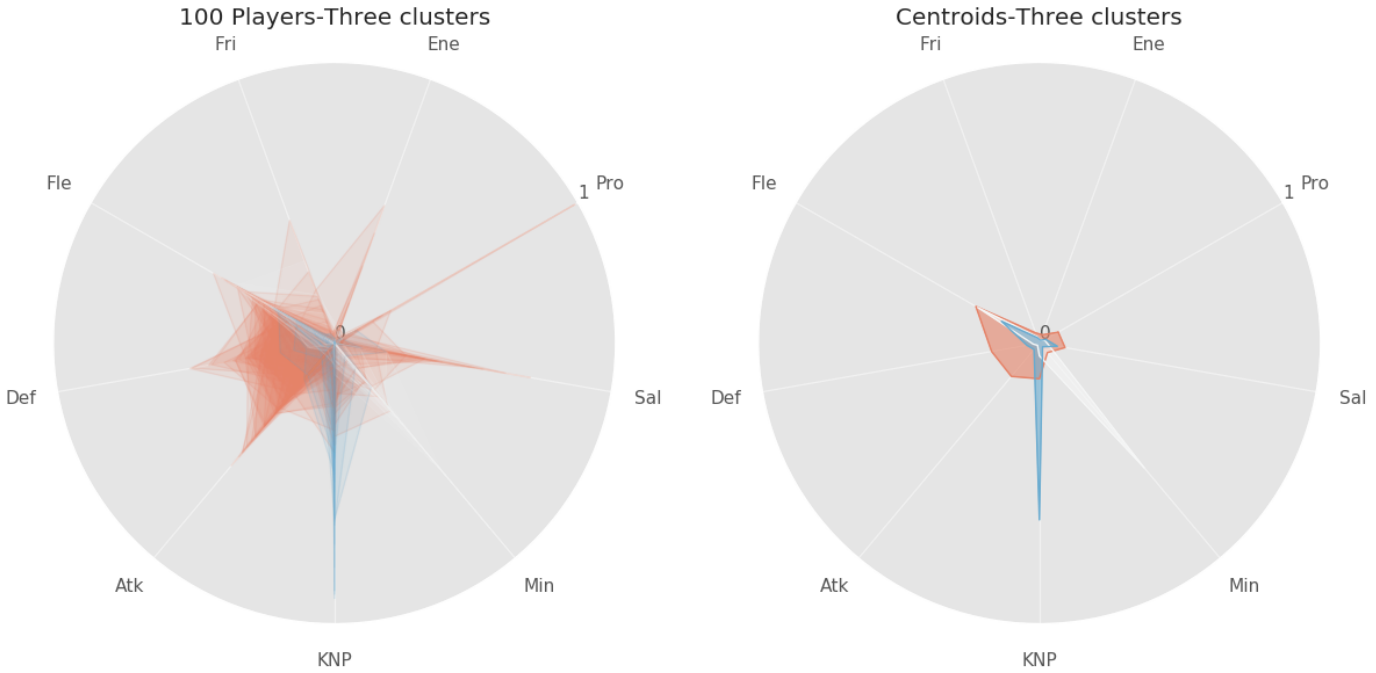


Figure 7.6: Results for a three-cluster grouping of player profiles: (left) Sample of 100 random players and (right) center of the clusters. The colors correspond to the different clusters. The coordinates correspond with: Production (Pro), Salvaging (Sal), Mining (Min), Kill NPC (KNP), Engaged Attack (Atk), Engaged Defense (Def), Fleet (Fle), Added As Positive (Fri), and Added As Negative (Ene).

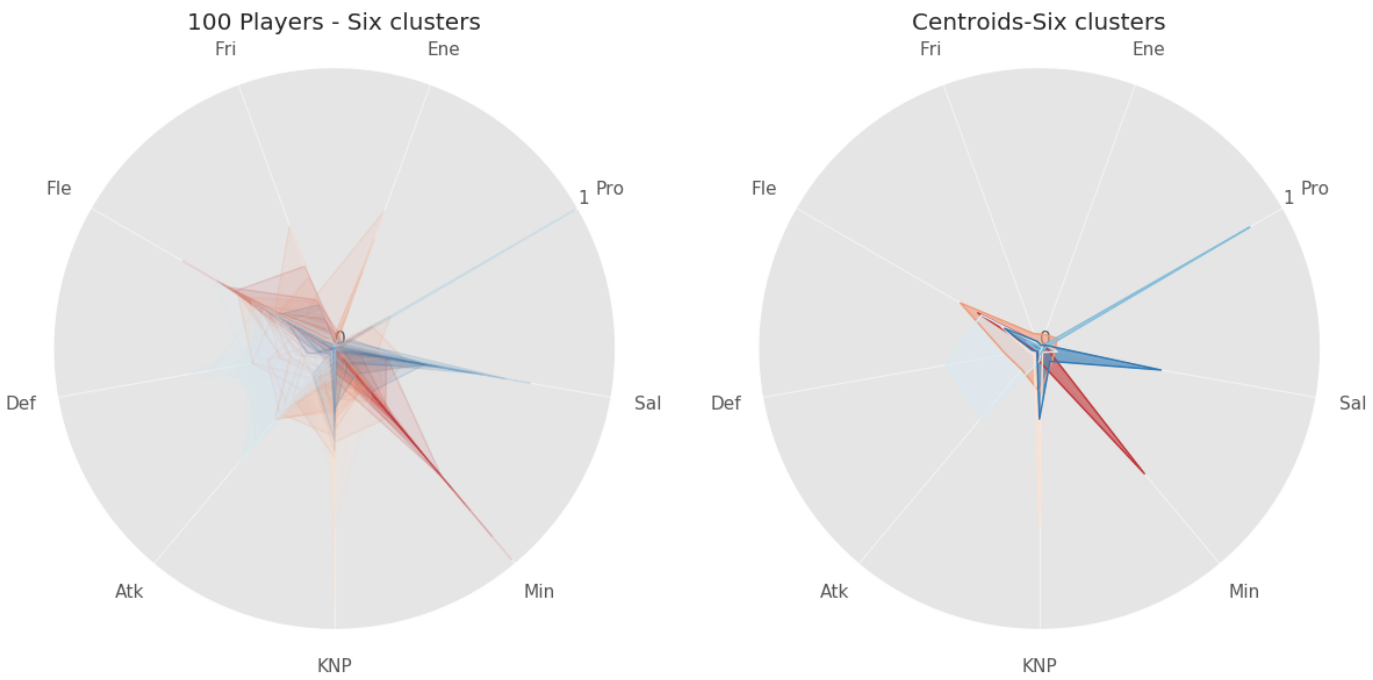


Figure 7.7: Results for a six-cluster grouping of player profiles: (left) sample of 100 random players and (right) center of the clusters. The colors correspond to the different clusters. The coordinates correspond to: Production (Pro), Salvaging (Sal), Mining (Min), Kill NPC (KNP), Engaged Attack (Atk), Engaged Defense (Def), Fleet (Fle), Added As Positive (Fri), and Added As Negative (Ene).

This algorithm split the players successfully into groups by their activity in-game, allowing the developers to target specific groups for advertising or pools. Before a patch that changes the in-game rules for fighting and can disturb the Player vs. Player combat system, the developers can survey the players that belong to the C1 Cluster, while players in clusters A1, B2 and B3 can be surveyed prior to changes in the distribution of the natural resources. In addition, this fragmentation allows us to study the time evolution of the different profiles in the game.

7.3 Conclusions

In this chapter, we compile different works about the socioeconomic aspects of EVE Online. The different sections plant seeds for more in-depth study in EVE and other MMORPGs as social and economic laboratories.

In the first part, we study the activity in EVE online performed by players daily and compare the days of the week. We consider the market activity (number of market transactions each day) as well as motion in-game (the number of jumps) to find a clear difference between workdays and weekdays. Players are more active during weekends than workdays and the difference is statistically significant. The activity each day has different underlying distributions, with an average increase of 20% in the number of transactions and 50% in the jumps.

In the second part, we use the activity performed by the different characters in-game to categorize them using a clustering algorithm. We found different possible partitions using hierarchical clustering. With a low number of clusters, it starts with a group of generalist characters that can be split into smaller and more specialized clusters. With these clusters, we can categorize characters for survey targeting or further studies of the EVE population.

Part III

Summary and Outlook

8

Summary and Outlook

The dissertation presented here can be subdivided into two parts connected by the sources of the data. In the first section, we collect the work concerning a statistical physics model for social interactions, focusing on international political networks. This model is first presented in chapter 2, expanded in chapter 3, and its properties are studied in chapter 4. In the second section, we assemble a set of works about different socioeconomic activities in a virtual world. In chapter 5, we study connections between real and virtual worlds. Next, we study price formation in chapter 6. Finally, we plant a seed for future works in chapter 7. Although the different sections and the chapters study different research questions with different perspectives, they are connected by the use of data from a virtual world.

In our first chapter, studying the international relationship network, we propose a methodology to implement social balance theory using Boltzmann-Gibbs statistical physics. Using a generic Hamiltonian with three-edge, two-edge, and one-edge interactions and using the degeneracy of the each type of triad, we found that the interactions have characteristic strength parameters that can be extracted from a model/data comparison. We have tested the model with two datasets from the virtual world EVE Online and two datasets from the real world, the Cold War and the Middle East. The data that we analyzed indicate that there is no clear tendency toward an increased occupation of the balanced triads as time progresses and have been reported in other systems [44]. However, we can capture this effect through the introduction of a finite temperature.

In the second chapter on international relationships, we extend the previous model with the option of creating and destroying edges. The Standard structural balance builds on active (i.e., "+" and "−") edges; however, it corresponds to a tiny fraction. With the addition of inactive edges, we can analyze the political networks as complete networks and facilitate a mean-field approach. We have used both the Hamiltonian model as well as its mean-field approximation in the analysis, with three sets of time-series data for the relationships in political networks. We use two sets of data from the virtual world EVE Online and one set from the Cold War era. We found that the cost of creating a connection is higher than any other interaction, locating the ground state of the system at a set of nodes with exclusively inactive connections, in contrast to previous models which took universal utopia as the minimal energy system state. With the transition probabilities, we found a fair model-data agreement for the EVE data. For the Cold War, the edge attributes are somewhat static, and it is difficult for the model to describe the data. With the mean-field approximation, we found a linear scaling between a non-trivial function for the average magnetization and activation, and the data provided evidence for the predicted linear scaling.

The last chapter for the section on international networks condenses a more in-depth analysis of the Hamiltonian developed in this section. The Hamiltonian works with a generalized version of the basic social balance, that can accommodate not only the difference incidence inside balance found in real-world systems but also includes other common models, such as the Ising model for magnetic interaction. In this chapter, we focus on the study of the behavior of the Hamiltonian without applying it to data. Using a mean-field approximation and Montecarlo simulations, we found that this Hamiltonian presents phase transitions whose existence is controlled by the relationship between the model parameters. The results of both methods agree qualitatively in the set of parameters and temperatures studied, but the phase transition was a bit displaced between both techniques.

In the first chapter of the second section, which focused on the economy of the virtual world, we connect in-game actions and player profiles with the performance and the socioeconomic situations in their countries. In chapter 5 of this section, we find evidence of correlations that could indicate that the real world can penetrate into a virtual world (EVE Online). More precisely, we found a negative connection between the aggressiveness out of the game (measured as the Global Terrorist Index and Global Peace Index) with more friendly profiles, as players are perceived more as a friend and less as an enemy and they are more likely to be dragged into combats rather than starting them. We also found correlations in market behavior and some

economic measures that players can feel directly, like unemployment, the exchange rate of the local currency vs. the dollar, and price inflation. These correlations support the hypothesis that players bring the socioeconomic status of their country to the way they interact in the market. With the evidence of interconnection and permeability between the virtual and real worlds, we reinforce the idea that they can be used as a tool for real-world studies.

In the second economic chapter, we study the pricing of manufacturing in a virtual world. In chapter 6 of this section, we use different hypotheses represented as separate models and we perform a series of linear regressions to test which models best explain the prices. These prices are dynamic and emerged from the market conducted by the players. We found a definite connection between the prices of a manufactured item and the cost of the components that can explain almost 70% of the variability. Through the introduction of other market measures and information on the production process, we increase the performance sharply (R^2 around 0.91). We also study the volatility close to changes in the production rules (patches) and find a heterogeneous response. Before some patches, the coefficient of determination drops, but for others, it increases. Proximity to changes in the rules could affect the market; however, this effect is not clear due to the peculiarities around each expansion. The difference in what information is available pre-patch and in the information quality is a factor that would affect the market, and the results of this last analysis are inconclusive.

In the last chapter of the economic section and the final chapter of this dissertation, we collect different works about the socioeconomic aspects in EVE Online to plant seeds for potential further in-depth studies in EVE or other massively-multiplayer online role-playing games (MMORPG). First, we examine the daily activities in the market and the mobility of EVE players and find a clear difference between workdays and weekends, players being more active in the latter. This shift is statistically significant and would affect any time analysis in EVE online. We also use the activity profiles for each character in the game to classify them using a clustering algorithm. With different sensitivities, we found that we can find a group of generalist characters or diverse specialist characters. With these clusters, we can categorize characters for survey targeting or further studies of the EVE population.

8.1 Outlook

In this dissertation, we present several works that can be used as a bedrock for the use of statistical physics in social networks and the use of virtual

worlds as a socioeconomic laboratory. It is presented as self-consistent, but various sections can be easily expanded for a better understanding and improvement of the outcomes. In this final section, we will point out some possible lines of research based on this dissertation.

For the first part of this dissertation, we can expand the applicability of the statistical physics of social balance. The Hamiltonian presented in the first section of this work has proven valuable for the analysis of the international relationship network (virtual and real) but it still can be applied to other systems where social balance plays an essential role in their evolution. Exploring social networks with our model could result in a strong argument in favor of or against these models. Additionally, the model can be tested in other fields where social balance could be applied, such as economic networks, the stock market, and neuron activity. The results of this approximation are unclear but exciting.

Aside from the direct application of the Hamiltonian, this Hamiltonian has unusual behavior. As described in chapter 3, the three-body and one-body interactions scale distinctly with the system extension. The system size affects the balance of terms in the Hamiltonian. This effect could be altered with minimal changes in the Hamiltonian and could lead to more suitable versions of the model for different systems. As an example, changing the two sums into one that scales with the number of triads would create a system that scales consistently. This new version of the Hamiltonian would shift the balance between the creation of a link vs. the triadic forces and would create systems with low activation. However, this would also affect the mean-field approximation and would need to be re-calculated.

For the economic section of this thesis, some consequences of the results could be further explored with a little more data. In chapter 5, we found a connection between the real and virtual world for aggression and market behavior. However, one might propose testing this idea with other aspects of the virtual and real world. One aspect could be mobility, to check the relationship between different scales of human mobility (international/national/urban) with in-game mobility. In addition, the production in EVE relies on the transportation of materials from the extraction point to production centers, and then to distribution. This logistic line could be similar to real-world transport networks, but the similarity between these two processes is still unproven.

The pricing analysis, as presented in chapter 6, could work as a starting point for a more in-depth study of virtual world pricing. Separation of markets by region, taking into account regional exports and imports, and better differentiation of high and low technology could further illuminate the process of pricing. Additionally, the effect of rule changes in EVE Online is

still unclear, but a historical study of the information available in the forums around past expansions and a further understanding of pricing could clarify the volatility around these updates.

Finally, the last chapter presents preliminary work that could work as auxiliary material for future EVE analysis. The segmentation and classification of players in EVE could lead to a comparison of the change in population roles over time with the different expansions, as well as a connection with real-world economic sectors. In the real world, we found that developed countries have a small fraction of their population in the primary sector. One could test whether the fraction of players in the EVE alliances doing production and mining depends on the "richness" of their alliance.

The use of video game data could help to complement other analysis of the real world. In the next years, in the age of big data, new real-world data will be liberated. This big data could displace virtual world data as share some of their characteristics: precise and clear. However, there are elements of virtual world data that big data will found it difficult to replace. First, the virtual world is more accessible to experiment with: performing a test of human mobility in a virtual world could be more straightforward that convince a city hall to implement it in a city. Second, we only have a timeline in the real world: the study of global economic systems is connected with their time, and we only have a series of events to study — the same stands for historical perspectives in the political networks. Virtual worlds can be used to present an alternative sequence of events to explore new theories.

Overall, this dissertation is a framework for later work using virtual worlds as a social laboratory. The use of virtual laboratories presents unique traits and it is a fantastic opportunity to collect highly detailed data in a diversity of situations. The players that populate these virtual worlds are the same people that populate the real world, and their behavior, obsessions, and way of thinking are carried from the real world to the virtual.

9
Samenvatting

Het hier gepresenteerde proefschrift bestaat uit twee delen, elk gekenmerkt door de bron van de gegevens. In het eerste deel betreft onderzoek verricht op een statistisch–fysisch model voor sociale interacties, gericht op internationale politieke netwerken. Dit model wordt in een eerste instantie beschreven in hoofdstuk 2, verder uitgebreid in hoofdstuk 3 en de eigenschappen ervan worden bestudeerd in hoofdstuk 4. In het tweede deel bekijken we verschillende sociaal-economische activiteiten in een virtuele wereld. In hoofdstuk 5 bestuderen we de verbanden tussen echte en virtuele werelden. Vervolgens onderzoeken we prijsvorming in hoofdstuk 6. Ten slotte geven we een aanzet voor toekomstige ontwikkelingen in hoofdstuk 7. Hoewel de verschillende secties en hoofdstukken verschillende onderzoeksvragen met verschillende perspectieven bestuderen, zijn ze verbonden door het gebruik van gegevens uit een virtuele wereld.

In ons eerste hoofdstuk, dat het internationale relatienetwerk bestudeert, stellen we een methode voor om de theorie van sociale balans te implementeren met behulp van de statistische fysica van Boltzmann–Gibbs. Met behulp van een generieke Hamiltoniaan met drie–rand, twee–rand en Één–rand interacties en met de ontaarding van elk type triade, vonden we dat de interacties karakteristieke sterkteparameters hebben die geëxtraheerd kunnen worden uit een vergelijking tussen model en gegevens. We hebben het model getest met twee datasets uit de virtuele wereld EVE Online en twee datasets uit de echte wereld, de Koude Oorlog en het hedendaagse Midden-Oosten. De gegevens die we hebben geanalyseerd, geven aan dat er geen

duidelijke neiging is tot een toegenomen bezetting van de gebalanceerde triaden naarmate de tijd vordert, hetgeen reeds is aangetoond in andere systemen [44]. We kunnen dit effect echter vastleggen door de introductie van een eindige temperatuur.

In het tweede hoofdstuk over internationale relaties breiden we het vorige model uit met de mogelijkheid om randen te creëren en te vernietigen. Het standaard structurele saldo is een functie van actieve (d.w.z. " + " en " − ") randen; dit vertegenwoordigt echter slechts een klein aandeel van alle relaties. Met de toevoeging van inactieve randen kunnen we de politieke netwerken als complete netwerken analyseren en een gemiddeld–veldbenadering mogelijk maken. We hebben zowel het Hamiltoniaanse model als de gemiddeld–veldbenadering in onze analyse gebruikt, met drie sets tijdreeksgegevens voor de relaties in politieke netwerken. We gebruiken twee sets gegevens uit de virtuele wereld EVE Online en Één set uit het tijdperk van de Koude Oorlog. We hebben geconstateerd dat de kosten van het maken van een verbinding hoger zijn dan elke andere interactie, omdat de grondtoestand van het systeem zich bevindt op een reeks knooppunten met uitsluitend inactieve verbindingen, in tegenstelling tot eerdere modellen die universele utopie als toestand van minimale energie beschouwden. Met de overgangskansen hebben we een redelijke overeenkomst tussen model en data gevonden voor de EVE–gegevens. Voor de Koude Oorlog zijn de randattributen enigszins statisch en heeft het model het moeilijk om de gegevens te beschrijven. Met de gemiddeld–veldbenadering vonden we een lineaire schaling tussen een niet–triviale functie voor de gemiddelde magnetisatie en activering, en de gegevens verschaften bewijs voor de voorspelde lineaire schaling.

Het laatste hoofdstuk van het gedeelte over internationale netwerken bevat een meer diepgaande analyse van de Hamiltoniaan die we in dit gedeelte hebben ontwikkeld. De Hamiltoniaan werkt met een veralgemeende versie van het sociale basisevenwicht, dat niet alleen de verschillende incidenties van evenwicht in echte systemen kan opvangen, maar ook andere gemeenschappelijke modellen omvat, zoals het Ising-model voor magnetische interactie. In dit hoofdstuk richten we ons op de studie van het gedrag van de Hamiltoniaan zonder het op gegevens toe te passen. Met behulp van een gemiddeld–veldbenadering en Monte–carlosimulaties hebben we ontdekt dat deze Hamiltoniaan faseovergangen bevat waarvan het bestaan wordt bepaald door het verband tussen de modelparameters. De resultaten van beide methoden komen kwalitatief overeen binnen het scala aan parameters en temperaturen die hier zijn onderzocht, maar de fase–overgang was enigzins verschoven tussen beide technieken.

In het eerste hoofdstuk van het tweede deel, dat zich richtte op

de economie van de virtuele wereld, leggen we een verband tussen in–game acties en spelersprofielen enerzijds, met de prestaties en de sociaal–economische situaties in hun landen anderzijds. In hoofdstuk 5 van dit gedeelte vinden we aanwijzingen voor correlaties die erop kunnen wijzen dat de echte wereld kan doordringen in een virtuele wereld (EVE Online). Om preciezer te zijn: we vonden een negatief verband tussen de agressiviteit uit het spel (gemeten als de zogenaamde Global Terrorist Index en Global Peace Index) met vriendelijkere profielen, omdat spelers meer als een vriend en minder als een vijand worden gezien en het waarschijnlijker is dat ze worden meegesleept in gevechten in plaats van ze te starten. We hebben ook correlaties gevonden in marktgedrag en enkele economische maatregelen die spelers direct kunnen voelen, zoals werkloosheid, de wisselkoers van de lokale valuta tegenover de dollar en prijsinflatie. Deze correlaties ondersteunen de hypothese dat spelers de sociaaleconomische status van hun land met zich meedragen bij de manier waarop ze op de markt communiceren. Dit bewijs van interconnectie en permeabiliteit tussen de virtuele en echte werelden, staaft het idee dat ze kunnen worden gebruikt als een hulpmiddel voor studies in de echte wereld.

In het tweede economische hoofdstuk bestuderen we de prijsstelling van productie in een virtuele wereld. In hoofdstuk 6 van dit gedeelte gebruiken we verschillende hypothesen die worden voorgesteld als afzonderlijke modellen en voeren we een reeks lineaire regressies uit om te testen welke modellen de prijzen het beste verklaren. Deze prijzen zijn dynamisch en kwamen voort uit de markt gevoerd door de spelers. We hebben een duidelijk verband gevonden tussen de prijzen van een vervaardigd product en de kosten van de onderdelen die bijna 70% van de variabiliteit kunnen verklaren. Door de invoer van andere marktgrootheden en informatie over het productieproces verbeteren we de performantie sterk (R^2 ongeveer 0.91). We bestuderen ook de volatiliteit in tijden van plotse veranderingen in de productieregels (door patches in het spel) en vinden een heterogene reactie. Voor sommige patches daalt de bepalingscoëfficiënt, maar voor anderen neemt deze toe. Het nabij zijn van wijzigingen in de regels kan de markt beïnvloeden; dit effect is echter niet duidelijk vanwege de eigenaardigheden bij elke uitbreiding. Het verschil in informatie die vooraf beschikbaar is en in de informatiekwaliteit zijn factoren die de markt zouden kunnen beïnvloeden, en de resultaten van deze laatste analyse zijn niet doorslaggevend.

In het laatste hoofdstuk van het economische gedeelte en het laatste hoofdstuk van dit proefschrift verzamelen we verschillende werken over de sociaal–economische aspecten in EVE Online om een aanzet te geven voor mogelijk diepgaander onderzoek in EVE of andere massively multiplayer online role–playing games (MMORPG). Eerst onderzoeken we de dagelijkse

activiteiten op de markt en de mobiliteit van EVE-spelers en vinden we een duidelijk verschil tussen werkdagen en de weekends, waarbij spelers actiever zijn bij die laatste. Deze verschuiving is statistisch significant en zou elke tijdsanalyse in EVE online beïnvloeden. We gebruiken ook de activiteitsprofielen voor elk personage in het spel om ze te classificeren met behulp van een clusteralgoritme. Met verschillende gevoeligheden hebben we geconstateerd dat we een groep generieke personages of diverse gespecialiseerde personages kunnen vinden. Met deze clusters kunnen we personages categoriseren voor enquêtetargeting of verdere studies van de EVE-populatie.

References

[1] John Q. Stewart. *The Development of Social Physics.* American Journal of Physics, 18(5):239–253, May 1950.

[2] William Warntz. *Transportation, Social Physics, And The Law Of Refraction.* The Professional Geographer, 9(4):2–7, July 1957.

[3] Mark Buchanan. *The Social Atom: Why the Rich Get Richer, Cheaters Get Caught, and Your Neighbor Usually Looks Like You.* Bloomsbury USA, New York, first edition edition, June 2007.

[4] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. *Statistical Physics of Social Dynamics.* Reviews of Modern Physics, 81(2):591–646, May 2009.

[5] Marc Barthelemy. *The Statistical Physics of Cities.* Nature Reviews Physics, 1(6):406–415, June 2019.

[6] H. E. Stanley, V. Afanasyev, L. A. N. Amaral, S. V. Buldyrev, A. L. Goldberger, S. Havlin, H. Leschhorn, P. Maass, R. N. Mantegna, C. K. Peng, P. A. Prince, M. A. Salinger, M. H. R. Stanley, and G. M. Viswanathan. *Anomalous Fluctuations in the Dynamics of Complex Systems: From DNA and Physiology to Econophysics.* Physica A: Statistical Mechanics and its Applications, 224(1):302–321, February 1996.

[7] Rosario N. Mantegna and H. Eugene Stanley. *Introduction to Econophysics by Rosario N. Mantegna.* /core/books/introduction-to-econophysics/6A2727FE42578790E6E1021B7955EE30, November 1999.

[8] Tobias A. Huber and Didier Sornette. *Can There Be a Physics of Financial Markets? Methodological Reflections on Econophysics.* The European Physical Journal Special Topics, 225(17):3187–3210, December 2016.

[9] Susan Hanson. *The Importance of the Multi-Purpose Journey to Work in Urban Travel Behavior.* Transportation, 9(3):229–248, October 1980.

[10] Balázs Cs. Csáji, Arnaud Browet, V.A. Traag, Jean-Charles Delvenne, Etienne Huens, Paul Van Dooren, Zbigniew Smoreda, and Vincent D. Blondel. *Exploring the Mobility of Mobile Phone Users.* Physica A: Statistical Mechanics and its Applications, 392(6):1459 – 1473, 2013.

[11] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R. James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J. Ramasco, Filippo Simini, and Marcello Tomasini. *Human Mobility: Models and Applications.* Physics Reports, 734:1–74, March 2018.

[12] Thomas C. Schelling. *Dynamic Models of Segregation.* The Journal of Mathematical Sociology, 1(2):143–186, July 1971.

[13] Robert Axelrod. *The Dissemination of Culture: A Model with Local Convergence and Global Polarization.* Journal of Conflict Resolution, 41(2):203–226, April 1997.

[14] Martin A. Nowak and David C. Krakauer. *The Evolution of Language.* Proceedings of the National Academy of Sciences, 96(14):8028–8033, July 1999.

[15] Vittorio Loreto, Andrea Baronchelli, and Andrea Puglisi. *Mathematical Modeling of Language Games.* In Stefano Nolfi and Marco Mirolli, editors, Evolution of Communication and Language in Embodied Agents, pages 263–281. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[16] Michael Batty. *The Size, Scale, and Shape of Cities.* Science, 319(5864):769–771, February 2008.

[17] Xiao-Yong Yan, Chen Zhao, Ying Fan, Zengru Di, and Wen-Xu Wang. *Universal Predictability of Mobility Patterns in Cities.* Journal of The Royal Society Interface, 11(100):20140834, November 2014.

[18] Shu-Heng Chen, Chia-Ling Chang, and Ye-Rong Du. *Agent-Based Economic Models and Econometrics.* The Knowledge Engineering Review, 27(2):187–219, April 2012.

[19] Serge Galam and Jean-Daniel Zucker. *From Individual Choice to Group Decision-Making.* Physica A: Statistical Mechanics and its Applications, 287(3):644–659, December 2000.

[20] Victor M. Yakovenko. *Monetary Economics from Econophysics Perspective.* The European Physical Journal Special Topics, 225(17):3313–3335, December 2016.

[21] Gallotti Riccardo, Bazzani Armando, and Rambaldi Sandro. *Towards a Statistical Physics of Human Mobility.* International Journal of Modern Physics C, 23(09):1250061, July 2012.

[22] Didier Sornette. *Physics and Financial Economics (1776–2014): Puzzles, Ising and Agent-Based Models.* Reports on Progress in Physics, 77(6):062001, May 2014.

[23] A. Drăgulescu and V. M. Yakovenko. *Statistical Mechanics of Money.* The European Physical Journal B, 17(4):723–729, October 2000.

[24] A. Drăgulescu and V.M. Yakovenko. *Evidence for the Exponential Distribution of Income in the USA.* The European Physical Journal B - Condensed Matter and Complex Systems, 20(4):585–589, April 2001.

[25] Mark Newman. *Networks.* OUP Oxford, Oxford, United Kingdom ; New York, NY, United States of America, edición: 2 edition, 26 de julio de 2018.

[26] Jean-Charles Delvenne and Anne-Sophie Libert. *Centrality Measures and Thermodynamic Formalism for Complex Networks.* Phys. Rev. E, 83(4):046117, April 2011.

[27] Giulio Cimini, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and Guido Caldarelli. *The Statistical Physics of Real-World Networks.* Nature Reviews Physics, 1(1):58–71, January 2019.

[28] A. J. Tatem, D. J. Rogers, and S. I. Hay. *Global Transport Networks and Infectious Disease Spread.* In Simon I. Hay, Alastair Graham, and David J. Rogers, editors, Advances in Parasitology, volume 62 of *Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications*, pages 293–343. Academic Press, January 2006.

[29] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J. Ramasco, and Alessandro Vespignani. *Multiscale Mobility Networks and the Spatial Spreading of Infectious Diseases.* Proceedings of the National Academy of Sciences, 106(51):21484–21489, 2009.

[30] Piergiuseppe Morone and Richard Taylor. *Knowledge Diffusion Dynamics and Network Properties of Face-to-Face Interactions*. Journal of Evolutionary Economics, 14(3):327–351, July 2004.

[31] Mandy Haggith, Ravi Prabhu, Carol J. Pierce Colfer, Bill Ritchie, Alan Thomson, and Happyson Mudavanhu. *Infectious Ideas: Modelling the Diffusion of Ideas across Social Networks*. Small-scale Forest Economics, Management and Policy, 2(2):225–239, May 2003.

[32] S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas. *Diffusion Dynamics on Multiplex Networks*. Physical Review Letters, 110(2):028701, January 2013.

[33] Réka Albert, István Albert, and Gary L. Nakarado. *Structural Vulnerability of the North American Power Grid*. Phys. Rev. E, 69(2):025103, February 2004.

[34] Olaf Sporns, Giulio Tononi, and Rolf Kötter. *The Human Connectome: A Structural Description of the Human Brain*. PLOS Computational Biology, 1(4), September 2005.

[35] Stephen P. Borgatti, Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. *Network Analysis in the Social Sciences*. Science, 323(5916):892–895, 2009.

[36] Jeffrey Travers and Stanley Milgram. *An Experimental Study of the Small World Problem*. Sociometry, 32(4):425–443, 1969.

[37] Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts. *An Experimental Study of Search in Global Social Networks*. Science, 301(5634):827–829, 2003.

[38] Réka Albert, Hawoong Jeong, and Albert-László Barabási. *Diameter of the World-Wide Web*. Nature, 401(6749):130–131, September 1999.

[39] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley. *Classes of Small-World Networks*. Proceedings of the National Academy of Sciences, 97(21):11149–11152, 2000.

[40] R. I. M. Dunbar. *Coevolution of Neocortical Size, Group Size and Language in Humans*. Behavioral and Brain Sciences, 16(4):681–694, December 1993.

[41] Bruno Gonçalves, Nicola Perra, and Alessandro Vespignani. *Modeling Users' Activity on Twitter Networks: Validation of Dunbar's Number*. PLOS ONE, 6(8):e22656, August 2011.

[42] Bethany L. Goldblum, Andrew W. Reddie, Thomas C. Hickey, James E. Bevins, Sarah Laderman, Nathaniel Mahowald, Austin P. Wright, Elie Katzenson, and Yara Mubarak. *The Nuclear Network: Multiplex Network Analysis for Interconnected Systems.* Applied Network Science, 4(1):36, December 2019.

[43] Serge Galam. *Fragmentation versus Stability in Bimodal Coalitions.* Physica A: Statistical Mechanics and its Applications, 230(1):174–188, August 1996.

[44] Patrick Doreian and Andrej Mrvar. *Structural Balance and Signed International Relations.* Journal of Social Structure, 16:2, 2015.

[45] Jeffrey Hart. *Symmetry and Polarization in the European International System, 1870 - 1879: A Methodological Study.* Journal of Peace Research, 11(3):229–244, September 1974.

[46] Galina Vinogradova and Serge Galam. *Global Alliances Effect in Coalition Forming.* The European Physical Journal B, 87(11):266, November 2014.

[47] Fritz Heider. *Attitudes and Cognitive Organization.* The Journal of Psychology, 21(1):107–112, January 1946.

[48] Dorwin Cartwright and Frank Harary. *Structural Balance: A Generalization of Heider's Theory.* Psychological Review, 63(5):277–293, 1956.

[49] Giuseppe Facchetti, Giovanni Iacono, and Claudio Altafini. *Computing Global Structural Balance in Large-Scale Signed Social Networks.* Proceedings of the National Academy of Sciences, 108(52):20953–20958, December 2011.

[50] Ernesto Estrada and Michele Benzi. *Walk-Based Measure of Balance in Signed Networks: Detecting Lack of Balance in Social Networks.* Physical Review E, 90(4):042802, October 2014.

[51] Seth A. Marvel, Steven H. Strogatz, and Jon M. Kleinberg. *Energy Landscape of Social Balance.* Physical Review Letters, 103(19):198701, November 2009.

[52] Hongzhong Deng, Peter Abell, Ofer Engel, Jun Wu, and Yuejin Tan. *The Influence of Structural Balance and Homophily/Heterophobia on the Adjustment of Random Complete Signed Networks.* Soc. Networks, 44:190 – 201, 2016.

[53] Kyu-Min Lee, Euncheol Shin, and Seungil You. *An Economic Model of Friendship and Enmity for Measuring Social Balance in Networks.* Physica A, 488:205 – 215, 2017.

[54] Wikipedia. *Corrupted Blood Incident.* https://en.wikipedia.org/wiki/Corrupted_Blood_incident, September 2019.

[55] Blizzard Entertainment. *World of Warcraft.* https://worldofwarcraft.com, 2019.

[56] Ran D. Balicer. *Modeling Infectious Diseases Dissemination Through Online Role-Playing Games.* Epidemiology, 18(2):260, March 2007.

[57] Eric T Lofgren and Nina H Fefferman. *The Untapped Potential of Virtual Game Worlds to Shed Light on Real World Epidemics.* The Lancet Infectious Diseases, 7(9):625–629, September 2007.

[58] Michael Szell, Renaud Lambiotte, and Stefan Thurner. *Multirelational Organization of Large-Scale Social Networks in an Online World.* Proceedings of the National Academy of Sciences, 107(31):13636–13641, August 2010.

[59] Clare Chambers. *How Virtual Are Virtual Economies? An Exploration into the Legal, Social and Economic Nature of Virtual World Economies.* Computer Law & Security Review, 27(4):377–384, August 2011.

[60] Michael Szell, Roberta Sinatra, Giovanni Petri, Stefan Thurner, and Vito Latora. *Understanding Mobility in a Social Petri Dish.* Scientific Reports, 2:457, June 2012.

[61] Bayer & Szell. *Pardus.* https://www.pardus.at/, 2003.

[62] CCP. *EVE Online.* https://www.eveonline.com/, 1997.

[63] CPP_Quant. *Player Age Distribution in EVE Online.* https://www.reddit.com/r/dataisbeautiful/comments/2p2s7m/player_age_distribution_in_eve_online_oc/, February 2014.

[64] Ernst Ising. *Beitrag zur Theorie des Ferromagnetismus.* Zeitschrift für Physik, 31(1):253–258, February 1925.

[65] Stephen G. Brush. *History of the Lenz-Ising Model.* Reviews of Modern Physics, 39(4):883–893, October 1967.

[66] M. Blume, V. J. Emery, and Robert B. Griffiths. *Ising Model for the {λ} Transition and Phase Separation in {$He3$-$He4$} Mixtures.* Phys. Rev. A, 4(3):1071–1077, September 1971.

[67] H. W. Capel. *On the Possibility of First-Order Phase Transitions in Ising Systems of Triplet Ions with Zero-Field Splitting.* Physica, 32(5):966 – 988, 1966.

[68] Andres M. Belaza, Kevin Hoefman, Jan Ryckebusch, Aaron Bramson, Milan van den Heuvel, and Koen Schoors. *Statistical Physics of Balance Theory.* PLOS ONE, 12(8):e0183696, August 2017.

[69] Andres M. Belaza, Jan Ryckebusch, Aaron Bramson, Corneel Casert, Kevin Hoefman, Koen Schoors, Milan van den Heuvel, and Benjamin Vandermarliere. *Social Stability and Extended Social Balance—Quantifying the Role of Inactive Links in Social Networks.* Physica A: Statistical Mechanics and its Applications, 518:270–284, March 2019.

[70] Guido Van Rossum and Fred L Drake Jr. *Python Tutorial.* Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995.

[71] Travis E Oliphant. *A Guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.

[72] Eric Jones, Travis Oliphant, Pearu Peterson, et al. *SciPy: Open Source Scientific Tools for Python.* http://www.scipy.org/, 2001.

[73] Skipper Seabold and Josef Perktold. *Statsmodels : Econometric and Statistical Modeling with Python.* In 9th Python in Science Conference, 2010.

[74] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. *Scikit-Learn: Machine Learning in Python.* Journal of Machine Learning Research, 12(Oct):2825–2830, 2011.

[75] John D. Hunter. *Matplotlib: A 2D Graphics Environment.* Computing in Science & Engineering, 9(3):90–95, May 2007.

[76] Wes McKinney et al. *Data Structures for Statistical Computing in Python.* In Proceedings of the 9th Python in Science Conference, volume 445, pages 51–56. Austin, TX, 2010.

[77] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. *Exploring Network Structure, Dynamics, and Function Using NetworkX*. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, Proceedings of the 7th Python in Science Conference, pages 11 – 15, Pasadena, CA USA, 2008.

[78] Tiago P. Peixoto. *The Graph-Tool Python Library*. figshare, 2014.

[79] Aaron Meurer, Christopher P Smith, Mateusz Paprocki, Ondřej Čertík, Sergey B Kirpichev, Matthew Rocklin, AMiT Kumar, Sergiu Ivanov, Jason K Moore, Sartaj Singh, et al. *SymPy : Symbolic Computing in Python*. PeerJ Computer Science, 3:e103, 2017.

[80] Norman P Hummon and Patrick Doreian. *Some Dynamics of Social Balance Processes: Bringing Heider Back into Balance Theory*. Social Networks, 25(1):17–49, January 2003.

[81] J. Leskovec, D. Huttenlocher, and J. Kleinberg. *Signed Networks and in Social and Media*. In CHI 2010: Machine Learning and Web Interactions April 10–15, 2010, Atlanta, GA, USA, 2010.

[82] Jürgen Lerner. *Structural Balance in Signed Networks: Separating the Probability to Interact from the Tendency to Fight*. Social Networks, 45:66–77, March 2016.

[83] T. Antal, P. L. Krapivsky, and S. Redner. *Dynamics of Social Balance on Networks*. Physical Review E, 72(3):036121, September 2005.

[84] Hugo Saiz, Jesús Gómez-Gardeñes, Paloma Nuche, Andrea Girón, Yolanda Pueyo, and Concepción L. Alados. *Evidence of Structural Balance in Spatial Ecological Networks*. Ecography, 40(6):733–741, 2017.

[85] Peter Abell. *Structural Balance in Dynamic Structures*. Sociology, 2(3):333–352, September 1968.

[86] James A. Davis. *Clustering and Structural Balance in Graphs*. Human Relations, 20(2):181–187, May 1967.

[87] T. Antal, P. L. Krapivsky, and S. Redner. *Social Balance on Networks: The Dynamics of Friendship and Enmity*. Physica D: Nonlinear Phenomena, 224(1):130–136, December 2006.

[88] Peter Abell and Mark Ludwig. *Structural Balance: A Dynamic Perspective*. The Journal of Mathematical Sociology, 33(2):129–155, March 2009.

[89] P. Gawronski, P. Gronek, and K. Kulakowski. *The Heider Balance and Social Distance.* Acta Physica Polonica B, 36(8):2549–2558, 2005.

[90] Vincent Antonio Traag, Paul Van Dooren, and Patrick De Leenheer. *Dynamical Models Explaining Social Balance and Evolution of Cooperation.* PLOS ONE, 8(4):1–7, April 2013.

[91] Steve Pressé, Kingshuk Ghosh, Julian Lee, and Ken A. Dill. *Principles of Maximum Entropy and Maximum Caliber in Statistical Physics.* Rev. Mod. Phys., 85(3):1115–1141, July 2013.

[92] The Economist. *Friends and Foes: Rifts in the Middle East.* The Economist, December 2015. Published: The Economist.

[93] "Wikipedia". *Military Intervention against ISIL — Wikipedia, The Free Encyclopedia".* https://en.wikipedia.org/wiki/Military_intervention_against_ISIL, April 2016.

[94] *The Correlates of War.* http://correlatesofwar.org, 2016.

[95] Douglas M Gibler. *"International Military Alliances, 1648-2008.".* CQ Press, 2009.

[96] Glenn Palmer, Vito D'Orazio, Michael Kenwick, and Matthew Lane. *"The MID4 Data Set: Procedures, Coding Rules, and Description.".* Conflict Management and Peace Science, 2015.

[97] Aaron Bramson, Kevin Hoefman, Milan van den Heuvel, Benjamin Vandermarliere, and Koen Schoors. *Measuring Propagation with Temporal Webs.* In Temporal Network Epidemiology, pages 57–104. Springer, 2017.

[98] P. Singh, S. Sreenivasan, B. K. Szymanski, and G. Korniss. *Competing Effects of Social Balance and Influence.* Phys. Rev. E, 93(4):042306, April 2016.

[99] Michael Szell and Stefan Thurner. *Measuring Social Dynamics in a Massive Multiplayer Online Game.* Social Networks, 32(4):313–329, October 2010.

[100] H. T. Diep, Miron Kaufman, and Sanda Kaufman. *Dynamics of Two-Group Conflicts: A Statistical Physics Model.* Physica A: Statistical Mechanics and its Applications, 469:183–199, March 2017.

[101] Miron Kaufman, Hung T. Diep, and Sanda Kaufman. *Sociophysics of Intractable Conflicts: Three-Group Dynamics.* Physica A: Statistical Mechanics and its Applications, 517:175–187, March 2019.

[102] Haifeng Du, Xiaochen He, Jingjing Wang, and Marcus W. Feldman. *Reversing Structural Balance in Signed Networks.* Physica A, 503:780 – 792, 2018.

[103] Alec Kirkley, George T. Cantwell, and M. E. J. Newman. *Balance in Signed Networks.* arXiv:1809.05140, 2018.

[104] Tiago P. Peixoto. *Inferring the Mesoscale Structure of Layered, Edge-Valued, and Time-Varying Networks.* Phys. Rev. E, 92(4):042807, October 2015.

[105] M. Blume. *Theory of the First-Order Magnetic Phase Change in UO_2.* Phys. Rev., 141(2):517–524, January 1966.

[106] P. V. Santos, F. A. da Costa, and J. M. de Araújo. *The Random Field Blume-Capel Model Revisited.* J. Magn. Magn. Mater., 451:737 – 740, 2018.

[107] Leo P. Kadanoff. *More Is the Same; Phase Transitions and Mean Field Theories.* Journal of Statistical Physics, 137(5):777, September 2009.

[108] Emanuel Costabile, Marcio A. Amazonas, J. Roberto Viana, and J. Ricardo de Sousa. *Study of the First-Order Transition in the Spin-1 Blume–Capel Model by Using Effective-Field Theory.* Phys. Lett. A, 376(45):2922 – 2925, 2012.

[109] Christoph Mark, Claus Metzner, Lena Lautscham, Pamela L. Strissel, Reiner Strick, and Ben Fabry. *Bayesian Model Selection for Complex Dynamic Systems.* Nat. Commun., 9(1):1803, 2018.

[110] F. Rabbani, Amir H. Shirazi, and G. R. Jafari. *Mean-Field Solution of Structural Balance Dynamics in Nonzero Temperature.* Physical Review E, 99(6):062302, June 2019.

[111] Leo P. Kadanoff. *More Is the Same; Phase Transitions and Mean Field Theories.* J. Stat. Phys., 137(5):777, September 2009.

[112] Roberta Sinatra and Michael Szell. *Entropy and the Predictability of Online Life.* Entropy, 16(1):543–556, January 2014.

[113] Benedikt Fuchs and Stefan Thurner. *Behavioral and Network Origins of Wealth Inequality: Insights from a Virtual World.* PLOS ONE, 9(8):e103503, August 2014.

[114] Bernat Corominas-Murtra, Benedikt Fuchs, and Stefan Thurner. *Detection of the Elite Structure in a Virtual Multiplex Social System by Means of a Generalised K-Core.* PLOS ONE, 9(12):e112606, December 2014.

[115] Juho Hamari and Lauri Keronen. *Why Do People Buy Virtual Goods: A Meta-Analysis.* Computers in Human Behavior, 71:59–69, June 2017.

[116] Kevin Hoefman, Aaron Bramson, Koen Schoors, and Jan Ryckebusch. *The Impact of Functional and Social Value on the Price of Goods.* PLOS ONE, 13(11):e0207075, November 2018.

[117] Marcel Martončik and Ján Lokša. *Do World of Warcraft (MMORPG) Players Experience Less Loneliness and Social Anxiety in Online World (Virtual Environment) than in Real World (Offline)?* Computers in Human Behavior, 56:127–134, March 2016.

[118] Fan Zhang and David Kaufman. *Massively Multiplayer Online Role-Playing Games (MMORPGs) and Socio-Emotional Wellbeing.* Computers in Human Behavior, 73:451–458, August 2017.

[119] Ryan Perry, Anders Drachen, Allison Kearney, Simone Kriglstein, Lennart E. Nacke, Rafet Sifa, Guenter Wallner, and Daniel Johnson. *Online-Only Friends, Real-Life Friends or Strangers? Differential Associations with Passion and Social Capital in Video Game Play.* Computers in Human Behavior, 79:202–210, February 2018.

[120] Stefan Thurner, Michael Szell, and Roberta Sinatra. *Emergence of Good Conduct, Scaling and Zipf Laws in Human Behavioral Sequences in an Online World.* PLOS ONE, 7(1):e29796, January 2012.

[121] Marcus Carter. *Emitexts and Paratexts: Propaganda in EVE Online.* Games and Culture, 10(4):311–342, July 2015.

[122] Wu-chang Feng, David Brandt, and Debanjan Saha. *A Long-Term Study of a Popular MMORPG.* In Proceedings of the 6th ACM SIGCOMM Workshop on Network and System Support for Games, NetGames '07, pages 19–24, New York, NY, USA, 2007. ACM.

[123] Sherwin Rosen. *Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition.* Journal of Political Economy, 82(1):34–55, January 1974.

[124] Daniel Müllner. *Modern Hierarchical, Agglomerative Clustering Algorithms.* arXiv:1109.2378 [cs, stat], September 2011.

[125] Institute for Economics & Peace. *Global Terrorism Index 2015.* Institute for Economics & Peace, 2015.

[126] Institute for Economics & Peace. *Global Peace Index 2015.* Institute for Economics & Peace, 2015.

[127] Alberto Abadie. *Poverty, Political Freedom, and the Roots of Terrorism.* American Economic Review, 96(2):50–56, 2006.

[128] Eva Godovicová. *The Relationship between Terrorism and Poverty in the Scholarly Debates.* Societas et Res Publica; Trnava, 1(3):75–85, 2012.

[129] World Bank. *World Bank Open Data.* https://data.worldbank.org/, 2019.

[130] Fischer Black and Myron Scholes. *The Pricing of Options and Corporate Liabilities.* Journal of Political Economy, 81(3):637–654, May 1973.

[131] Mark Bils and Yongsung Chang. *Understanding How Price Responds to Costs and Production.* Carnegie-Rochester Conference Series on Public Policy, 52:33–77, June 2000.

[132] Karl Claxton, Andrew Briggs, Martin J. Buxton, Anthony J. Culyer, Christopher McCabe, Simon Walker, and Mark J. Sculpher. *Value Based Pricing for NHS Drugs: An Opportunity Not to Be Missed?* BMJ, 336(7638):251–254, January 2008.

[133] Shutao Cao, Wei Dong, and Ben Tomlin. *Pricing-to-Market, Currency Invoicing and Exchange Rate Pass-through to Producer Prices.* Journal of International Money and Finance, 58:128–149, November 2015.

[134] Francisco Barillas and Jay Shanken. *Comparing Asset Pricing Models.* The Journal of Finance, 73(2):715–754, 2018.

[135] R. W. Ferrero, S. M. Shahidehpour, and V. C. Ramesh. *Transaction Analysis in Deregulated Power Systems Using Game Theory.* IEEE Transactions on Power Systems, 12(3):1340–1347, August 1997.

[136] H. Yaiche, R. R. Mazumdar, and C. Rosenberg. *A Game Theoretic Framework for Bandwidth Allocation and Pricing in Broadband Networks*. IEEE/ACM Transactions on Networking, 8(5):667–678, October 2000.

[137] Maryam Esmaeili, Ghazaleh Allameh, and Taraneh Tajvidi. *Using Game Theory for Analysing Pricing Models in Closed-Loop Supply Chain from Short- and Long-Term Perspectives*. International Journal of Production Research, 54(7):2152–2169, April 2016.

[138] Mihály Ormos and Dávid Zibriczky. *Entropy-Based Financial Asset Pricing*. PLOS ONE, 9(12):e115742, December 2014.

[139] Andreas Hinterhuber. *Towards Value-Based Pricing—An Integrative Framework for Decision Making*. Industrial Marketing Management, 33(8):765–778, November 2004.

[140] Yan Zhang and J. H. Huang. *Cost-Based Pricing Model with Value-Added Tax and Corporate Income Tax for a Supply Chain Network*. Applied Mathematical Modelling, 38(1):168–180, January 2014.

[141] Takayuki Mizuno, Wataru Souma, and Tsutomu Watanabe. *The Structure and Evolution of Buyer-Supplier Networks*. PLOS ONE, 9(7):e100712, July 2014.

[142] Richard C. Marston. *Pricing to Market in Japanese Manufacturing*. Journal of International Economics, 29(3):217–236, November 1990.

[143] Yung Y. Yang and Min Hwang. *Price Behavior in Korean Manufacturing*. The Review of Economics and Statistics, 76(3):461–470, 1994.

[144] Professor Jaewoo Lee. *Pricing-To-Market In Korean Manufacturing Exports*. International Economic Journal, 9(4):1–12, December 1995.

[145] Vittorio Corbo and Paul D. McNelis. *The Pricing of Manufactured Goods during Trade Liberalization: Evidence from Chile, Israel, and Korea*. The Review of Economics and Statistics, 71(3):491–499, 1989.

[146] Thorsten Lehnert. *Asset Pricing Implications of Good Governance*. PLOS ONE, 14(4):e0214930, April 2019.

[147] Yan Zhang and Guoping Xia. *Short-Run Cost-Based Pricing Model for a Supply Chain Network*. International Journal of Production Economics, 128(1):167–174, November 2010.

[148] Mathieu Parenti, Philip Ushchev, and Jacques-François Thisse. *Toward a Theory of Monopolistic Competition.* Journal of Economic Theory, 167:86–115, January 2017.

[149] Herbert A. Simon. *Rationality in Psychology and Economics.* The Journal of Business, 59(4):S209–S224, 1986.

[150] Kelvin J. Lancaster. *A New Approach to Consumer Theory.* Journal of Political Economy, 74(2):132–157, April 1966.

[151] César A. Hidalgo and Ricardo Hausmann. *The Building Blocks of Economic Complexity.* Proceedings of the National Academy of Sciences, 106(26):10570–10575, June 2009.

[152] Eric Kemp-Benedict. *An Interpretation and Critique of the Method of Reflections.* https://mpra.ub.uni-muenchen.de/60705/, December 2014.

[153] Hamparsum Bozdogan. *Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions.* Psychometrika, 52(3):345–370, September 1987.

[154] A. Kolmogorov. *Sulla Determinazione Empirica Di Una Lgge Di Distribuzione.* Inst. Ital. Attuari, Giorn., 4:83–91, 1933.

[155] N. Smirnov. *Table for Estimating the Goodness of Fit of Empirical Distributions.* The Annals of Mathematical Statistics, 19(2):279–281, June 1948.

[156] Ruck Thawonmas, Ji-Young Ho, and Yoshitaka Matsumoto. *Identification of Player Types in Massively Multiplayer Online Games.* In Proc. the 34th Annual Conference of International Simulation And Gaming Association (ISAGA2003), Chiba, Japan, pages 893–900, 2003.

[157] Ji-Young Ho and Ruck Thawonmas. *Episode Detection with Vector Space Model in Agent Behavior Sequences of MMOGs.* In Proceedings of Future Business Technology Conference, pages 47–54, 2004.

[158] Amund Tveit, Øyvind Rein, Jørgen V. Iversen, and Mihhail Matskin. *Scalable Agent-Based Simulation of Players in Massively Multiplayer Online Games.* In In Proceedings of the 8th Scandinavian Conference on Artificial Intelligence, Frontiers in Artificial Intelligence and Applications. IOS. Press, 2003.

[159] Keinosuke Fukunaga and Larry D. Hostetler. *The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition.* IEEE Trans. Information Theory, 21:32–40, 1975.

[160] D. Comaniciu and P. Meer. *Mean Shift: A Robust Approach toward Feature Space Analysis.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5):603–619, May 2002.

[161] Joe H. Ward Jr. *Hierarchical Grouping to Optimize an Objective Function.* Journal of the American Statistical Association, 58(301):236–244, March 1963.