



universität  
wien

# DISSERTATION

Titel der Dissertation

Statistical physics approaches to  
large-scale socio-economic networks

Verfasser

Dipl.-Ing. Michael Szell

angestrebter akademischer Grad

Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien, 2011

Studienkennzahl lt. Studienblatt: A 791 411

Dissertationsgebiet lt. Studienblatt: Dr.-Studium der Naturwissenschaften (Physik)

Betreuer: Univ. Prof. DDr. Mag. Stefan Thurner

Univ. Prof. Dr. Mag. Christoph Dellago



# Contents

<b>Outline</b>	<b>8</b>
<b>1 Introduction: The need for a data-driven social science</b>	<b>9</b>
<b>2 Statistical physics frameworks for social systems</b>	<b>13</b>
2.1 Human beings as particles? . . . . .	13
2.1.1 From disorder to order . . . . .	14
2.1.2 Mean-field approximation and the role of topology . . . . .	15
2.2 Complex Systems . . . . .	17
2.2.1 Self-organization and power laws . . . . .	17
2.2.2 Breakdown of classical entropy . . . . .	18
2.2.3 Anomalous diffusion . . . . .	18
2.3 Network theory . . . . .	19
2.3.1 Random networks . . . . .	19
2.3.2 Scale-free networks . . . . .	20
2.3.3 Models of growing networks . . . . .	21
2.3.4 Rewiring and models of non-growing networks . . . . .	22
2.3.5 Notation and statistical tools . . . . .	22
<b>3 Quantification of socio-economic behavior using an online society</b>	<b>31</b>
3.1 Establishing a socio-economic laboratory . . . . .	31
3.1.1 Problems with data . . . . .	31
3.1.2 The research potential of Massive Multiplayer Online Games .	32
3.1.3 Previous data-driven work on online worlds . . . . .	33
3.1.4 Employing the self-developed online world Pardus . . . . .	34
3.2 Overview and history of Pardus . . . . .	38
3.3 The game universes . . . . .	39
3.4 The data analyzed . . . . .	41
3.5 Players and census of characters . . . . .	42
3.5.1 Age and nationality . . . . .	42
3.5.2 Lifetimes of characters . . . . .	42
3.5.3 Gender of characters . . . . .	43
3.6 Life within Pardus . . . . .	43
3.6.1 Social interaction . . . . .	43

3.6.2	Economy . . . . .	45
<b>4</b>	<b>On the statistical nature of social actions and reactions</b>	<b>53</b>
4.1	Human behavioral sequences . . . . .	53
4.2	Transition probabilities . . . . .	55
4.3	World lines . . . . .	57
4.4	Motifs, entropy and Zipf law . . . . .	60
4.5	Discussion of behavioral sequences . . . . .	62
<b>5</b>	<b>Mobility in a society: Diffusion, long-time memory, and borders</b>	<b>64</b>
5.1	Game universe properties . . . . .	66
5.2	Data set description . . . . .	67
5.3	Basic features of motion . . . . .	68
5.4	Mobility reveals socio-economic clusters . . . . .	69
5.5	A long-term memory model . . . . .	75
5.6	Discussion of mobility of players . . . . .	80
<b>6</b>	<b>Measuring social dynamics with socio-economic networks</b>	<b>82</b>
6.1	Network extraction . . . . .	82
6.2	Evolution of networks . . . . .	84
6.2.1	Testing for preferential attachment . . . . .	84
6.2.2	Network densification . . . . .	87
6.2.3	Measuring non-trivial network properties . . . . .	88
6.3	Testing classical sociological hypotheses . . . . .	90
6.3.1	Confirmation of the Weak ties hypothesis . . . . .	90
6.3.2	Confirmation of triadic closure . . . . .	93
6.3.3	Confirmation of the Dunbar number . . . . .	99
6.3.4	Inconclusive social balance dynamics . . . . .	99
6.4	Two categories of enemies . . . . .	101
6.5	Differences in network types . . . . .	103
6.6	Summary of social network dynamics . . . . .	103
<b>7</b>	<b>Multi-relational organization of societies: Network–network interactions</b>	<b>105</b>
7.1	Nature of the various networks . . . . .	109
7.2	Network–network interactions . . . . .	110
7.3	Large-scale empirical test of structural balance . . . . .	114
7.4	A network evolution model of signed triadic closure . . . . .	121
7.5	Discussion of the multi-relational organization of societies . . . . .	122
<b>8</b>	<b>Summary and outlook</b>	<b>123</b>
	<b>Bibliography</b>	<b>127</b>

# Abstract

In the past decade a variety of fields has been explored by statistical physicists, leading to an increase of our quantitative understanding of various systems composed of many interacting elements, such as social systems. However, an empirical quantification of human behavior on a societal level has so far proved to be tremendously difficult due to problems in data availability, quality and ways of acquisition. In this doctoral thesis we compile for the first time a large-scale data set consisting of practically all actions and properties of 350,000 odd participants of an entire human society interacting in a self-developed Massive Multiplayer Online Game, over a period of five years. We describe this social system composed of strongly interacting players in the game in three consecutive levels. In a first step, we examine the *individuals* and their behavioral properties over time. A scaling and fluctuation analysis of action-reaction time-series reveals persistence of the possible actions and qualitative differences between “good” and “bad” players. We then study and model the diffusion process of human mobility occurring within the “game universe”. We find subdiffusion and a power-law distributed preference to return to more recently visited locations. Second, on a higher level, we use network theory to quantify the topological structure of *interactions* between the individuals. We focus on six network types defined by direct interactions, three of them with a positive connotation (trade, friendship, communication), three with a negative one (enmity, attack, punishment). These networks exhibit non-trivial statistical properties, e.g. scale-free topology, and evolve over time, allowing to test a series of long-standing social-dynamics hypotheses. We find qualitative differences in evolution and topological structure between positive and negative tie networks. Finally, on a yet higher level, we consider the *multiplex network* of the player society, constituted by the coupling of the single network layers. We quantify interactions between different networks and detect the non-trivial organizational principles which lead to the observed structure of the system and which have been observed in real human societies as well. Our findings with the multiplex framework provide evidence for the half-century old hypothesis of structural balance, where certain frustrated states on a microscopic level tend to be avoided. Within this setup we demonstrate the feasibility for generating novel scientific insights on the nature of collective human behavior in large-scale social systems.

# Kurzfassung

Die statistische Physik erforschte im letzten Jahrzehnt eine Fülle von wissenschaftlichen Gebieten, was zu einem besseren quantitativen Verständnis von verschiedenen, aus vielen Elementen bestehenden Systemen, z.B. von sozialen Systemen, geführt hat. Eine empirische Quantifizierung von menschlichem Verhalten auf gesellschaftlichem Niveau hat sich allerdings bisher als sehr schwierig erwiesen, wegen Problemen bei der Gewinnung und Qualität von Daten. In dieser Doktorarbeit erstellen wir zum ersten mal einen umfangreichen über fünf Jahre gesammelten Datensatz, der praktisch alle Aktionen und Eigenschaften der 350.000 Teilnehmer einer gesamten menschlichen Gesellschaft aus einem selbstentwickelten Massive Multiplayer Online Game enthält. Wir beschreiben dieses aus stark wechselwirkenden Spielern bestehende soziale System in drei Ebenen. In einem ersten Schritt analysieren wir die *Individuen* und deren Verhalten im Verlauf der Zeit. Eine Skalen- und Fluktuationsanalyse von Aktions-Reaktions-Zeitreihen enthüllt Persistenz der möglichen Aktionen und qualitative Unterschiede zwischen “guten” und “schlechten” Spielern. Wir untersuchen danach den Diffusionsprozess der im Spieluniversum stattfindenden Bewegungen der Individuen. Wir finden Subdiffusivität und eine durch ein Potenzgesetz verteilte Präferenz zu kürzlich besuchten Orten zurückzukehren. Zweitens, auf der nächsthöheren Ebene, verwenden wir Netzwerktheorie um die topologische Struktur der *Interaktionen* zwischen Individuen zu quantifizieren. Wir konzentrieren uns auf sechs durch direkte Interaktionen definierte Netzwerke, drei davon positiv (Handel, Freundschaft, Kommunikation), drei negativ (Feindschaft, Attacke, Bestrafung). Diese Netzwerke weisen nichttriviale statistische Eigenschaften auf, z.B. skaleninvariante Topologie, und entwickeln sich in der Zeit, was uns erlaubt eine Reihe von Hypothesen über sozialdynamische Phänomene zu testen. Wir finden qualitative Unterschiede zwischen positiven und negativen Netzwerken in Evolution und Struktur. Schließlich untersuchen wir das *Multiplex-Netzwerk* der Spielergesellschaft, das sich aus den einzelnen Netzwerk-Schichten zusammensetzt. Wir quantifizieren Interaktionen zwischen verschiedenen Netzwerken und zeigen die nichttrivialen Organisationsprinzipien auf die auch in echten menschlichen Gesellschaften beobachtet wurden. Unsere Erkenntnisse liefern Belege für die Hypothese der strukturellen Balance, die eine Vermeidung von gewissen frustrierten Zuständen auf mikroskopischem Niveau postuliert. Mit diesem Aufbau demonstrieren wir die Möglichkeit der Gewinnung neuartiger wissenschaftlicher Erkenntnisse über die Natur von kollektivem menschlichen Verhalten in großangelegten sozialen Systemen.

# Acknowledgements

Foremost, I would like to express my sincere gratitude to my main advisor Stefan Thurner for the continuous support of my PhD studies and research, for his guidance, enthusiasm, and the sharing of his knowledge and practical skills on a multitude of scientific and organizational levels from which I have learned immensely. I also thank my second advisor Christoph Dellago and all colleagues involved at the Physics center for their dedication and the smooth admission at University of Vienna. Further, I would like to thank my colleagues Peter Klimek, Manfred Pöchacker, Rudolf Hanel, Peter Vavra, Manuel Schölling, Roberta Sinatra, Marco D'Errico, Sebastian Poledna, as well as Anita Wanjek, with whom I have spent a most enjoyable time at Medical University of Vienna. Moreover, I am grateful to my collaborators Renaud Lambiotte and Giovanni Petri for having provided me a wonderful time in London, as well as to Roberta Sinatra and Vito Latora for their extraordinary hospitality in Catania. I would also like to thank all the other international colleagues and friends for their support and for invitations and accommodations for collaborations, conferences, workshops and summer schools throughout Europe and the US, and for their kind feedback and advice on scientific matters. I further thank Werner Bayer for his patience and help concerning data gathering and cleaning, as well as for providing computing power for some resource-intensive calculations. Last but not least I thank my parents Katalin and Andreas for their constant support of my endeavors, in particular of my personal development and scientific career. Köszönöm!

I gratefully acknowledge financial support from the European Cooperation in Science and Technology (COST) Action MP0801 Physics of Competition and Conflicts, from the Austrian Science Fund Fonds zur Förderung der wissenschaftlichen Forschung (FWF) P19132 and P23378, from EU FP7 project INSITE, as well as from the Österreichische Forschungsgemeinschaft (ÖFG) for covering travel expenses. I also thank the Santa Fe Institute for the opportunities and open environment offered during the Complex Systems Summer School 2010, where ideas to a part of this thesis originated.

# Outline

This doctoral thesis is a compilation of Refs. [203–205, 208], work already published or in review at the time of this writing, containing the research efforts over my three years stay at the Section for Science of Complex Systems, Medical University of Vienna. These works were performed in collaboration with Stefan Thurner (Medical University of Vienna), Renaud Lambiotte and Giovanni Petri (Imperial College London), Roberta Sinatra and Vito Latora (University of Catania).

The thesis is organized as follows. We start motivating the aims and goals of this work in Chapter 1, elaborating the need for a data-driven social science. Chapter 2 reviews the valuable contribution of statistical physics towards this issue, and gives a literature review with focus on complex systems and network theory. Here we also list the mathematical and statistical tools of network theory used subsequently. In Chapter 3 we expose the background and problems in the understanding of large-scale human behavior, and motivate the goal of establishing a socio-economic laboratory. Compiled primarily from Ref. [203], we present and describe in detail the experimental setting, i.e. the environment of a self-developed Massive Multiplayer Online Game together with its users, a society of human players consisting of over 350,000 individuals. Results are presented in subsequent Chapters, where we first focus on the “building blocks” of the society, the individuals and their actions, then on the network topology of their interactions, and finally on the interplay between the different interaction networks. Beginning with the findings of Ref. [208] in Chapter 4, we characterize the statistical nature of actions and reactions of individuals. Further, we give an analysis on the mobility of the individuals in their network-shaped virtual universe in Chapter 5, based on Ref. [204], measuring and modeling the subdiffusive nature of their movements. Chapter 6 contains results assembled from Ref. [205], where three different interaction and relation networks between the subjects are analyzed in detail, with both static and dynamic aspects. In Chapter 7 we present the final results, based on Ref. [203], where six different types of socio-economic networks are analyzed not only as single entities, but *in conjunction*. There, we study network–network interactions and the multi-relational organizational structure of the socio-economic system, with an application in social balance theory. We conclude the thesis and give an outlook in Chapter 8.

# Chapter 1

## Introduction: The need for a data-driven social science

Quantification of collective human behavior and social dynamics poses a unique, century old challenge. It is remarkable to some extent that mankind knows more about dynamics of subatomic particles than it knows about the dynamics of human groups. The reason for this situation is that the establishment of a fully experimental and quantifiable social science of group dynamics is tremendously complicated by two factors: First, unlike many problems in the natural sciences, dynamics of societies constitute a *complex system*, characterized by strong and long-range interactions, which are in general not treatable by traditional mathematical methods and physical concepts, see Section 2.2. Second, data is of comparably poor availability and quality [140, 216]. Evidently it is much harder to obtain data from social systems than from repeatable experiments on (non-complex) physical systems. Despite these severe problems, it is nevertheless paramount to arrive at a better understanding of collective human behavior. Only recently it became evident in the context of economics and finance, which costs are associated to misconceptions of human collective behavior. The by far most devastating catastrophes in the history of mankind are caused by social, man-made phenomena [111]: “Let us illustrate these risks by some numbers: World War I caused more than 15,000,000 victims, and World War II even 60,000,000 fatalities. The latter generated costs of 1,000 billion 1944 US\$ and destroyed 1710 cities, 70,000 villages, 31,850 industrial establishments, 40,000 miles of railroad, 40,000 hospitals, and 84,000 schools. Moreover, the world has seen many lossful wars ever since. The current financial and economic crises triggered an estimated loss of 4-20 trillion US\$. Climate change is expected to cause natural disasters, conflicts for water, food, land, migration, social and political instability. The related reduction of the world gross domestic product is expected to amount to 0.6 trillion US\$ per year or more.” Other collective social phenomena can have similar devastating effects on society, such as the amplification of epidemics spreading through human mobility [55]. Considering these examples, one could in fact say ‘The major risks are social’. If the dynamics behind collective behavior are going to remain as poorly understood as they are today, without the ability to generate

statements with predictive value, any attempts of rationally managing crises related to human groups in its various forms will remain illusionary.

A large social system is a *complex system*, featuring non-linear interactions between its constituents, the individuals making up the system, leading to emergent macroscopic properties. Complex systems have the fundamental property of strong interactions between its constituents, a feature which often allows the representation of the system as a network, where nodes represent individuals, and links are the one-to-one relations or interactions between them [44, 56, 83, 176]. These networks typically have a time-varying topology, i.e. the network structure is dynamically dependent on the particular configuration of node properties, or generalized “spins” [27, 31]. As already noted by Ettore Majorana 70 years ago [148], social systems (and social networks defined by the interactions and relations within these systems) can be treated well statistically, and may be well approached with methods and tools from statistical physics [16, 44]. In particular, the modern field of network science can be built around the concept of the (statistical) ensemble, see Section 2.3. A vast number of theoretical models have been introduced to explain large-scale social phenomena, including the derivation of network Hamiltonians [27, 32]. At the same time, 20th century sociology has produced a large number of observations and hypotheses, however usually only on a qualitative level and on small-scale groups of individuals, due to the tremendous difficulties in acquisition of sufficient and reliable data. Due to recent advances in information and communication technologies, the increased capabilities to store and process large quantities of data from such systems make it now possible to subject the proposed models and hypotheses to experimental tests, and to quantify behavioral processes in human socio-economic systems with extraordinary high statistical significance, increasing our understanding of collective social phenomena.

The present doctoral thesis provides an experimental validation of a set of hypotheses following approaches motivated by statistical physics, based on data extracted from a set of more than 350,000 human individuals living in a self-developed online world, ‘Pardus’ [2], forming a closed social system. Complete information on the properties and interactions of the participants of this online environment is available in the form of massive log-files and high frequency time-series. Individuals are engaged in a multitude of social relations and are pursuing basic economic activities to sustain a “living” in their virtual game-universe. This environment can be seen as a large-scale “socio-economic laboratory”, a gigantic agent-based model in which the agents are human beings [77, 94]. In particular the thesis follows the subsequent paths of research:

1. Test experimentally to what extent mean-field assumptions in previously formulated models [44] are justified. There is a growing number of indications that in strongly interacting systems, such as social systems, heterogeneous patterns of interactions seem to prevail [8, 44].
2. Calibration and quantification of microscopic interactions (couplings) of the constituents on different interaction networks. In particular, network analysis

---

of the (evolving) structure of interactions [205].

3. Multiplex analysis of interactions between different networks, quantification of the organizational structure of the society [203].
4. Study and modeling of diffusion processes occurring in the system [204].
5. Scaling analysis of action–reaction time-series with fluctuation analysis and anomalous diffusion techniques [208], as in Ref. [201].

For the first time the available detailed measurements of a large-scale social system allow to test existing models with the precision of physical experiments, and to improve these models if needed. For a historical placement of this new, unprecedented way of analyzing social systems, Table 1.1 gives an overview of the capabilities of social network data sets in the past and present. In this table, “Era I” corresponds to the 20th century pre-information age, where data gathering of social networks was conducted in an “analog” way only by e.g. interviews or questionnaires [162, 215]. “Era II” corresponds roughly to the first decade of the 21st century, in which, additionally to previous methods, a number of specific large-scale data sets have become available to the scientific community due to the capabilities of data storage and collection, from means of digital communication technologies such as the internet or mobile phones [140, 216]. “Era III” marks the time thereafter, in which the development of virtual, large-scale “socio-economic laboratories” such as Pardus becomes feasible, opening up a novel way for conducting social science [13, 91, 188, 189, 203, 205]. Data sets from Era II typically do not feature multiple relation types. By this we mean for example friendship or family ties, enmity, trade, important for understanding the fundamentally multi-layered functioning of social systems, see Chapter 7. While the number of subjects has reached the maximum possible scale in Era II, i.e. planetary orders of magnitude [143], due to lack of rich link data the full complexity of a society still cannot be studied there. Given that the legal and ethical conditions are met, socio-economic laboratories also allow for the possibility of experiments. If these conditions are too restrictive, at the very least societies can be studied in multiple environments with different sets of rules or constraints. Online games often naturally feature this possibility when the game environment has to be split up into independent game-universes or so-called *shards* due to technical or performance reasons [45]. This allows the straightforward implementation of parallel worlds, in which slightly different sets of rules may exist, perfectly controlled by the experimenter/game administrator. With such parallel universes, multiple “histories” of human societies can be tracked and compared simultaneously, allowing to fine-tune or test for robustness single parameters, for example in the mechanisms of the financial system.

Table 1.1: Sketch of a tabular history of the typically available capabilities of data sets in the analysis of social systems, in particular in social network analysis. “Era I”, “Era II”, and “Era III” mark the 20th century, the first decade of the 21st century, and the subsequent time, respectively, in which data sets from social systems provide different capabilities for researchers. “Rich node data” stands for the availability of various data on the subjects, such as sex, age, ZIP code, etc. “Rich link data” refers to the availability of multiple relation types, such as friendship or family ties, enmity, etc., important for understanding the fundamentally multi-layered aspect of social systems, see Chapter 7. “High frequency” concerns the availability of longitudinal data, usually time-stamped to the second in information-age data sets. “Societal scope” stands for the possibility to gather an all-encompassing view of a human society, requiring both rich node and link data, and a large number of subjects.

	Era I	Era II	Era III
Typical number of subjects $N$	$10^1 - 10^2$	$10^3 - 10^9$	$10^3 - 10^9$
Rich node data	yes	yes	yes
Rich link data	yes	no	yes
High frequency	no	yes	yes
Societal scope	no	no	yes
Avoidance of interviewer effects	no	yes	yes
Experiments	yes	no	yes
Setup costs (per individual)	very high	practically zero, but data has to be available	low

# **Chapter 2**

## **Statistical physics frameworks for social systems**

Statistical physics is the branch of physics which accounts for the understanding of systems composed of many elements. At its base is the idea that a macroscopic system made up of a large number of objects, e.g. atoms, is feasible to be understood only from a statistical point of view. Analyzing the properties of each of its constituents would practically not be possible. This approach allows to understand the macroscopic properties that characterize the system and can be actually measured in experiments. A classic example is gases: it is not possible to study the individual motions of  $10^{23}$  particles, but the state of one mole of gas can be fully characterized by observing its pressure, temperature and the volume occupied, since the universal law behind its behavior, i.e. the ideal gas law, is well-grounded and derivable from statistical physics.

Given its very general framework and its applicability outside the realm of traditional physics, in the recent years researchers have successfully explored a variety of fields with statistical physics methodology, such as in biology, computer science, medicine, or in the social sciences. The main goal of a statistical physics approach of social dynamics is to understand regularities on large scales as collective effects of the local interactions between single individuals, considered as relatively simple entities [44].

In the following sections of this Chapter, we review a non-exhaustive compilation of the statistical physics frameworks and paradigms used for describing social systems and for modeling collective social phenomena. We introduce the concept of complex systems in Section 2.2, and focus in particular on complex network theory, one of the core themes of this thesis, in Section 2.3.

### **2.1 Human beings as particles?**

In social systems, the constituents are not particles but humans. Human beings are obviously more complex “entities” than atomic particles and feature complicated

individual patterns of behavior, being the result of a multitude of physiological or psychological processes<sup>1</sup>. The traditional approach in the social sciences has tended to view the psychological complexities of human individuals as necessary to consider when trying to understand human behavior in any form [16]. However, formulating a deterministic scientific framework to exactly describe the processes behind *individual* behavior or interactions amounts to a close to hopeless task – it is not possible to model social agents without significant simplifications.

Despite these complications on the *individual* level, efforts to model human society using the methods of statistical physics by means of e.g. agent based models have in the recent years provided ample reason to assume that a number of collective behavioral phenomena can be understood on the basis of strongly simplified interaction rules [16].

### 2.1.1 From disorder to order

Many decisions of humans are the outcomes of simple imitation or learning processes, in which individuals influence each other, and as a consequence come dynamically to common agreements . If all my friends favor the same political faction, it is likely that I will also have a tendency to favor it because I am influenced by the opinions of my friends. Different variations of opinion dynamics processes have been incorporated by various models: i) the voter model [183], in which only two neighbors influence each other in a time step, ii) the majority rule [89, 132, 135], where each member of the group adopts the state of the majority of its neighbors, or iii) the Axelrod model of cultural dynamics [12], in which each individual has a set of beliefs and can become “more similar” to others through repeated interactions. The emergence of agreements, or more generally of ordered social configurations, can result from subtle effects in the local influences between the individuals. In the famous Schelling model of urban segregation [195], the slightest bias of individuals to prefer a neighborhood where the majority of the immediate neighbors are of the same “race” can create global patterns of racial segregation.

Considering that the transitions between disordered and ordered configurations has traditionally been of high interest in statistical physics [118], it is not surprising that quantitative physical models have been applied increasingly to social dynamics phenomena. The Ising model is a paradigmatic example in which such transitions occur. In a system of  $N$  spins (agents)  $s_i$ , having the values  $+1$  or  $-1$ , each spin is energetically pushed to align with the spins of its nearest neighbors. The total energy is given by

$$H = -\frac{1}{2} \sum_{(i,j)} s_i s_j, \quad (2.1)$$

with the sum running over the pairs of nearest neighbor spins [44]. The average magnetization  $m = \frac{1}{N} \sum_i \langle s_i \rangle$  starts taking a non-zero value as soon as the tem-

---

<sup>1</sup>For example, human behavior includes adaptive expectation, which has motivated the famous *Lucas critique* on economic policies [147].

perature  $T$  of the system decreases below a critical temperature  $T_c$  (angle brackets denote averages over different realizations). Besides its relevance for traditional physical phenomena, such as for ferromagnets, the Ising model is well suited for modeling phenomena in social dynamics and has been applied in this context in a large number of studies [44]. Analogous phenomena of phase transitions can take place in social systems, where ordered domains of small size grow through a coarsening process, with their global statistical features remaining unchanged over time. The Hamiltonian in equation (2.1) can be extended to incorporate nontrivial topologies, for example network Hamiltonians [31], to account for more complex and realistic interaction structures between elements. We study the related phenomenon of social balance in Chapter 7, in which particular configurations of local building blocks (triads constituted by positive or negative relations) are expected and found to be underrepresented in the social system due to reasons which are analogous to energy minimization processes.

Adding to the theoretical work, empirical evidence has revealed the existence of astonishing global regularities in social systems [40] becoming visible when studied *statistically*, on a large scale. Although human individuals typically feature strong interactions with only a limited number of “interaction partners” (and no interactions with the vast majority of all the others), system-wide transitions between disorder and order have been observed, such as the emergence of consensus about a specific issue [44], urban segregation [195], or transitions in traffic or crowd dynamics [112]. Statistical physics makes it possible to understand the properties of these collective social phenomena because in most situations they do not depend on the exact microscopic details of the processes involved [16, 44]. Rather, for many questions it is enough to consider only the most important features of single individuals in models, and sometimes only higher level features such as symmetries, dimensionality, or conservation laws play a relevant role for the global behavior. However, to generate quantitative statements, and to relate the statistical laws to the microscopic properties of the system, these models need to be calibrated with empirical data measured from real social systems, which is usually a very difficult task, see Section 3.1, and constitutes one of the aims of this thesis.

### 2.1.2 Mean-field approximation and the role of topology

The most basic assumption on the microscopic interactions between the entities of a system is to suppose that all of them exert an influence on all others. This is true in many physical systems, such as electromagnetic or gravitational systems. Such a simple approximation can also be applied to a human society, where in a *well-mixed population* each individual has the same likelihood of interacting with any other individual. This assumption is common in areas such as evolutionary game theory which deals with competing strategies in a population, and which allows to significantly simplify adaptation and selection processes to a system of first order differential equations. In this approach, the state of a population is given by the vector  $x = (x_1, x_2, \dots, x_n)$  on the  $n$ -simplex  $S_n$ , where  $x_i$  is the relative frequency

of strategy  $i$ , i.e.  $\sum_i x_i = 1$  and  $0 \leq x_i \leq 1$  for all  $i$ . Every strategy has a *fitness* value, contained in a payoff matrix  $A$ , which determines its success against other strategies. The evolution of the strategies in the system is then tractable in a special input-output equation, the replicator equation

$$\dot{x}_i = x_i [(Ax)_i - x \cdot Ax], \quad (2.2)$$

where a strategy  $i$  gains  $(Ax)_i$  but loses the average payoff in the population of  $x \cdot Ax = \sum_i x_i (Ax)_i$  [197]. The replicator equation is a special case of a master equation, a concept which has also been used in the so-called *sociodynamics* approach to model social systems [110]. In all these models, the results are independent of the spatial configuration, as the “mean-field approximation” assumes that the dynamics can be determined by studying interactions between “typical” interaction partners, a notion which is equivalent to the concept of representative agent in economics [113]. Although the bulk of the literature on game-theoretic models focuses on well-mixed populations, there is an increasing number of works studying these interactions on complex topologies [202]. However, contrary to established opinions in the field, systematic simulations have shown that the underlying spatial structure and the particular update rules used can have a strong influence on the outcome of these games [186].

Trading off for a certain amount of mathematical simplicity, a more realistic approach to handle social systems is given by the assumption that individuals are not connected to *all* others, but only to a finite and constant fraction of the whole population. In practice this assumption implies a regular topology such as a lattice, where each individual locally interacts with a fixed number of neighbors. Regular lattices are the basis for the Ising model, for many agent-based models, and for most statistical physics models of opinion dynamics. Also in these models there is a recent tendency for considering more realistic topologies [44].

The recent shift of focus from the study of dynamics in well-mixed populations or on regular topologies to irregular topologies, is caused by more and more empirical evidence that the structure of social and economic relations between humans is typically of a *heterogeneous* form [44, 140]. This fact allows to represent the structure of human interactions as a complex network, see Section 2.3. Moreover, human beings do not stay their entire life in the same position but can be highly mobile, see Chapter 5. Also, their social and economic surroundings can change over time, so that the underlying interaction networks change dynamically, see Section 2.3. These additional, dynamic degrees of freedom often lead to the dynamic emergence of measurable macroscopic phenomena which are the focus of study in the framework of complex systems.

## 2.2 Complex Systems

The exploding number of publications in the past decade on diverse natural phenomena using a statistical physics approach has increased our understanding of systems of different nature, such as of complex social systems [16, 44]. Complex systems are commonly understood as systems composed of a large number of elementary units which as a whole exhibit properties not obvious from the properties of the individual parts. The microscopic interactions in the system lead to the *emergence* of macroscopic properties. Typically strong interactions between the elementary units of a complex system allow to represent them as networks, a core theme of this thesis which we cover in Section 2.3.

Complex systems are abundant in nature, including human and animal societies. Typical examples include [34]: ant or termite colonies, in which tasks are allocated without central control, flocks of birds or fish swarms moving in coordination without a leader, vehicle traffic on highways or in cities, pedestrian crowds, stock markets, and in general any complete social or economic systems constituted by humans.

The following subsections contain some common properties related to complex systems which are related to the core of the work.

### 2.2.1 Self-organization and power laws

The concept of self-organization and of self-organized criticality (SOC) was first described in the so-called *sandpile model* [14]. In this classic model, the dynamical response of an idealized pile of sand to small random perturbation is studied. The perturbations lead to avalanches with scale-free distributions in space and time (“ $1/f$  noise”). This phenomenon is not an exceptional case but may be the underlying concept for scaling in a wide class of systems with extended degrees of freedom. Power law distributions,  $P(k) \sim k^{-\gamma}$ , have in the recent years been found to be abundant in a wide range of complex systems, including social, economic, or ecological systems [168]. In some social systems, this phenomenon is called *Zipf’s law* after the linguist George Kingsley Zipf, who found scale-free distributions in word frequencies of natural languages [226]. We report similar findings in Chapter 4, where we study “words” in the action sequences of individuals. The distribution of city sizes is another classical example termed after Zipf. In economic systems, power laws in income or wealth distributions are known as *Pareto distributions* after the economist Vilfredo Pareto, who observed this phenomenon over a century ago [175]. Power law distributions also imply that extreme events such as stock market crashes occur much more frequently than expected from normal distributions [109]. Throughout this thesis we find several examples for power laws in the social system studied. Reference [53] gives a good overview of the statistical methods for making accurate parameter estimates for power-law data, based on maximum likelihood methods and the Kolmogorov-Smirnov statistic. More important than exactly characterizing the empirical distributions in social systems is however the development of new methods for validating the models that have been proposed to explain those power laws.

## 2.2.2 Breakdown of classical entropy

Non-interacting and weakly interacting statistical systems in which the number of particles is large can be perfectly described by thermodynamics. Complex systems on the other hand are typically characterized by strong and long-range interactions, which can fundamentally change their macroscopic properties as a function of system size or its degrees of freedom [102]. Here, a classical thermodynamic description is usually not possible and traditional concepts of entropy, i.e. Boltzmann-Gibbs-Shannon entropy, cannot be applied. Because of this, there is presently ongoing research on non-extensive entropies suited for adequately characterizing such strongly interacting statistical systems. For example, when the axiom of extensivity was dropped in a recent, axiomatic approach [102, 103], a multitude of entropies, potentially suited for long-range (non-ergodic) systems, could be brought into a general framework, with some of them representing possible candidates for application to complex systems. Measuring the Shannon  $n$ -tuple redundancy in behavioral sequences in Chapter 4, we find evidence indicating that Boltzmann-Gibbs-Shannon entropy might not be an extensive quantity in this case, indicating strong structure in behavioral time series.

## 2.2.3 Anomalous diffusion

Diffusion processes in complex systems often do not follow Gaussian statistics [155]. Physical systems in which non-trivial transport phenomena play a role include turbulent flows, phase-space motion in chaotic dynamics, or transport in highly heterogeneous media such as porous materials or gels. Usually these anomalies are related to the presence of non-homogeneous topological features such as a fractal structure [225]. In complex systems, instead of Brownian motion, i.e. a linear time dependence of the mean square displacement,  $\sigma^2(t) \sim t$ , often anomalous diffusion is found, which typically features a non-linear, power law growth of the mean square displacement,

$$\sigma^2(t) \sim t^\nu, \nu \neq 1. \quad (2.3)$$

This functional form is strongly connected to the breakdown of the classical central limit theorem due to heavy-tail distributions or long-range correlations. In this case the systems often feature non-Markovian time evolution, as a typical manifestation of non-local temporal phenomena [155]. In social systems, anomalous diffusion has been observed for example in human mobility [192, 199, 214, 218], which can be caused by the use of different means of transportation [15, 101] or by heterogeneous human activity patterns, leading to broad spatial jump or waiting time distributions. We report a similar result on the mobility of human-controlled avatars in an online game in Chapter 5, where we measure long-range temporal correlations in the movement patterns of players in the game. We show that taking into account such correlation phenomena is essential for modeling the observed statistical features of movement.

## 2.3 Network theory

“Networks are all around us, and we are ourselves, as individuals, the units of a network of social relationships of different kinds and, as biological systems, the delicate result of a network of biochemical reactions. Networks can be tangible objects in the Euclidean space, such as electric power grids, the Internet, highways or subway systems, and neural networks. Or they can be entities defined in an abstract space, such as networks of acquaintances or collaborations between individuals.” [33]

Accompanying the recent availability of massive data sets on human activities in social systems, methodologies to handle these data are being improved continuously, with the field of network science leading the way. A network viewpoint emphasizes that the behavior of a complex system is shaped by the strong interactions among its constituents and offers the possibility to analyze systems of very different nature within a unified, mathematically well-tractable framework. Social network analysis, the field of network analysis which focuses on relationships among social entities and which started to develop in the early 1920’s [215], has experienced a revival with several new concepts being introduced or refined, such as scale-free networks [18]. Methods and algorithms are being constantly advanced to allow dealing with further challenges, e.g. the dynamical, longitudinal (time-varying) facet of the torrent of data we are now confronted with [33].

Network theory has its origins in graph theory, going back to Euler and the famous “Seven bridges of Königsberg” problem which he resolved in 1735. In its modern form graph theory has been developed by mathematicians since the 1950’s, most famously as one of the first by the Hungarian mathematicians Erdős and Rényi. Graph theory concerns the combinatorial or probabilistic study of the abstract mathematical objects of graphs (for a formal definition see Section 2.3.5), often as single entities. Modern network theory however is based on the study of graphs as realizations of a very large number of graphs sharing the same macroscopic properties, i.e. a *network ensemble*. One of the most important macroscopic properties in this context is the degree distribution, or the exponent of the underlying power law, see the following subsections. We make use of this paradigm for example in Chapter 6, where we approach the empirically measured networks as realizations of the ensembles of equally sized networks with the same degree distributions to calculate triad significance profiles, or in Chapter 7 where the network ensemble is made up of networks with identical topology but randomly reshuffled signs. Network theory in today’s form was introduced primarily by statistical physicists at the end of the 1990’s [19, 20, 217] and has since then significantly contributed to the views in which researchers approach and think about complex systems [18, 216].

### 2.3.1 Random networks

Graph theory in its modern form started with Erdős and Rényi studying probabilistically the properties of graphs as a function of the number of their random connections [33]. They proposed a model in which graphs with  $N$  nodes and  $L$  links

are generated randomly [78], known as Erdős-Rényi (ER) graphs or ER-networks. In this model, (undirected) links are added randomly to a set of isolated  $N$  nodes, until the network has reached  $L$  links (multiple and self-links are prohibited). One ER-network is the result of a (usually huge) number of realizations, all of which form an ensemble of ER-networks with the same macroscopic properties. This model has a strong analogy to the canonical ensemble [33]. In another classic model, ER-networks are created by adding links to all pairs of nodes with a probability  $0 < p < 1$ . This process defines again an ensemble of ER-networks, in which however the probability to find a network with exactly  $L$  links is not 1, but  $p^L(1 - p)^{N(N-1)/2-L}$  [36]. This second model has a strong analogy to the grand-canonical ensemble, and in the limit  $N \rightarrow \infty$ , where the average degree  $\bar{k}$  is fixed, both models coincide [33]. It is noteworthy that within the network framework, phase transitions occur at certain critical values. For example, a second order phase transition occurs at a critical probability  $p_c = 1/N$ , corresponding to the critical average degree  $\bar{k}_c = 1$ . If  $p$  lies below  $p_c$ , the network will exhibit no giant component (for  $N \rightarrow \infty$ ), and for  $p > p_c$  the giant component emerges with probability 1 (for a definition of giant component see Section 2.3.5).

The distribution of the degrees of all nodes in a network,  $P(k)$ , is a fundamental property which has been under particular study since it makes it possible to identify underlying and universal principles common in a variety of different networks. In a random network, the degree distribution is a Poissonian for large  $N$  and fixed  $\bar{k}$ ,

$$P(k) \sim e^{-\bar{k}} \frac{\bar{k}^k}{k!}. \quad (2.4)$$

Note that this distribution is very narrowly focused around the average degree  $\bar{k}$  – almost all nodes in the network have a degree of  $\bar{k}$  or a value very close to  $\bar{k}$ . Due to lack of data, for a very long time it was assumed that most real networks display this highly homogeneous degree distribution. However, when more and more measurements on large networks became available in the late 1990's, it turned out that most real networks have fat-tailed degree distributions, many of them showing approximate scale-free degree distributions.

### 2.3.2 Scale-free networks

In contrast to previous expectations, most networks found in the real world do not display Poissonian degree distributions such as ER-networks, but heavy-tailed, often scale-free distributions characterized by power laws:

$$P(k) \sim k^{-\gamma}, \quad (2.5)$$

where the exponent  $\gamma$  often varies between 2 and 3. Networks with this degree distribution have a diverging second moment  $\sigma^2 = \overline{k^2} - \bar{k}^2$  and are therefore not straightforward to treat with traditional analytical methods. These networks are called scale-free because power laws have the same functional form at all scales [168].

A heavy-tailed degree distribution emphasizes that interactions between the component of a system are far from homogeneous as in regular lattices or ER-networks, and that nodes are topologically not equivalent. The presence of *hubs*, i.e. of nodes with a very high number of connections, is typical here. Besides the degree distribution, further evidence for this heterogeneity and for the nontrivial connection patterns can be found when considering the average nearest neighbour degree  $k^{\text{nn}}$  or the clustering coefficient  $C$  as a function of degree  $k$  (for formal definitions see Section 2.3.5). In the presence of homogeneous, therefore uncorrelated interaction patterns, both these functions  $k^{\text{nn}}(k)$  and  $C(k)$  are constant. However, as has been shown in many real networks, and as we also present evidence in the case of social networks in Chapter 6, these functions are usually not constant. An increasing (decreasing) function  $k^{\text{nn}}(k)$  implies a tendency for high-degree nodes to be connected to other high-degree nodes (assortativity) or to be not connected to high-degree nodes (disassortativity) [165]. Networks with non-trivial topological features as described above are called complex networks.

In the past decade a large number of real networks has been uncovered to exhibit scale-free degree distributions and heterogeneous connection patterns, ranging from the autonomous system spanning the internet and the world wide web to networks defined by scientific citations [53], or biological, metabolic networks or networks of actors who have played in the same films [33]. Some networks from a variety of fields have in common similar power law exponents or growth processes, giving hope to find *universalities*, i.e. common features underlying a multitude of different network types in systems of different nature.

### 2.3.3 Models of growing networks

With the discovery that most networks are heterogeneous and have fat-tailed, often scale-free, degree distributions, the need arose to find models explaining the growth of the networks which develop these properties. One of the first was the Barabási-Albert (BA) model, which as an essential ingredient incorporates the mechanism of *preferential attachment* [19]. The idea behind preferential attachment is that nodes with higher degrees – the more “popular” nodes – tend to attract new links with a higher rate than “unpopular”, low-degree nodes. In particular, a BA network is constructed by beginning with a small number  $m_0$  of isolated nodes, and then adding nodes one by one having  $m(\leq m_0)$  links that connect them to  $m$  nodes already present in the system. The probability  $\Pi$  that a new node is connected to an existing node  $i$  depends linearly on the degree of that node,  $k_i$ , namely  $\Pi(k_i) = k_i / \sum_j k_j$ . As can be shown analytically, a BA network constructed as above reaches a scale-invariant state, with a degree distribution following a power law with exponent 3. It is further possible to modify the model to produce networks with different exponents [69].

Other network evolution models producing fat-tailed degree distributions include, for example, an attachment mechanism with local walks [213], or a number of generalizations of the original BA model [33]. Most of these models have in common growth only, i.e. the sole addition of links, but no deletions. As we however show

in Chapter 7, there can exist a high churn of links in real networks, defined as the number of removed links divided by the number of newly added links, necessitating the implementation of new models accounting for this observation.

### 2.3.4 Rewiring and models of non-growing networks

Network evolution models in which the number of links remains constant are called rewiring models, which in some cases can be solved exactly [79]. One classic example of a rewiring model is the ‘small-world’ network model [217]. In this model, a parameter  $\varepsilon$  regulates the ‘randomness’ of the network: If  $\varepsilon = 0$ , the network is regular, if  $\varepsilon = 1$ , the network is an ER-network. For special intermediate values of  $\varepsilon$ , the model is able to produce networks with low average path length  $\bar{g}$  (the so-called *small-world* phenomenon [156]) and with relatively high clustering coefficient  $C$  (for definitions see Section 2.3.5). These two properties are typical for many empirical networks, such as networks of film actors or biological networks [217].

More recent non-growing network models include a “gas-like” mechanism [209]. In this constant-size network process, “inelastic” interactions occur between pairs of nodes, realized by mergers of pairs of nodes into new single nodes. Merged nodes then obtain the union of all links of the previously separate nodes. With each merger, a chemostat condition leads to a newly created node, randomly linked to the network. This network model can be shown to be a generalization of the model developed by [131], which introduced merging and regeneration processes of nodes earlier. Another model of this type generalizes network dynamics to include both rewiring and growth modes, which depend on the internal topology as well as on a metric imposed by the space they are embedded in [207]. This model associates a fixed energy cost to every link and leads to a degree distribution which can be seen as an energy distribution of the type emerging within non-extensive statistical mechanics.

### 2.3.5 Notation and statistical tools

#### Graph

In mathematical terms, networks are described as *graphs* [68, 215]. An undirected graph  $\mathcal{G} = (\mathcal{N}, \mathcal{L})$  is defined as a pair of sets, the node set  $\mathcal{N}$  containing all nodes  $i$  and the link set  $\mathcal{L}$  containing unordered pairs  $l_{ij} := \{i, j\}$  denoting those nodes which are connected by an undirected link (edge). A directed graph (*digraph*) has a link set  $\mathcal{L}$  which contains *ordered* pairs  $l_{ij} := (i, j)$  marking nodes which are connected by a directed link (arc) going from  $i$  to  $j$ . The expressions  $N, L$  denote cardinalities of the respective sets. A graph is called *complete* if connections between all pairs of nodes exist.

#### Symmetrization

The *symmetrization* or *reflexive closure* of a digraph  $\mathcal{G} = (\mathcal{N}, \mathcal{L})$  is constructed as follows: Start with  $\mathcal{G}^* = (\mathcal{N}, \mathcal{L}^*)$ , where  $\mathcal{L}^*$  is an empty link set, and for all pairs

of nodes  $i$  and  $j$  add the undirected link  $l_{ij}$  to  $\mathcal{L}^*$  if the directed link  $l_{ij} \in \mathcal{L}$  or if  $l_{ji} \in \mathcal{L}$ .

### Weighted graph

In unweighted graphs all links are treated equally. A *weighted graph* is a generalization in which the weight  $w_{ij}$  of a link  $l_{ij}$  may take any non-zero real value, usually  $\in \mathbb{R}$ .

### Dyad

A dyad is a (sub)graph consisting of two nodes. A directed dyad can be a *null dyad* (no links), *asymmetric* (one link, going in one direction), or *mutual* (two links, one going in one direction and the other going in the opposite one).

### Signed graph

A *signed digraph* is a pair  $(\mathcal{G}, \sigma)$ , where  $\mathcal{G} = (\mathcal{N}, \mathcal{L})$  is a digraph and  $\sigma: \mathcal{L} \rightarrow \{-1, +1\}$  is a sign function assigning each directed link a binary value, e.g. in the context of social networks denoting positive or negative relationship [65]. We write  $s_{ij}$  short for  $\sigma(l_{ij})$ , and set  $s_{ij} := 0$  when  $l_{ij}$  does not exist.

Every signed digraph has a valency matrix  $V$  with entries  $v_{ij}$  defined as [105]:

$$\begin{aligned} v_{ij} = o & \quad \text{if } s_{ij} = s_{ji} = 0 \\ v_{ij} = p & \quad \text{if } s_{ij} + s_{ji} > 0 \\ v_{ij} = n & \quad \text{if } s_{ij} + s_{ji} < 0 \\ v_{ij} = a & \quad \text{otherwise.} \end{aligned} \tag{2.6}$$

These entries correspond to null ( $o$ ) dyads, to dyads with only positive ties ( $p$ ), to dyads with only negative ties ( $n$ ), and to dyads with one positive and one negative tie ( $a$  for ambivalent relationship), respectively.

### Degree

In an undirected graph the *degree*  $k_i$  of a node  $i$  is the number of links connecting to it. All  $k_i$  nodes which are directly linked to  $i$  are called (nearest) *neighbors* of  $i$ . A node with degree 0 has no neighbors and is called *isolated*. We denote the average degree of all nodes in a network by  $\bar{k}$ . In a directed graph the in-degree  $k_i^{\text{in}}$  of a node  $i$  is the number of its incoming links, the out-degree  $k_i^{\text{out}}$  the number of its outgoing links. We denote the average degree of all nearest neighbors of a node  $i$  by  $k_i^{\text{nn}}$ . We denote the average degree of all nearest neighbors of all nodes as a function of degree  $k$  by  $k^{\text{nn}}(k)$ .

The measure  $\rho(k^{\text{in}}, k^{\text{out}})$ , where  $\rho$  is the Pearson correlation coefficient, evaluates the correlations between in-degree and out-degree around the same node,

giving a measure of the deviation of a directed network from a Eulerian network, i.e.  $\rho(k^{\text{in}}, k^{\text{out}}) = 1$  only for a Eulerian graph, namely  $k_i^{\text{in}} = k_i^{\text{out}}$  for each node  $i$ .

### Geodesic and diameter

In an undirected graph, the *geodesic* or *shortest path*  $g_{ij}$  of two nodes  $i$  and  $j$  is the smallest number of links one needs to get from  $i$  to  $j$ . If a graph is disconnected, i.e. there exist at least two non-empty sets of non-connected nodes (called *components*), geodesics between all nodes of different components are set to  $\infty$ . We denote the average geodesic, or characteristic path length, of a graph by  $\bar{g}$ . The average geodesic of a random graph is  $\bar{g}_r \approx \ln N / \ln \bar{k}$  [68]. The *diameter*  $g_{\max}$  of a network is its largest geodesic.

### Clustering coefficient

The *clustering coefficient*  $C_i$  of node  $i$  in an undirected graph is the ratio between the number  $y_i$  of links between its  $k_i$  neighbors and the number of all possible links  $k_i(k_i - 1)/2$  between them,

$$C_i := \frac{2y_i}{k_i(k_i - 1)}. \quad (2.7)$$

The network's clustering coefficient  $C$  is the average over all clustering coefficients,  $C = (1/N) \sum_i C_i$ . A random graph's clustering coefficient  $C_r$  is given by  $C_r = \bar{k}/N$  [68].

### Efficiency

Global efficiency of an unweighted network  $\mathcal{G}$  with  $N$  nodes is defined as

$$E_{\text{glob}}(\mathcal{G}) := \frac{2}{N(N-1)} \sum_{i \neq j \in \{1, \dots, N\}} g_{ij}^{-1}. \quad (2.8)$$

Global efficiency  $E_{\text{glob}}$  can be thought of as a measure how efficiently information is exchanged over a network, given that all nodes are communicating with all other nodes concurrently. Local efficiency  $E_{\text{loc}}$ , as a measure of a system's fault tolerance is defined as

$$E_{\text{loc}}(\mathcal{G}) := \frac{1}{N} \sum_{i \in \{1, \dots, N\}} E_{\text{glob}}(\mathcal{G}_i), \quad (2.9)$$

where  $\mathcal{G}_i$  is the graph of all neighbors of node  $i$  (not containing  $i$ ). Both values  $E_{\text{glob}}$  and  $E_{\text{loc}}$  are in the interval  $[0, 1]$ . Note that global efficiency is a reasonable approximation for the inverse geodesic in unweighted graphs; local efficiency is a good approximation for the clustering coefficient when most local networks  $\mathcal{G}_i$  are not sparse [139].

## Reciprocity

*Reciprocity* measures the tendency of individuals to reciprocate connections, i.e. the creation of mutual instead of asymmetric dyads [215]. Following [116], a naive reciprocity index can be defined by

$$R := \frac{L}{L^*} - 1, \quad (2.10)$$

where  $L^*$  is the number of undirected links in the reflexive closure of the digraph<sup>2</sup>. Values of  $R = 0$  and  $R = 1$  stand for no mutual dyads and mutual dyads only, respectively. Reciprocity may also be quantified by defining the fraction

$$r^* := \frac{L^{\leftrightarrow}}{L}, \quad (2.11)$$

where  $L^{\leftrightarrow} \equiv 2(L - L^*)$  counts the number of directed links in all mutual dyads of the digraph. Due to conceptual problems with  $r^*$ , [90] we use the following reciprocity index

$$\rho := \frac{r^* - \bar{a}}{1 - \bar{a}} \in [\rho_{\min}, 1], \quad (2.12)$$

with  $\bar{a} := \frac{L}{N(N-1)}$  measuring the ratio of observed to possible directed links, and  $\rho_{\min} := -\frac{\bar{a}}{1-\bar{a}} \in [-1, 0]$  for  $\bar{a} \leq 1/2$  (the expression  $\rho_{\min}$  makes sense for  $\bar{a} \leq 1/2$ . Otherwise it is not possible to have  $L^{\leftrightarrow} = 0$ ). The index  $\rho$  allows to distinguish between *reciprocal* ( $\rho > 0$ ), *areciprocal* ( $\rho = 0$ ) and *antireciprocral* ( $\rho < 0$ ) networks. Further,  $\rho$  enables a clear ordering of networks independent of link density which is not possible with  $r^*$  [90].

## Assortativity

Assortative mixing coefficients are the Pearson correlation coefficients of the degrees at either ends of a link [165]:

$$r = \frac{\overline{k_{\text{to}}k_{\text{from}}} - \overline{k_{\text{to}}}\overline{k_{\text{from}}}}{\sqrt{\overline{k_{\text{to}}^2} - \overline{k_{\text{to}}}^2} \sqrt{\overline{k_{\text{from}}^2} - \overline{k_{\text{from}}}^2}} \in [-1, 1]. \quad (2.13)$$

Bars denote averages,  $k_{\text{to}}$  and  $k_{\text{from}}$  index the (in-, out- or undirected) degrees of nodes at the beginning and end of links, respectively. Following [116] we measure assortativity  $r_{\text{undir}}$  in the reflexive closures as well as coefficients for all four combinations of in- and out degrees in the directed networks:  $r_{\text{inin}}$ ,  $r_{\text{inout}}$ ,  $r_{\text{outin}}$ ,  $r_{\text{outout}}$ . A positive degree-degree correlation coefficient indicates assortativity, i.e. the tendency of nodes with high (low) degrees connecting to nodes with high (low) degrees, a negative correlation means disassortativity, i.e. the tendency of nodes with high (low)

<sup>2</sup>The factor 2 of equation (1) in [116] is dropped since we identify pairs of directed links of mutual dyads with single undirected links in the construction of the reflexive closure.

degrees connecting to nodes with low (high) degrees. Assortativity (disassortativity) implies an increasing (decreasing) function  $k^{\text{nn}}(k)$ .

### Bridge

A *bridge* is a link which, when removed, increases the amount of disconnected components in the graph by one [215]. A link is a *local bridge* of degree  $a$  if its removal causes its endpoints  $i$  and  $j$  to have geodesic  $g_{ij} = a$  [97].

### Overlap (of neighboring nodes)

The overlap of two neighboring nodes measures the number of neighbors common to both of them. We adopt the definition used in [172],

$$O_{ij} := \frac{m_{ij}}{(k_i - 1) + (k_j - 1) - m_{ij}} \in [0, 1], \quad (2.14)$$

where  $m_{ij}$  is the number of neighbors common to both nodes  $i$  and  $j$ . A value of  $O_{ij} = 0$  (1) corresponds to an empty (identical) common neighborhood of nodes  $i$  and  $j$ . The Overlap  $O_{ij}$  is not to be confused with measures of network–network overlaps as defined later.

### Betweenness

Link betweenness centrality, short *link betweenness* or *load*, is defined for an undirected link  $l_{ij}$  by

$$b_{ij} := \sum_{n_e \in V} \sum_{n_f \in V \setminus \{n_e\}} \frac{\theta_{ef}(l_{ij})}{\theta_{ef}}, \quad (2.15)$$

where  $\theta_{ef}(l_{ij})$  is the number of geodesics between  $n_e$  and  $n_f$  that contain  $l_{ij}$ , and  $\theta_{ef}$  is the total amount of geodesics between  $n_e$  and  $n_f$  [172]. Betweenness can be viewed as a measure of “*traffic*” if e.g. all pairs of nodes exchange information at the same rate [68].

### Giant component and largest connected component

In graphs with infinitely many nodes one observes the emergence of a *giant component* when crossing a percolation threshold [68]. The emerging giant component is the only component holding infinitely many nodes. In finite graphs, we call the component having the highest number of nodes *largest connected component*. We denote the fraction of nodes being in the largest connected component by  $\Gamma$ .

### Triad

A *triad* is a (sub)graph consisting of three nodes. In a digraph there exist 16 isomorphism classes of triads [105]. We adopt the notation of [159] for the 13 connected

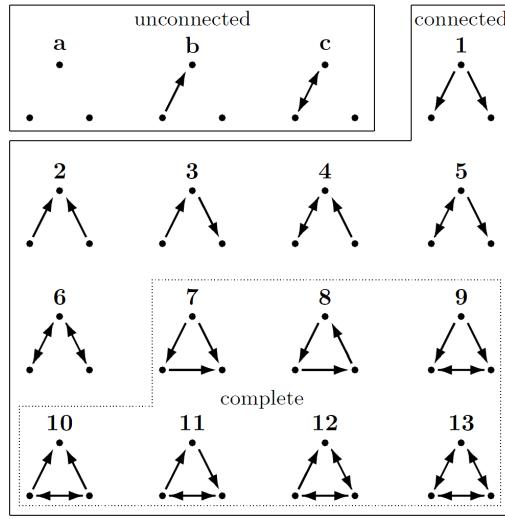


Figure 2.1: The 16 isomorphism classes of triads and their ids.

classes and label the unconnected ones by  $a$ ,  $b$  and  $c$ , see Fig. 2.1. Within the group of connected triad classes, seven are complete<sup>3</sup>.

### Triad significance profile

The *triad significance profile* (TSP) is the vector of statistical significances of each connected triad class compared to random networks drawn from the  $U(X_{*+}, X_{+*}, M^*)$  distribution, i.e. of random networks having identical in/out degrees and equally likely numbers of mutual dyads for each node [159, 185]. Statistical significance of a triad class  $i$  is measured by the  $Z$  score

$$Z_i = \frac{(N_i^{\text{real}} - \bar{N}_i^{\text{rand}})}{\text{std}(N_i^{\text{rand}})}, \quad (2.16)$$

where  $N_i^{\text{real}}$  is the frequency of occurrence of the triad class in the considered network, and  $\bar{N}_i^{\text{rand}}$  and  $\text{std}(N_i^{\text{rand}})$  are the average frequency of occurrence and the standard deviation in an ensemble of random networks drawn from  $U(X_{*+}, X_{+*}, M^*)$ . The TSP is the normalized vector of all 13  $Z$  scores,

$$\text{TSP}_i = \frac{Z_i}{\left(\sum_{i=1}^{13} Z_i^2\right)^{1/2}}. \quad (2.17)$$

Note that the TSP emphasizes the *relative* significances of triad classes, constituting an appropriate comparison parameter for networks of arbitrary sizes [159].

---

<sup>3</sup>We write connected short for weakly connected, i.e. every two nodes are joined by a semipath [105]. We write completeness short for completeness of the reflexive closure, i.e. a directed triad is complete if it contains no null dyads, it is incomplete otherwise.

## Clustering similarity

To quantify the relative goodness of the partitions obtained from community detection applied to various network-related matrices in Chapter 5, we make use of three measures of clustering similarity: the Fowlkes and Mallows index  $\mathcal{F}$  [85], the Rand's criterion  $\mathcal{R}$  [179] and the normalized information variation (NVI) [58], as defined below.

Consider a set of nodes  $\mathcal{T}$  of cardinality  $n$  and two partitions  $\mathcal{C}$  and  $\mathcal{C}'$  of  $\mathcal{T}$ , then the set of all un-ordered pairs of elements of  $\mathcal{T}$  is the union of the sets [38, 153]:

$t_{11}$  is the set of pairs the same community under  $\mathcal{C}$  and  $\mathcal{C}'$ ;

$t_{01}$  is the set of pairs not in the same community under  $\mathcal{C}$  but under the same community in  $\mathcal{C}'$ ;

$t_{10}$  is the set of pairs in the same community under  $\mathcal{C}$  but not under the same community in  $\mathcal{C}'$ ;

$t_{00}$  is the set of pairs not in the same community under  $\mathcal{C}$  and  $\mathcal{C}'$ ;

and  $n_{11}$ ,  $n_{01}$ ,  $n_{10}$ ,  $n_{00}$  are their respective cardinalities (and  $n_{11} + n_{01} + n_{10} + n_{00} = n(n - 1)/2$ ). The  $\mathcal{F}$  and  $\mathcal{R}$  indices are then given by:

$$\mathcal{F} = \frac{n_{11}}{\sqrt{(n_{11} + n_{10})(n_{11} + n_{01})}}, \quad \mathcal{R} = \frac{2(n_{11} + n_{00})}{n(n - 1)}, \quad (2.18)$$

which are essentially two ways of quantifying how well the partitions match pairs of nodes. Therefore a perfect match between two partitions will have  $\mathcal{F}, \mathcal{R} = 1$ . The *Variation of Information* (VI) is a measure based on information theoretical concepts and represents the informational distance between two partitions. Therefore, if the VI is large, the two partitions are very dissimilar. The VI of a partition is bounded by  $\log_2 n$ , hence it is possible to normalize it, obtaining the *Normalized Variation of Information* ( $\text{NVI} \in (0, 1)$ ):

$$\text{NVI}(\mathcal{C}, \mathcal{C}') = \frac{\mathcal{VI}(\mathcal{C}, \mathcal{C}')}{\log_2 n}, \quad (2.19)$$

where

$$\mathcal{VI}(\mathcal{C}, \mathcal{C}') = \mathcal{H}(\mathcal{C}) + \mathcal{H}(\mathcal{C}') - 2\mathcal{I}(\mathcal{C}, \mathcal{C}'). \quad (2.20)$$

The terms in equation (2.20) are the entropy  $\mathcal{H}(\mathcal{C})$  of partition  $\mathcal{C}$  and the mutual information between two partitions  $\mathcal{C}$  and  $\mathcal{C}'$  [38]:

$$\mathcal{H}(\mathcal{C}) = - \sum_{i=1}^k P(i) \log_2 P(i), \quad \mathcal{I}(\mathcal{C}, \mathcal{C}') = \sum_{i=1}^k \sum_{j=1}^l P(i, j) \log_2 \frac{P(i, j)}{P(i)P(j)}, \quad (2.21)$$

where  $P(i) = \frac{|C_i|}{n}$  is the probability that an element of  $\mathcal{T}$  chosen at random belongs

to community  $C_i \in \mathcal{C}$ , and  $P(i, j) = \frac{|C_i \cap C'_j|}{n}$  the probability that an element belongs to  $C_i \in \mathcal{C}$  and to  $C'_j \in \mathcal{C}'$ .

### Rank distribution

Constructing the rank of elements in a data set is the transformation in which numerical values are replaced by their rank when the data are sorted. In a network, nodes can be ranked for example by their degree, where the node with largest degree has rank 1, the second largest has rank 2, etc. Nodes with the same degree have the same rank; the difference to the subsequent rank is the number of nodes which shared the previous rank. For example if there are three nodes with degree 45 and rank 10, nodes with degree 44 have rank 13. We denote the rank of a quantity  $a$  by  $\text{rk}(a)$ . The rank distribution is then the probability distribution of ranks, ordered by rank.

### Multiplex network

A *multiplex network* is a set of networks, each spanned by a different relation, all defined on the same set of nodes. This superposition of networks is also called multi-relational, multi-modal or multivariate network in the literature [215]. We denote different network types and their measures with Greek indices. The standard way to represent a network is through its adjacency matrix  $A_{ij}$  which, for an unweighted, undirected network, is a symmetric matrix whose elements  $A_{ij}$  are equal to one if there is a link between  $i$  and  $j$ , and zero otherwise. To incorporate the existence of multiple relations (multiplex networks), it is common to define the tensor  $A_{ij;\alpha}$ , sometimes called super-sociomatrix [215]. This tensor has dimension  $N \times N \times R$ , where  $N$  is the total number of nodes and  $R$  is the number of different link types between the same set of agents. For a fixed value of  $\alpha$ ,  $A_{ij;\alpha}$  is the adjacency matrix of the network defined by link type  $\alpha$ . By construction, the properties of each network can be obtained from its adjacency matrix  $A_{ij;\alpha}$ . For instance, the degree  $k_{i;\alpha}$  of a node  $i$  is given by  $\sum_j A_{ij;\alpha}$ , the total number of links in network  $\alpha$  is  $L_\alpha = \sum_i k_{i;\alpha}/2$ . The number of paths of length  $n$  between nodes  $i$  and  $j$  is given by  $(A_\alpha^n)_{ij}$ . A whole new layer of complexity opens once the interplay between different sorts of networks is considered. From a mathematical point of view, multiplexity can be revealed by coupling different adjacency matrices.

For a combined analysis of a whole multiplex network we define an *envelope network* which is composed of the set of all links of *all* interaction types. In the envelope network, a link from  $i$  to  $j$  exists if it exists in at least one of the relational networks.

### Measures for multiplex networks

The statistical properties of different network types can be studied as separate entities using the following notations and generalized measures.  $N_\alpha$  is the number of nodes in the network type  $\alpha$ , and  $L_\alpha^{\text{dir(undir)}}$  is the number of (un)directed links.

Reciprocity is labeled by  $r_\alpha$ , and  $\rho(k_\alpha^{\text{in}}, k_\alpha^{\text{out}})$  is the correlation of in- and out-degrees within the  $\alpha$  network. Average degree, clustering coefficient, and clustering coefficient with respect to the corresponding random graph are marked by  $\bar{k}_\alpha$ ,  $C_\alpha$  and  $C_\alpha/C_\alpha^{\text{rand}}$ , respectively.

### Network–network overlap

The (link) overlap  $O_{\alpha\beta} = \frac{1}{2} \sum_{ij} A_{ij;\alpha} A_{ij;\beta}$  between the graphs  $\alpha$  and  $\beta$  counts the number of links they have in common. This overlap measures between different types of networks is not to be confused with the overlap  $O_{ij}$  between neighbours as defined earlier. Similarly, the multiplicity  $m_{ij} = \sum_\alpha A_{ij;\alpha}$  of a link between  $i$  and  $j$  counts the number of different links between these nodes. The Jaccard coefficient  $J_{\alpha\beta}$  between two different sets of links  $\alpha$  and  $\beta$  quantifies the interaction between the two networks by measuring the tendency that links simultaneously are present in both networks.  $J_{\alpha\beta}$  is a similarity score between two sets of elements and is defined as the size of the intersection of the sets divided by the size of their union [212],

$$J_{\alpha\beta} \equiv |\alpha \cap \beta| / |\alpha \cup \beta|. \quad (2.22)$$

Related similarity measures, such as the cosine similarity measure can be designed.

### Network–network correlators

For two random variables  $X$  and  $Y$  with mean values  $\bar{X}$  and  $\bar{Y}$ , and standard deviations  $\sigma_X$  and  $\sigma_Y$ , the correlation coefficient  $\rho(X, Y)$  is defined as

$$\rho(X, Y) = \frac{E[(X - \bar{X})(Y - \bar{Y})]}{\sigma_X \sigma_Y} \in [-1, 1]. \quad (2.23)$$

For example, the *reciprocity* coefficient  $r$ , as defined earlier, is the correlation coefficient between the transposed entries of the adjacency matrix of a directed graph,  $X \equiv A_{ij}$ ,  $Y \equiv A_{ji}$  [90]. We introduce two related coefficients to compare different network types  $\alpha$  and  $\beta$ . First, we consider  $\rho(k_\alpha, k_\beta)$ , to evaluate the correlations between the degrees of a node in the two different graphs  $\alpha$  and  $\beta$ , and second,  $\rho(\text{rk}(k_\alpha), \text{rk}(k_\beta))$ , which is calculated the same way as  $\rho(k_\alpha, k_\beta)$ , with the difference that not degrees but ranks of degrees are used.

# **Chapter 3**

## **Quantification of socio-economic behavior using an online society**

### **3.1 Establishing a socio-economic laboratory**

#### **3.1.1 Problems with data**

Usually the bottleneck in quantitative approaches to the analysis of social systems is the availability and quality of data. An increasing number of massive data sets – tracking behavioral “fingerprints” humans leave throughout their lives via mobile phones, online applications, etc. – has become available in the past decade, significantly advancing our scientific understanding of human behavior [140]. While in the 20th century measuring social relations was possible basically only via questionnaire-based methods, involving rather small numbers of individuals, in the order of 10–100, recent studies incorporate data sets of human social systems on up to planetary scales [143]. These large data volumes allow to treat socio-economic behavior of humans by means of statistical physics [44]. A major goal of this data-driven approach to social sciences is to understand the underlying mechanisms of the microscopic interactions between large numbers of humans and their connections in order to understand emerging macroscopic features. An aim is to eventually predict extreme events, and to uncover universal patterns which may emerge in systems of different nature.

While this new data-driven approach to the social sciences has improved our understanding of a number of *specific* aspects of societal phenomena, such as spread of epidemics [176], an *all-embracing* picture of social systems is far from being available. Especially since human society is a complex system, see Section 2.2, full information of its *surroundings*, contexts or boundaries, together with the interactions between these boundaries and the system itself is needed to understand the system properly [216].

Despite these extreme difficulties in data requirements, the analysis of *small-scale* social systems has a tradition in the social sciences [76, 152, 174, 215]. Concerning large-scale studies, recently there have been significant achievements in understanding a number of massive social networks on a quantitative basis, such as the cell

phone communication network [138, 172, 172], features of the world-trade network [115], email networks [169], the network of financial debt [37] and the network of financial flows [136]. The integration of various dynamical networks of an entire society has so-far been beyond the scope of any realistic data source. However, with the increasing availability of vast amounts of electronic fingerprints people leave throughout their lives, this situation is about to change. Online sources are capturing more and more aspects of life, boosting our understanding of collective human behavior [140, 145]. Regarding data acquisition it is essential not only to record decisions of individual humans but also the simultaneous state of their surroundings. Further, in any data-driven science the observed system should not be significantly perturbed through the act of measurement. In social science experiments subjects usually are fully aware of being observed – a fact that might strongly influence their behavior. Finally, data acquisition in the social sciences becomes especially tiresome on group levels, see e.g. the tremendous efforts which have been undertaken in a classic experiment by Newcomb et al. [162], in which a group of 17 students was congregated, observed, and questioned over several weeks. Traditional methods of social science such as interviews and questionnaires do not only require a large amount of time and resources to deliver statistically meaningful assertions, but may introduce well-known biases [41]. To many it might seem clear that social sciences can not overcome these problems, and that therefore social sciences would always remain on a lower quantitative and predictive level than the natural sciences.

#### 3.1.2 The research potential of Massive Multiplayer Online Games

The issue of taking simultaneous, unobtrusive measurements on a large number of subjects together with their *complete* surroundings, might appear in a radically more positive light when considering Massive Multiplayer Online Games (MMOGs) [13, 45]. These online platforms provide a fascinating way of simultaneously observing thousands of individuals engaged in a multitude of social and economic activities, they allow to conduct *complete* measurements of socially interacting humans, and provide data at rates comparable to physical experiments. Remarkably, one of the largest collective human activities on the planet is the playing of online games. Currently more than a hundred million people worldwide play MMOGs – the well-known game *World of Warcraft* alone has more than ten million subscribers as of today [6]. MMOGs exhibit such an enormous success due to offering their players possibilities to experience alternative or second lives, not only providing (virtual) *economic* opportunities, but also a huge variety of possible *social interactions* among players. Many MMOGs provide rich virtual environments facilitating socialization and interactions on group levels [45, 223, 224]. Motivation of players to participate in MMOGs are highly heterogeneous, ranging from establishing friendships, gain of respect and status within the virtual society, to the fun of destroying the hard work of other players. Besides economical and social interactions, modern MMOGs also offer a component of exploration, e.g. players can explore their ‘physical’ environment, such as specific features of their universe, ‘biological’ details of virtual monsters, etc., and share their

findings within ‘specialist’ communities.

From a scientific point of view online games provide a tool for understanding collective human phenomena and social dynamics on an entirely different scale [13, 46]. In these games *all* information about *all* actions taken by *all* players can be easily recorded and stored in log-files at practically no cost. This quantity of data has been unthinkable in the traditional social sciences where sample sizes often do not exceed several dozens of questionnaires, school classes or students in behavioral experiments. In MMOGs on the other hand, the number of subjects can reach several hundred thousands, with millions of recorded actions<sup>1</sup>. These actions of individual players are known in conjunction with their surroundings, i.e. the circumstances under which particular actions or decisions were taken. This offers the unique opportunity to study a complex social system: conditions under which individuals take decisions can in principle be controlled, the specific outcomes of decisions can be measured. In this respect social science is on the verge of becoming a fully experimental science [140] which should increasingly become capable of making a great number of repeatable and eventually falsifiable statements about collective human behavior, both in a social and economical context.

Another advantage over traditional ways of data acquisition in the social sciences is that players of MMOGs do not consciously notice the measurement process<sup>2</sup>. These ‘social experiments’ practically do not perturb or influence the sample. Moreover MMOGs not only open ways to explore sociological questions, but – if economic aspects are part of the game (as it is in many MMOGs) – also to study economical behavior of groups. Here again economical actions and decisions can be monitored for a huge number of individual players within their social and economical contexts. This means that MMOGs offer a natural environment to conduct *behavioral economics* experiments, which have been of great interest in numerous small-scale surveys, see e.g. [88, 114]. It becomes possible to study the *socio-economic unit* of large online game societies.

### 3.1.3 Previous data-driven work on online worlds

Since widespread usage of the internet is a necessity for individuals to devote themselves to MMOGs or other online social platforms, scientific research on these environments did not start much earlier than a decade ago. This was roughly the time when internet usage started to become mainstream in large parts of the world and the increased capabilities of data transmission allowed the establishment of well-frequented online worlds [22, 91]. Accordingly, the literature on social research in MMOGs is relatively few and far between, although in the past very few years a boom

---

<sup>1</sup>The concept of *sample* has to be reconsidered here, since within the system all data of all individuals is available and a “sample” can by default span the whole system. The question of which individuals from the total world population join the system in the first place may become important.

<sup>2</sup>Players are informed that data is logged for scientific purposes and give their consent, prior to their participation in the game.

of scientific articles on the issue could be noticed [13, 154]. Because the researchers who study these platforms almost never have access to the raw game data bases of the companies which own the game servers, usually only data can be accessed which is publicly available to players of the game. Researchers who want to study such social systems online, therefore need to create accounts just as any other player to conduct interviews or experiments, or they need to generate scripts or plug-ins to the game to harvest the publicly available data [13]. Because of this, especially data-driven studies on MMOGs are rare. Also, in relation to the total amount of data stored on the game servers, data gained this way is relatively poor and incomplete, and may suffer from the same methodological issues involved in the collection of data in classical social studies [221]. However, even with such fragmentary and possibly biased data, a few researchers have been able to conduct expressive social or economic studies. One of these studies was conducted by Castranova on testing an economic law of elasticity in a virtual world [47]. Others, such as Yee and collaborators, have conducted a census of online players [223], studied groups and their social dynamics in online games [71, 72, 126, 219], or analyzed the motivations of players from a psychological point of view [224]. It was further shown by Chesney et al. [50] that small-scale social laboratory settings can be recreated in *Second Life*, where individuals play the Ultimatum Game or conduct other game-theoretic experiments well-known in behavioral finance. While all of these studies typically describe specific aspects of certain online games or have the potential to make classical behavioral studies simpler or less costly, the lack of large-scale data sets does not allow for a quantitative, statistically significant treatment of social systems on a societal level, in which the power of the statistical physics approach could reasonably come into play to uncover potentially “universal” features of human behavior.

Due to a multitude of legal issues and the large amount of work involved in data extraction, it is usually extremely difficult to receive raw data on player activities from game companies. To our knowledge, this case has happened only once, with the online game EVE online [43], leading to a study on social norms [210]. Other works in this spirit were done by Grabowski et al. [96], who conducted a series of social network studies using data from the private server of a Lineage II shard. On the other hand, a research group around Castranova and Ross has recently developed the browser-based online game “Greenlandgame”, for the purpose of conducting socio-economic experiments in virtual worlds with focus on resource management and territorial conflicts [188, 189].

#### 3.1.4 Employing the self-developed online world Pardus

The only other example we are aware of, in which the developers of an online game coincide with the researchers who conduct the studies on the game’s population, is the case of this thesis and the articles compiled within [203–205, 208]. In the past years we have recorded practically all actions of all players taken in the self-developed, proprietary MMOG Pardus [2] which is online since 2004. Pardus is an open-ended game with a worldwide player base of almost 400,000 individuals. Players reside

### 3.1. Establishing a socio-economic laboratory

All my helpers have stopped black-marketing data.

**Anyone black-marketing drugs, I would appreciate a PM if you're willing to log all your trades. It's pretty simple. 😊**

In particular, I'd like to have someone with any amount of sneakiness and/or haggling, and also non-TSS members.

Additionally, I'd prefer it if someone would peer-review one item in my work. I derived a formula that should compute the average number of drug trades that can occur \*before\* the black market closes, including traps and bribes, given the percent chance that the BM stays open.

I'll outline the method I used, to make it simpler to verify:

1. I used a special-case discrete negative binomial distribution ( $r=1$ ) to represent the number of trades, " $k$ ", before it closes once (hence  $r=1$ ), if " $p$ " is the probability of the BM staying open:  

$$f(k) = (1 - p) \cdot p^k$$
- so as an example, if the BM stays open for 60% of trades, the chance that the 2nd trade will be the last trade before the BM closes is  $0.4 \cdot 0.6^2$ , or about 14.4%. But this only gave me the probability that for trial number " $k$ ", the BM would close on the following trade.
2. To calculate the average number of trades that occur before the BM closes, I would have to sum all the probabilities starting from trade zero (since the BM can shut down on the first trade) until they add up to 50%. That will be the average number of trades before the BM closes since 50% of the time it will close before that, and 50% of the time it will close after that. This then, represents the summation, and " $s$ " represents the trade \*before\* the BM closes:  

$$(1 - p) \cdot \sum_{k=0}^s p^k = 0.50$$
3. But I want a closed-form equation that I don't have to iteratively sum every time I log more data. This is a variation on the first equation [here](#) that I used to remove the summation for a closed-form solution:  

$$\sum_{k=m}^s p^k = \frac{p^m - p^{s+1}}{1 - p}$$
4. And finally, substituting that for the summation (the " $1-p$ " terms cancel), simplifying, and solving for " $s$ " yielded:  

$$s = \frac{\ln 0.5}{\ln p} - 1$$

Again, " $s$ " is the average number of trades over many illegal BM trades you can expect before the BM closes on the "next" trade. But it's more useful to know on which trade the BM can be expected to shutdown since when the BM shuts down, the drugs are still sold for money successfully.

To find out on which trade the BM will most often shut down, just remove the "- 1" at the end. Plug in the chance that the BM will stay open for " $p$ ", and solve for " $s$ " in this equation:  

$$s = \frac{\ln 0.5}{\ln p}$$

I know that by using Bernoulli trials I'm assuming that each trade is independent of the rest, but at the moment it's an assumption I'm willing to make. That's the only fault I can see here... Please, someone double check my logic 😊

EDIT: found two mistakes, it should be closer to correct.

Figure 3.1: One example illustrating the rich exploratory life of players in the Pardus society. This forum entry, posted on November 27th 2010 in the Pardus forums, shows a “scientific method” proposed by one of the players, who encourages his fellow players to gather data and to peer-review his work on a particular market mechanism in the game (“I’d prefer it if someone would peer-review one item in my work.”).

and act within a virtual, persistent futuristic universe and make up their own goals. Most players invent and develop their virtual social lives without constraints by the game setup. The game’s environmental topology is given but can be manipulated by the players to some extent. Players self-organize within groups and subgroups, claim territories, decide to go to war, etc., completely on their own accounts. Players typically participate in the game for several weeks to several years, see Section 3.5. One illustrative example for the richness and complexity of the Pardus society is given by the continuous attempts of players to explore and understand the laws governing their universe. Figure 3.1 shows a “scientific method” proposed by one of the individuals in the game forum, who encourages his fellow players to gather data and to peer-review his work on a particular market mechanism in the game.

Players of Pardus characteristically engage in various economic activities to increase their wealth (non-convertible game money): There are numerous possibilities for jobs, such as mining and processing basic resources from the environment, trade,

production, assembly and consumption of commodities, etc. Economic life is embedded in a production tree which provides a basic framework for player-created industries. Trade occurs following simple ‘rules’ within dynamic and demand-oriented virtual markets constituted by groups of players, see Section 3.6.2. Social life within Pardus is based on means of communication with fellow players in various forms, such as chat, forum, private messages, which allow the establishment of e.g. friendships or hostile relations, see Section 3.6.1. There are a number of ways to publicly display one’s ‘status’ within the virtual society: Purchase of expensive status symbols, such as space ships, earning of medals of honor for war efforts or for defeating outlaws, etc. These possibilities are not only well used, but constitute an important psychological driving force for many players.

Given the complete data set from the Pardus game, one can identify four major directions of possible research.

**1. Network analysis.** It is possible to directly access the dynamics of several types of social networks such as dynamics of friend networks, networks of enemies, or communication networks. Especially the latter offer a way to directly relate findings in the game with real-world communication networks, such as a data set of cell phone calls which has been recently analyzed from a network perspective [138, 172]. While there exists some insight into real-world friend networks in the literature, e.g. of the Facebook community [93], there is practically no knowledge of topology and dynamics of enemy networks [137]. Since the time resolution of our data is accurate to one second, it becomes possible to study time courses of global network properties. Using such longitudinal network data it can be understood if and how communities show aging effects, such as *densification*, i.e. shrinking diameters and growing average degrees. This phenomenon has been observed in societies and online communities [142, 144], as well as in the evolution of scientific fields and cities [28, 29]. The vast majority of social network studies analyze single or at best small numbers of network snapshots. Important exceptions include time resolved studies of an internet dating community [116], the analysis of a university email network [134], of the web of scientific coauthorships [164, 167, 182], as well as several large-scale networks of various types [144].

Network growth and re-linking processes can be directly studied and compared to well-known models, such as e.g. the preferential attachment model [19] or static relinking models [207]. Preferential attachment dynamics of real-world networks have been verified in a few recent studies [60, 123, 142].

**2. Testing traditional social-dynamics hypotheses.** The Pardus data allows for direct empirical testing of long-standing hypotheses on social network dynamics, such as the *Hypothesis of triadic closure* [97, 180], the *Weak ties hypothesis* [97], or the *Hypothesis of social balance* [65, 104].

For quantification purposes we employ network measures such as *betweenness centrality* [86] and *overlap* which measures how often a given pair of nodes has links to other common nodes [172]. To our knowledge, no longitudinal measurements of large-scale signed networks exist as of today. One well-known social network study on monks in a monastery can be found in the classic literature [191], as well as

a modern long-time survey of social dynamics in classrooms [127]. These are first attempts of systematic social balance experiments, however being far from conclusive due to limited data and small scales (10 to 100 nodes), and a low number of samples (about  $\approx 10$  observations). Further, the extent of *reciprocity* and *Triad significance profiles* [157] together with their dynamics can be directly accessed from the game data. To understand microscopic changes in social network dynamics, transition rates between dyadic and triadic structures can be measured – yielding parameters needed e.g. for calibrating agent-based models of social network dynamics, as e.g. in [10]. So far these transition rates could only be assumed by model builders and have never been measured in actual societies.

**3. Economic analysis.** The fact that all players are engaged in economic activities allows for statistically significant measurements of *wealth and income distributions* which can be compared to real economies [49, 70, 222]. The process of price formation – and more generally preference relations – for all goods and services in the game can be observed within the social and economical context of players. All prices of all goods in the game are recorded with a time resolution of one second and can be analyzed with respect to ‘stylized’ facts in real prices, in a straightforward fashion. These ‘stylized’ facts, such as volatility clustering, fat-tailed return distributions, squared auto-correlation decays, etc., are known for traded goods in the real world [57]. Further, co-evolution dynamics, i.e. the evolution of economic properties of players (e.g. wealth) as a function of their local social networks, and vice versa, their social evolution as a function of their economic network, can be extracted from the data. The theoretical literature on co-evolving networks is relatively sparse [30, 31]; to our knowledge there exist practically no measurements on this issue so far.

**4. Group formation and dynamics – gender and country aspects.** Players have the possibility to create and join ‘alliances’ (communities) in the game, which allow them to streamline ideas, join forces for common projects, or coordinate aims and beliefs. The dynamics, formation, interaction, and disappearance of these cohesive communities can be investigated readily in the data. Players choose to have a male or female character allowing all data to be partitioned into female and male networks. This offers the possibility to search for behavioral differences in ‘networking’ patterns between female and male players. Further, economic productivity and communication patterns can be analyzed gender-specifically. The same holds for country specifics of players.

In this work we focus on the first two of the above directions: analysis of complex network structures/dynamics and testing sociological hypotheses. Along these lines we establish further evidence that online game communities may serve as a model for real world communities. It is not obvious *a priori* that a population of online players is a representative sample of real-world societies [220]. However, several recent studies are providing evidence that human behavior on a collective level is remarkably robust, meaning that statistical differences of real-world communities and game-societies are often marginal [124, 126].

## 3.2 Overview and history of Pardus

Pardus [2] is a browser-based MMOG in a science-fiction setting, open to the public and played since September 2004. A browser-based MMOG is characterized by a substantial number of users playing together in the same virtual environment connected by an internet browser. The game's manual states [4]: "Pardus can be categorized as a Massive Multiplayer Online Browser Game (MMOBG), a subcategory of a Massive Multiplayer Online Game (MMOG). This means that thousands of people from all over the world are playing together in the same fictional universe at the same time over the internet. Because Pardus is also about roleplaying, it is a subcategory of a Massive Multiplayer Online Roleplaying Game (MMORPG) too. Since 'MMOBRPG' would be a bit too complicated, we left the 'RP' out." For a detailed categorization of online games see [22, 45, 154].

In Pardus every player owns an account with one *character* per game universe; players are forbidden to operate multiple accounts. A character is a pilot owning a spacecraft with a certain cargo capacity, roaming the virtual universe trading commodities, socializing, and much more, "to gain wealth and fame in space" [5]. The main component of Pardus consists of trade simulation with a society of players heavily driven by social factors such as friendship, cooperation or competition. Conflictual relations may result in aggressive acts such as attacks, fights, revenge, even destruction of another player's means of production or transportation. Under certain conditions, hostile acts may degenerate into large-scale conflicts between different factions of players – wars. There is no explicit 'winning' in Pardus as there is no inherent set of goals nor allowed or forbidden 'moves' (with a few exceptions mainly concerning decent language and behavior towards fellow players). Pardus is a *virtual world* or *synthetic world* with a gameplay based on socializing and role-playing, with interaction of player characters with others and with non-player characters as its core elements [45].

### Action Points – the unit of time

Every game action carried out by a player (trade, travel, etc.) costs certain amount of so-called *Action Points* (APs). These points can not exceed a maximum of 6,100 APs per character. For characters owning less APs than their maximum, every six minutes 24 APs are automatically regenerated, i.e. 5,760 APs per day. Once a player's character is out of APs, she has to wait for being able to play on. As a result the typical Pardus player logs in once a day to spend all her APs on several activities within a few minutes (for each character/universe). This makes APs, the game's unit of time, the most valuable factor: Those players who use their APs most efficiently can experience the fastest progress or earn the highest profits. Social activities such as chatting (see Section 3.6.1) do not consume APs. Highly involved players usually spend a lot more real time on the game's features of socialization as well as on planning and coordinating their future moves than on actually spending their APs.

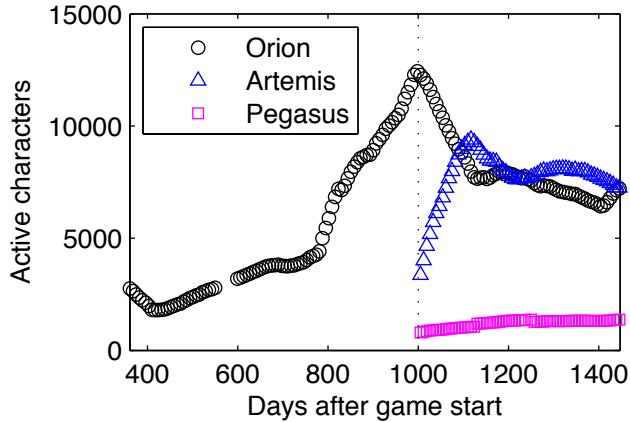


Figure 3.2: Evolution of number of active characters in the game universes. The large increase of players in Orion between days  $\approx 800$  and 1,000 is due to advertisement campaigns after October 2006. At day 1,000 (dotted line) the Artemis and Pegasus universes were opened. Some thousand players abandoned their Orion characters focusing on their new Artemis characters. This explains the mirrored development in these two universes after day 1,000.

### 3.3 The game universes

There exist three separate game universes: Orion, Artemis, and Pegasus. Presently Pardus is actively played by  $\approx 13,000$  players, over 380,000 have registered so far. Orion and Artemis each inhabit  $\approx 6,000$  active characters, Pegasus with its  $\approx 1,000$  characters is for paying customers only. We count existing characters who have mastered the game’s tutorial environment as active (characters which are inactive for 120 days get deleted automatically, see Section 3.5). Figure 3.2 depicts the evolution of universe populations for the time range where data is available, see Section 3.4. The majority plays the game for free, paying members receive *Premium accounts* which bestow them with additional features not available to users with *Free account* status (such as the possibility of character creation in the Pegasus game universe). Orion was opened on September 14<sup>th</sup> 2004, Artemis and Pegasus 1,000 days later, on June 10<sup>th</sup> 2007. Between universes it is impossible to move, trade, or exchange game money. The universes are independent<sup>3</sup>.

Orion has the longest history but has some missing data at its beginning, see Section 3.4. Orion has converged to social structures and ties which are relatively constant over time, as seen in a number of measures (not shown). In contrast, Artemis and Pegasus can be observed from their start on, and measuring the evolution of social networks from the time of their “birth” is possible. Here we do not face the problem of the ‘missing past’ [144]. Since Pegasus is only accessible to paying

<sup>3</sup>This is not entirely correct since some players have openly revealed their identities, i.e. they have disclosed which characters they are controlling in different game universes. It is not clear how many attempts have been made to copy existing social ties between universes. Although it is discouraged, it may happen that e.g. vendettas between players who are aware of their mutual identities in different universes are carried out within more than one universe.

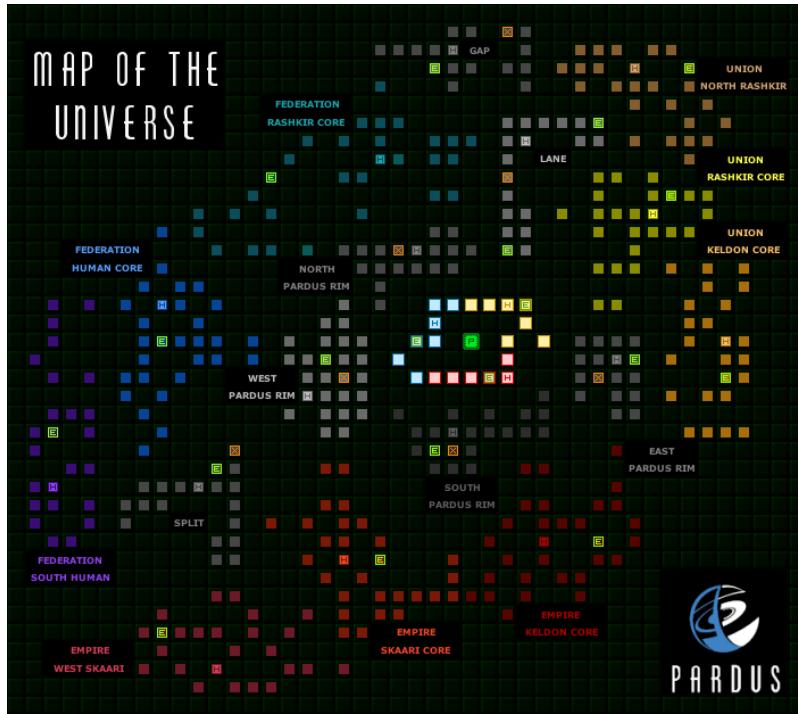


Figure 3.3: Map of a Pardus universe. Each colored square represents a sector, consisting of  $15 \times 15$  fields on average. Colors indicate cluster membership. A cluster is a number of sectors, usually around 20, pre-defined as a political region. Nearby sectors are connected by wormholes (not shown). See Chapter 5 for a detailed account on clusters and the topology of the universe.

customers of the game, it has much less players. It is due to these reasons that most parts of this work consider only data from Artemis. However, since character creation in Orion and Artemis is free but not so in Pegasus, the former universes display considerably higher fluctuations in player numbers and activity rates than the latter.

Note that the Pardus universes are virtual worlds, permanently evolving without a scheduled end. They are far from being in equilibrium due to often added new and changed game features (especially in the game's official test phase before 2006-10-01), due to fluctuations in player numbers, or due to large-scale collective actions of players, etc. The distinction between transient phase and steady state may be hard to determine for certain network properties, due to the large amount of strong mutual influences naturally present in systems of high complexity.

Space in Pardus is two-dimensional. The topology of all universes is equivalent, however there is the possibility for players to block regions of space from being entered by other players, see Chapter 5. Each game universe is divided into 400 *sectors*. Figure 3.3 shows the map of the Pardus universe on which the sectors are displayed as colored squares. Each sector consists of  $15 \times 15$  *fields* on average. Fields are the smallest units of space and are displayed as square images in-game. They form a square grid on which continuous ship movement is possible by clicking on the desired destination field within the *space chart*. This chart is a  $7 \times 7$  fields cut-out of

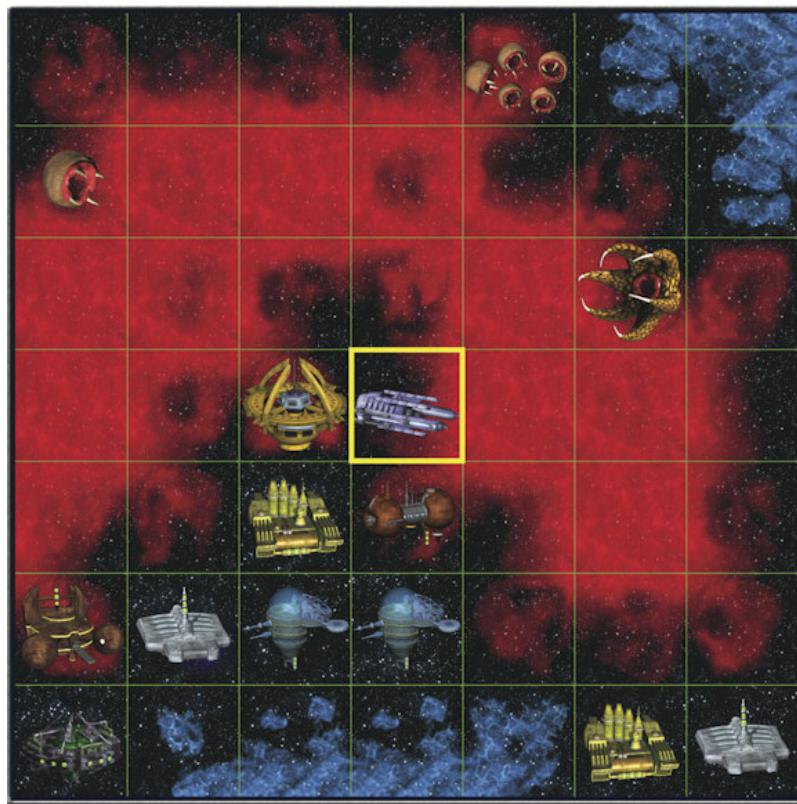


Figure 3.4: Space chart. A cut-out of  $7 \times 7$  fields of the universe (green lines, not shown in-game) visible to every player on his navigation screen; the current position is in the central field (yellow box). By clicking on another field, the ship moves to that location and all fields are updated so the ship is again located in the center.

the universe visible to every player with their current position located on the central field, see Fig. 3.4. A sector's boundary is impenetrable; moving between nearby sectors is possible by tunneling through field objects called *wormholes*. A collection of nearby  $\approx 20$  sectors is called a *cluster*. The typical spatial range of activities of a character is usually confined to one cluster for several weeks or longer. In Chapter 5 we give a more detailed account on socio-economic and topological properties of the universe due to its importance for mobility.

## 3.4 The data analyzed

Daily database backups recorded at 05:32 GMT<sup>4</sup> are available from 2005-09-09. The day 2005-09-09 is the 360<sup>th</sup> after 2004-09-14, i.e. the 360<sup>th</sup> day after Orion was opened. Backups from the following dates are not available due to unknown reasons: 2006-03-24 to 2006-04-23, 2006-04-28 to 2006-04-30, 2006-10-24 to 2006-10-26, 2007-03-20, 2007-05-10, 2007-09-21, 2008-02-09, 2008-06-09. Since we have complete data

<sup>4</sup>This time was chosen for the daily backup and maintenance scheduler because it is the time of lowest player activity.

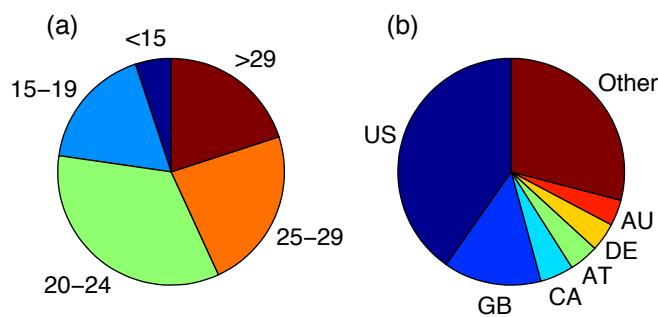


Figure 3.5: Distribution of player age (a), nationality (b) in 2005.

only for Artemis and Pegasus (with the exception of three days), and because Artemis has more active characters than Pegasus, in this work we refer only to Artemis data unless explicitly stated otherwise. For clarity only weekly data points are shown in all time evolution plots of Chapter 6, except for Figs. 6.5 and 6.11. The range of days from which data was considered varies from Chapter to Chapter due to the different times when these Chapters were developed. The legal department of the Medical University of Vienna has attested the innocuousness of the used anonymized data.

## 3.5 Players and census of characters

### 3.5.1 Age and nationality

In a poll taken in the Pardus forums in the beginning of 2005 the age of players was assessed. From a total of 255 votes, 5% reported their age to be less than 15 years, 18% between 15–19, 34% between 20–24, 23% between 25–29, and 20% are older than 29 years, Fig. 3.5 (a). The distribution of player nationalities can be estimated by technical means and reads approximately as follows: United States (US) 40%, United Kingdom (GB) 14%, Canada (CA) 5%, Austria (AT) 4%, Germany (DE) 4%, Australia (AU) 4%, Other 29%, Fig. 3.5 (b).

### 3.5.2 Lifetimes of characters

Characters are automatically deleted after an inactivity (not logging in) period of 120 days. Additionally, every player has the option to delete her account or single characters at any time. Rarely, it happens that accounts get deleted due to breaking of game rules, such as the operation of multiple accounts. We call all deletions which are not due to inactivity *self-induced*. Figure 3.6 shows the cumulative distribution of character lifetimes (in days) of all 16,980 characters who existed for at least one day, but not on the last one. If a character's lifetime lies before day 120 (dotted line), her deletion could have been self-induced only. If a character's lifetime is longer than 120 days, her deletion was either self-induced or automatic due to inactivity. These

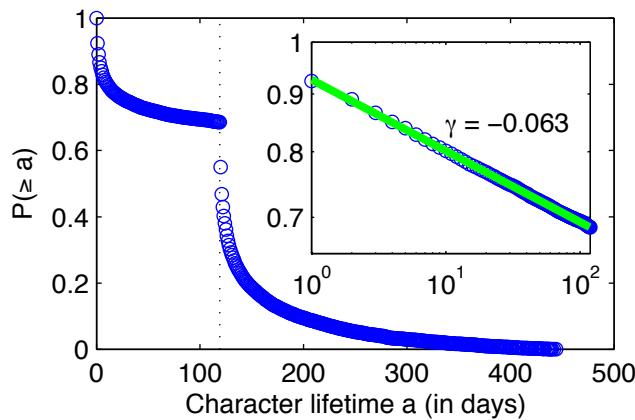


Figure 3.6: Cumulative distribution of character lifetimes (in days) of all 16,980 characters who existed at least for one day but not on the last one. The dotted line marks day 120. The inset shows the distribution of lifetimes of  $< 120$  days in a double-logarithmic scale.

two regimes are evident in the figure: the regime of only self-induced deletion follows a power-law with exponent  $\gamma = -0.063$ . In the second regime, the distribution is neither a power-law nor an exponential probably due to the overlay of the two deletion schemes.

Of all characters, 7.6% have a lifetime of 0 days, i.e. they delete themselves on the first day of their existence. At least 31.4% of all deletions are self-induced, and  $\approx 13\%$  of all characters become inactive after their first day.

### 3.5.3 Gender of characters

When signing up for the first time, players have to choose between a male and female character; this decision is irrevocable. Depending on gender a male or female *avatar* (profile-like picture) of characters is displayed in certain places in the game. In the Artemis universe,  $\approx 90\%$  of all characters are male.

## 3.6 Life within Pardus

### 3.6.1 Social interaction

#### Means of communication

There are three ways for players to communicate inside the game, facilitating social activity. Players may use these facilities independently from game-mechanic states (such as their ship's location within the universe, their wealth, etc.):

1. *Chat*: Pardus offers several built-in chat channels per universe where players can simultaneously communicate with many others. Chat entries scroll up and disappear; thus the chat is well-suited only for temporary talks.

### 3. Quantification of socio-economic behavior using an online society

---

2. *Forum*: In the forum, messages, called *posts*, can consist of several lines and stay for a long time. This enables more thorough discussions than in chat. Posts are organized within *threads* which correspond to a topic. There are universe-specific subforums as well as global subforums which can be accessed from all universes.
3. *Private message*: Within a universe it is possible to send private messages (PMs) to any other player; this action and the PM's content is only seen by sender and receiver – a system similar to email. When a PM is sent the receiver gets immediately notified in a status bar. PMs always have exactly one recipient. Presently a daily total of  $\approx 10,000$  PMs are exchanged within Pardus.

### Alliances and factions

Players are able to create or join alliances, which are entities with the purpose to combine groups of individual players allowing them to operate with joined forces for their possibly common aims. A player can be member of not more than one alliance at the same time and may leave or be kicked out by its leader anytime. Each alliance has a private chat channel; members are able to send *alliance-PMs* to all other members at once. Alliances can be interpreted as loose analogies to firms or societies. Although there is no explicit limitation, they consist of up to about two hundred players, a value not far from the Dunbar number of 150 [73], a number that also appears in the maximum number of friends in social networks, see Section 6.3.3.

Besides alliances there exist three *factions*, pre-defined groups which can be joined by players independent of alliance membership. Doing so yields certain benefits such as the possibility to gain access to faction technology: ships and equipment. The majority of all players are members of a faction; as alliance membership, faction membership is unique, meaning that at one point in time a player can only be member of one faction. Sometimes two factions are conducting war where affiliated players receive certain incentives for joining the (armed) conflict with their opponents.

### Friends and enemies

For a small amount of APs, players can mark others as *friends* or *enemies*. This can be done for any reason. The marked characters are added to the markers personal *friends* or *enemies list*. Additionally, every player has a personal *friend of* and *enemy of* list, displaying all players who have marked them as friend or enemy, respectively. When being marked or un-marked as friend or enemy the affected player immediately receives an informative system message. It is only possible to mark someone either as friend or as enemy, but not both simultaneously. Besides names and online statuses being displayed in the PM contacts for quick access, the friends and enemy list serve game-mechanic purposes: friends/enemies are automatically or optionally included/excluded for certain actions. For example, enemies of building

or starbase owners are not able to use the services offered in the respective places (see Section 3.6.2).

We stress that friend/enemy lists and friend of/enemy of lists are *completely private*, meaning that no one except the marking and marked players have information about ties between them. It is not possible to see second degree neighbors (e.g. friends of friends) or the number of ties another player has<sup>5</sup>. Note that this is in contrast to many online social networking services such as Facebook, where usually second degree neighbors and number of friends are visible. Thereby Pardus' system does not introduce potentially strong biases concerning accumulation of friends (some users may tend to accumulate friends for the main purpose of increasing their publicly visible number of friends [93]). Our data thus represents a more realistic social situation, in the sense that social ties are not immediately accessible but need to be discovered by e.g. communication with or careful observation of others.

Besides character names and online status being displayed on every player's personal PM contacts page for quick access, the friends and enemies lists serve game-mechanic purposes: friends/enemies are automatically or optionally included/excluded for certain actions. For example, enemies of building owners are not able to use the services offered in the respective places. Note that friend and enemy markings need not necessarily denote *affective* friendships or enmity, they rather indicate a certain degree of cooperative or uncooperative stance motivated by affective and/or cognitive incentives. However, we assume these two motives to coincide to a great extent, e.g. it seems highly unlikely that someone marked as enemy/friend due to rational considerations at the same time constitutes the affective opposite of friend/enemy within the game (and vice versa). PMs as well as friend and enemy relations can be displayed as networks, Fig. 3.7, see also Section 6.1.

### 3.6.2 Economy

Besides social interactions, a variety of different economic activities form the basis of life within Pardus. While he have not yet analyzed economic aspects of the game in detail, it is important to describe the rich economic environment for an understanding of the complexity of the socio-economic system in Pardus. The currency of game money is the so-called *credit*. This money is not convertible to real money. Every player starts her life with 5,000 credits. Since most assets needed for making progress – such as ships, ship equipment, buildings – are traded in credits, it is of basic interest to earn money during the game (the richest players currently possess hundreds of millions of credits).

There exist over 30 commodity types, such as food, water, energy, or metal. Five of these can be mined (see Mining below) and are therefore called *resources*. In contrast, we label all other commodities *luxuries*. The basic indivisible unit in which

---

<sup>5</sup>On 2008-08-24 the so-called *profile* feature of the game was extended, allowing players with Premium accounts to publicly display their numbers of friends or enemies. Since this feature was introduced at the very end of our last measured data (2008-09-01) and only a negligible proportion of players are making use of it, it can be regarded as irrelevant here.

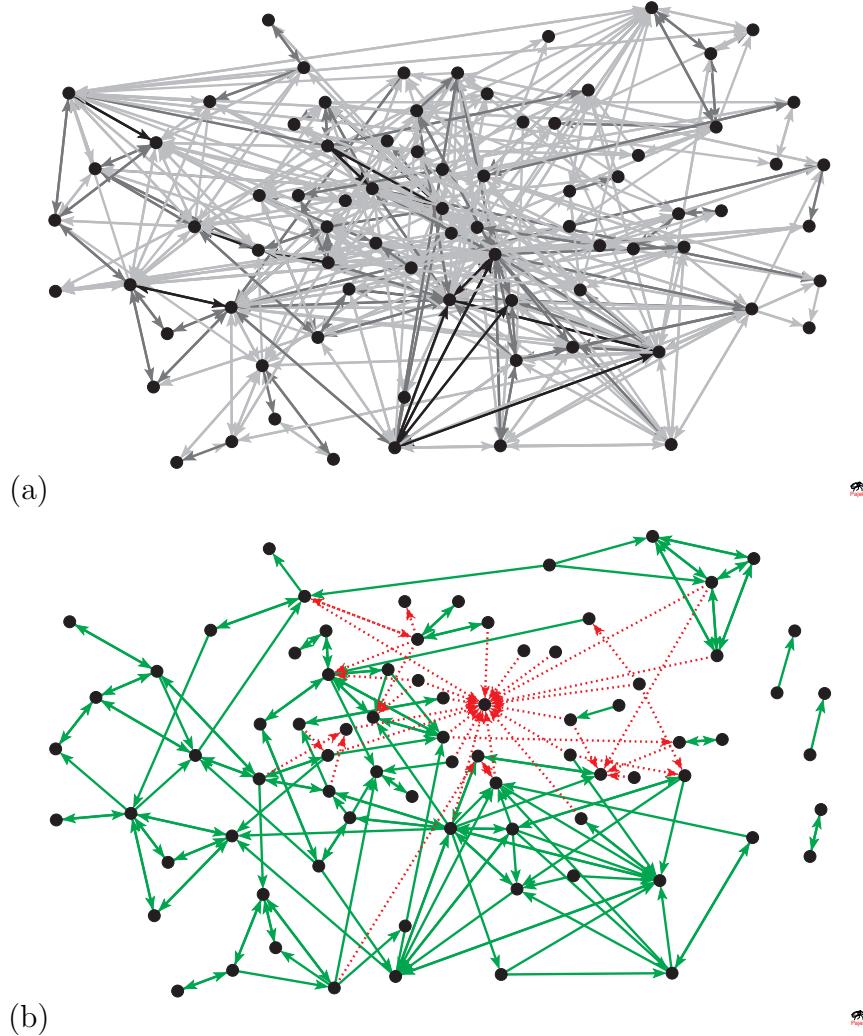


Figure 3.7: (a) Accumulated PM communications over 445 days between 78 randomly selected individuals who existed on the first and last day. Link colors of light gray, gray, and black correspond to 1–10, 11–100 and 101–1000 PMs sent, respectively. (b) Friend (green, solid) and enemy (red, dashed) relations on day 445 between the same individuals. See our Youtube channel <http://www.youtube.com/user/complexsystemsvienna> for animated time evolutions of these networks.

the amount of commodities as well as storage and cargo capacity is measured is the *ton*.

A variety of techniques for generating revenues are practiced in Pardus:

- **FWE Run**

The game's primary trade centers are *planets* and *starbases* (both subsequently referred to as *bases*), one or two to be found in most sectors. These are immobile and indestructible entities provided by the game with a practically unlimited amount of credits available. Bases offer essential services such as the repair of ship equipment or access to local trade markets. Because of these essential

features, most activities in the game take place on or in the neighborhood of bases. For being able to provide certain services in a reasonable scope, bases need to have reached a certain extent of *population*. The population of a base is an integer value, usually between 500 and 200,000, which is also directly related to the capacity and spending power of the base, see below. Generally, a base's population grows when it is provided with enough commodities of a specific sort by players. In case these upkeep-commodities are not delivered, a base loses population. Every three hours the upkeep is automatically checked, followed by the corresponding calculation and execution of population shift (this recurring event is called *tick*). Thus, it is a basic long-time interest of most players to regularly provide upkeep to the bases they tend to interact with.

This interest is fueled naturally by the opportunity for profit: Prices on bases are exclusively determined by local supply and demand. For every commodity type  $i$  available for trade on base  $k$ , there exists a balance value  $b_{i,k}(s_k)$ , dependent on commodity type and the base's current population  $s_k$ <sup>6</sup>. The number  $b_{i,k}$  is a negative (positive) integer and indicates how many tons of commodities will be consumed (produced) at the next tick. The base's storage capacity  $m_{i,k}$  for commodity type  $i$  is defined by:

$$m_{i,k} := |b_{i,k}| \cdot 20 + 20. \quad (3.1)$$

Finally, the commodity's price  $c_{i,k}$  is determined via the exponential relation

$$c_{i,k} := \left\lfloor c_{i,0} \cdot 10^{-\frac{a_{i,k}}{m_{i,k}}} \right\rfloor, \quad (3.2)$$

where floor brackets denote the floor function,  $c_{i,0}$  refers to pre-defined, globally constant basic prices, and  $a_{i,k}$  denotes the presently stocked amount of the respective commodity type. Equation (3.2) corresponds to sale prices. For purchase prices,  $c_{i,0}$  is replaced with  $0.8 \cdot c_{i,0}$ , i.e. by a single base commodities are always purchased for a 20% lower price than sold.

Every tick, this dynamic system of supply and demand creates economic disequilibria throughout the universe: Due to changed amounts  $a$ , prices of consumed commodities go up, while the prices of produced commodities go down. As a trader, closing this price gap by shuttling to and from bases with full loads of the appropriate kinds of commodities immediately after a tick can yield high profits. Since planets produce food and water but need energy in large quantities for upkeep, for starbases vice versa, the widely practiced trade technique involving these commodities has been coined Food-Water-Energy Run (FWE Run) by the Pardus community. Note the dynamic, self-regulating nature of

---

<sup>6</sup>The latter dependence is non-linear and not the same for all commodities. Values for  $|b|$  usually range between 1 and 1000, where luxuries typically have low balance values, while resources and upkeep-commodities have high balance levels.

this process: Where many players practice the FWE Run, price gaps close quickly. In sparsely inhabited regions however, where there is less competition, more or easier profits can be achieved. Thus, the system is naturally regulated by diffusion of certain player activities over the game universe.

- **Long-range trade**

Since the sectors which connect different clusters contain no planets and are designed to take slightly more APs to travel through, profitable FWE Runs are usually confined to sectors within the same cluster; i.e. the FWE Run is an intra-cluster trade technique. There are however commodities which must be hauled through several clusters in order to gain a reasonable profit. The “indecent” version of this inter-cluster trade consists of four kinds of *body parts* available in the black markets of bases, the “decent” version consists of three kinds of *jewels* available at special places. Each brand is produced in certain clusters only and purchased by non-player characters (NPCs) in remote clusters for relatively high prices - the farther away, the higher the payout in general. Since purchasing body parts leads to a slight reduction of a base’s population, the trade of these commodities is despised by those players who wish to see their bases prosper. There is a mechanism which makes it possible to find out the last trader of body parts on a base, allowing every member of the society to trace and punish the offenders if desired.

- **Bulletin board missions**

On every base, the game automatically generates missions which are offered to players on a local bulletin board. These are assignments limited in time having destination within the cluster – e.g. deliver 10 tons of packages to a planet in a nearby sector within 60 minutes – which among other benefits generate a payout by the base when fulfilled. There are also special missions to destinations outside of the cluster.

- **Mining**

Almost every field contains an amount  $r$  of one of the natural resources which regenerate once every day. By using APs, every player can mine resources without any additional costs, e.g. for the purpose to sell them to bases or to other players. When mining, a certain percentage of the available field resources is gathered. Again, note the self-regulating nature: In remote sectors where field resources are abundant, players can achieve high profits with relatively little effort. In highly populated regions however, money is hard to earn this way since a lot of miners have already drained substantial fractions of the possibly available resources<sup>7</sup>.

---

<sup>7</sup>The derogatorily used term *stripminer* was coined by players to refer to myopic or ignorant miners who make their earnings by draining natural resources as much as possible. Regularly, heated ecological debates are raised in the Pardus forums about the moral badness of stripminers and about joint efforts to antagonize them.

- **Climbing the production tree**

Usually luxuries are processed out of resources or less valuable luxuries in player-owned buildings. For example, a brewery manufactures expensive liquor out of cheap energy, water, food, and chemical supplies. Every player has the possibility to construct a small number of such buildings. On each field, only one building or base may reside. Buildings are immobile and act as production sites as well as outlets: The owner may freely specify limits and prices for consumed and produced commodities and may set her building to open trade. In this case, every player is able to carry on commerce with the building at any time, to provide upkeep or to purchase its output for the pre-set prices. Analogously to bases, buildings tick every six hours at which times upkeep is consumed and output produced. The credits a building has available coincides with the owner's credits – whenever there is trade happening in a building, the owner's credits are affected directly, independent of her (i.e. her ship's) current location. If a player is out of credits, it is not possible to sell commodities to her buildings.

Only by collaboration of players may high profits be achieved: Pardus' production tree has multiple layers of interdependencies [3], requiring several players to coordinate the construction of local industry chains. Once established, it is up to the owners to provide their buildings with upkeep themselves or to let others do this work. A method of trade is practiced in densely industrialized sectors, consisting of players flying from low-level to high-level buildings one by one, reusing the purchased commodity outputs as upkeep stock for the next outlet. These suppliers can profit only if building owners arrange reasonable prices. This is usually in the owners' interest since delegated supply work allows the owner to spend APs for other activities.

Most end-products, i.e. the produced luxuries at the top of the production tree which cannot be reused as upkeep in other outlets, are usable commodities. For example, *drugs* may be consumed to gain APs, *robots* are used to repair a ship's armor, or *droid modules* can be installed on buildings to provide powerful defenses. Therefore, the motivation for developing industries is to some part based on the Pardus society's tangible needs.

- **Player starbases**

Every player has the option to build one player-owned starbase (PSB). The fields on which PSB construction is possible have certain distance requirements, e.g. a Chebyshev distance of at least 15 to fields containing a planet or a starbase, and a distance of at least 5 to wormholes. Because of these requirements, there is only a limited amount of  $\approx 400$  PSBs that can exist in a universe at the same time. As NPC starbases<sup>8</sup>, PSBs tick every 3 hours. Unlike buildings, PSBs have separate credit pools, e.g. the owner has to deposit

---

<sup>8</sup>In cases where it is necessary to distinguish entities in the game not owned by players from player owned ones, the former receive the prefix *NPC* (for non-player character).

credits in her PSB if it should be able to buy commodities through open trade. Also, prices do not dynamically adjust to supply and demand, see Eq. (3.2), but can be freely set as flat rates by the owner similar to buildings.

A PSB has the following sources of income: every tick, credits equal to 20% of its population are earned as taxes. Successful bulletin board missions taken by players on the starbase pay 10% of their rewards. Any ship repairs done at the starbase pay by default 11% of their cost to the starbase – the owner is able to set higher or lower repair surcharges. Commodity trade and production/sale of ships and equipment are further ways to supplement starbase income. Ships and equipment are produced out of commodities and credits, where the credit requirement is a fraction of the (globally constant) basic ship price at NPC bases. Hence by setting a sale price for constructed ships and equipment between production cost and basic price, both seller and buyer profit. Competition between ship and equipment constructors drives sale prices down, occasionally below production costs, e.g. to lure new players into joining the constructor's alliance with the bargain of receiving a ship for a low price or even for free.

- **Pardus PSB subsidy**

Players receive a daily subsidy from their faction if they maintain a faction-aligned PSB in the special Pardus cluster, which is only accessible to Premium Account characters. This enforces competition within this region.

- **Ship-to-ship trade**

If players meet on the same field, they may “donate” any amount of their cargo hold or available credits to the trade partner. Remote donations are not possible<sup>9</sup>.

- **Raid**

Evil-minded players have the possibility to raid ships or buildings, taking away commodities by force. Currently social pressure, player organization and building defenses are however so highly developed that raid is rarely practiced. The situation may have been different in the universes' early build-up phases.

- **Bounty hunting**

A bounty is an amount of credits placed by a player on another player; everyone has the possibility to place bounties, for any reason. Anyone who destroys another player's ship receives the sum of bounties which were placed on her head

---

<sup>9</sup>By using indirect methods of money transaction it is however very well possible to remotely transfer credits – for example by taking advantage of the fact that a player's credits coincide with the credits available in her buildings. A player therefore does not need to meet with another player directly, but credits between two players can be transferred remotely by using the trade feature of buildings. It is remarkable that there have been attempts by some players to organize banks by specifically utilizing this function in declared buildings, completely altering their purpose originally intended by the game developers.

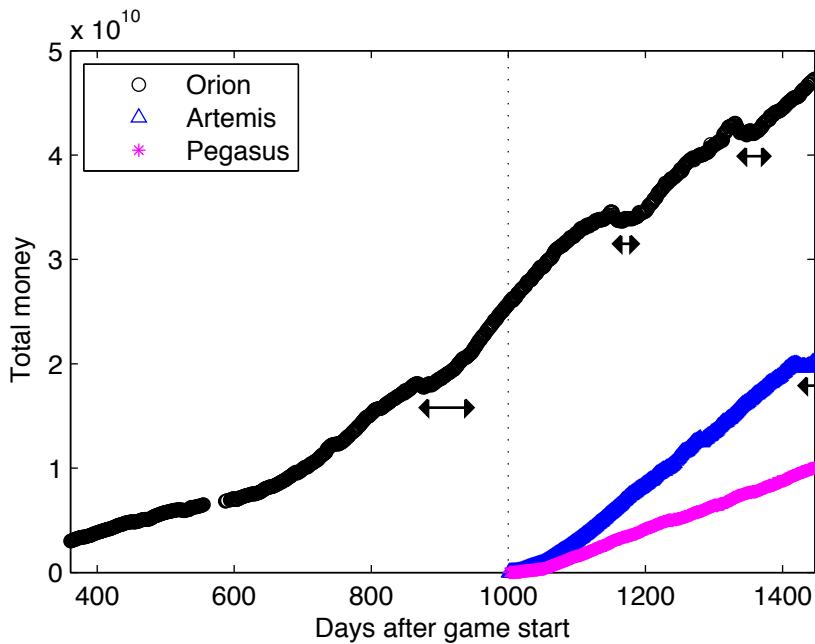


Figure 3.8: Evolution of total money owned by players. Arrows mark all times of war between two factions, where large amounts of credits are spent for military efforts or annihilated due to destructions of player starbases. The dotted line at day 640 marks the opening of the Artemis and Pegasus universes; gaps appear due to unavailable data.

and thereby cancels them. A “shot down” player loses some of her progress<sup>10</sup> but otherwise continues playing the game as before. On every base there are global bounty boards displaying all existing bounties and bounty reasons.

- **Scamming**

A rarely practiced but very profitable if well-executed method, which is not supported by the game system per se, is scamming, i.e. tricking other players into giving you money by means of deception or by earning their trust. Although their short-time profits may be huge, scammers are usually “out-laws” for the rest of their lives within the game’s society.

We make the following observation: all presented classes of traders (FWE-runners, miners, suppliers, long-range traders) as well as classes of players defined by other in-game activities (fighter, bounty hunter, hacker, . . . ) are not pre-determined explicitly, but arise out of a potential for profit of some kind. Any player acts as she pleases according to her own perception, the game’s framework facilitates certain ways of (continuous) action however, leading to the evolution of a complex socio-economic system with emerging collective phenomena, such as wars or the formation of economic cartels<sup>11</sup>. In case of traders the profit is of financial nature, other classes of players may be motivated by different values such as social (or anti-social) prestige.

<sup>10</sup>The loss of progress consists of one piece of ship equipment, as well as a fraction of the player’s so-called *skill* and *experience* points which are needed in combat or for constructing buildings.

<sup>11</sup>Note that the game’s official phase of public testing lasted for two years. During this time,

### 3. Quantification of socio-economic behavior using an online society

---

“Economic realism” becomes apparent on the following level: Money is generally earned to be utilized *as a means to an end*, e.g. for being invested into more prestigious ships or into military efforts to prevail over one’s enemies. Figure 3.8 shows how the total money owned by players grows over time except for times of war, where player spend large amounts of credits to destroy their enemies. Therefore, players are driven by tangible motivations to make profits, in contrast to models where arbitrary utility functions are assumed to be maximized by virtual agents.

Through these mechanics players find themselves strongly encouraged to take part in the struggles of every day economic life, as known from the real world: collaboration, competition, cartelization, fraud, etc. We plan to analyze the Pardus economy in detail in further works.

---

the developers had to undertake several acts of balancing and fine-tuning to the system to make every dynamic game feature as well-adjusted as possible in terms of equalization. For example, there are three competing factions with different strengths and weaknesses which may be joined by players. Balancing their features or adding new ones to achieve similar levels of attractivity while maximizing incomparability was several years of work. Insofar some tuning had to be done, however the system found its stationary states surprisingly fast every time.

# **Chapter 4**

## **On the statistical nature of social actions and reactions**

Before analyzing the *structure* of human relations and interactions, as expressed by single and interacting socio-economic networks, see subsequent Chapters 6 and 7, respectively, in this Chapter we focus on the statistical nature of sequences of human behavioral actions in the virtual universe of the previously introduced MMOG Pardus. We do so to understand the constituents of the system first, i.e. their statistical properties of performing actions and reactions. In this first approach we do not consider the topology of the networks spanned by these actions.

Sequential behavioral data is available on the scale of an entire society, which is in general impossible to obtain. The unique data of the online game Pardus [2] allows to unambiguously track all actions of all players over long time periods. We focus on the stream of eight types of actions which are translated into an 8-letter alphabet. This *code* of actions of individual players is then analyzed by means of standard timeseries approaches as have been used, for example, in DNA sequence analyzes [177, 201]. The data set considered in this Chapter contains practically all actions of all players of Pardus, over 1,238 days. From all sequences of all 34,055 Artemis players who performed or received an action at least once within 1,238 days, we removed players with a life history of less than 1000 actions, leading to the set of the most active 1,758 players which are considered throughout this Chapter. This Chapter is based on Ref. [208].

### **4.1 Human behavioral sequences**

Using the data set from our online game, it is possible to extract multiple social relationships between a fixed set of humans [203]. We consider eight different actions every player can execute at any time. These are communication (C), trade (T), setting a friendship link (F), removing an enemy link (forgiving) (X), attack (A), placing a bounty on another player (punishment) (B), removing a friendship link (D), and setting an enemy link (E). While C, T, F and X can be associated with

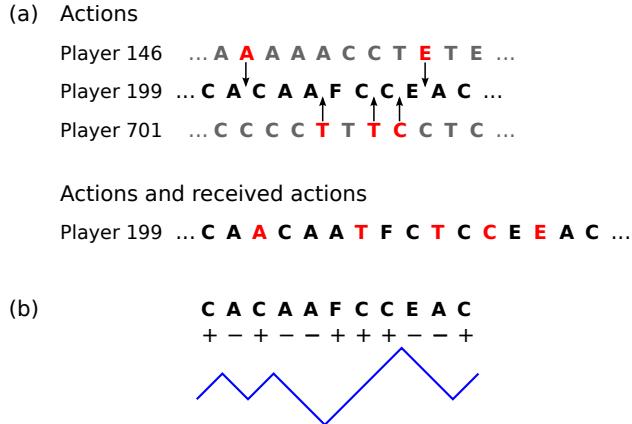


Figure 4.1: (a) Short segment of action sequences of three players,  $A^{146}$ ,  $A^{199}$ , and  $A^{701}$ . Some actions of players 146 and 701 are directed toward player 199. This results in a sequence of received-actions for player 199,  $R^{199} = \{\dots ATTCE \dots\}$ . The combined sequence of actions (originated from - and directed to) player 199,  $C^{199}$ , is shown in the last line; red letters mark actions from others directed to player 199. (b) Schematic illustration showing the definition of a binary walk in ‘good-bad’ action space (good-bad ‘world line’). A positive action (C, T, F or X) means an upward move, a negative action (A, B, D and E) is a downward move. Good people have rising world-lines.

*positive* (good) actions, A, B, D and E are hostile or *negative* (bad) actions. We classify communication as positive because only a negligible part of communication takes place between enemies [205]. Segments of action sequences of three players (146, 199 and 701) are shown in the first three lines of Fig. 4.1 (a).

We consider three types of sequences for any particular player. The first is the stream of  $N$  consecutive actions  $A^i = \{a_n | n = 1, \dots, N\}$  which player  $i$  performs during his ‘life’ in the game. The second sequence is the (time-ordered) stream of actions that player  $i$  receives from all the other players in the game, i.e. all the actions which are directed towards player  $i$ . We denote by  $R^i = \{r_n | n = 1, \dots, M\}$  received-action sequences. Finally, the third sequence is the time-ordered combination of player  $i$ ’s actions and received-actions, which is a chronological sequence from the elements of  $A^i$  and  $R^i$  in the order of occurrence. The combined sequence we denote by  $C^i$ ; its length is  $M + N$ , see also Fig. 4.1 (a). The  $n$ th element of one of these series is denoted by  $A^i(n)$ ,  $R^i(n)$ , or  $C^i(n)$ . We do not consider the actual time between two consecutive actions which can range from milliseconds to weeks, rather we work in ‘action-time’.

If we assign  $+1$  to any positive action C, T, F or X, and  $-1$  to the negative actions A, B, D and E, we can translate a sequence  $A^i$  into a symbolic binary sequence  $A_{\text{bin}}^i$ . From the cumulative sum of this binary sequence a ‘world line’ or ‘random walk’ for player  $i$  can be generated,  $W_{\text{good-bad}}^i(t) = \sum_{n=1}^t A_{\text{bin}}^i(n)$ , see Fig. 4.1 (b). Similarly, we define a binary sequence from the combined sequence  $C^i$ , where we assign  $+1$  to an executed action and  $-1$  to a received-action. This sequence we call  $C_{\text{bin}}^i$ , its cumulative sum,  $W_{\text{act-rec}}^i(t) = \sum_{n=1}^t C_{\text{bin}}^i(n)$  is the ‘action-receive’ random-walk or world line. Finally, we denote the number of actions which occurred during a day

Table 4.1: First row: total number of actions by all players (with at least 1000 actions) in the Artemis universe of the Pardus game. Further rows: first 4 moments of  $r_Y(d)$ , the distribution of the log-increments of the  $N_Y$  processes (see text). Approximate log-normality is indicated. The large values of kurtosis for T and A result from a few extreme outliers.

	D	B	A	E	F	C	T	X
$\sum_{d=1}^{1238} N_Y(d)$	26,471	9,914	558,905	64,444	82,941	5,607,060	393,250	20,165
mean	0.002	0.001	0.004	-0.002	-0.002	0.000	0.003	0.002
std	1.13	0.79	0.54	0.64	0.35	0.12	0.28	0.94
skew	0.12	0.26	0.35	0.08	0.23	0.11	1.00	-0.01
kurtosis	3.35	3.84	6.23	3.67	3.41	3.76	13.89	3.72

in the game by  $N_Y(d)$ , where  $d$  indicates the day and  $Y$  stands for one of the eight actions.

The number of occurrences of the various actions of all players over the entire time period is summarized in Tab. 4.1 (first line). Communication is the most dominant action, followed by attacks and trading which are each about an order of magnitude less frequent. The daily number of all communications, trades and attacks,  $N_C(d)$ ,  $N_T(d)$  and  $N_A(d)$  is shown in Fig. 4.2 (a), (b) and (c), respectively. These processes are reverting around a mean,  $R_Y$ . All processes of actions show an approximate Gaussian statistic of its log-increments,  $r_Y(d) = \log \frac{N_Y(d)}{N_Y(d-1)}$ . The first 4 moments of the  $r_Y$  series are listed in Tab. 4.1. The relatively large kurtosis for  $T$  and  $A$  results from a few extreme outliers. The distribution of log-increments for the  $N_C$ ,  $N_T$  and  $N_A$  timeseries are shown in Fig. 4.2 (d). The lines are Gaussians for the respective mean and standard deviation from Tab. 4.1. As maybe the simplest mean-reverting model with approximate lognormal distributions, we propose

$$N_Y(d) = N_Y(d-1)^{\rho_Y} e^{\xi(d)} R_Y^{(1-\rho_Y)}, \quad (4.1)$$

where  $\rho_Y$  is the mean reversion coefficient,  $\xi(d)$  is a realization of a zero mean Gaussian random number with standard deviation  $\sigma_Y$ , and  $R_Y$  is the value to which the process  $N_Y(t)$  reverts to.  $\sigma$  is given by the third line in Tab. 4.1.

## 4.2 Transition probabilities

With  $p(Y|Z)$  we denote the probability that an action of type  $Y$  follows an action of type  $Z$  in the behavioral sequence of a player.  $Y$  and  $Z$  stand for any of the eight actions, executed or received (received is indicated by a subscript  $r$ ). In Fig. 4.3 (a) the transition probability matrix  $p(Y|Z)$  is shown. The  $y$  axis of the matrix indicates the action (or received-action) happening at a time  $t$ , the probabilities for the actions (or received-actions) that immediately follow are given in the corresponding horizontal place.

This transition matrix specifies to which extent an action or a received action of a player is influenced by the action that was done or received at the previous

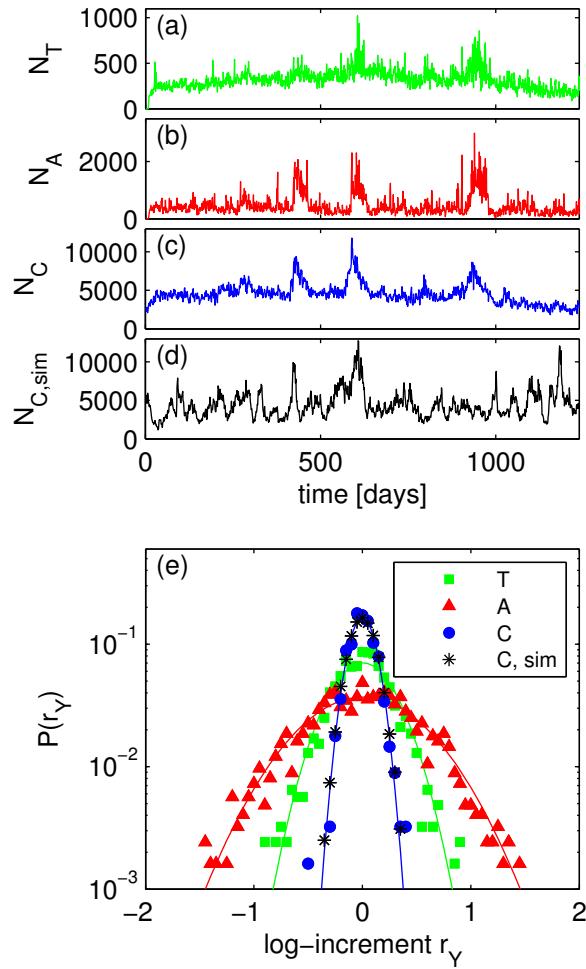


Figure 4.2: Timeseries of the daily number of (a) trades, (b) attacks, (c) communications in the first 1238 days in the game. Clearly a mean reverting tendency of three processes can be seen. (d) Simulation of a model timeseries, Eq. (4.1), with  $\rho = 0.94$ . We use the values from the  $N_C$  timeseries,  $R = 4000$ , and standard deviation  $\sigma = 0.12$ . Compare with the actual  $N_C$  in (c). The only free parameter in the model is  $\rho$ . Parameters are from Tab. 4.1. Mean reversion and log-normality motivate the model presented in Eq. (4.1). (e) The distributions of log-increments  $r_Y$  of the processes and the model. All follow approximate Gaussian distribution functions.

time-step. In fact, if the behavioral sequences of players had no correlations, i.e. the probability of an action, received or executed, is independent of the history of the player's actions, the transition probability  $p(Y|Z)$  simply is  $p(Y)$ , i.e. to the probability that an action or received action  $Y$  occurs in the sequence is determined by its relative frequency only. Therefore, deviations of the ratio  $\frac{p(Y|Z)}{p(Y)}$  from 1 indicate correlations in sequences. In Fig. 4.3 (b) we report the values of  $\frac{p(Y|Z)}{p(Y)}$  for actions and received actions (received actions are indicated with the subscript  $r$ ) classified only according to their positive (+) or negative (-) connotation. In brackets we report the Z-score with respect to the uncorrelated case. We find that the probability to perform

a good action is significantly higher if at the previous time-step a positive action has been received. Similarly, it is more likely that a player is the target of a positive action if at the previous time-step he executed a positive action. Conversely, it is highly unlikely that after a good action, executed or received, a player acts negatively or is the target of a negative action. Instead, in the case a player acts negatively, it is most likely that he will perform another negative action at the following time-step, while it is highly improbable that the following action, executed or received, will be positive. Finally, in the case a negative action is received, it is likely that another negative action will be received at the following time-step, while all other possible actions and received actions are underrepresented. The high statistical significance of the cases  $P(-|-)$  and  $P(-r|-r)$  hints toward a high persistence of negative actions in the players' behavior, see below.

Another finding is obtained by considering only pairs of received actions followed by performed actions. This approach allows to quantify the influence of received actions on the performed actions of players. For these pairs we measure a probability of 0.02 of performing a negative action after a received positive action. This value is significantly lower compared to the probability of 0.10 obtained for randomly reshuffled sequences. Similarly, we measure a probability of 0.27 of performing a negative action after a received negative action. Note that this result is not in contrast with the values in Fig. 4.3 (b), since only pairs made up of received actions and performed actions are taken into account.

## 4.3 World lines

The world lines  $W_{\text{good}-\text{bad}}^i$  of good-bad action sequences are shown in Fig. 4.4 (a), the action-reaction world lines in Fig. 4.4 (b).

As a simple measure to characterize these world lines we define the slope  $k$  of the line connecting the origin of the world line to its end point (last action of the player). A slope of  $k = 1(-1)$  in the good-bad world lines  $W_{\text{good}-\text{bad}}$  indicates that the player performed only positive (negative) actions. The slope  $k^i$  is an approximate measure of ‘altruism’ for player  $i$ . The histogram of the slopes for all players is shown in Fig. 4.4 (b), separated into good (blue) and bad (red) players, i.e. players who have performed more good than bad actions and vice versa. The mean and standard deviation of slopes of good, bad, and all players are  $\bar{k}^{\text{good}} = 0.81 \pm 0.19$ ,  $\bar{k}^{\text{bad}} = -0.40 \pm 0.28$ , and  $\bar{k}^{\text{all}} = 0.76 \pm 0.31$ , respectively. Simulated random walks with the same probability 0.90 of performing a positive action yield a much lower variation,  $\bar{k}^{\text{sim}} = 0.81 \pm 0.01$ , pointing at an inherent heterogeneity of human behavior. For the combined action-received-action world line  $W_{\text{act}-\text{rec}}$  the slope is a measure of how well a person is integrated in her social environment. If  $k = 1$  the person only acts and receives no input, she is ‘isolated’ but dominant. If the slope is  $k = -1$  the person is driven by the actions of others and does never act nor react. The histogram of slopes for all players is shown in Fig. 4.4 (e). Most players are well within the  $\pm 45$  degree cone. Mean and standard deviation of slopes of good, bad,

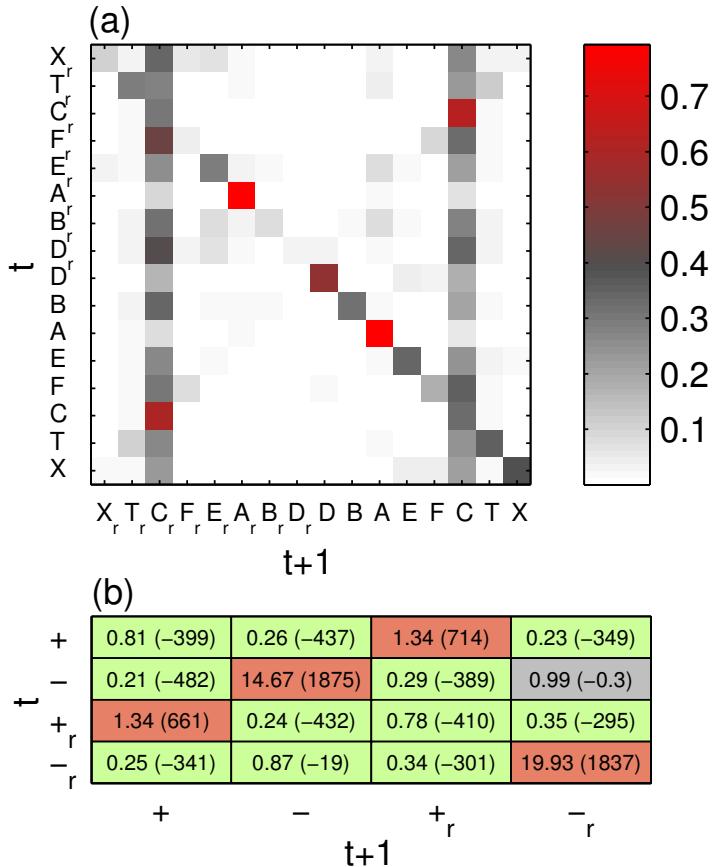


Figure 4.3: (a) Transition probabilities  $p(Y|Z)$  for actions (and received-actions)  $Y$  at a time  $t+1$ , given that a specific action  $Z$  was executed or received in the previous time-step  $t$ . Received-actions are indicated by a subscript  $r$ . Normalization is such that rows add up to one. The large values in the diagonal signal that human actions are highly clustered or repetitive. Large values for  $C \rightarrow C_r$  and  $C_r \rightarrow C$  reveal that communication is a tendentially anti-persistent activity – it is more likely to receive a message after one sent a message and vice versa, than to send or to receive two consecutive messages. (b) The ratio  $\frac{p(Y|Z)}{p(Y)}$ , shows the influence of an action  $Z$  at a previous time-step  $t$  on a following action  $Y$  at a time  $t+1$ , where  $Y$  and  $Z$  can be positive or negative actions, executed or received (received actions are indicated by the subscript  $r$ ). In brackets, we report the  $Z$ -score (significance in number of standard deviations) in respect to a sample of 100 randomized versions of the data set. The cases for which the transition probability is significantly higher (lower) than expected in uncorrelated sequences are highlighted in red (green). Receiving a positive action after performing a positive action is highly overrepresented, and vice versa. Performing (receiving) a negative action after performing (receiving) another negative one is also highly overrepresented. Performing a negative action has no influence on receiving a negative action next. All other combinations are strongly underrepresented, for example after performing a negative action it is very unlikely to perform a positive action with respect to the uncorrelated case.

and all players are  $\bar{k}^{\text{good}} = 0.02 \pm 0.10$ ,  $\bar{k}^{\text{bad}} = 0.30 \pm 0.19$ , and  $\bar{k}^{\text{all}} = 0.04 \pm 0.12$ , respectively. Bad players are tendentially dominant, i.e. they perform significantly more actions than they receive. Simulated random walks with equal probabilities for up and down moves for a sample of the same sequence lengths, we find again a much

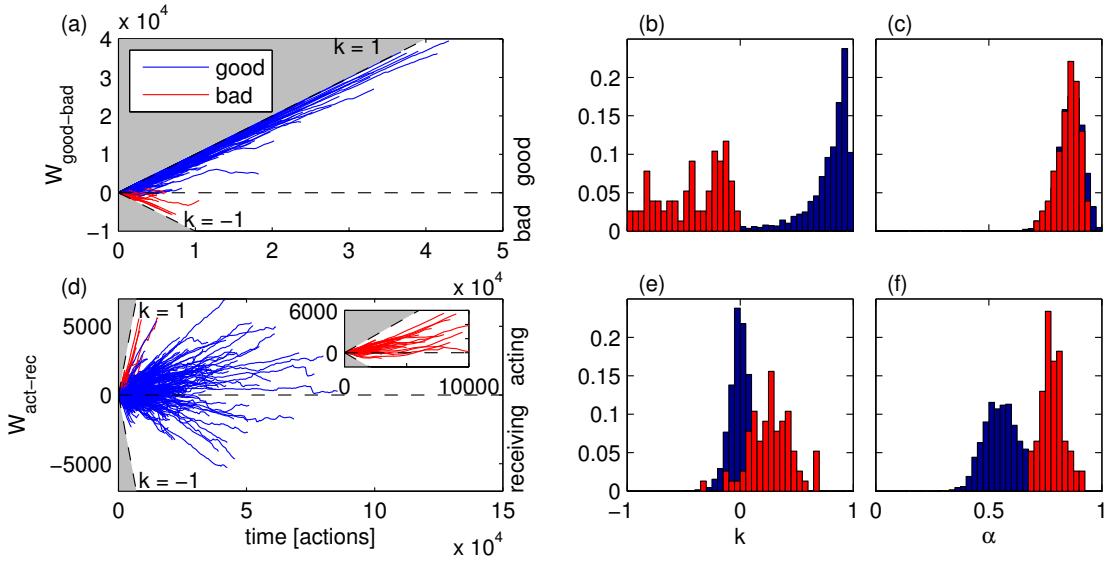


Figure 4.4: (a) World lines of good-bad action random walks of the 1,758 most active players, (b) distribution of their slopes  $k$  and (c) of their scaling exponents  $\alpha$ . By definition, players who perform more good (bad) than bad (good) actions have the endpoints of their world lines above (below) 0 in (a) and only fall into the  $k > 0$  ( $k < 0$ ) category in (b). (d) World lines of action-received random walks, (e) distribution of their slopes  $k$  and (f) of their scaling exponents  $\alpha$ . The inset in (d) shows only the world lines of bad players. These players are typically dominant, i.e. they perform significantly more actions than they receive. In total the players perform many more good than bad actions and are strongly persistent with good as well as with bad behavior, see (c), i.e. actions of the same type are likely to be repeated.

narrower distribution with slope  $\bar{k}^{\text{sim}} = 0.00 \pm 0.01$ .

As a second measure we use the mean square displacement of world lines to quantify the persistence of action sequences. The mean square displacement  $M^2$  of a world line  $W$  is defined as

$$M^2(\tau) = \langle (\Delta W(\tau) - \langle \Delta W(\tau) \rangle)^2 \rangle, \quad (4.2)$$

where  $\Delta W(\tau) \equiv W(t + \tau) - W(t)$  and  $\langle \cdot \rangle$  is the average over all  $t$ . The asymptotic behavior of  $M(\tau)$  yields information about the ‘persistence’ of a world line.  $M(\tau) \propto \tau^{\frac{1}{2}}$  is the pure diffusion case,  $M(\tau) \propto \tau^\alpha$  with scaling exponent  $\alpha \neq \frac{1}{2}$  indicates persistence for  $\alpha > \frac{1}{2}$ , and anti-persistence for  $\alpha < \frac{1}{2}$ . Persistence means that the probability of making an up(down) move at time  $t + 1$  is larger(less) than  $p = \frac{1}{2}$ , if the move at time  $t$  was an up move. For calculating the exponents  $\alpha$  we use a fit range of  $\tau$  between 5 and 100. We checked from the mean square displacements of single world lines that this fit range is indeed reasonable.

The histogram of exponents  $\alpha$  for the good-bad random walk, separated into good (blue) and bad (red) players, is shown in Fig. 4.4 (c), for the action-received-action world line in (f). In the first case strongly persistent behavior is obvious, in the second there is a slight tendency towards persistence. Mean and standard deviation for the good-bad world lines are  $\alpha_{\text{good-bad}} = 0.87 \pm 0.06$ , for the action-

received actions  $\alpha_{\text{act-rec}} = 0.59 \pm 0.10$ . Simulated sequences of random walks have – as expected by definition – an exponent of  $\alpha_{\text{rnd}} = 0.5$ , again with a very small standard deviation of about 0.02. Figure 4.4 (a) also indicates that the lifetime of players who use negative actions frequently is short. The average lifetime for players with a slope  $k < 0$  is  $2528 \pm 1856$  actions, compared to players with a slope  $k > 0$  with  $3909 \pm 4559$  actions. The average lifetime of the whole sample of players is  $3849 \pm 4484$  actions.

## 4.4 Motifs, entropy and Zipf law

By considering all the sequences of actions  $A^i$  of all possible players  $i$ , we have an ensemble which allows to perform a motif analysis [198]. We define an  $n$ -string as a subsequence of  $n$  contiguous actions. An  $n$ -motif is an  $n$ -string which appears in the sequences with a probability higher than expected, after lower-order correlations have been properly removed. In particular, we define  $n$ -strings as a subsequence of  $n$  contiguous actions. With an alphabet size of 8 (number of action types), across the entire ensemble  $8^n$  different  $n$ -strings can appear, each of them occurring with a different probability. The frequency, or observed probability, of each  $n$ -string can be compared to its expected probability of occurrence, which can be estimated on the basis of the observed probability of lower order strings, i.e. on the frequency of  $(n - 1)$ -strings. For example, the expected probability of occurrence of a 2-string  $(A_t, A_{t+1})$  is estimated as the product of the observed probability of the single actions  $A_t$  and  $A_{t+1}$ ,

$$p^{\text{exp}}(A_t, A_{t+1}) = p^{\text{obs}}(A_t)p^{\text{obs}}(A_{t+1}). \quad (4.3)$$

Similarly, the probability of a 3-string  $(A_t, A_{t+1}, A_{t+2})$  to occur can be estimated as

$$p^{\text{exp}}(A_t, A_{t+1}, A_{t+2}) = p^{\text{obs}}(A_t, A_{t+1})p^{\text{obs}}(A_{t+2}|A_{t+1}), \quad (4.4)$$

where  $p^{\text{obs}}(A_{t+2}|A_{t+1})$  is the conditional probability to have action  $A_{t+2}$  following action  $A_{t+1}$ . By definition of conditional probability, one has

$$p^{\text{obs}}(A_{t+2}|A_{t+1}) = \frac{p^{\text{obs}}(A_{t+1}, A_{t+2})}{p^{\text{obs}}(A_{t+1})} \quad (4.5)$$

(see [198] for details). An  $n$ -motif in the ensemble is then defined as a  $n$ -string whose observed probability of occurrence is significantly higher than its expected probability. This definition of motif is not to be confused with *network motifs*, i.e. overrepresented configurations of subgraphs such as certain triad classes [159].

We computed the observed and expected probabilities  $p^{\text{obs}}$  and  $p^{\text{exp}}$  for all  $8^2 = 64$  2-strings and for all  $8^3 = 512$  3-strings, focusing on those  $n$ -strings with the highest ratio  $\frac{p^{\text{obs}}}{p^{\text{exp}}}$ . Higher orders are statistically not feasible due to combinatorial explosion. We find that the 2-motifs in the sequences of actions  $A$  are clusters of same letters: BB, DD, XX, EE, FF, AA with ratios  $\frac{p^{\text{(obs)}}}{p^{\text{(exp)}}} \approx 169, 136, 117, 31, 15, 10$ ,

respectively. This observation is consistent with the previous first-order observation that actions cluster. The most significant 3-motifs however are (with two exceptions) the palindromes: EAX, DAF, DCD, DAD, BGB, BFB, with ratios  $\frac{p(\text{obs})}{p(\text{exp})} \approx 123, 104, 74, 62, 33, 32$ , respectively. The exceptions disappear when one considers actions executed on the same screen in the game as equivalent, i.e. setting or removing friends or enemies: F, D, E, X. This observation hints towards processes where single actions of one type tend to disrupt a flow of actions of another type.

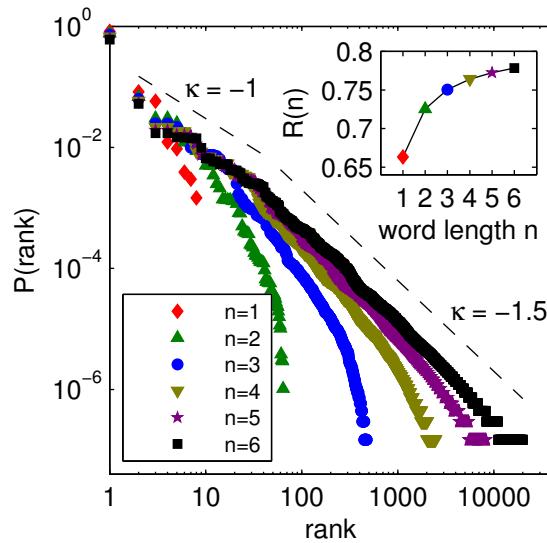


Figure 4.5: Rank ordered probability distribution of 1 to 6 letter words. Slopes of  $\kappa = -1$  and  $\kappa = -1.5$  are indicated for reference. The inset shows the Shannon  $n$ -tuple redundancy as a function of word length  $n$ .

Finally, we partition the action sequences into  $n$ -strings ('words'). Fig. 4.5 shows the rank distribution of word occurrences of different lengths  $n$ . The distribution shows an approximate Zipf law [226] (slope of  $\kappa = -1$ ) for ranks below 100. For ranks between 100 and 25,000 the scaling exponent approaches a smaller value of about  $\kappa \sim -1.5$ . The Shannon  $n$ -tuple redundancy (see e.g. [177, 201]) for symbol sequences composed of 8 symbols (our action types) is defined as

$$R^{(n)} = 1 + \frac{1}{3n} \sum_i^{8^n} P_i^{(n)} \log_2 P_i^{(n)}, \quad (4.6)$$

where  $P_i^{(n)}$  is the probability of finding a specific  $n$ -letter word. Uncorrelated sequences yield an equi-distribution,  $P_i = 8^{-n}$ , i.e.  $R^{(n)} = 0$ . In the other extreme of only one letter being used,  $R^{(n)} = 1$ . In Fig. 4.5 (inset)  $R^{(n)}$  is shown as a function of sequence length  $n$ . Shannon redundancy is not a constant but increases with  $n$ . This indicates that Boltzmann-Gibbs entropy might not be an extensive quantity for action sequences [102].

## 4.5 Discussion of behavioral sequences

The analysis of human behavioral sequences as recorded in a massive multiplayer online game shows that communication is by far the most dominant activity followed by aggression and trade. Communication events are about an order of magnitude more frequent than attacks and trading events, showing the importance of information exchange between humans. It is possible to understand the collective timeseries of human actions of a particular type ( $N_Y$ ) with a simple mean-reverting log-normal model. On the individual level we are able to identify organizational patterns of the emergence of good overall behavior. Transition rates of actions of individuals show that positive actions strongly induces positive *reactions*. Negative behavior on the other hand has a high tendency of being repeated instead of being reciprocated, showing the ‘propulsive’ nature of negative actions. However, if we consider only reactions to negative actions, we find that negative reactions are highly overrepresented. The probability of acting out negative actions is about 10 times higher if a person received a negative action at the previous timestep than if she received a positive action. The action of communication is found to be of highly reciprocal ‘back-and-forth’ nature. The analysis of binary timeseries of players (good-bad) shows that the behavior of almost all players is ‘good’ almost all the time. Negative actions are balanced to a large extent by good ones. Players with a high fraction of negative actions tend to have a significantly shorter life. This may be due to two reasons: First because they are hunted down by others and give up playing, second because they are unable to maintain a social life and quit the game because of loneliness or frustration. We interpret these findings as empirical evidence for self organization towards reciprocal, good conduct within a human society. Note that the game allows bad behavior in the same way as good behavior but the extent of punishment of bad behavior is freely decided by the players.

Behavior is highly persistent in terms of good and bad, as seen in the scaling exponent ( $\alpha \sim 0.87$ ) of the mean square displacement of the good-bad world lines. This high persistence means that good and bad actions are carried out in clusters. Similarly high levels of persistence were found in a recent study of human behavior [81]. A smaller exponent ( $\alpha \sim 0.59$ ) is found for the action–received-action timeseries.

Finally we split behavioral sequences of individuals into subsequences (of length 1-6) and interpret these as behavioral ‘words’. In the ranking distribution of these words we find a Zipf law to about ranks of 100. For less frequent words the exponent in the rank distribution approaches a somewhat smaller exponent of about  $\kappa \sim -1.5$ . From word occurrence probabilities we further compute the Shannon  $n$ -tuple redundancy which yields relatively large values when compared for example to those of DNA sequences [177, 201]. This reflects the dominance of communication over all the other actions. The  $n$ -tuple redundancy is clearly not a constant, reflecting again the non-trivial statistical structure of behavioral sequences.

After having analyzed and characterized the nature of actions of the individuals, we consider their motions and diffusive properties through their universe in the following Chapter 5, before we shift our viewpoint to the topology of *interaction*

*networks* in Chapter 6, and to the correlations *between* these structures, Chapter 7.

# **Chapter 5**

## **Mobility in a society: Diffusion, long-time memory, and borders**

Understanding the statistical patterns of human mobility, predicting trajectories and uncovering the mechanisms behind human movements [21] is a considerable challenge with important practical applications to traffic management [98, 108], planning of urban spaces [149, 190], epidemiology [15, 55, 120, 176], information spreading [160, 173], and geo-marketing [122, 178]. In the last years, advanced digital technologies have provided huge amounts of data on human activities, allowing to extract information on human movements. For instance, observations of banknote circulation [39, 206], mobile phone records [95], online location-based social networks [193, 194], GPS location data of vehicles [25], or radio frequency identification traces [21, 48, 190], have been used as proxies for human movements. These studies have provided insights into several aspects of human mobility, uncovering distinct features of human travel behavior such as scaling laws [39, 199], predictability of trajectories [200], and impact of motion on disease spreading [15, 26, 55, 120]. However, from a comparative analysis of the different works it emerges clearly that a “unified theory” of human mobility is still outstanding, since results, even on some very basic features of the motion, often appear to be contrasting [21]. One example is the measured distribution of human trip lengths in various types of transportation: some studies agree that mobility is generally characterized by fat-tailed distributions of trip lengths [39, 199], while others report exponential or binomial forms [21, 25, 190]. The discrepancies arise due to the different mobility data sets used, where mobility is indirectly inferred from some specific human activity in a particular context. For instance, mobile phone records typically provide location information only when a person uses the phone [199], while radio frequency identification traces like the ones of Oyster cards<sup>1</sup> in the London subway [190] only log movements based on public transportation systems. Analyses of these data sets can then result in a possibly biased view of the underlying mobility processes. Furthermore, most of the analyzed data sets have poor information on how socio-economic factors influence human mobility

---

<sup>1</sup>Oyster cards are a form of electronic ticketing used on public transport services within the London region of England.

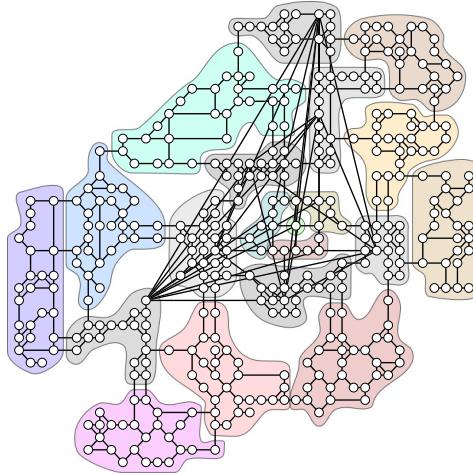


Figure 5.1: The universe map of the massive multiplayer online game *Pardus*. The universe can be represented as a network with  $N = 400$  nodes, called sectors (playing the role of cities), and  $L = 1160$  links. Sectors are organized into 20 different regions, called clusters, shown in the figure as different color-shaded areas. There is no explicit set of goals in the game. Players are free to interact in a number of ways to e.g. increase their virtual wealth or status. Players move between sectors to interact with other players, e.g. to trade, attack, wage war, or to explore the virtual world.

patterns. More generally, the lack of an all-encompassing record set with positional raw data including complete information on the socio-economic context and on the behavior of all members of a human society, has so far limited the possibilities for a comprehensive exploration of human mobility.

Using positional data of the players in the game universe<sup>2</sup>, in combination with other socio-economic information from the game, we uncover various fundamental features of mobility, and we provide a possible description of the mechanisms causing the observed anomalous diffusion. First, we find the emergence of different spatial scales, due to the strong tendency of the players to limit their economic activities to some specific areas over long time periods and to avoid crossing the borders between different areas. Making use of this observation, we propose an efficient method to identify socio-economic regions by means of community detection algorithms based solely on the measured movement dynamics. Our second result unveils the driving mechanism behind the movement patterns of players: Locations are visited in a specific order, leading to strong long-term memory effects which are essential to understand and reproduce the observed trajectories. Finally, we provide large-scale evidence that neglecting either of these spatial or temporal constraints may obstruct the possibility of understanding the processes behind human mobility.

---

<sup>2</sup>Whenever we address the position or the movement of ‘a player’ in the game universe, this is meant as a short form for referring to the virtual avatar which is uniquely associated to and controlled by the player. This abbreviation is consistent with the tendency of players to identify themselves with their avatars.

## 5.1 Game universe properties

The topology of the universe can be represented as a network with 400 nodes, called *sectors*, embedded in a two-dimensional space, the so-called *universe map* shown in Fig. 5.1. Each sector is like a city where players can have social relations (establish new friendships, make enemies and wage wars), and entertain economic activities (trade and production of commodities), see Section 3.6.2. Typically, sectors adjacent on the universe map, as well as a few far-apart sectors, are interconnected by links which allow players to move from sector to sector. At any point in time, each sector is typically attended by a large number of players. The network is sparse and, similarly to other spatial networks, is not a small world, see Section 2.3.4. It has a characteristic path length  $\bar{g} = 11.89$  and a diameter  $g_{\max} = 27$ , which means that, on average, players have to move through a non-negligible number of sectors to traverse the universe.

The universe network is sparse, displaying an average degree of  $\bar{k} = 2.9$  with  $N = 400$  nodes and  $L = 1160$  links. It is highly clustered (clustering coefficient  $C = 0.089$ , which is 12 times larger than  $C_r$ , the average clustering coefficient for corresponding random networks, i.e. networks having the same number of nodes and the same average degree as the universe network), but is not small-world [33]: the characteristic path length  $d_{\max} = 11.89$  is more than twice  $L_r$ , the average path length on a random graph with the same number of nodes [68] and the same average connectivity, and relative to the number of nodes the diameter  $g_{\max} = 27$  is large. This means that, on average, players have to move through a non-negligible number of sectors to traverse the universe. However, in a rough estimation, it takes only around 1,500 APs to traverse a typical cluster, and approximately 10,000 (around 2 days worth of APs) to traverse the whole universe. With a few exceptions, the universe is a planar network, i.e. it has no intersecting links. In this sense it is similar to Euclidean networks such as street networks, with the difference that the length of links has no effect in Pardus. Locally, the universe is lattice-like, which is reflected in the relatively high local efficiency  $E_{\text{loc}} = 0.80$  [139] and the narrow degree distribution (almost all sectors have degree 2, 3, or 4).

Besides the clusters, which are pre-defined, there is the possibility for players to build obstacles called ‘military outposts’ blocking almost any region of space from being entered by single or groups of other players. This option adds player-driven ‘political borders’ to the natural ones, just as borders of nations often constitute significant limitations to mobility of human beings. Such player-built obstacles can be taken down by other players through the use of (usually organized) hostile force. Considering the political dynamics of military outposts is beyond the scope of this work. Note, however, that often groups of players construct military outposts at the borders of clusters, possibly further adding to the mobility-hindering effect of clusters (see below).

The sectors have been originally organized by the developers of the game into 20 different *clusters*, which are perceived by the players as different political or socio-economic regions such as countries. For example, a player who is member of a

---

$N$	400
$L$	1160
$\bar{k}$	2.9
$C$	0.089
$C/C_r$	12.33
$L$	11.89
$\bar{g}/\bar{g}_r$	2.11
$E_{loc}$	0.80
$E_{glob}$	0.03
$g_{max}$	27

Table 5.1: Network properties of the Pardus universe: number of nodes  $N$ , number of (undirected) links  $L$ , average degree  $\bar{k}$ , clustering coefficient  $C$ , clustering coefficient to corresponding coefficient of random graph  $C/C_r$ , average geodesic  $L$ , average geodesic to corresponding average geodesic of random graph  $\bar{g}/\bar{g}_r$ , local efficiency  $E_{loc}$ , global efficiency  $E_{glob}$ , diameter  $g_{max}$ .

political faction in the game is provided some game-relevant protection in all clusters which are controlled by the faction, and has the opportunity of social promotion when accomplishing certain tasks within these clusters. Each cluster is shown in Fig. 5.1 with a different background color. All clusters contain about 20 sectors each, with the exception of the central cluster, consisting of just one sector, and its surrounding three clusters having only 6-7 sectors. Sectors belonging to the same cluster are geographically close on the map, meaning that the distance between any two sectors in the same cluster is small, with an average distance around 3. Players typically have a “home cluster” where they focus their socio-economic activities over long time periods. Occasionally, they also move to sectors belonging to other clusters in order to explore the universe, to relocate their home (migrate), or during extreme game events such as wars.

## 5.2 Data set description

As in the other Chapters of this work, we focus on the *Artemis* game universe. To make sure we only consider active players, we select all players who exist in the game between the days 200 and 1,200. We discard the first 200 days because social networks between players of *Pardus* have shown aging effects in the beginning of the universe, i.e. there seems to exists a transient phase in the development of the society, possibly affecting mobility, which we would like to avoid considering [205]. This cut selects 1,458 players active over a time-period of 1,000 days. The field-ID (position within a sector) of these players is logged every day at 05:35 GMT. From these field-IDs, we select the corresponding sector-IDs, i.e. the player positions on the nodes of the universe network, leaving us with a  $1,458 \times 1,000$  mobility matrix containing 1,458 sequences each consisting of 1,000 sector-IDs. Note that we use daily snapshots of position data since we are interested in the long-time mobility

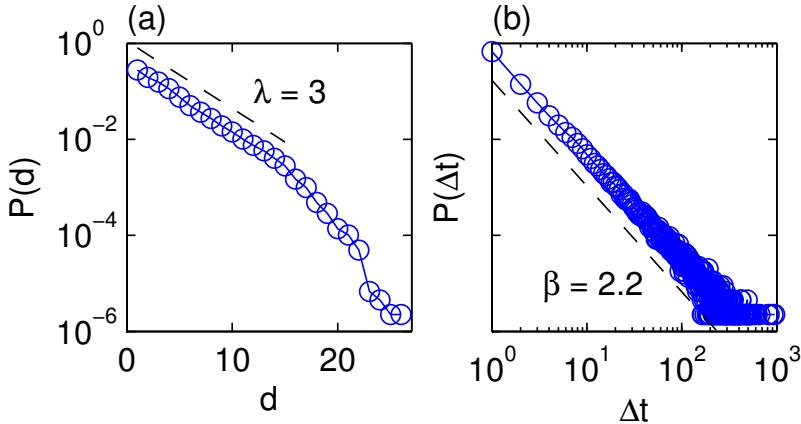


Figure 5.2: Distribution of jump distances and of waiting times. With each player a time series consisting of the sector positions over 1000 days is associated. A *jump* is said to occur when the sector position in the time series changes from one day to the following. The length  $d$  of a jump is measured in terms of graph distance and can take an integer value between 1 and  $g_{\max} = 27$ , the diameter of the network. (a) The probability distribution of jump distances is reported in a semi-log plot. For  $d \leq 15$ , the distribution follows an exponential  $P(d) \sim e^{-\frac{d}{\lambda}}$  with a characteristic length  $\lambda \approx 3$ . Players can also remain in the same sector for several days, without moving to other sectors. We define as waiting time  $\Delta t$  the number of consecutive days a player spends in one sector. (b) We show the probability distribution of waiting times  $\Delta t$  in a log-log plot, which is well fitted by a power-law  $P(\Delta t) \sim \Delta t^{-\beta}$ , with  $\beta \approx 2.2$ .

and not in the detailed paths the players take during the (typically) few minutes of their daily navigations. The effects of this ‘systematic coarse-graining’ seem modest compared to biases occurring for example in mobile phone data, where location information is usually only available at the heterogeneously distributed times when a mobile phone is used, and locations, hence distances, have to be triangulated via heterogeneously scattered mobile phone towers [200]. These issues inherent in previously studied mobile phone data may become resolved in future studies, considering recent disclosures that exact location data (GPS) is constantly recorded by a number of mobile phone devices [1].

### 5.3 Basic features of motion

A *jump* occurs when a player’s sector position changes from one day to the following. The associated length  $d$  of a jump is measured in terms of graph distance, an integer value between 1 and  $g_{\max} = 27$ . The probability distribution of jump distances, computed for all players over the whole observation period, is reported in Figure 5.2 (a). For  $d \leq 15$ , the distribution is well-fitted by an exponential:

$$P(d) \sim e^{-\frac{d}{\lambda}}, \quad (5.1)$$

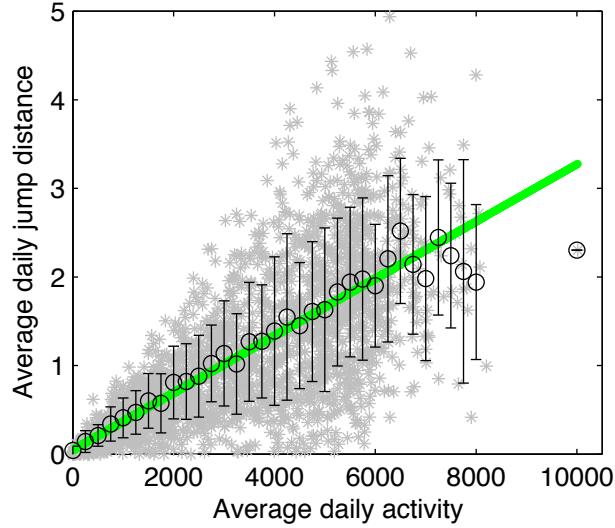


Figure 5.3: Average daily jump distance versus average daily activity as measured in Action Points (APs) spent, for each player. Circles and error bars are binned averages with standard deviations. The green line is a least-squares fit, making a trend visible between mobility and activity. High activity is necessary, but not sufficient, for high mobility: in order to have a high mobility, a player needs to spend many APs (there are no data points in the top left part), but there exists a number of players with low mobility who spend a high number of APs (there are data points in the lower right part).

with a characteristic jump length  $\lambda \approx 3$ . The existence of a typical travel distance, as also recently found in other mobility data [25, 190], is maybe related to the use of a single transportation mode in *Pardus* [133]. This allows to disentangle the intrinsic heterogeneity of the players from the effects due to the presence of different means of transportation [15], which might be the cause of the scale-free distributions found in mobile phone or other mobility data sets [39, 95]. It has in fact been suggested that power laws in distance distributions of movement data may emerge from the coexistence of different scales [21, 101].

In some cases, players stay in the same sector for a number of consecutive days. For instance, 11 of the 1,458 considered players, although being active in the game, never jump within the entire observation period. On average, a player does not change sector in approximately 75% of the days. To better characterize the motion, we computed the waiting times  $\Delta t$  (measured in terms of number of days) between all pairs of consecutive jumps, over all players. The distribution of these waiting times, shown in Fig. 5.2 (b) follows a power-law distribution

$$P(\Delta t) \sim \Delta t^{-\beta}, \quad (5.2)$$

with an exponent  $\beta \approx 2.2$ , in agreement with other recent measurements on human dynamics [17], where a slope of  $\beta \approx 2$  was reported. In addition, we found that the average waiting times of individual players are distributed as a power-law. This implies a strong heterogeneity in the motion of different players, which is related to

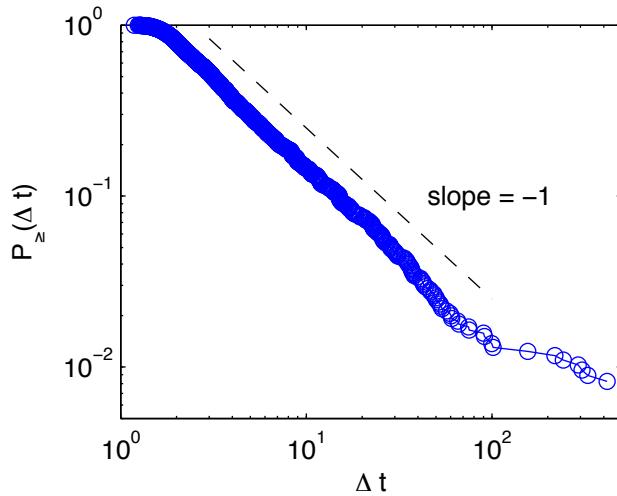


Figure 5.4: Cumulative distribution of average waiting times  $\overline{\Delta t}$  of single players. The dashed line is a power law with slope  $-1$ , which fits well the distribution. The fat-tailed nature of this distribution shows the heterogeneity in movement behaviour of players.

the heterogeneity in their general activity.

In fact, Fig. 5.3 shows the average daily jump distance versus the average daily number of APs spent for each player. The number of APs used per day is distributed heterogeneously, i.e. some players are much more active than others. In general, the more APs a player uses, the more she travels. However, high activity is necessary but not sufficient for high mobility: in order to travel far, a player needs to spend many APs (corresponding to a lack of data points in the top left part of the figure), but there exists a number of players with low mobility who spend a high number of APs (there exist data points in the lower right part of the figure).

Another quantity which indicates the extent of activity in mobility is the average waiting time  $\overline{\Delta t}$  of a single player, i.e. how many days a player stays on a sector on average. The cumulative distribution of these average waiting times of all players, Fig. 5.4, displays a power law with slope  $-1$ , showing further the heterogeneity of player activity with respect to mobility, motivating the transformation from real-time  $t$  to jump-time  $\tau$  for determining the distribution of first return times  $P_{\leftarrow}(\tau)$  (see below).

## 5.4 Mobility reveals socio-economic clusters

Mobility patterns are influenced by the presence of the socio-economic regions in the network, highlighted in colors in Fig. 5.1. The typical situation is illustrated in Fig. 5.5 (a), with jumps within the same cluster being preferred to jumps between sectors in different clusters. In order to quantify this effect, we report in Fig. 5.5 (b), blue circles, the observed number of jumps of length  $d$  within the same cluster, divided by the total number of jumps of length  $d$ . This ratio is a decreasing function of the distance  $d$ , and reaches zero at  $d = 12$ , since no sectors at such distance do

belong to the same cluster. As a null model we report the fraction of sector pairs at distance  $d$  which belong to the same cluster, see red squares in the same figure. The significant discrepancy between the two curves indicates that players indeed tend to avoid crossing the borders between clusters. For example, a jump of length  $d = 8$  from one sector to another sector in the same cluster is expected only in 3% of the cases, while it is observed in about 20% of the cases.

Now, the propensity of a player to spend long time periods within the same cluster might be simply related to the topology of the network, as in the case of random walkers whose motions are constrained on graphs with strong community structures [82]. Nodes belonging to the same cluster are in fact either directly connected or are at short distance from one another. This proximity is reflected in the block-diagonal structure of the adjacency matrix  $A$  and of the distance matrix  $D$ , respectively shown in Fig. 5.6 (a) and (b). We have therefore checked whether the presence of the socio-economic clusters originally introduced by the developers of the game can be derived solely from the structure of the network. For this reason we adopted standard community detection methods based on the adjacency and on the distance matrix [11, 166]. The results, reported respectively in Fig. 5.6 (d) and (e), show that detected communities deviate significantly from the clusters, implying that in our online world the socio-economic regions cannot be recovered merely from topological features. In comparison we considered the player transition count matrix  $M$ , shown in Fig. 5.6 (c), which displays a similar block-diagonal structure as  $A$  and  $D$ , but with the qualitative difference that it contains *dynamic* information on the system. The entry  $m_{ij}$  of the transition count matrix  $M$  is equal to the number of times a player's position was on sector  $i$  and then, on the following day, on sector  $j$ . This number is cumulated for all players. The entry  $\pi_{ij}$  of the transition probability matrix  $\Pi$  corresponds to the probability that a player moves to a sector  $j$  given that on the previous day the player's location was sector  $i$ . It reads:  $\pi_{ij} = \frac{m_{ij}}{\sum_l m_{il}}$ , where  $m_{ij}$  is the number of observed player movements from sector  $i$  to sector  $j$ , and the sum over  $l$  is over all sectors of the universe. The matrix  $\Pi$  has the property that the entries of each row sum to one. Figure 5.6 (f) shows that community detection methods applied to the transition count matrix  $M$  reveal almost perfectly all the socio-economic areas of the universe. This finding demonstrates that mobility patterns contain fundamental information on the socio-economic constraints present in a social system. Therefore, a community detection algorithm applied to raw mobility information, as the one proposed here, is able to extract the underlying socio-economic features, which are invisible to methods based solely on topology.

In the remainder of this section, we give a detailed treatment of adopted community detection methods and measures. Since the Pardus universe is divided into a number of *clusters*, defining political, independent units in the game universe, we have shown that the mobility patterns of players are influenced by such borders. At the same time, the topology of the Pardus universe itself might affect the mobility patterns. In order to investigate the importance of these two elements, one needs to compare the topological modules that can be extracted from the adjacency matrix  $A$

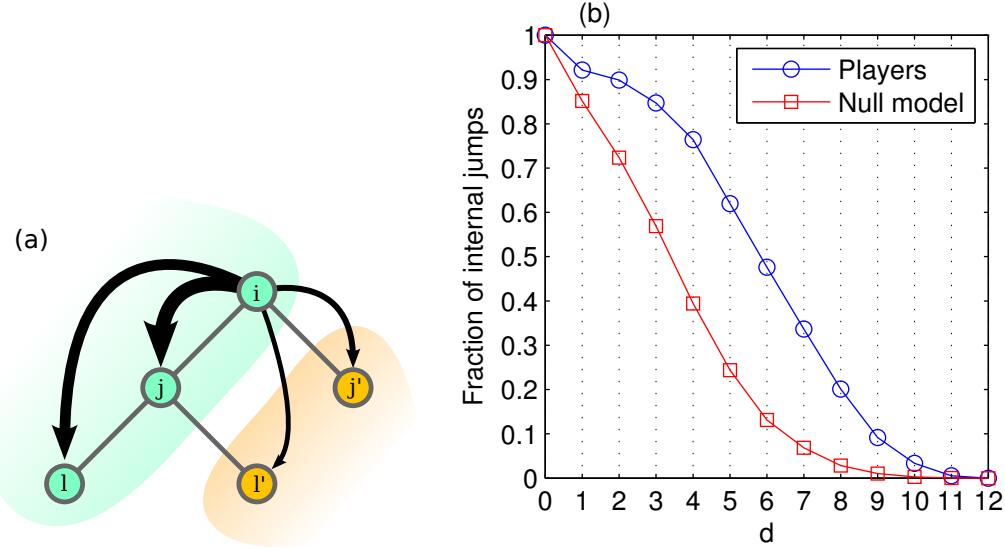


Figure 5.5: Influence of socio-economic clusters on mobility. (a) Sketch of jump patterns from a sector  $i$  to sectors within the same cluster,  $j$  and  $l$ , and to sectors in a different cluster,  $j'$ ,  $l'$ . Although sectors  $j'$  and  $l'$  have the same graph distance from sector  $i$  as sectors  $j$  and  $l$  respectively, transitions across cluster border have smaller probabilities. (b) Quantitative evidence of the tendency of players to avoid crossing borders. Red squares show the null model, i.e. the fraction of all pairs of sectors at a given distance  $d$  being in the same cluster. Blue circles show the fraction of measured jumps leading into the same cluster, per distance. Coincidence of the two curves would indicate that clusters have no effect on mobility. Clearly this is not the case – there is a strong tendency of players to avoid crossing the borders between clusters.

or distance matrix  $D$ , with the dynamical communities emerging from the collective movement behaviour of players.

At the sector level, the Pardus universe is a directed weighted network with  $L = 1,160$  links and  $N = 400$  nodes. The majority of links are *wormholes* ( $\sim 95\%$ ), mutual links that connect nearby nodes (see Fig. 5.1) and have a small traveling cost (in terms of APs). The long-range links in Fig. 5.1 instead represent *X-holes* and *Y-holes*. Players moving along such links incur significantly higher traveling costs than in the case of wormholes. Since X/Y-holes may be only scarcely used in-game, in addition to studying the complete directed weighted adjacency matrix,  $A$ , we also study the adjacency matrix  $A^{\text{wh}}$  where X/Y-holes were removed, yielding a symmetric and unweighted network. Finally, we consider the weighted network  $D^{\text{inv}}$ , defined element-wise as  $d_{ij} = d(i,j)^{-1} \forall i \neq j$  and  $d_{ii} = 0 \forall i$ , where  $d(i,j)$  is the shortest path distance on the Pardus network.

The player dynamics was studied at the aggregate level through the transition count matrix  $M$  and the normalized transition matrix  $\Pi$ . Each element  $m_{ij}$  of  $M$  corresponds to the total number of times a player was found at position  $i$  at time  $t$  and at position  $j$  at time  $t+1$ . The transition matrix  $\Pi = (\pi_{ij})$  is obtained by row-normalizing  $M$  so that  $\pi_{ij} = \frac{m_{ij}}{\sum_l m_{il}}$ . Hence, for all rows  $i$ ,  $\sum_j \pi_{ij} = 1$  and  $\Pi$  is a well-defined probability matrix for the transitions between pair of nodes in the

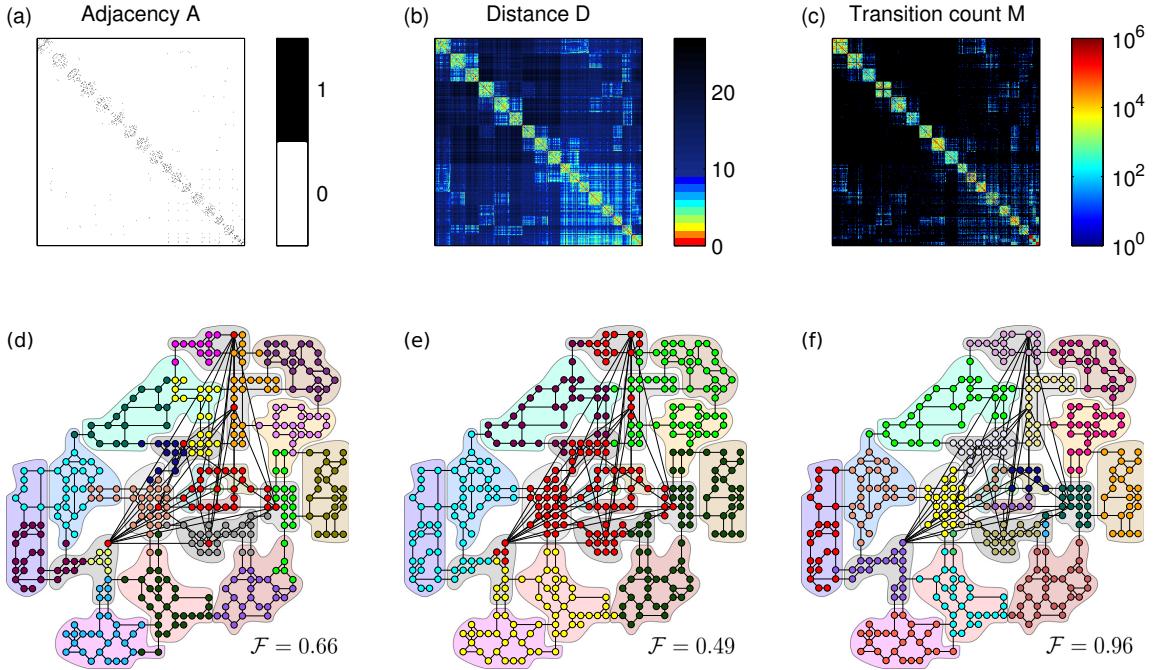


Figure 5.6: Extracting communities from network topology and from mobility patterns. (a) The adjacency matrix  $A$  of the universe network, (b) the matrix  $D$  of shortest path distances, and (c) the matrix  $M$  of transition counts of player jumps. Each of the three matrices contains  $400 \times 400$  entries, whose values are color-coded. Sector IDs are ordered by cluster, resulting in the block-diagonal form of the three matrices. We have used modularity-optimization algorithms to extract community structures from the information encoded in the three matrices. Different node colors represent the different communities found, while the 20 different color-shaded areas indicate the predefined socio-economic clusters as in Fig. 5.1. The displayed Fowlkes and Mallows index  $\mathcal{F} \in [0, 1]$  quantifies the overlap of the detected communities with the predefined clusters. The closer  $\mathcal{F}$  is to 1, the better the match, see the detailed treatment of measures in the text. (d) Although information contained in the adjacency matrix  $A$  allows to find 18 communities, a number close to the real number of clusters, the communities extracted do not correspond to the underlying color-shades areas ( $\mathcal{F} = 0.66$ ). (e) Extracting communities from the distance matrix  $D$  only results in 6 different groups ( $\mathcal{F} = 0.49$ ). (f) The 23 communities detected using the transition count matrix  $M$  reproduce almost perfectly the real socio-economic clusters ( $\mathcal{F} = 0.96$ ), with only a few mismatched nodes detected as additional clusters.

network. Note that for both  $M$  and  $\Pi$  the diagonal elements can be significantly different from zero and therefore the resulting networks display self-loops. Moreover, both matrices  $M$  and  $\Pi$  correspond to directed, weighted networks, and therefore can be thought as representing flows across the networks. For completeness, we also define the symmetrized versions of the matrices above, namely the symmetrized jump matrix  $M^{\text{symm}} = (M + M^T)/2$ ,  $\Pi$  and the symmetrized transition matrix  $\Pi^{\text{symm}} = (\Pi + \Pi^T)/2$ . The corresponding weighted networks are undirected and represent a first coarse-graining of the information contained in the dynamical flows. It is thus interesting to compare these two to understand how much information is lost in the coarse-graining.

We performed community detection algorithms by optimizing modularity [11, 166]. To ensure consistency, we checked the results under different heuristics and repeated detections [130, 184]. Figure 5.7 shows the communities extracted from the  $M$ ,  $M^{\text{symm}}$ ,  $\Pi$  and  $\Pi^{\text{symm}}$  matrices. The colored hulls are included for comparison and indicate the Pardus cluster to which each sector belongs. For comparison, in Fig. 5.8 we plot the communities obtained from the topological quantities, namely the directed weighted adjacency matrix  $A$ , the undirected unweighted matrix  $A^{\text{wh}}$  and the inverse distance matrix  $D^{\text{inv}}$ . One can easily see that the communities extracted from the transition matrices appear to reproduce much better the cluster structure as opposed to the topological communities.

Note also that the partitions obtained for the dynamical transition matrices contain communities composed of a single node. Although unusual in community detection, this result is consistent with the mobility patterns. In fact, we measure the positions of players at the same time every day. Then, the presence of non zero values on the diagonal of  $M$ ,  $M^{\text{symm}}$ ,  $\Pi$  and  $\Pi^{\text{symm}}$  simply means that there is a positive probability for a player to be found again on the same node after 24 hours, implying that the player either stayed still on the node or traveled but came back to its original position within 24 hours. These self-loops are responsible for the presence of single-node communities in the dynamical matrices and for their absence in the topological ones, where there are no self-loops.

We find a different number of communities for different matrices, necessitating the use of community overlap measures, to quantify which result is the closest to the actual Pardus cluster structure. To quantify the relative goodness of the partitions obtained from the various matrices, we calculate three measures of clustering similarity: the Fowlkes and Mallows index  $\mathcal{F}$  [85], the Rand's criterion  $\mathcal{R}$  [179] and the normalized information variation (NVI) [58], see Section 2.3.5.

Table 5.2 reports the values obtained for the studied matrices. The values of the Fowlkes-Mallows and Rand indices for the dynamical communities are much closer to 1 than the ones for the topological communities. The result is confirmed also by the NVI values, where we measured very small values for the dynamical partitions, indicating that player mobility follows closely the Pardus cluster structure.

It could be argued that this similarity emerges from the topological structure of the network. However, we also found a substantial difference between the topological partitions and the actual Pardus clusters, one order of magnitude larger than the differences between dynamical partitions and the Pardus clusters. Therefore, the clusters of Pardus genuinely influence the motions of players, and this influence is *not* an effect resulting from the topological features of the network, or at least not significant. Moreover, this result is robust under different measures of player movement, as shown by the remarkable stability of the values of the clustering similarity measures for the other dynamical cases,  $M^{\text{symm}}$ ,  $\Pi$  and  $\Pi^{\text{symm}}$ , which are close to the ones obtained for  $M$ , see Table 5.2.

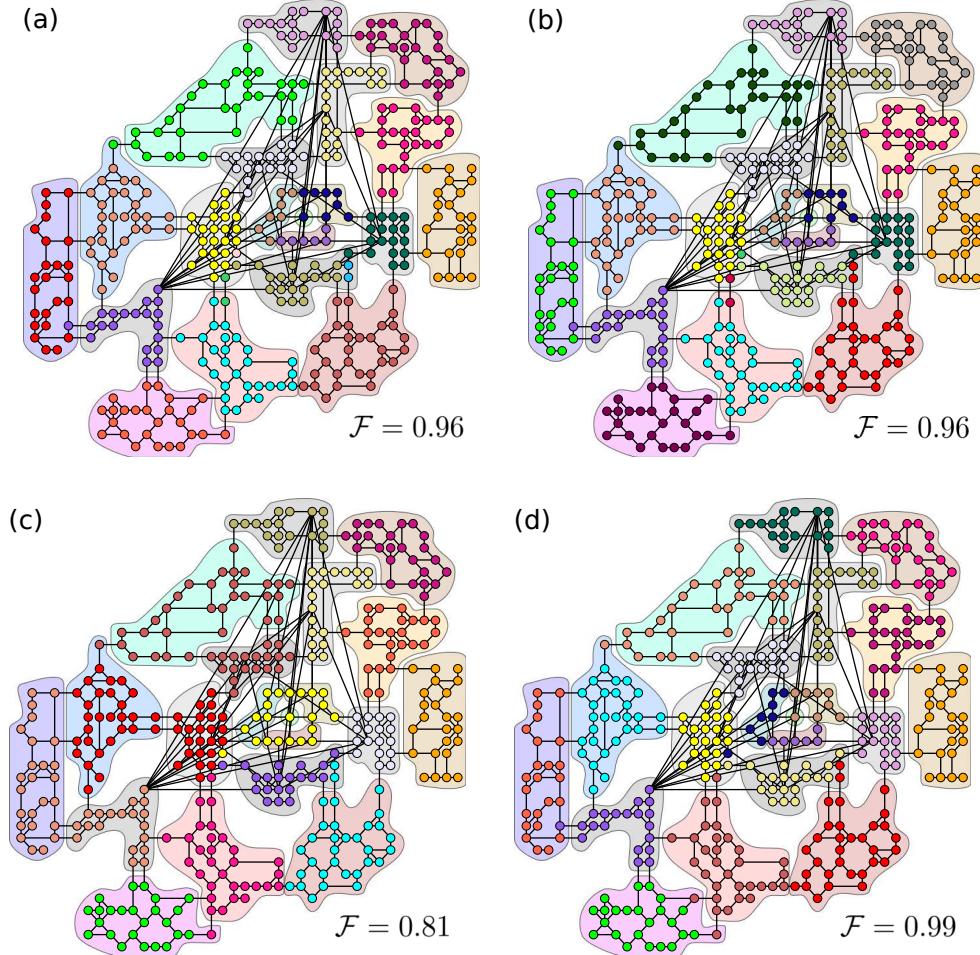


Figure 5.7: Extracting communities from mobility patterns. Communities found for (a) the jump matrix  $M$ , (b) the symmetrized jump matrix  $M^{\text{symm}} = (M + M^T)/2$ , (c) the transition matrix  $\Pi$  and (d) the symmetrized transition matrix  $\Pi^{\text{symm}} = (\Pi + \Pi^T)/2$ . Different node colors represent the different communities found, while the 20 different color-shaded areas indicate the predefined socio-economic clusters as in Fig. 5.1. The communities found through the information of motions reproduces well the bulk of the Pardus cluster structure, with a few exceptions along borders where some nodes are assigned to wrong clusters. The Fowlkes and Mallows index  $\mathcal{F}$  is close to 1 for all detected partitions, reflecting the good match. For more measures, see Table 5.2.

## 5.5 A long-term memory model

In order to characterize the diffusion of players over the network, we have computed the mean square displacement (MSD) of their positions,  $\sigma^2(t)$ , as a function of time. The MSD is defined as  $\sigma^2(t) = \langle (r(T+t) - r(T))^2 \rangle$ , where  $r(T)$  and  $r(T+t)$  are the sectors a player occupies at times  $T$  and  $T+t$  respectively, and where  $(r(T+t) - r(T))$  denotes the distance between the two sectors. The average  $\langle \cdot \rangle$  is performed over all windows of size  $t$ , with their left boundaries going from  $T=0$  to  $T=1,000-t$ , and over all the 1,458 players in the data set. If  $\sigma^2$  has the form  $\sigma^2(t) \sim t^\nu$  with an exponent  $\nu < 1$ , the diffusion process is subdiffusive, in the case  $\nu > 1$  it is super-diffusive. An exponent of  $\nu = 1$  corresponds to classical Brownian

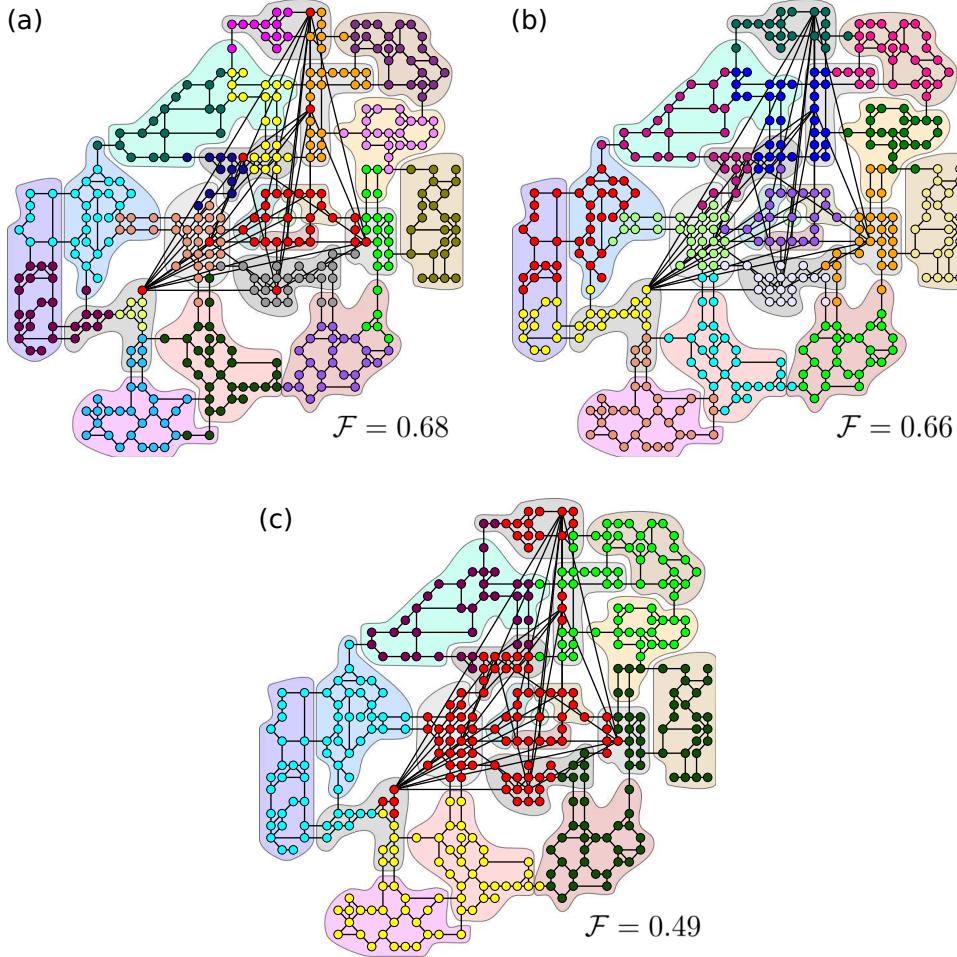


Figure 5.8: Extracting communities from topological information. Communities found for (a) the adjacency matrix  $A$ , (b) the adjacency matrix  $A^{wh}$  in which the X/Y-holes were removed yielding an undirected unweighted network, and (c) the distance matrix  $D^{\text{inv}}$ . Different node colors represent the different communities found, while the 20 different color-shaded areas indicate the predefined socio-economic clusters as in Fig. 5.1. The partitions obtained from the adjacency matrices produce communities that cross over the borders between clusters and therefore do not recover the clusters well. This is particularly evident in the case of  $D^{\text{inv}}$  where only 6 communities are found. The Fowlkes and Mallows index  $\mathcal{F}$  is not close to 1 for all detected partitions, reflecting the bad match. For more measures, see Table 5.2.

motion [155, 218].

Results reported in Fig. 5.9 (a) indicate that, for long times, the MSD increases as a power-law:

$$\sigma^2(t) \sim t^\nu, \quad (5.3)$$

with an exponent  $\nu \approx 0.26$ . This anomalous subdiffusive behavior is not a simple effect of the topology of the Pardus universe. In fact, as shown in Fig. 5.9 (b), gray stars, the simulation of ordinary random walks on the same network produces a standard diffusion with an exponent  $\nu \approx 1$  up to  $t \approx 100$  days, and then a rapid saturation effect which is not present in the case of the human players. Insights from the previous section suggest that the anomalous diffusion behavior might be related

Matrix	Network Properties	$n_{comm}$	$\mathcal{F}$	$\mathcal{R}$	NVI
Clusters	—	20	1	1	0
$A$	directed, unw.	18	0.678	0.963	0.179
$A^{wh}$	undirected, unw.	17	0.655	0.957	0.180
$D^{inv}$	directed, weight.	6	0.489	0.864	0.271
$M$	directed, weight.	23	0.963	0.996	0.025
$M^{\text{symm}}$	symmetrized, weight.	22	0.957	0.995	0.026
$\Pi$	directed, weight.	14	0.812	0.973	0.075
$\Pi^{\text{symm}}$	symmetrized, weight.	19	0.999	0.993	0.036

Table 5.2: Overview of community detection results for the studied matrices. From left to right, the columns correspond to: the studied matrix, the properties of the corresponding network, the number of communities found  $n_{comm}$ , the scores for the Fowlkes-Mallows index  $\mathcal{F}$  [85], the adjusted Rand's criterion  $\mathcal{R}$  [179] and, finally, the normalized information variation (NVI) [58]. For reference, the first row contains the values for the Pardus cluster structure. The closer the indices  $\mathcal{F}$  and  $\mathcal{R}$  are to 1, and the closer the NVI is to 0, the better a found partition matches the clusters. The values reported clearly indicate that the Pardus cluster structure is faithfully reproduced by the player mobility. On the other hand, the topological, non-dynamic properties (e.g. adjacency matrix, distance matrix) produce partitions that are very different from the Pardus cluster structure.

to the tendency of players to avoid crossing borders. We have therefore considered a Markov model in which each walker moves from a current node  $i$  to a node  $j$  with a transition probability  $\pi_{ij} = m_{ij} / \sum_l m_{il}$ , where  $m_{ij}$  is the number of jumps between sector  $i$  and sector  $j$ , as expressed by the transition count matrix  $M$  of Fig. 5.6 (c). The probabilities  $\pi_{ij}$  are the entries of the transition probability matrix  $\Pi$ , which contains all the information on the day-to-day movement of real players, such as the preference to move within clusters, the length distribution of jumps, as well as the tendency to remain in the same sector. Despite this detailed amount of information used (the matrix  $\Pi$  has 160,000 elements), the Markov model fails to reproduce the asymptotic behavior of the MSD, see magenta diamonds in Fig. 5.9 (b). Since the model considers only the position of the individual at its current time to determine its position at the following time, deviations from empirical data appear presumably due to the presence of higher-order memory effects. For this reason we have considered the recently proposed preferential return model [199] which incorporates a strong memory feature. The model is based on a reinforcement mechanism which takes into account the propensity of individuals to return to locations they visited frequently before. This mechanism is able to reproduce the observed tendency of individuals to spend most of their time in a small number of locations, a tendency which is also prevalent in the mobility behavior of Pardus players, see Fig. 5.10.

However, the implementation of the preferential return model on the Pardus universe network is not able to capture the scaling patterns of the MSD, as shown in Fig. 5.9 (b). The reason is that in the model the probability for an individual to move to a given location does not depend on the current location, nor on the order of previously visited locations. Instead, we observe that individuals tend to return with higher probability to sectors they have visited recently and with lower probability to

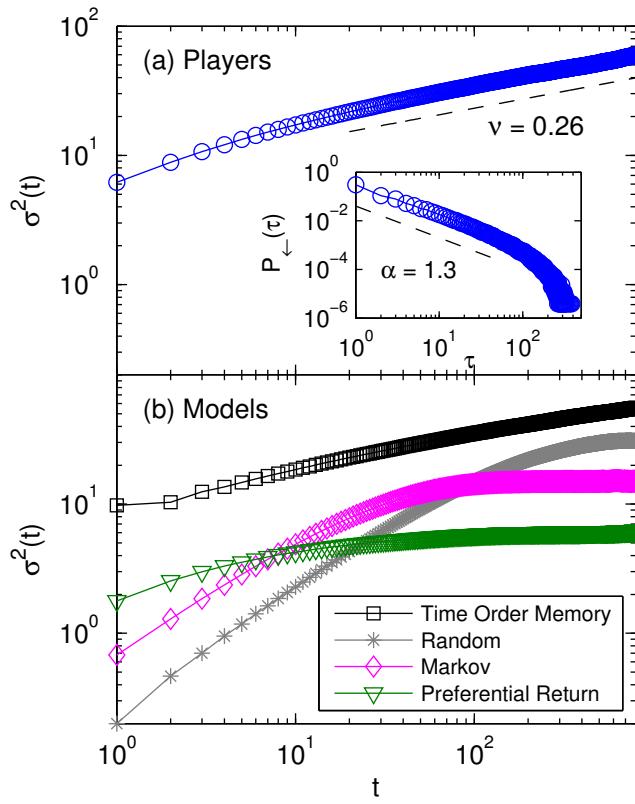


Figure 5.9: Diffusion scaling in empirical data and simulated models. (a) The mean square displacement (MSD) of the positions of players follows a power relation  $\sigma^2(t) \sim t^\nu$  with a subdiffusive exponent  $\nu \approx 0.26$ . The inset shows the average probability  $P_{\leftarrow}(\tau)$  for a player to return after  $\tau$  jumps to a sector previously visited. The curve follows a power law  $P_{\leftarrow}(\tau) \sim \tau^{-\alpha}$  with an exponent of  $\alpha \approx 1.3$  and an exponential cutoff. We report, for comparison, (b) the MSD for various models of mobility. For random walkers and in the case of a Markov model with transition probability  $\pi_{ij} = m_{ij} / \sum_j m_{ij}$  we observe an initial diffusion with an exponent  $\nu \approx 1$  and then a rapid saturation of  $\sigma^2(t)$ , due to the finite size of the network. A preferential return model also shows saturation and does not fit the empirical observed scaling exponent  $\nu$ . Conversely, a model with long-time memory (Time Order Memory) reproduces the exponent almost perfectly. Such a model makes use of the empirically observed  $P_{\leftarrow}(\tau)$  while the Markov model and the preferential return model over-emphasize preferences to locations visited long ago and does not recreate the empirical curve well. Curves are shifted vertically for visual clarity.

sectors visited a long time before. Consequently a sector that has been visited many times but with the most recent visit dating back one year has a lower probability to be visited again than a sector that has been visited just a few times but with the last visit dating back only e.g. one week.

To highlight this mechanism we measured the return time distribution in the jump-time series, which is the transformation of the time-series of daily sector IDs occupied by the players from real-time to jump-time, in order to be able to compare time-series of different length and to focus on the movements between sectors. An

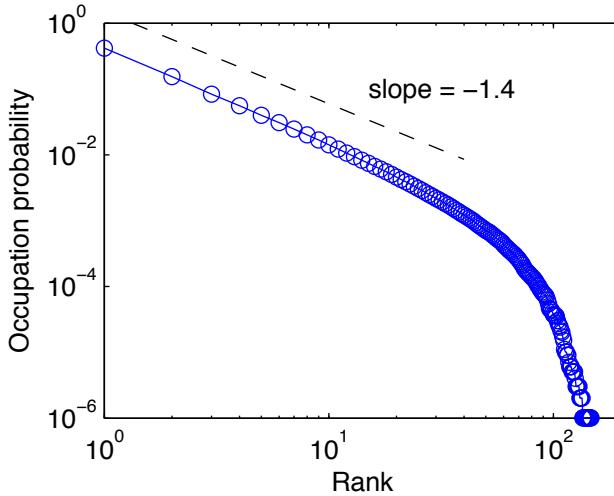


Figure 5.10: Distribution of occupation probabilities, averaged over the distributions of all single players, ordered by rank. The dashed line is a power law with slope  $-1.4$ , which fits well the distribution. Single players may have different locations which they prefer, however the averaged distribution over these single, rank-ordered occupation probabilities shows a fat tail, and therefore the common pattern of significantly preferring the stay in certain locations over other locations.

example of this conversion is provided: a time series

$$[5, 5, 5, 32, 32, 104, 5, 5, 104, 104, 104, 32, 337, 337, 32 \dots]$$

becomes in jump-time

$$[5, 32, 104, 5, 104, 32, 337, 32, \dots].$$

We denote jump-time by the greek letter  $\tau$ , that is, at jump-time  $\tau$  a player has performed exactly  $\tau$  jumps. We use  $\tau$  in the computation of the first return time distribution. In the hypothetical time series of sectors  $[5, 32, 104, 5, 104, 32, 337, 32]$  a first return to a sector lying  $\tau = 1$  jumps back happens 2 times (104, 5, 104 and 32, 337, 32), for  $\tau = 2$  this happens once (5, 32, 104, 5), for  $\tau = 3$  also once (32, 104, 5, 104, 32). Hence, in this example,  $P_{\leftarrow}(1) = 0.5$ ,  $P_{\leftarrow}(2) = P_{\leftarrow}(3) = 0.25$ . The sum over all  $P_{\leftarrow}(\tau)$  is equal to 1.

In particular, we extracted the probability  $P_{\leftarrow}(\tau)$  for an individual to return again (for the first time) to the currently occupied sector after  $\tau$  jumps. As shown in the inset of Fig. 5.9 (a), we found that the return time distribution reads

$$P_{\leftarrow}(\tau) \sim \tau^{-\alpha}, \quad (5.4)$$

with an exponent  $\alpha \approx 1.3$ . We used this information for constructing a model which takes into account the higher re-visiting probability of recently explored locations. In this way we capture the long-term scaling properties of movements. Exactly these asymptotic properties are relevant for issues of epidemics spreading or traffic management.

This “Time Order Memory” (TOM) model incorporates a power-law distribution of first return times, together with a power-law distribution of waiting times and an exponential distribution of jump distances, as those observed empirically in Fig. 5.2. We show below that these ingredients are sufficient to reproduce the subdiffusive behavior reported in Fig. 5.9 (a). The model works as follows: an individual stands still in a given sector for a number of days drawn from the waiting time distribution, Eq. (5.2). Then, the individual jumps. There are two possibilities: (i) with a probability  $v$  she returns to an already visited sector, (ii) with the probability  $1 - v$  she jumps to a – so far – unexplored sector. In case (i), one of the previously visited sectors is chosen according to Eq. (5.4). In the exploration case (ii), the individual draws a distance  $d$  from the distance distribution, Eq. (5.1), and jumps to a randomly selected, unexplored sector at that distance. The model has four parameters. The parameters  $\lambda$ ,  $\beta$  and  $\alpha$  of equations (5.1), (5.2) and (5.4) respectively, are fixed by the data. Further, averaging over all jumps and players, the probability of returning to an already visited location is  $v \approx 0.83$ . Similarly to the measured data, the MSD of the TOM model, black squares in Fig. 5.9 (b), exhibits no saturation effects and displays an exponent  $\nu_{\text{TOM}} = 0.23 \pm 0.02$  (the standard error was calculated over an ensemble of 50 times 100 realizations) in full agreement with the exponent observed for the players.

## 5.6 Discussion of mobility of players

The slope of  $\nu = 0.26$  and the lack of saturation of the MSD of the players over the whole observation period shows the substantial level of subdiffusivity in the motions of individuals, consistent with previous findings [155, 192, 199, 214, 218]. However, the mere tendency of individuals to return to already visited locations is not sufficient to understand these subdiffusive properties of the MSD, but it is fundamental to consider a mechanism that takes into account the temporal order of visited locations, as achieved by the TOM model. Moreover, the TOM model is realistic in the sense that, in contrast to Markov models, it takes into account the tendency of individuals to develop a preference for visiting certain locations. At the same time it allows for the possibility that a previously preferred location becomes not frequented anymore. This view provides an alternative to recently suggested reinforcement mechanisms in preferential return models [199]. The possibility for individuals to “change home” is relevant when the model should be able to account for migration, which is an important feature in the long-time mobility behavior of humans.

Finally, we discuss to which extent the findings from our “social petri dish” are valid also for human populations unrelated to the game. Previous analyses of human social behavior in Pardus [203, 205] have shown agreement with well-known socio-logical theories and with properties on comparable behavioral data. Examining the preference of players to move *within* socio-economic regions is of obvious importance for understanding the role of political or socio-economic borders on the movement

and migration of humans, where the presence of borders has a strong influence on mobility [138, 163, 181, 206]. In the next Chapter, we analyze the evolution and topology of networks of social relations between these acting, re-acting, and mobile individuals.

# **Chapter 6**

## **Measuring social dynamics with socio-economic networks**

In this Chapter we focus on the analysis of complex network structures/dynamics and on testing classic sociological hypotheses. Along these lines we establish further evidence that online game communities may serve as a model for real world communities. It is not obvious *a priori* that a population of online players is a representative sample of real-world societies [220]. However, several recent studies are providing evidence that human behavior on a collective level is remarkably robust, meaning that statistical differences of real-world communities and game-societies are often marginal [124, 126].

### **6.1 Network extraction**

We represent all measured networks as digraphs with nodes representing the characters of players. Note that we do not consider isolated nodes, i.e. characters having no PM communication or friend/enemy relations. In the following we use ‘link’ short for ‘directed link’.

#### **Private messages – communication networks**

The first set of networks is extracted by considering all PM communications on a weekly timescale. Within the timeframes  $[d-6, d]$  for all days  $d > 6$  all PMs between all characters (who exist over these timeframes) were used to define the PM network at day  $d$ : a weighted link pointing from node  $n_i$  to node  $n_j$  is placed if character  $i$  has sent at least one PM to character  $j$  within a given week. Weights correspond to the total number of PMs sent within this week. Figure 3.7 (a) illustrates a subgraph of PM networks of accumulated PM communications over all 445 days between 78 randomly selected characters.

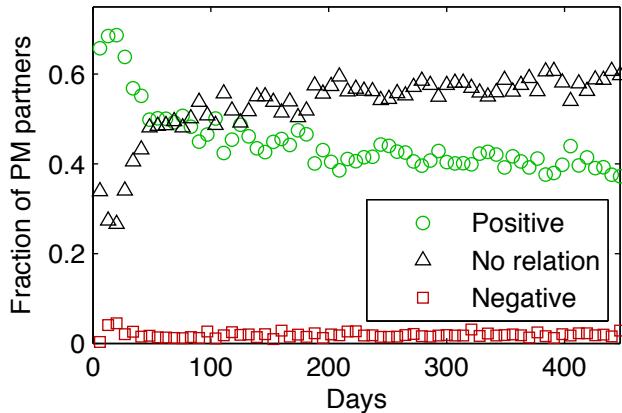


Figure 6.1: Fraction of PM partners per relation class.

### Friends and enemies

Friend and enemy markings constitute the second and third sets of extracted networks: A link is placed from  $n_i$  to  $n_j$  if character  $i$  has marked character  $j$  as friend/enemy. Note that friend/enemy markings exist until they are removed by players (or as long as the related players exist), while PM networks are constructed through an accumulating process. Friend and enemy networks are unweighted, since it is not possible mark friends/enemies more than once.

Since links of friend and enemy networks never coincide (it is not possible to mark someone as both friend and enemy), we can consider the union of friend-and enemy networks as signed networks. Figure 3.7 (b) illustrates a subgraph of the friend/enemy network of day 445. Note the intense cliquishness/reciprocity of friends and the strong enemy *in-hub*. We show below that these features are typical for these networks.

### Friend and enemy relations between PM partners

A connection between PM networks and friend/enemy networks can be made visible by partitioning all pairs of characters  $\{n_i, n_j\}$  into four classes of friend and/or enemy relations, corresponding to the possible valency matrix entries  $v_{ij}$  of a signed digraph. These classes  $o, p, n, a$  correspond to dyads without friend/enemy ties ( $o$ ), dyads with asymmetric or mutual friend markings ( $p$ ), dyads with asymmetric or mutual enemy markings ( $n$ ), dyads with one friend and one enemy marking ( $a$ ). Figure 6.1 depicts the fraction of all PM partners as partitioned into these classes. Relations of class  $a$  are not shown since they almost never appear. From the fact that the majority ( $> 95\%$ ) of PM partners consists of positively related characters ( $\approx 40\%$  on the last day) and of characters having no friend or enemy relation ( $\approx 58\%$  on the last day), we expect a much stronger correlation of PM networks with friend networks than with enemy networks.

	PMs			Friends			Enemies		
	day 50	day 150	day 445	day 50	day 150	day 445	day 50	day 150	day 445
$N$	2,466	2,461	2,879	2,712	3,709	4,313	1,253	2,161	2,906
$L$	10,705	9,773	16,272	15,367	21,563	31,929	4,468	11,077	21,183
$\bar{k}$	5.25	4.80	6.77	6.85	7.36	9.79	6.69	9.77	13.77
$C$	0.24	0.19	0.17	0.34	0.28	0.25	0.02	0.03	0.03
$C/C_r$	112.14	99.10	74.08	133.30	143.32	109.52	3.47	5.87	6.13
$\bar{g}$	4.45	4.71	4.11	4.78	4.45	3.97	3.99	3.64	3.38
$\bar{g}/\bar{g}_r$	0.94	0.95	0.99	1.16	1.08	1.08	1.06	1.08	1.11
$E_{loc}$	0.29	0.23	0.23	0.38	0.34	0.32	0.02	0.04	0.05
$E_{glob}$	0.23	0.22	0.25	0.19	0.22	0.25	0.25	0.28	0.32
$\rho$	0.79	0.79	0.80	0.79	0.73	0.68	0.12	0.09	0.11
$r_{C(k)}$	-0.03	-0.02	-0.01	0.19	0.08	-0.09	0.02	-0.02	-0.00
$r_{undir}$	-0.13	-0.17	-0.04	0.06	-0.06	-0.00	-0.19	-0.23	-0.24
$\Gamma$	0.981	0.991	0.987	0.929	0.952	0.973	0.954	0.975	0.992

Table 6.1: Network properties at days 50, 150, and the last day 445, for PM, friend and enemy networks.

## 6.2 Evolution of networks

We measure the time evolution of the following basic network properties: number of nodes  $N$ , directed links  $L$  and average degree  $\bar{k}$ , relative size of largest connected component  $\Gamma$ , average geodesic  $\bar{g}$ , clustering coefficient  $C$ , as well as the comparison values  $\bar{g}/\bar{g}_r$  and  $C/C_r$ , see Section 2.3.5. Average degrees, geodesics and clustering coefficients are measured on the reflexive closures of the networks. Geodesics and clustering coefficients were calculated using the `MatlabBGL` package<sup>1</sup>, which efficiently implements standard procedures such as Johnson’s algorithm for finding all geodesics in sparse graphs. The measured network properties are displayed in Fig. 6.2, values of days 50, 150, and 445 are shown in Table 6.1. We use some of these properties to test for the preferential attachment model in Section 6.2.1, to check for network densification in Section 6.2.2, and we report several observations on other non-trivial network features in Section 6.2.3.

### 6.2.1 Testing for preferential attachment

The model of preferential attachment (PA) assumes that nodes which link to a network for the first time preferably attach to nodes with high degrees, i.e. to ‘popular’ nodes [19], see Section 2.3.3. In the extracted directed networks of friends (enemies) we assume this popularity (‘disdain’) being well expressed by the in-degree. We call characters who get connected to a network for the first time *newcomers*. To test

<sup>1</sup>We used version 4.0.

[http://www.stanford.edu/~dgleich/programs/matlab\\_bgl](http://www.stanford.edu/~dgleich/programs/matlab_bgl).

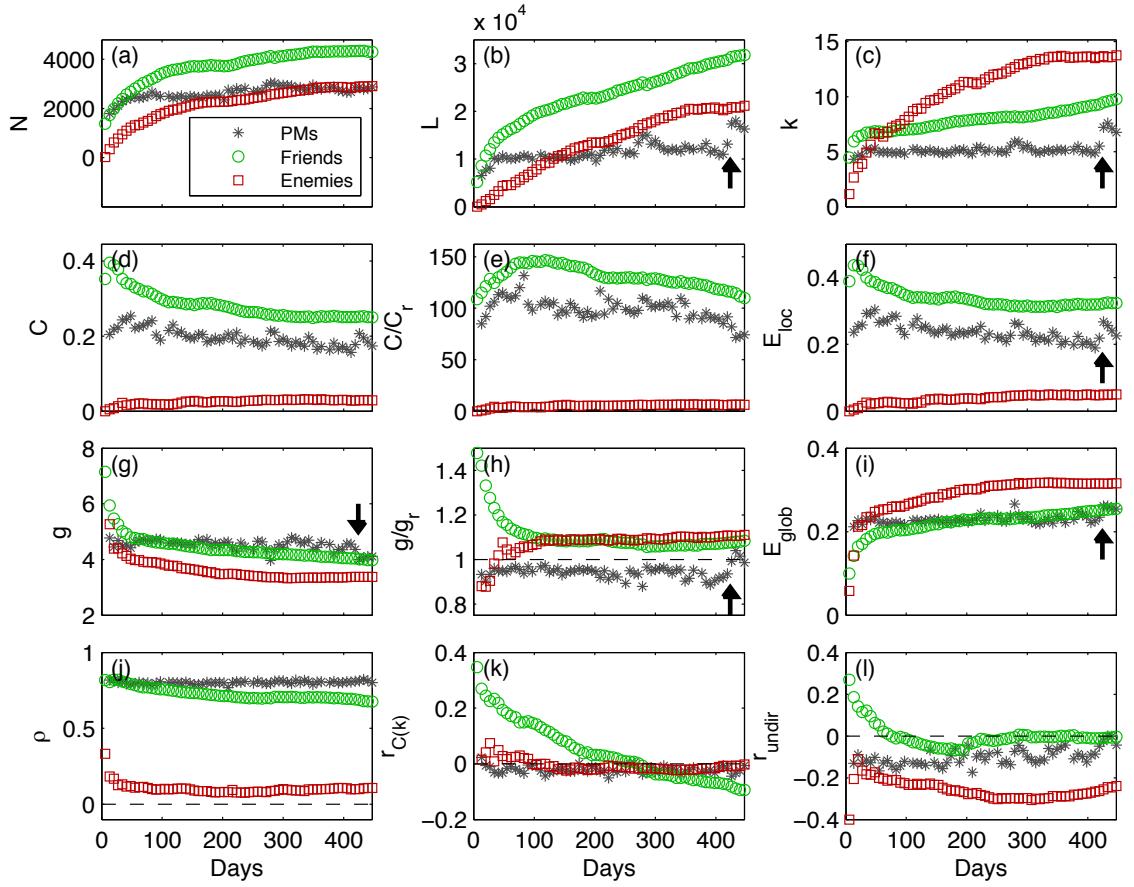


Figure 6.2: Network properties: (a) number of nodes  $N$ , (b) number of (directed) links  $L$ , (c) average degree  $\bar{k}$ , (d) clustering coefficient  $C$ , (e) clustering coefficient  $C$  divided by clustering coefficient of corresponding random graph  $C_r$ , (f) local efficiency  $E_{\text{loc}}$ , (g) average geodesic  $\bar{g}$ , (h) average geodesic  $\bar{g}$  divided by average geodesic of corresponding random graph  $\bar{g}_r$ , (i) global efficiency  $E_{\text{glob}}$ , (j) reciprocity  $\rho$ , (k) assortative mixing coefficient (of clustering in undirected network)  $r_{C(k)}$ , (l) assortative mixing coefficient (in undirected network)  $r_{\text{undir}}$ . Arrows mark the outbreak of an in-game war at day 422.

whether evolutions of present networks display a PA bias we measure in-degrees of characters who are marked by newcomers as friend (enemy). Whenever there exists a link  $l_{ij}$  on day  $d+1$  which has not existed on the previous day  $d$  we say that a (one-day) *link event* has taken place between  $n_i$  and  $n_j$  on day  $d$ ; we call  $n_i$  the *source* and  $n_j$  the *destination* of this event.

If preferential attachment holds in its classical form, the following facts should be observed:

1. Linking probability  $P(k) \propto k^\alpha$ , with  $\alpha = 1$
2. Degree distribution follows a power-law  $P(k) \sim k^{-\gamma}$
3. Clustering coefficient  $C$  versus degree  $k$  is uniform

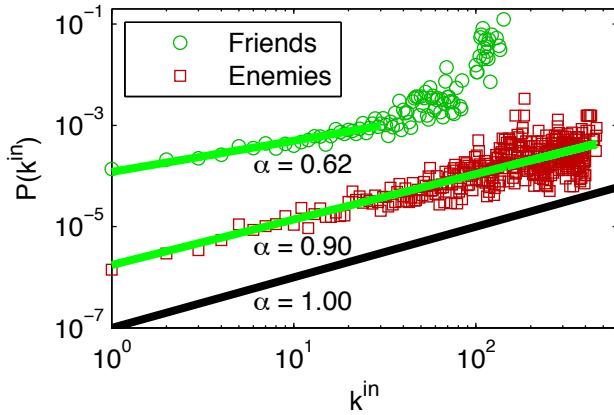


Figure 6.3: Empirical probability  $P(k^{\text{in}})$  for newcomers connecting to nodes with in-degree  $k^{\text{in}}$ . Data is used between days 200 and 400. The black line depicts slope  $\alpha = 1$  and indicates the linear dependence assumption needed in the PA model. Green lines denote least squares fits. Values for enemies are vertically displaced by a factor 0.1 for better visibility.

A linking probability  $P(k) \propto k^\alpha$ , with  $\alpha = 1$ , means that the probability  $P$  of a newcomer connecting to an existing node  $n_i$  with in-degree  $k_i^{\text{in}}$  is  $P(k^{\text{in}}) \propto (k^{\text{in}})^\alpha$  with  $\alpha = 1$ . Figure 6.3 shows  $P(k^{\text{in}})$  versus  $k^{\text{in}}$  for friend and enemy networks; all link events between newcomers and their destinations have been used from day 200 to 400. Least squares fits in double-logarithmic scale yield an exponent of  $\alpha = 0.62$  for friend markings with  $k^{\text{in}} < 30$ , but  $\alpha = 0.90$  for all enemy markings. We observe an increased upward bending for players having in-degrees larger than about 100, i.e. for very popular players. A similar effect of ‘super-preferentiality’ has been measured in [142] for LinkedIn, a social networking site for professional contacts. There an exponent  $\alpha = 0.6$  is reported. This first observation shows that the formation process behind friendship networks cannot be explained with classical PA. Actual preferential attachment has been measured in relatively few works [117, 123, 142]; the evolution of preferential attachment parameters was measured in [60].

Concerning the degree distribution, it does not follow a power-law in friend networks, see Fig. 6.4 (b). However, the situation is different for enemy networks, where the distribution of in-degrees follows an approximate power-law with exponent  $\gamma \approx 1$  (in the cumulative distribution), see Fig. 6.4 (c). Note that – while not clearly visible on the last day’s plot in Fig. 6.4 (c) – we find the distribution of out-degrees in enemy networks separated into two regimes following approximate power-laws with different exponents,  $\gamma \approx -0.6$  and  $\gamma \approx -2.5$ , respectively (in the cumulative distribution). This is the case for many days in the Artemis universe and also for most days in the Orion universe (not shown).

Finally, the clustering coefficient versus degree exhibits a clear downward trend in friend networks and PM networks, see Figs. 6.4 (d) and (e). The exponent of the clustering coefficient  $C(k)$  versus degree  $k$  is  $\gamma \approx -0.4$ . Note also the clearly negative value of  $r_{C(k)}$  at the last days in the friend networks, Fig. 6.2 (k). About the same

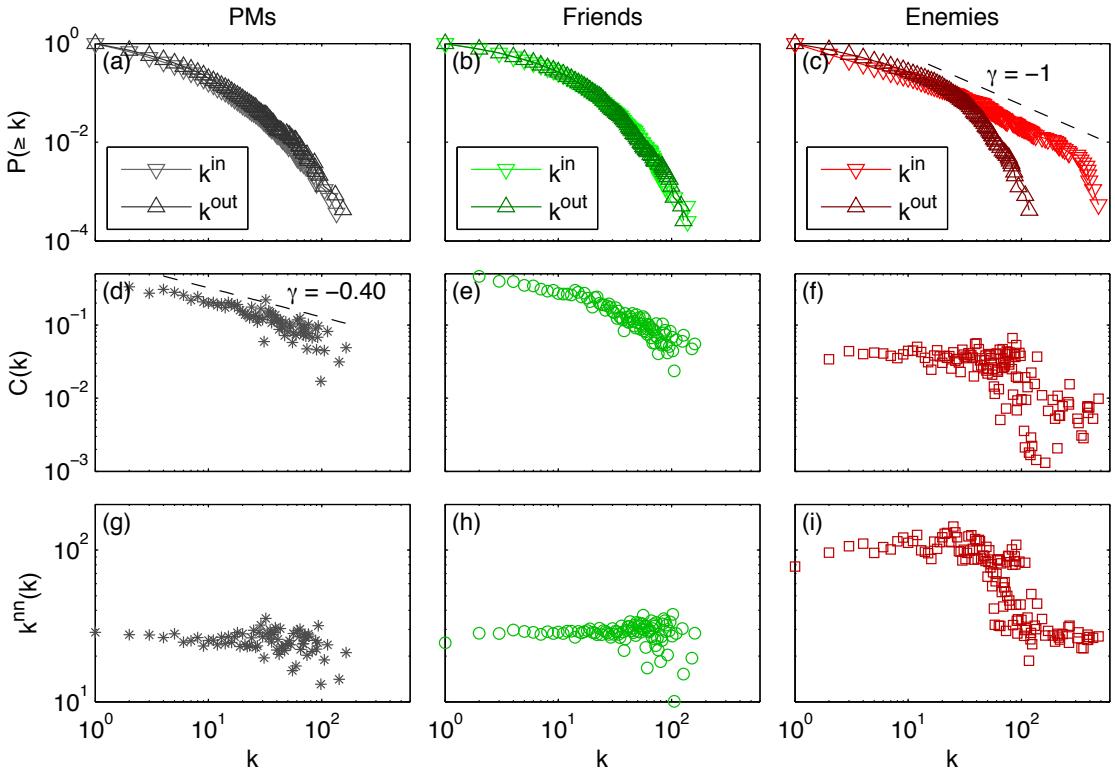


Figure 6.4: Cumulative degree distribution of (a) PM, (b) friend and (c) enemy networks; clustering coefficient  $C$  as a function of degree for the (d) PM, (e) friend and (f) enemy networks; nearest neighbor degree  $k_{nn}$  versus degree of the (g) PM, (h) friend, and (i) enemy networks. All distributions were taken at the last day.

exponent has been measured in a game-theoretic model on co-evolving networks [30]; an exponent of  $\gamma \approx -0.33$  was found in another large-scale social network [59]. The situation is again different in enemy networks: Clustering coefficients and degrees are to a large extent independent; deviations for large degrees can be explained by two different mechanics of marking enemies, see Section 6.4.

In conclusion the model of preferential attachment can not be applied for friend networks. The situation of the enemy networks might be closer to a PA mechanism, but also there the situation is more intricate.

## 6.2.2 Network densification

In Ref. [144] intriguing observations concerning universal features of growing real-world networks have been made. Their empirical observations apparently challenge two conventional assumptions of popular network models such as PA [19], namely constant average degrees and slowly growing network diameters:

1. *Shrinking diameters*: As networks grow their diameters decrease.
2. *Densification power-laws*: Over time, networks become more dense. Densi-

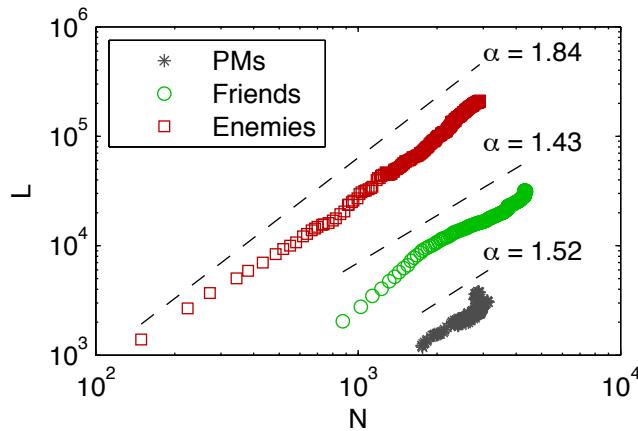


Figure 6.5: Links  $L$  versus nodes  $N$  for all timepoints. Values for enemies and PMs are shifted vertically for clarity.

fication – measured in number of edges versus number of nodes – follow a power-law. The average degree grows.

All growing networks measured in this work confirm the observations of growing average degrees and shrinking diameters (Fig. 6.2 (g) shows shrinking geodesics; we observe the same evolution for diameters and effective diameters as defined in [144] (not shown)). Merely the enemy network reaches a steady state shortly before day 400. Concerning power-law densification, even though our data does not allow for a statistically conclusive quantitative statement, visual inspection clearly reveals that growth is super-linear, Fig. 6.5. The  $L$  versus  $N$  curves for the most part have slopes between 1 and 2. Due to heavy fluctuations and limited extension over  $N$ , fits to power-laws are not reliable. Network densification has previously been studied under the name of *accelerated growth* [67, 68]. Growing average degrees were observed in all three time evolution studies of growing networks we are aware of [116, 144, 182]. From decreasing distances it follows that global efficiency increases, Fig. 6.2 (i). At the same time, evolution of local efficiency follows the evolution of clustering coefficients, Fig. 6.2 (f). PM networks have approximately the same local and global efficiency (around 0.25), in friend networks local efficiency (around 0.33) is higher than global efficiency (around 0.22), in enemy networks local efficiency is much lower (around 0.04) than global efficiency (around 0.28).

### 6.2.3 Measuring non-trivial network properties

We make the following observations on the evolving network measures shown in Fig 6.2 and Table 6.1 which have not been discussed in previous sections.

### Average geodesics close to random network value

For all networks the comparison parameter  $\bar{g}/\bar{g}_r$  lies well within the band  $[0.5, 2]$  as reported for various scale-free networks by [68], Fig. 6.2 (h). It fluctuates slightly above 1 for enemy and friend networks, growing for the former, decreasing for the latter. In PM networks  $\bar{g}/\bar{g}_r$  is slightly below 1 for most time points.

### Changing clustering coefficients

Clustering coefficients of friend networks decrease, those of enemy networks increase, Fig. 6.2 (d). Concerning  $C/C_r$ , values fall in friend and PM networks but grow in enemy networks, Fig. 6.2 (e).  $C/C_r$  curves of PM networks fall between the curves of enemy and friend networks. Decreasing clustering coefficients have been reported for co-authorship networks [182], online social networks [117], and appear in a model of growing social networks [125]. Note that  $C/C_r$  has a high value for friend networks ( $C/C_r > 100$ ), as is expected for most social networks of positive ties [170].

### Positive reciprocity

All networks are reciprocal, Fig. 6.2 (j). At the last day the PM network has a reciprocity of  $\rho \approx 0.80$ , the friend network has  $\rho \approx 0.68$  after having reached a maximum of  $\rho \approx 0.83$  in the first days. Excluding the first few days, the reciprocity indices of enemy networks lie around  $\rho \approx 0.1$ . The naive reciprocity  $R$  displays qualitatively similar behavior (not shown). Low reciprocities in enemy networks may be explained by deliberate refusal of reciprocation, to demonstrate aversion by lack of any response.

Holme et al. [116] report a naive reciprocity value around  $R = 0.4$  for a message network, a value much lower than our measured ones around  $R \approx 0.7$  in PM networks. We suspect two factors responsible for this discrepancy: A higher community coherence – i.e. more social pressure to respond – in Pardus, and a possibly high inactivity rate of users on the dating site. Probably for the same reason of community coherence, reciprocities  $\rho \approx 0.8$  of Pardus PM networks are well above  $\rho = 0.194$ , a value reported for messages in email networks [90].

### No assortativity

For PM networks, all five considered assortative mixing coefficients reach steady state values slightly below zero, Fig. 6.2 (l) (only plots of  $r_{\text{undir}}$  are shown). PM networks are therefore disassortative, i.e. a player who sends/receives PMs to/from players with many PM-partners displays a slight tendency of having few PM-partners and vice versa. For a possible explanation see [116] who attest this observation as being in contrast to collaboration networks, for which positive assortativity has been measured. There it has been claimed that in friend networks individuals are substitutable and negative mixing is optimal. The approximate steady state of friend networks displays no clear tendency towards assortativity or disassortativity after a

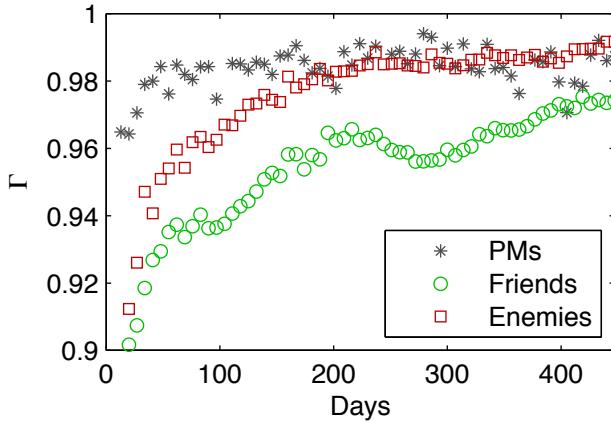


Figure 6.6: Fraction of nodes in the largest connected component  $\Gamma$ .

transient phase of falling assortativity. Note that by using an assortativity profile it is possible to uncover families of networks, similar to TSPs [84, 157].

### Structural change of PM networks due to times of war

On day 422 a war between a substantial number of players broke out in the game universe. A structural change of PM networks is identifiable, most clearly in the number of links  $L$ , average degree  $\bar{k}$ , average geodesic  $\bar{g}$ , local efficiency  $E_{\text{loc}}$  and global efficiency  $E_{\text{glob}}$ , see arrows in Fig. 6.2.

### Growing largest connected component

The fraction of nodes in the largest connected component,  $\Gamma$ , is growing in friend and enemy networks over almost all 445 days, Fig. 6.6. On the last day we find  $\Gamma \approx 0.973$  for friends and  $\Gamma \approx 0.992$  for enemies. The value for PM networks fluctuates around  $\Gamma \approx 0.985$ .

## 6.3 Testing classical sociological hypotheses

### 6.3.1 Confirmation of the Weak ties hypothesis

The Weak ties hypothesis of Granovetter is an important proposition of sociology and builds upon the assumption that “the degree of overlap of two individual’s friendship networks varies directly with the strength of their tie to one another” [97]. By Granovetter’s paradoxical formulation of the Weak ties hypothesis (“The Strength of Weak Ties”), weak ties (e.g. casual acquaintanceships) are proposed to be strong in the sense that they link communities in an essential way – i.e. they are *local bridges* of high degree – while strong ties (standing for e.g. good friendships) correspond to replacable intra-community connections. Under given social balance

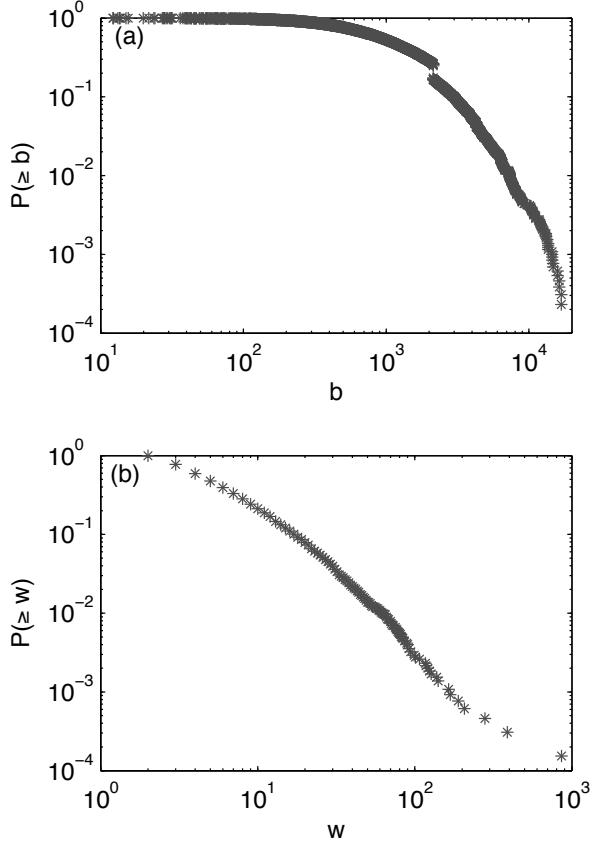


Figure 6.7: Cumulative distribution of (a) betweenness and (b) weight values in the mutual part of the largest connected component of the PM network at the last day.

assumptions, except for very unlikely conditions, “*no strong tie is a bridge*”, and “*all bridges are weak ties*” [97]. As an intuitive notion of *strength* of an interpersonal tie, Granovetter mentions “the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie”.

For validating the Weak ties hypotheses, see below, we measured overlap versus PM weight (number of exchanged PMs) and overlap versus betweenness in the largest connected component of the mutual part of the last PM network. For definitions of overlap and betweenness see Section 2.3.5, we report the cumulative distributions of these quantities in Fig. 6.7. We used this reduced network because it can be directly compared to [172]. Results on *full* PM networks and on PM networks accumulated over different time-spans are very similar however (not shown). Betweenness for all links was calculated with the algorithm provided in the `MatlabBGL` package.

Quantitatively the hypothesis concerning the connection between tie strength and overlap of friendship circles should manifest itself in an increasing function  $O(w)$  of overlap versus weight. This is clearly the case for PM networks, Fig. 6.8 (b), where

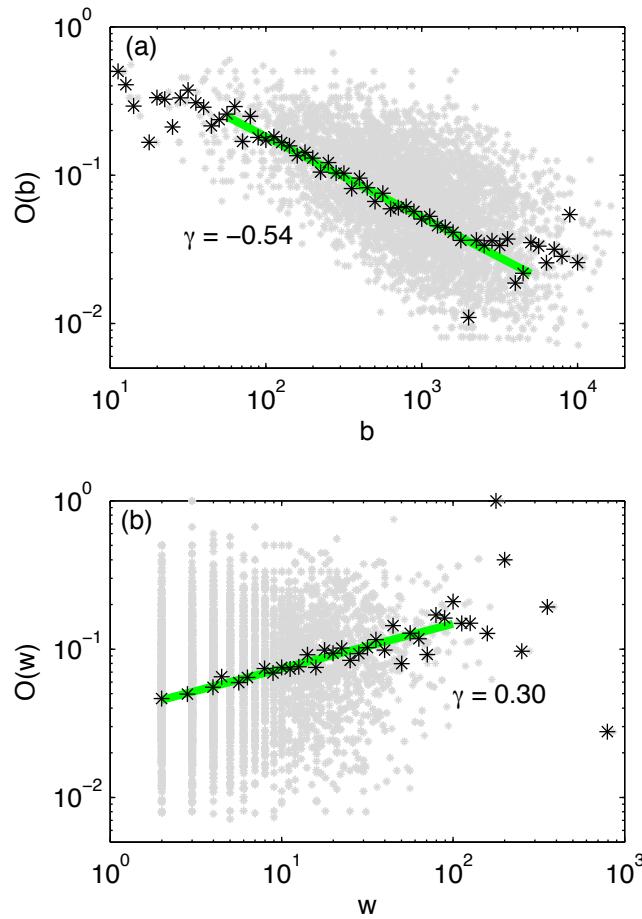


Figure 6.8: Overlap versus (a) betweenness and versus (b) weight in the mutual part of the largest connected component of the PM network at the last day. Light gray markers show individual overlap values of the links. Note that 2,336 links (out of 6,502) have overlap  $O = 0$ . Black markers denote logarithmically binned averages, green lines are least squares fits.

an approximate cube root law is suggested,

$$O(w) = w^{0.30} \approx \sqrt[3]{w}. \quad (6.1)$$

Note that our method does not encounter sampling issues as in e.g. [172]; in this sense our results are free of bias.

A direct way of testing the Weak ties hypothesis is to examine the correlation between betweenness and overlap. By the Weak ties hypothesis, the overlap  $O(b)$  as a function of betweenness should be decreasing. Our data strongly confirms this prediction, Fig. 6.8 (a) and suggests an inverse square root law. Logarithmically

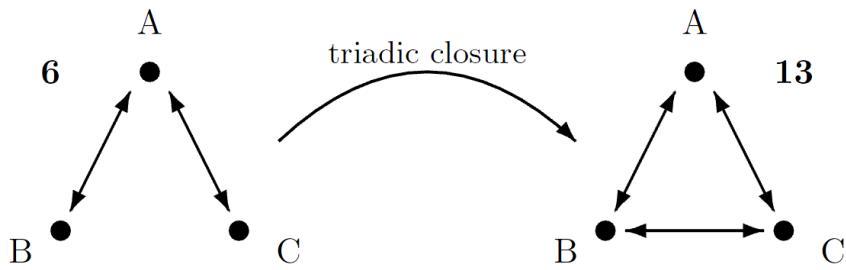


Figure 6.9: One transition representing triadic closure: id 6 → id 13.

binned average values lie on a line with slope  $\gamma \approx -0.54$ .

$$O(b) = b^{-0.54} \approx \frac{1}{\sqrt{b}}. \quad (6.2)$$

These results are in agreement with mobile phone call network data [172], and are robust across game universes and accumulation times. The Weak ties hypothesis has also previously been tested by [87] on a small-scale social network of biologists.

### 6.3.2 Confirmation of triadic closure

In this section we give large-scale quantitative evidence for another important prediction of [97], the hypothesis of *triadic closure*. This conjecture follows balance considerations of Ref. [107] and reads as follows: In a social network in which there exist weak and strong (or no) ties between individuals, “the triad which is most *unlikely* to occur, [...] is that in which A and B are strongly linked, A has a strong tie to some friend C, but the tie between C and B is absent” [97]. The phenomenon of triadic closure [180] states that individuals are driven to reduce this cognitive dissonance, Fig. 6.9.

#### Triad significance profiles (implicit evidence)

If triadic closure – as formulated by Granovetter [97] – holds, it causes the triad in which there exist strong ties between all three subjects A, B and C to appear in a higher than expected frequency. The statistical tool we use to quantify this possible effect is the triad significance profile (TSP), see Section 2.3.5. Comparing TSPs of various types of networks can reveal “superfamilies” of evolved or designed networks which have similar local structures in common [157].

For calculating the TSP, we draw random networks from the  $U(X_{*+}, X_{+*}, M^*)$  distribution by using the same switching algorithm and Monte Carlo method as [157, 159]. We use the same program `mfinder`<sup>2</sup>. See [158, 185] for details. We calculate the TSPs for all three network types at the last day with the following

<sup>2</sup>We used version 1.2.

<http://www.weizmann.ac.il/mcb/UriAlon>.

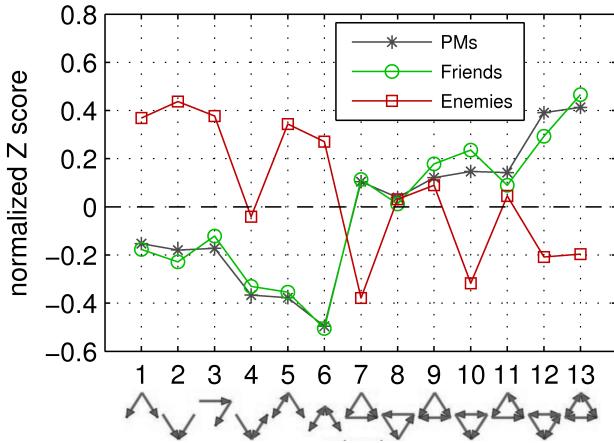


Figure 6.10: Triad significance profiles for the three network types (day 445).

parameters: 100 random networks, each generated by performing  $Q \cdot L$  switches, where  $Q$  is drawn uniformly from  $\{100, \dots, 200\}$ . Resulting TSPs are displayed in Fig. 6.10.

Following the considerations of Granovetter about tie strength we identify the concept of weak/strong ties with asymmetric/mutual dyads in our digraphs. Translated into our formalism the hypothesis reads: “In friend networks, triad class 6 should have smallest  $Z$  score, triad class 13 should have highest  $Z$  score”. In other words, triad class 13 should be the network’s strongest three-node *motif*, triad class 6 should be its strongest three-node *antimotif* [159]. More generally, if we focus on completeness, we expect underrepresentation of the incomplete triad classes 1–6 and overrepresentation of the complete triad classes 7–13. Note that quantitative evidence for triadic closure given this way is implicit at best, since the overrepresentation (underrepresentation) of triad class 13 (6) does not explain *how* or *if* there is a direct connection in the evolution of these triad classes. We give explicit evidence for triadic closure in the subsequent section by measuring triad transition dynamics.

The question about the reverse situation, networks of negative ties, has been raised in the outlook of [97] but, to our knowledge, has never been measured on large scales. Following the same social balance considerations we expect reversed roles of completeness: Instead of the *absence* of a completing third link, its *presence* should cause cognitive dissonance (however, note that a complete triad with only negative links may be seen as ambiguous concerning balance [61, 64]). Thus triad classes 1–6 should be overrepresented, triad classes 7–13 underrepresented in enemy networks.

For friend and PM networks, excellent agreement is found with Granovetter’s prediction: Triad class 6 has the minimum, class 13 a maximal  $Z$  score, Fig. 6.10. Our findings further coincide with the TSPs of the superfamily of social and hyperlink networks found in [157] and with TSPs of other social networks [99]. Concerning enemy networks, we observe confirmation of our reverse hypothesis to a large extent:

Most enemy  $Z$  scores have opposite signs of those in friend networks. Note the exceptions: triad id 4 is not clearly overrepresented, ids 9 and 11 are not clearly underrepresented. The circular triad (id 8) should be considered an exceptional or ‘neutral’ class, having no clear tendency in all network types.

**An attempt at explaining the order of  $Z$  scores.** When facing the TSP of the PM and friend networks in Fig. 6.10 in this ordering of triad ids, one cannot help but notice an approximate monotonic decrease of  $Z$  scores within the ordered set  $(1, 2, 3, 4, 5, 6)$  and an approximate monotonic increase within the ordered set  $(7, 8, 9, 10, 11, 12, 13)$  of triad ids. Recall that we adopted the notation of [157, 159] who apparently (or coincidentally?) number the connected triad classes by first considering completeness, and as a second criterion increasing naive reciprocity  $R$ . If a triad class  $i$  is complete, we write  $c(i) = 1$ , else  $c(i) = 0$ . The reciprocities of all triad classes are  $R(1) = R(2) = R(3) = R(7) = R(8) = 0$ ,  $R(9) = R(10) = R(11) = \frac{1}{3}$ ,  $R(4) = R(5) = \frac{1}{2}$ ,  $R(12) = \frac{2}{3}$ ,  $R(6) = R(13) = 1$ . These two types of orderings can be understood as total orders of all considered triad classes:  $\{1, 2, 3, 4, 5, 6\} <_c \{7, 8, 9, 10, 11, 12, 13\}$ <sup>3</sup> for completeness and  $\{1, 2, 3, 7, 8\} <_R \{9, 10, 11\} <_R \{4, 5\} <_R 12 <_R \{6, 13\}$  for reciprocity. The lexicographical order  $<_{c,R}$  induced by relations  $<_c$  and  $<_R$ , i.e. the latter takes effect should the former yield equivalence, applies to the ordering of [157, 159] as follows:  $\{1, 2, 3\} <_{c,R} \{4, 5\} <_{c,R} 6 <_{c,R} \{7, 8\} <_{c,R} \{9, 10, 11\} <_{c,R} 12 <_{c,R} 13$ .

We now order all triad ids by increasing  $Z$  score of the friendship network of the Orion universe of day 1087,  $\mathbf{O}^+$ , and attempt to explain the emergence of this specific order by incorporating the two above and two additional well-founded sociological properties. It is important to stress the speculative nature of this explanation. Besides, a certain variance of error in resulting TPSs is expected, since they were derived via a Monte Carlo method which is approximative by definition.

One property not considered so far is transitivity, for which “strong, statistical evidence [was provided, that it] is a very important structural tendency in social networks” [215]. An ordered triad involving actors  $i$ ,  $j$ , and  $k$  is *transitive* (T) if the existence of arcs  $(n_i, n_j)$  and  $(n_j, v_k)$  implies the existence of  $(n_i, v_k)$ . If either of the two conditions of this statement is not met (i.e.  $(n_i, n_j)$  and/or  $(n_j, v_k)$  does not exist), the triad is called *vacuously transitive* (VT). A triad isomorphism class is T if all its six realizations of ordered triads are T or VT [215]. In our notation, triad classes  $\{7, 9, 10, 13\}$  are T,  $\{1, 2\}$  are VT,  $\{3, 4, 5, 6, 8, 11, 12\}$  are intransitive (IT). We introduce this new property via the function  $t$  where  $t(i) = -1$  for an IT,  $t(i) = 0$  for a VT, and  $t(i) = 1$  for a T class  $i$ . Note that we expect IT (T) classes to have a negative (positive) impact on  $Z$  score. A second property is the (for clarity of order *negative*) amount of close friends disagreeing which we term  $f$ . This value is  $f(6) = -2$ ,  $f(4) = f(11) = f(12) = -1$ , for all other triad classes it is 0 [215].

---

<sup>3</sup>For two triad classes  $i$  and  $j$  we write  $i <_E j$  short for  $E(i) < E(j)$  and  $i =_E j$  short for  $E(i) = E(j)$ , where  $E$  is a function which – based on a property – assigns each triad class a numerical value. If  $i$  and  $j$  are sets of triad ids, the relation  $x <_E y$  holds for all pairs of elements  $x \in i$  and  $y \in j$  and the equivalences  $x_1 =_E x_2$  and  $y_1 =_E y_2$  hold for all pairs of elements  $x_1, x_2 \in i$  and  $y_1, y_2 \in j$ .

$i$	6	5	4	2	1	3	8	11	7	9	10	12	13
$c(i)$													
$R(i)$	0	0	0	0	0	0	1	1	1	1	1	1	1
$t(i)$	-1	-1	-1	0	0	-1	-1	-1	1	1	1	-1	1
$f(i)$	-2	-1	0	0	0	0	0	-1	0	0	0	-1	0
$TSP_i(\mathbf{O}^+)$	-0.45	-0.37	-0.34	-0.27	-0.24	-0.15	0.02	0.11	0.16	0.20	0.24	0.34	0.38
$TSP_i(\mathbf{A}^+)$	-0.49	-0.39	-0.29	-0.26	-0.19	-0.15	0.01	0.09	0.12	0.15	0.29	0.27	0.45
$TSP_i(\mathbf{P}^+)$	-0.51	-0.38	-0.28	-0.27	-0.19	-0.13	0.02	0.08	0.12	0.15	0.26	0.25	0.47
$TSP_i(\mathbf{A}^{\text{PM}})$	-0.52	-0.37	-0.32	-0.19	-0.07	-0.13	0.02	0.13	0.06	0.07	0.19	0.37	0.50
$TSP_i(\mathbf{P}^{\text{PM}})$	-0.56	-0.34	-0.29	-0.14	-0.07	-0.11	0.02	0.11	0.04	0.08	0.14	0.34	0.54

Table 6.2: Triad ids with sociologically relevant properties ( $c$  for completeness,  $R$  for reciprocity,  $t$  for transitivity,  $f$  for negative number of friends disagreeing) ordered by increasing Z scores of the friendship network of the Orion universe of day 1087,  $\mathbf{O}^+$ .

Whenever  $f(i) < 0$ , an additional tendency towards lower Z score is expected.

In this new order, Table 6.2 depicts all considered attributes. By generalizing Granovetter's considerations, in the group of incomplete triads  $\{1, 2, 3, 4, 5, 6\}$  relation  $<_{c,R}$  is supposed to order by monotonically decreasing Z score, within  $\{7, 8, 9, 10, 11, 12, 13\}$  by monotonically increasing Z score. For  $\mathbf{O}^+$ , this is true for all classes except for pair  $\{7, 11\}$ ; an additional exception is pair  $\{10, 12\}$  for the friendship networks of the Artemis and Pegasus universes,  $\mathbf{A}^+$  and  $\mathbf{P}^+$  respectively. For these pairs, note that while reciprocity of one class is higher than the other one's, both transitivity and close friends disagreeing are inconsistent with reciprocity (i.e.  $i <_R j$  but  $i >_t j$  and  $i >_f j$ ). This observation suggests that the effect of reciprocity may be “outweighed” by these two properties. Further note that using above considerations, we are not able to explain the clear ordering of classes  $2 <_Z 1 <_Z 3$  in friend networks. For the properties considered these triad classes are too similar as to allow substantiated differentiation<sup>4</sup>.

**Time evolution of TPSs.** In the previous paragraphs of this Section, triad significance profiles of one day were analyzed. By measuring evolutions of TSP, we are able to confirm the robustness of the results. Figure 6.11 shows the time evolution of TSPs in PM, friend, and enemy networks, each day measured with the same parameters as in the previous section. For visual clarity, all single Z score evolutions were smoothed with a moving average filter using a time window of 7 days. Z scores of PM networks stay constant, see Fig. 6.11 (a). For friend networks the order of Z scores stays relatively constant except for the pairs  $\{2, 4\}$  and  $\{10, 12\}$ , which switch order abruptly at day  $\approx 290$ , Fig. 6.11 (b). At this time, a new game feature was introduced in Pardus, which allowed players to join *syndicates*. Because of this a number of players reconsidered their friend and enemy relations. Besides these abrupt changes, some trends are discernable, such as the slow decrease of triad classes 1 and 3 or the increase of triad classes 9 and 11. In enemy networks, Fig. 6.11 (c), TSPs undergo heavy fluctuations but settle into three groups of triad classes after some hundred days:  $\{1, 2, 3, 5, 6\}$  being highly overrepresented ( $Z$  score  $> 0.2$ ),  $\{4, 8, 9, 11\}$  being neither clearly over- nor underrepresented ( $Z$  score between  $-0.1$  and  $0.1$ ), and  $\{7, 10, 12, 13\}$  being highly underrepresented ( $Z$  score  $< -0.15$ ). In conclusion, TSPs may need some hundred days to reach an approximate steady state. On the other hand it is apparent from Fig. 6.11 (b) that as social networks evolve, their microscopic structures do not always stay completely constant. Sudden jumps in the TSP trajectories signal abrupt global systemic changes.

### Triad transition rates (explicit evidence)

To obtain insight in *triad dynamics* and to test for triadic closure explicitly, we directly count all transformations of all triads in friend networks, enabling us to deduce empirical transition probabilities from one type to another. We measure

---

<sup>4</sup>Note that there are several nuances of transitivity. For example, triad class 3 has 5 VT realizations and only 1 IT one [215]. Thus there is only the smallest possible difference between classes 1, 2, and 3 by the considered properties.

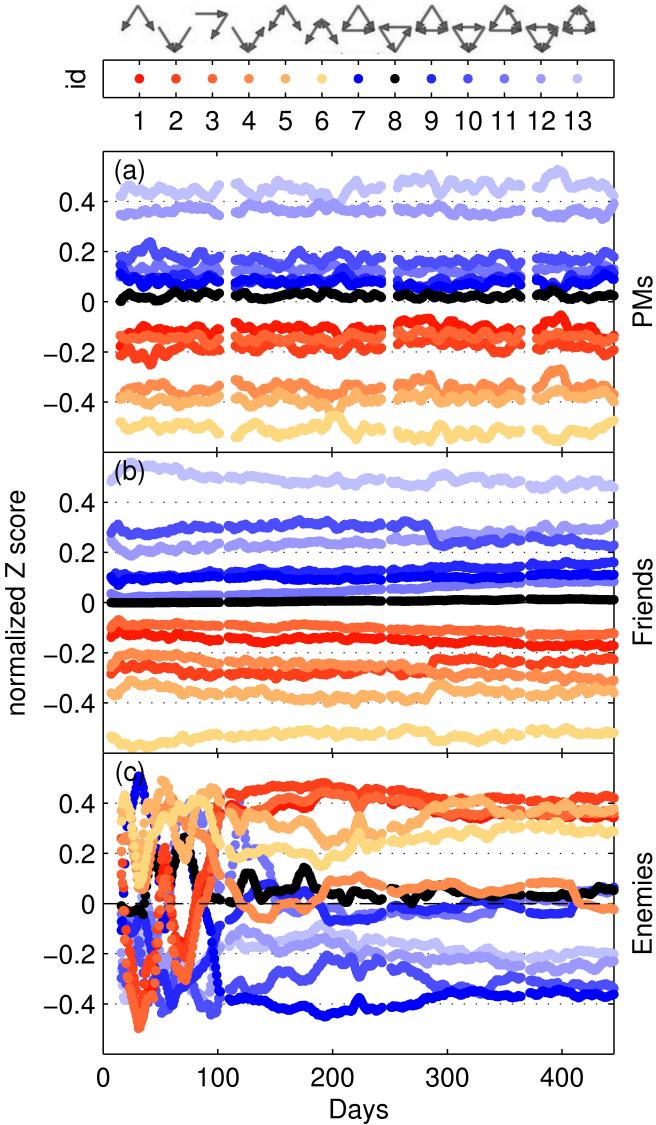


Figure 6.11: Evolution of triad significance profiles for (a) PM, (b) friend, and (c) enemy networks. A moving average filter with a time window of 7 days was used for smoothing.

$13 \times 16$  matrices  $\Pi^d$  of day- $d$ -to-day-( $d + 50$ ) triad transition probabilities for each day  $d \in \{150, \dots, 200\}$ . This matrix is constructed by using an algorithm adapted from [23]. Its entries  $\pi_{ij}^d$  are the empirical probabilities that triads of class  $i$  on day  $d$  become triads of class  $j$  on day  $d + 50$ . Note that the sum of each row of  $\Pi^d$  equals 1. Values between differing rows are not directly comparable due to a highly heterogeneous triad census. For example, the census of all connected triads of day 200 reads (20503, 19872, 13286, 36737, 48320, 58137, 1510, 20, 1209, 4862, 453, 4788, 7887), ordered by increasing triad id. The number of unconnected triads, especially null triads, is much larger namely  $\binom{N}{3}$  minus the number of connected triads; in our case this amounts to an order of magnitude of  $\sim 10^{10}$ . Similarly, we define the

matrices  $\mathbf{K}^d$  containing empirical transition counts  $k_{ij}^d$  of triads of class  $i$  becoming class  $j$ . For both matrices only those triads are counted in which all three of the involved characters still exist on day  $d + 50$ .

We denote the matrix of element-wise time averages of  $\boldsymbol{\Pi}^d$  and  $\mathbf{K}^d$  over all considered days by  $\boldsymbol{\Pi}$  and  $\mathbf{K}$ , respectively. Entries  $k_{ij}$  of matrix  $\mathbf{K}$  are empirical average 50-day transition counts of triads changing from class  $i$  to class  $j$ , entries  $\pi_{ij}$  of matrix  $\boldsymbol{\Pi}$  the corresponding transition probabilities. Figures 6.12 (a) and (b) show  $\boldsymbol{\Pi}$  and  $\mathbf{K}$ . Figure 6.12 (c) displays the matrix of asymmetries in 50-day transition counts  $\mathbf{K}$  for friend networks, i.e.  $\mathbf{K} - \mathbf{K}^T$ . Similar matrices for enemy networks have very different entries (not shown), matrices for PM networks were not calculated.

We first discuss transitions between the groups of incomplete and complete connected triads, i.e. on values in the center right and lower center zones of  $\mathbf{K}$ . According to the hypothesis of triadic closure,  $k_{6,13}$  should contain high values, its counterpart  $k_{13,6}$  lower ones. As apparent from Fig. 6.12 (b), this is the case:  $k_{6,13} = 305.5 > 22.6 = k_{13,6}$ . This result is directly visualized in Fig. 6.12 (c), which depicts the matrix  $\mathbf{K} - \mathbf{K}^T$ . The darker a square  $ij$ , the higher the outflow  $i \rightarrow j$  compared to the inflow  $j \rightarrow i$ ; the lighter a square  $ij$ , the higher the inflow  $j \rightarrow i$  compared to the outflow  $i \rightarrow j$ .

In general, we measure more incomplete  $\rightarrow$  complete transitions between connected triad classes (upper right zone of  $\mathbf{K}$ ) than vice versa (lower center zone of  $\mathbf{K}$ ). On the other hand, some exceptions can be identified, for example  $k_{5,13} = 20.0 < 39.2 = k_{13,5}$ . Whenever these exceptions appear they are comparably mild; again, observations are robust across game universes and time spans. Note that in [203] we measured transition rates of *signed* triadic closure, i.e. triadic closure in signed networks, and clear up dynamics in signed networks.

### 6.3.3 Confirmation of the Dunbar number

Out-degrees of networks studied here are limited by  $k^{\text{out}} \approx 150$ , Fig. 6.4 (a), (b), and (c). This number was conjectured by [73] to be a limiting number in group sizes of humans and human-like mammals, due to their limited cognitive capacities.

### 6.3.4 Inconclusive social balance dynamics

Social balance theory goes back to the cognitive balance considerations of [107]. A complete triad  $n_i n_j n_k$  is defined to be *balanced* if the product of signs  $s_{ij} s_{jk} s_{ki} = 1$ , and is *unbalanced* otherwise. Members of a balanced complete triad thus fulfill the following adage [10, 107]:

- a friend of my friend is my friend
- a friend of my enemy is my enemy
- an enemy of my friend is my enemy
- an enemy of my enemy is my friend

In physics the first statement corresponds to a ferromagnetic system, the other three to a ‘frustrated’ system. In graph theory, the concept of social, or structural,

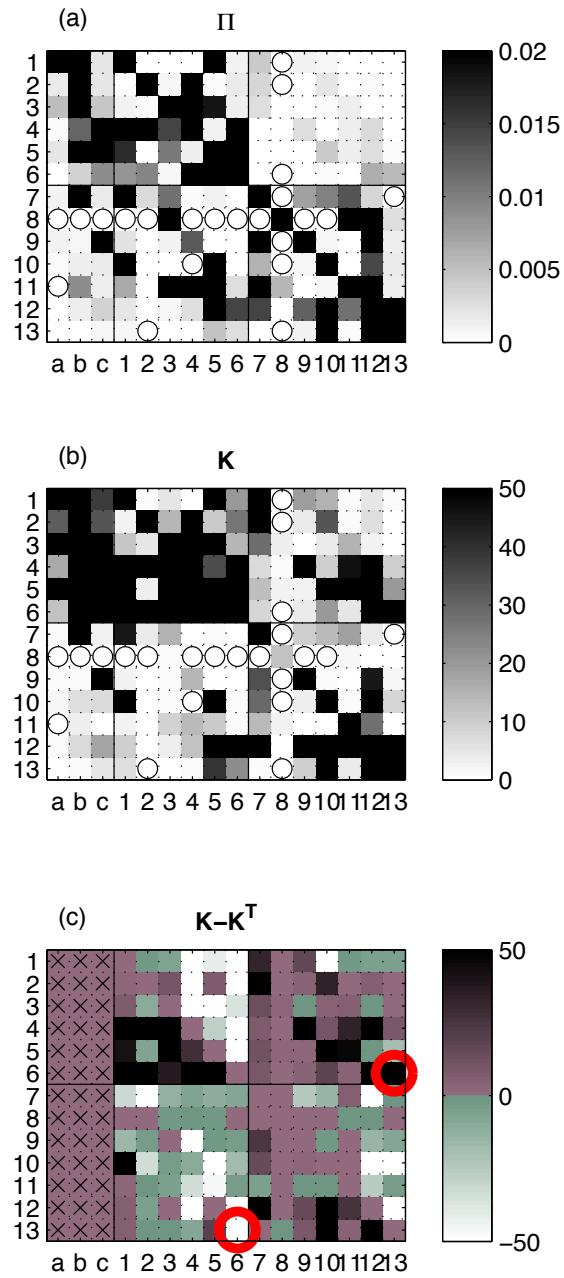


Figure 6.12: (a) Matrix  $\Pi$  of empirical average 50-day transition probabilities of triad classes in friend networks over the days 150 to 200. Circles mark transitions which never occurred. Black squares mark average transition probabilities  $\geq 0.02$ . (b) Matrix  $\mathbf{K}$  of empirical average 50-day transition counts of triad classes in friend networks. Black squares mark average transition counts  $\geq 50$ . (c) Matrix of asymmetries between empirical average 50-day transition counts  $\mathbf{K}$  of triad classes in friend networks. Black and white squares mark average transition counts with asymmetries  $\geq 50$ . Black crosses mark entries without data. Red circles mark the asymmetry of transitions between triad classes 6 and 13.

balance has been generalized to an arbitrary amount of subjects by [42]; hypotheses about the evolution of social balance have been conjectured [64, 66]. We measure the evolution of social balance by using optimizational partition algorithms implemented in **Pajek**<sup>5</sup>. Here we face three concrete problems: (i) *Algorithmic complexity*: Due to algorithmic complexity runtime diverges for large numbers of nodes (**Pajek** limits the number of nodes to 250). Thus for measurements of balance we are forced to select groups of characters. (ii) *Group selection*: Characters have different sign-up dates. This starts to matter for any selected group when long-time considerations are carried out. One gets inhomogeneous groups, where some characters have a long history of relations whereas others have not. (iii) *Growth of average degrees*: The permanent growth of average degrees, Fig. 6.2 (a), is inconsistent with the necessary assumption of constant degrees, such as taken as basis for the monastery study of [191] analyzed in [65]. This assumption is also needed in models and analytical work [10], where dynamics takes place only on complete graphs, i.e. graphs displaying a dichotomy of link types (positive or negative link) in contrast to the trichotomy (positive, negative or no link) of the Pardus networks.

Ignoring these three issues, our attempts to measure the evolution of social balance in several groups of characters at various time scales and cluster sizes yield no conclusive results. Neither an increase nor a decrease in balance could be observed. Measuring social balance of fixed groups over time seems inherently futile, for the following reason: Signs of links do almost never switch [203], so there may be no propagation of balance in social systems. This observation is diametrically opposed to aforementioned classical assumptions [10]. Also, system complexity is so high (factions, alliances, wars, ...), that simple analytical steady states (e.g. two internally positively – but among each other negatively connected sets of nodes) are impossible to reach. For a deeper discussion about these and several other problems related to social balance theory and inconsistencies in empirical findings see [64, 121].

Note that the friend relation is transitive, but not so the enemy relation<sup>6</sup>. By construction, complete triads in the reflexive closure of friend (enemy) networks are balanced (unbalanced). In enemy networks one therefore expects less complete triads and clustering coefficients  $C$  closer to their random graph values – i.e. smaller  $C/C_r$  – than in networks of friends. As apparent from Fig. 6.2 (e), this is fully confirmed.

## 6.4 Two categories of enemies

It has been suggested that identifying negative social tie mechanics is more important for gaining insight in social group dynamics than identifying mechanics of positive social ties [137]. Our measurements provide first steps toward this direction.

We observe that in-degrees in enemy networks grow much bigger ( $k^{\text{in}} \approx 500$ ) than in-degrees in friend and PM networks ( $k^{\text{in}} \approx 150$ ), Figs. 6.4 (b) and (c). Also, the assortativity coefficient  $r_{\text{undir}}$  of enemy networks is clearly negative, Fig. 6.2 (l). In

<sup>5</sup>We use version 1.24. [63, 65, 66]

<sup>6</sup>A binary relation  $R$  is transitive if  $xRy$  and  $yRz$  implies  $xRz$ .

the average neighbor degree  $k^{\text{nn}}$  versus degree  $k$  one observes two classes or types, Fig. 6.4 (i). The first class consists of characters with low degrees ( $k < 50$ ), all having an average neighbor degree of  $k^{\text{nn}} \approx 100$ . The second class of characters has a degree of  $k > 100$  and an average neighbor degree of  $k^{\text{nn}} \approx 28$ . Between these classes a sharp transition occurs. This impression is robust across time and game universes.

As mentioned in Section 6.2.1, on many individual days the distribution of out-degrees of enemy networks is separated into two regimes roughly following power-laws with markedly different exponents. We find clear qualitative differences between in- and out-degree distributions in enemy networks, in contrast to the other network types. Further, reciprocity is very low in enemy networks, as opposed to high reciprocity in the other network types, Fig. 6.2 (j). These measurements suggest two distinct mechanics of enemy marking dynamics at work:

1. *Private enemies*: A player who directly experiences a negative (asocial) action by another player, such as an attack of her building or a verbal insult, is likely to *immediately* react by marking the offender as enemy. If the two involved players keep this a private affair, only a local, dyadic vendetta without envolvement of more players may ensue.
2. *Public enemies*: Some players have a destructive personality or, more commonly, want to *try out* a destructive personality [45]. For this purpose they may (role)play evil characters, such as ‘pirates’. These players tend to take enjoyment in destroying other players’ work or see it as their ‘task’. Anonymity of the internet facilitates this behavior to some degree since possible social repercussions in real-world reputation are absent. For this reason these few individuals tend to cause a lot of offenses to a big number of players. If such a subject is identified by the community (players are very busy in using the forums to keep others up-to-date of the latest offenses), she may receive pre-emptive enemy markings, either by friends of offended friends or by otherwise non-involved players who happen to read the forums. This destructive behavior and the indirect marking mechanism leads to the emergence of ‘public enemies’, i.e. a few characters with a very high in-degree of enemy markings. The strength of positive social ties is likely to be boosted by people who share the same common enemies: “A world that includes self-proclaimed and loudly advertised Evil people running about represents a great boon to those who are hungry to fight for the Good. Without Evil people, who could be Good?” [45].

It is an open question to which extent negative social behavior in real society deviates from the behavior of humans in our game society. Due to lack of other high-frequency analyses on large-scale negative tie networks, it remains to be established whether the above findings can be referred to as ‘universal’.

## 6.5 Differences in network types

Different network types have different properties [157, 170], and show a different evolution of these properties. The clear asymmetries of evolutions of reciprocity, clustering per degree, and assortativity between friend and enemy networks are obvious, Fig. 6.2 (j), (k), and (l). These asymmetries originate from differences in the corresponding formation processes. Intuitively, someone either is your enemy or not – which is apparent to determine and declare –, but friendship comes in several shades of gray more likely being dependent on long-time social dynamics. This intuition is confirmed by [137]: “The evolution of negative relationships may be very different from positive relationships. Friendship development is viewed as a gradual process. According to social penetration theory [...], friendship development proceeds from superficial interaction in narrow areas of exchange to increasingly deeper interaction in broader areas. [...] Qualitative work indicates that negative relationship development is a much faster process that tends to lead to the other person being included in coarse-grained categories such as “rival” or “enemy”. Thus, the formation of negative relationships is not the mere opposite of the way that positive relationships form.” Properties of and interactions between several of the measured networks are discussed in detail in [203].

Further, we observe that reciprocities as well as assortative mixing coefficients need much more time to reach steady state in friend networks than in other networks. It seems to be clear why equilibria of the properties of PM networks are reached fast: Writing or responding to a PM is an administrative, immediately executable action to a large extent independent of social ties. On the other hand, choosing your friends or the reciprocation of friendships is a much more delicate operation, necessitating in-depth considerations about e.g. social balance.

Another issue is the possible ‘import’ of social ties. In a recent study of online social networks, the effect of a transition from assortative social networks to disassortative or non-assortative networks has been observed [117]. While there it could not be checked if networks are at their initial stage or not, here we are able to do so (for the Artemis and Pegasus universes). Our following hypothesis coincides with the reasoning of [117]: At the initial stage of social online networks, some people may ‘import’ their social relations from existing previous ones. This is possible since many players who previously played in Orion created characters in Artemis and Pegasus when these new universes opened. At later times (in our case 100 or more days), evolving new relations of online interaction take overhand and decrease assortativity, down to disassortativity. The same effect may lead to the observed changes in clustering coefficients, reciprocity, and other network properties.

## 6.6 Summary of social network dynamics

We explore novel possibilities of a quantification of human group-behavior on a fully empirical and falsifiable basis. We study network structure and its evolution of

several social networks extracted from a massive multiplayer online game data set. Practically all actions of all 300,000 players over a period of three years are available within one unique and coherent source. Players live a second economic life and are typically engaged in a multitude of social activities within the game. With this data we can show for the first time marked differences in the dynamics of friend and enemy dynamics. A detailed analysis of high-frequency log files focuses on three types of social networks and allows to subject a series of long-standing social-dynamics hypotheses to empirical tests with extraordinary precision. Along these lines we propose two social laws in communication networks, the first expressing betweenness centrality as the inverse square of the overlap, the second relating communication strength to the cube of the overlap. These laws not only provide strong quantitative evidence for the validity of the Weak ties hypothesis of Granovetter, they are also fully falsifiable. Our study of triad significance profiles confirms several well-established assertions from social balance theory. We find overrepresentation (underrepresentation) of complete (incomplete) triads in networks of positive ties, and vice versa for networks of negative ties. We measure empirical transition probabilities between triad classes and find evidence for triadic closure, again with unprecedented precision. We compare our findings with data from non-virtual human groups and conclude that online game communities should be able to serve as a model for a wide class of human societies. We demonstrate the realistic chance of establishing socio-economic laboratories which allow to measure dynamics of our kind at levels of precision so far only known from the natural sciences.

In recent work on an extended data set [203], compiled in the next Chapter, further non-trivial structures of social systems in our online world were uncovered, reasserting its potential for harnessing knowledge about the fundamental nature of the organization and dynamics of human society. There we first study the basic network properties of the newly added network types, relate them to the old ones, and finally analyze them *in conjunction* to uncover the interdependencies between the different network types, showing the nontrivial organization of the social system in Pardus.

# **Chapter 7**

## **Multi-relational organization of societies: Network–network interactions**

Human societies can be regarded as large numbers of locally interacting agents, connected by a broad range of social and economic relationships. These relational ties are highly diverse in nature and can represent e.g. the feeling a person has for another (friendship, enmity, love), communication, exchange of goods (trade) or behavioral interactions (cooperation or punishment). Each type of relation spans a social network of its own. A systemic understanding of a whole society can only be achieved by understanding these individual networks and how they influence and co-construct each other. The shape of one network influences the topologies of the others, as networks of one type may act as a constraint, an inhibitor, or a catalyst on networks of another type of relation. For instance, the network of communications poses constraints on the network of friendships, trading networks are usually constrained to positively connedted interactions such as trust, and networks representing hostile actions may serve as a catalyst for the network of punishments. A society can therefore be characterized by the superposition of its constitutive socio-economic networks, all defined on the same set of nodes. This superposition is usually called multiplex, multi-relational, multi-modal or multivariate network, see Fig. 7.1. The study of small-scale multiplex networks has a long tradition in the social sciences [215], and has been applied to areas such as homophily in social networks [152], the effect of combined interactions on an agent's behavior [76] and the non-trivial inter-relation between family and business networks [174]. Multiplexity is thought to play an important role in the organization of large-scale networks. For example the existence of different link types between agents explains the overlap of community structures observed in social networks, where nodes may belong to several communities, each associated to one different type of interaction [7, 80]. Methodological work on multiplex networks includes the development of multiplex community detection [161], clustering [196] and other network analysis algorithms [187]. The role of multiple relation types in measured social networks has recently

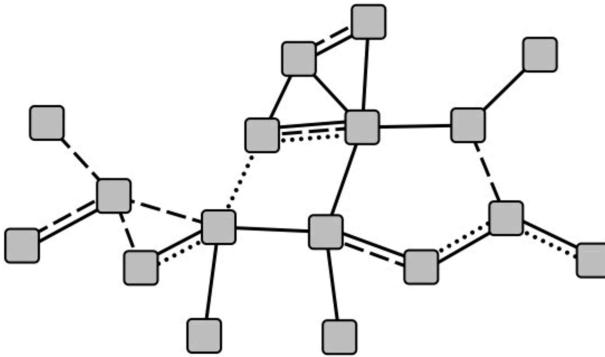


Figure 7.1: Multiplex networks consist of a fixed set of nodes connected by different types of links. This multi-relational aspect is usually neglected in the analysis of large social networks. In our MMOG data set, six types of social links can exist between any two players, representing their friendship or enmity relations, their exchanged private messages, their trading activity, their one-to-one aggressive acts against each other (attacks), and their placing of head-money (bounties) on other players as e.g. means of punishment.

been investigated across communication media [24], in an online game [119], as well as in ecological networks [92].

The increased use of large-scale data sets on specific human behaviors in the past decade [140] suffers from the drawback of a relatively coarse-grained representation of social processes taking place between individuals and of blindness in respect to the existence of different types of social interactions. For example in most works on email [169] or mobile phone networks [138, 173], the existence and weight of a link is determined by the volume of information exchanged between two individuals. Although nodes can be generally well characterized (age, sex, zip code, etc.), the corresponding type of interaction (e.g. family or work interaction), is usually unavailable in the data and can only be inferred from behavioral patterns [74]. Moreover, research on large social networks has focused on single types of interaction only, e.g. phone or email communication, and has ignored the wide spectrum of human interactions in real life [152]. Whenever inter-dependencies and feedbacks between multiple relational interactions are significant, an aggregate representation of the different network types, or the representation of one single type will lead to a biased and misleading characterization of the organization of the system.

In this Chapter we present an attempt toward fully characterizing the multiplex nature of a large-scale social system. To this end we analyze a data set from the Pardus society as in Chapter 6, extended with data on three types of one-to-one relations, see below. Having become extremely popular over the past years, there exists a multitude of large-scale online games – often played by thousands, sometimes even millions. These games offer the possibility to experience alternative lives in which players can engage in different types of social interactions, ranging from establishing friendships and economic relations to the formation of groups, alliances, fighting and even waging of war [205], see Chapter 3.

---

The data allows to identify the nature of one-to-one interactions between players; the topological properties of the corresponding networks – defined on the same set of agents – can be studied. We show in the following that different types of interactions are characterized by distinct connectivity patterns. Exploring the inter-dependence of the different networks reveals how multiplexity shapes the organization of the system at different levels, from the stability of local motifs to the global overlap between the networks. Moreover, the existence of positively and negatively connotated interactions between players, e.g. through declared friendship or enmity, allows to analyze the organization of the system from the point of view of signed networks [215]. Within this framework it becomes possible to experimentally verify structural balance [75], a long-standing theory in social psychology [107] proposed for understanding emergence of conflict and tension in social systems [42]. The central idea behind structural balance is that some configurations of signed motifs, i.e. local ‘building blocks’ of networks containing positive and/or negative ties, are socially and psychologically more stable than others and are therefore more likely to be present in human societies. By measuring the dynamics and abundance of signed triads (sets of three nodes connected by positive or negative links), we perform a large-scale validation of structural balance and provide insights indispensable for a realistic modeling of conflicts.

To our knowledge, the only large-scale data set incorporating multiple interactions is the Facebook network of Lewis et al. [145]. Pardus offers several advantages over this Facebook data. The Facebook network consists of three types of interactions between users: declared friendship relationships, picture friendships (being tagged in an online photo by a user), and dorm/roommate friendships. However, two of these three types of interactions lead to a tainted representation of the social system. First, the friendship network of Facebook is known to be biased by the visibility of the friends of a user on its webpage [93]. In Pardus, friend and enemy lists are *completely private*, meaning that no one except the marking and marked players have information about positive or negative ties between them. Our data thus represents a more realistic social situation, in the sense that social ties are not immediately accessible to the public but need to be found out by communication with or by careful observation of others. Second, dorm membership is obtained from the projection of a bipartite network. This procedure is known to distort the number of cliques in a network [171]. For this reason, we only focus on one-to-one interactions between players and discard indirect interactions such as the participation to a chat.

The data set considered in this Chapter contains practically all actions of all players of the MMOG Pardus, of the game *universe*, Artemis, in which  $N = 18,819$  players have interacted with at least one other player over the first 445 consecutive days of this universe’s existence. This data set contains longitudinal and relational data allowing for an almost complete and dynamical mapping of multiplex relations of an entire society. The data is free of interviewer effects since agents are not conscious of their actions being logged. Measurement errors which usually affect reliability of survey data [41], are practically absent. The longitudinal aspect of the

Table 7.1: Network properties. Properties of directed networks: number of nodes  $N_\alpha$  (connected to at least one link), number of directed links  $L_\alpha^{\text{dir}}$ , reciprocity  $r_\alpha$  and in-degree/out-degree correlation  $\rho(k_\alpha^{\text{in}}, k_\alpha^{\text{out}})$ . Greek indices mark network types. Properties of the corresponding undirected networks: number of undirected links  $L_\alpha^{\text{undir}}$ , average degree  $\bar{k}_\alpha$ , clustering coefficient  $C_\alpha$  and ratio to the corresponding random graph clustering  $C_\alpha/C_{\text{rand}}$ . The networks, when considered as separate entities, present distinct types of organization depending on the nature of the interactions. Positively (negatively) connoted links present high (low) values of  $r_\alpha$ ,  $\rho(k_\alpha^{\text{in}}, k_\alpha^{\text{out}})$  and  $C_\alpha$ .

	Positive ties			Negative ties			Envelope (all os)
	Friends	PMs	Trades	Enemies	Attacks	Bounties	
directed	$N_\alpha$	4,313	5,877	18,589	2,906	7,992	2,980
	$L_\alpha^{\text{dir}}$	31,929	185,908	796,733	21,183	57,479	5,096
	$r_\alpha$	0.68	0.84	0.57	0.11	0.13	0.20
	$\rho(k_\alpha^{\text{in}}, k_\alpha^{\text{out}})$	0.88	0.98	0.93	0.11	0.64	0.31
undirected	$L_\alpha^{\text{undir}}$	21,118	107,448	568,923	20,008	53,603	4,593
	$\bar{k}_\alpha$	9.79	36.57	61.21	13.77	13.41	3.08
	$C_\alpha$	0.25	0.28	0.43	0.03	0.06	0.01
	$C_\alpha/C_{\text{rand}}$	109.52	45.71	131.95	6.13	37.27	13.88
							109.93

data allows for the analysis of dynamical aspects such as the emergence and evolution of network structures. Finally, it is possible to extract multiple social relationships between a fixed set of humans. We focus on the following set of six types of one-to-one interactions between players: friendship and enmity relations, private message (PM) communication, trades, attacks, and revenge/punishment through head money (bounties). We label these networks by Greek indices:  $\alpha = 1$  refers to friendship networks,  $\dots, \alpha = 6$  to bounties. We focus on one-to-one interactions only (without projections as e.g. used in [145, 171]) and discard indirect interactions such as mere participation in a chat.

Friendship and enmity networks are taken as snapshots at the last available day 445. All other networks are aggregated over time, meaning that whenever a link existed within day 1 and 445, it is counted as a link. For simplicity, we use unweighted, directed networks. Further, we define undirected networks as follows: A link exists between nodes  $i$  and  $j$  if there exists at least one directional link between those nodes. We construct triads (motifs of three connected nodes [215]) from undirected links.

## 7.1 Nature of the various networks

Different types of connectivity patterns may signal different organization principles behind the formation of networks [9, 165]. Statistical properties of the six networks, when considered as separated entities, are collected in Table 7.1. There,  $N_\alpha$  is the number of nodes in the network  $\alpha$ ,  $L_\alpha^{\text{dir(undir)}}$  is the number of (un)directed links. Reciprocity is labeled by  $r_\alpha$ ,  $\rho(k_\alpha^{\text{in}}, k_\alpha^{\text{out}})$  is the correlation of in- and out degrees within the  $\alpha$  network. Average degree, clustering coefficient, and clustering coefficient with respect to the corresponding random graph are marked by  $\bar{k}_\alpha$ ,  $C_\alpha$  and  $C_\alpha/C_\alpha^{\text{rand}}$ , respectively. For definitions of the measures, see Section 2.3.5.

### Positive links are highly reciprocal, negative links are not

Table 7.1 shows that networks with a positive connotation (friendship, private messages (PMs) and trades) are strongly reciprocal [90] (see Section 2.3.5), in the sense that node pairs have a high tendency to form bi-directional connections, while networks with a negative connotation (enmity, attack and bounty) all show significantly smaller reciprocity. Low reciprocation in enemy networks may partially be explained by deliberate refusal of reciprocation to demonstrate aversion by total lack of response [205]. For attack networks, it may originate from the asymmetry in the strength of the players (a strong player is more likely to attack a weaker player to secure a win). Asymmetry in negative relations is confirmed in the correlations between node in-degrees and out-degrees. Positive links are almost balanced in the in- and out-degrees,  $\rho \sim 1$ , whereas negative links show an obvious suppression in doing to others what they did to you.

### Power-law degree distributions indicate aggressive actions

Studying cumulative in- and out-degree distributions, we find pronounced power-law distributions for aggressive behavior, i.e. attacking (out-degree for attacks), being declared an enemy (in-degree for enmity), and punishing/being punished (out- and in-degree for bounty). Power-laws are absent for positive (friendship, communication, trade) and passive links (being attacked), see Fig. 7.2. This discrepancy in degree distributions hints at qualitatively different link-growth/rewiring processes taking place in positive tie networks compared to the negative ones. For example, the classic network growth model of preferential attachment [19] leads to a power-law degree distribution. As we have shown in [205], the growth of enemy networks is well characterized by this model, but not the growth of friend networks.

### Positive links cluster

From Table 7.1 it is clear that positively connedted links show higher clustering coefficients than negatively connedted ones. High values of the clustering coefficient are intuitively plausible for positive interactions due to their cohesive nature and the benefits of dense sub-graphs for better performance [54]. The significantly lower values of clustering for negative values suggests that mechanisms such as triadic closure [180] are not dominant for negative interactions, and has its origin in the balance of signed motifs, see later.

The independent analysis of the different networks reveals distinct types of organization which depend on the nature of the links. It is crucial to account for these distinct topological properties in models for the dynamics of cooperation and conflict in human societies. To demonstrate the danger of not differentiating between types of interactions we include data on the “envelope network” in Table 7.1. Neglecting the nature of social ties and mixing different interactions (even within the same data set) may result in a gross mis-representation of the system, in this case at least by losing the typical low reciprocity and clustering observed in negative tie sub-networks.

For a detailed analysis of the time-evolution of single network properties on the same data set (first 445 days in the Artemis game universe), see [205]. There several ‘aging’ or ‘maturing’ effects were reported, such as a decrease of the clustering coefficient and reciprocity in friend networks over time.

## 7.2 Network–network interactions

Due to strong interactions between different social relations, a next level of complexity enters when considering the co-existence of different types of links [141]. From now on, we only focus on undirected versions of the networks. To quantify the resulting inter-dependencies between pairs of networks, we follow two approaches.

On one hand, we focus on the link-overlap between networks and calculate the Jaccard coefficient  $J_{\alpha\beta}$  between two different sets of links  $\alpha$  and  $\beta$ . The Jaccard co-

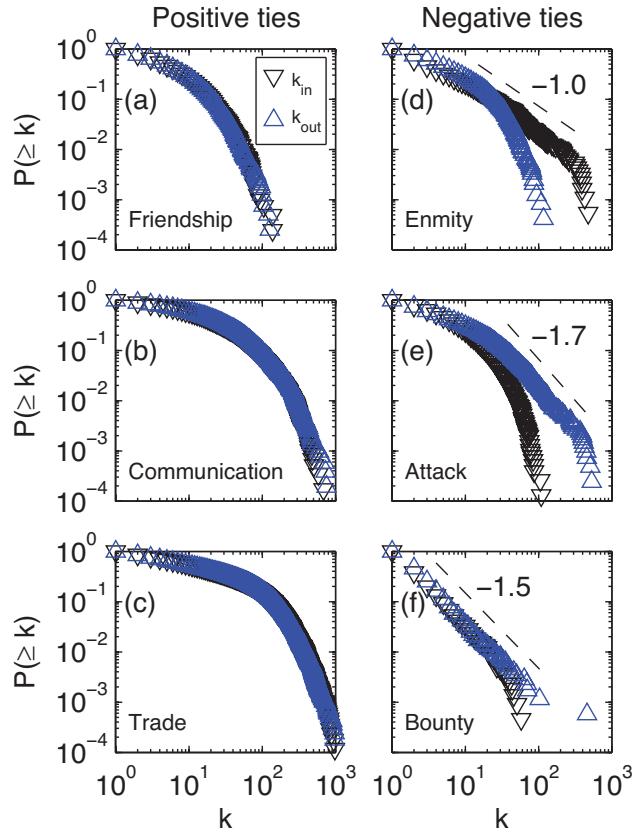


Figure 7.2: Cumulative in-degree and out-degree distributions for the six types of networks spanning the same set of agents: (a) friendship, (b) communication, (c) trade, (d) enmity, (e) attack, (f) bounty. Note the differences between in- and out-degree distributions and the presence of power-laws (with cutoffs) for negatively conned interactions (right column), which are absent for positive ties (left column). It is immediately clear that topological properties of social networks depend strongly on the nature of their ties. Ignoring this multi-relational composition can lead to loss of essential information.

efficient quantifies the interaction between two networks by measuring the tendency that links simultaneously are present in both networks, see Section 2.3.5.

On the other hand, we compute correlations  $\rho(k_\alpha, k_\beta)$  between node degrees in different networks, see Section 2.3.5. These coefficients measure to which extent degrees of agents in one type of network correlate with degrees of the same agents in another one. If  $\rho(k_\alpha, k_\beta)$  is close to 1, players who have many (few) links in network  $\alpha$  have many (few) links in network  $\beta$ . Note that both measures might be affected by different network sizes or average degrees. To account for this possibility, we additionally compute the correlations  $\rho(\text{rk}(k_\alpha), \text{rk}(k_\beta))$  between rankings of node degrees, where rk represents rank, see Section 2.3.5. Overlap and correlation quantities provide roughly complementary insights into the organization of social structures. In Fig. 7.3 for all pairs of networks the three measures are shown. Note that no obvious causal directions are implied and that all correlations are positive. From highest to smallest overlap (from left to right), Fig. 7.3 provides the following conclusions:

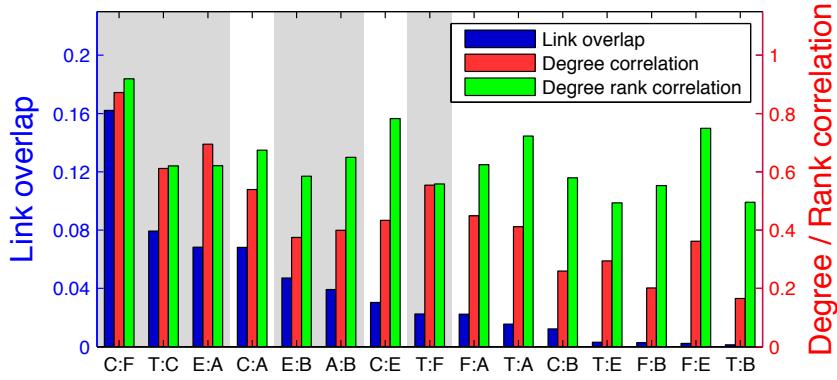


Figure 7.3: Link overlap (Jaccard coefficient), degree correlation  $\rho(k_\alpha, k_\beta)$  and degree rank correlation  $\rho(\text{rk}(k_\alpha), \text{rk}(k_\beta))$  for all pairs of networks (ordered by link overlap), with the notations E for Enmity, F for Friendship, A for Attack, T for Trade, C for Communication and B for Bounty. Pairs of equal connotation (positive-positive or negative-negative) are marked with a gray background. These pairs have high overlaps, while oppositely-connotated pairs have lower overlaps. The various relations are organized in a non-trivial way, suggesting that agents play very different roles in different relational networks.

*Communication–Friendship.* The pronounced overlap implies that friends tend to talk *with each other*. The equally pronounced correlation attests that players who communicate with many (few) others tend to have many (few) friends. The former result was already reported in [205], where a high fraction of communication partners was shown to be friends.

*Trade–Communication.* The high overlap shows that trade partners have a tendency to communicate with each other, while the high correlations shows a tendency of communicators being traders.

*Enmity–Attack.* The high overlap shows that enemies tend to attack each other, or that attacks are likely to lead to enemy markings. The high correlations imply that aggressors or victims of aggression tend to be involved into many enemy relations.

*Communication–Attack.* The relatively high overlap shows that there is a tendency for communication taking place between players who attack each other. The relatively high correlation implies that players who communicate with many (few) others tend to attack or be attacked by many (few) players. Aggression is not anonymous, but accompanied by communication.

*Enmity–Bounty and Attack–Bounty.* Similar to Enmity–Attack.

*Communication–Enmity.* Similar to Communication–Attack.

*Trade–Friendship.* Similar to Trade–Communication, however with a smaller

overlap. It is more difficult for traders to become friends than to just communicate.

*Friendship–Attack.* The low overlap shows that attacks tend to *not* take place between friends, or that fighting players do *not* tend to become friends. The relatively high correlations mean that players with many (few) friends attack or are attacked by many (few) others.

*Trade–Attack.* Similar to Friendship–Attack.

*Communication–Bounty.* Similar to Communication–Attack and Communication–Enmity, however with much smaller overlap and degree correlations.

*Trade–Enmity.* For this and all other interactions, overlap vanishes. Players who trade with each other almost never become enemies and vice versa.

*Friendship–Bounty.* Similar to Communication–Bounty.

*Friendship–Enmity.* The degree (rank) correlation is substantial, suggesting that players who are socially active tend to establish both positive as well as negative links. However, the vanishing overlap shows the absence of ambivalent relations. Friends are never enemies.

*Trade–Bounty.* This interaction shows the smallest values for all three properties, which could be due to substantial differences in network sizes. The relatively small correlation may suggest that players who are experienced in trade have a tendency to *not* act out negative sentiments by spending money on bounties.

The exact values of the two correlation measures have to be interpreted with some caution. High values might be biased by e.g. the time a player spent in the game or by ignoring link weights for the number of exchanged private messages or traded money. Nevertheless, low values of  $\rho(k_\alpha, k_\beta)$  indicate that hubs in one network are not necessarily hubs in another (see e.g. the Trade–Enmity case), suggesting that agents play very different roles in different relational networks. For example agents can be central for flows of information but peripheral for flows of goods [106]. In the following we present further relations between the above network–network measures and study their evolutions in time.

Correlating network–network measures reveals a strong relation between link overlap and degree correlation ( $\rho = 0.88$ , p-value  $< 10^{-5}$ ), see Fig. 7.4 (a). Pairs of networks of the same connotation have a higher overlap than oppositely-connotated pairs; a similar tendency for degree correlation is apparent. A correlation between link overlap and degree rank correlation is also present, however with lower significance ( $\rho = 0.63$ , p-value  $< 0.01$ ), see Fig. 7.4 (b). We mark pairs including a communication network as neutral, since messages may involve both positively or negatively connotated content.

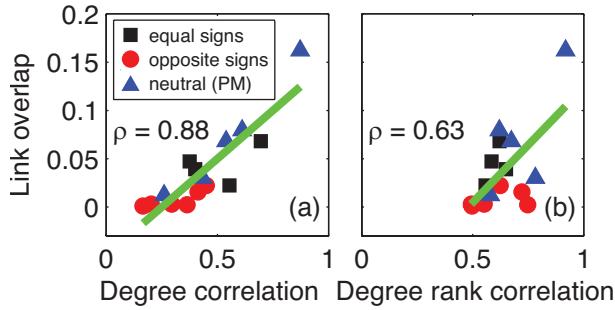


Figure 7.4: Correlation between the network–network measures of (a) link overlap and degree correlation ( $p$ -value  $< 10^{-5}$ ), (b) link overlap and degree rank correlation ( $p$ -value  $< 0.01$ ). Pairs of networks with equal connotation (positive-positive or negative-negative) have a higher overlap than oppositely-signed pairs, with a similar tendency for degree correlation. We mark pairs including a communication network as neutral, since messages may involve both positively or negatively connotated content.

To assess to what extent network–network properties of link overlap, degree correlation, and degree rank correlation change over time, we show these properties at days 150, 300, and 445 for all pairs of networks in Fig. 7.5. Here, accumulated networks, i.e. all except friendship and enmity networks, are accumulated over days 1 to 150, over days 1 to 300, and over days 1 to 445, respectively. Friendship and enmity networks are taken at these times. The number of players involved in the envelope network (i.e. in any relation) changes from 9,862 to 15,103 to 18,819 at these points in time, respectively. Changes are relatively small, except for degree correlations of pairs including bounty networks. Overlap values generally tend to decrease slightly over time.

### 7.3 Large-scale empirical test of structural balance

To analyze the nature of large-scale multiplex networks and to draw conclusions from our observations, we need to address an issue that is usually obsolete for experiments on small social systems. When considering different types of interactions between students of a class or diplomatic positions between countries, it is reasonable to assume that all agents in the network are aware of each other's existence. In large-scale social networks, in contrast, the absence of any type of link between two nodes may either correspond to the existence of an indifferent/neutral interaction, or to the absence of any past and present contact between both agents. The fact that agents only know a fraction of the total set of agents is typical of sparse networks and originates from the finite capacities of its nodes, i.e. agents have limited time, resources, and cognitive capacities. In the Pardus networks this finiteness is affirmed by the observation that out-degrees of friendship and enmity networks have an upper bound, limited by the Dunbar number of about 150 [205], presumed as a natural limit for social ties humans are able to sustain [73]. The average degree  $\bar{k}_\alpha$  is well below

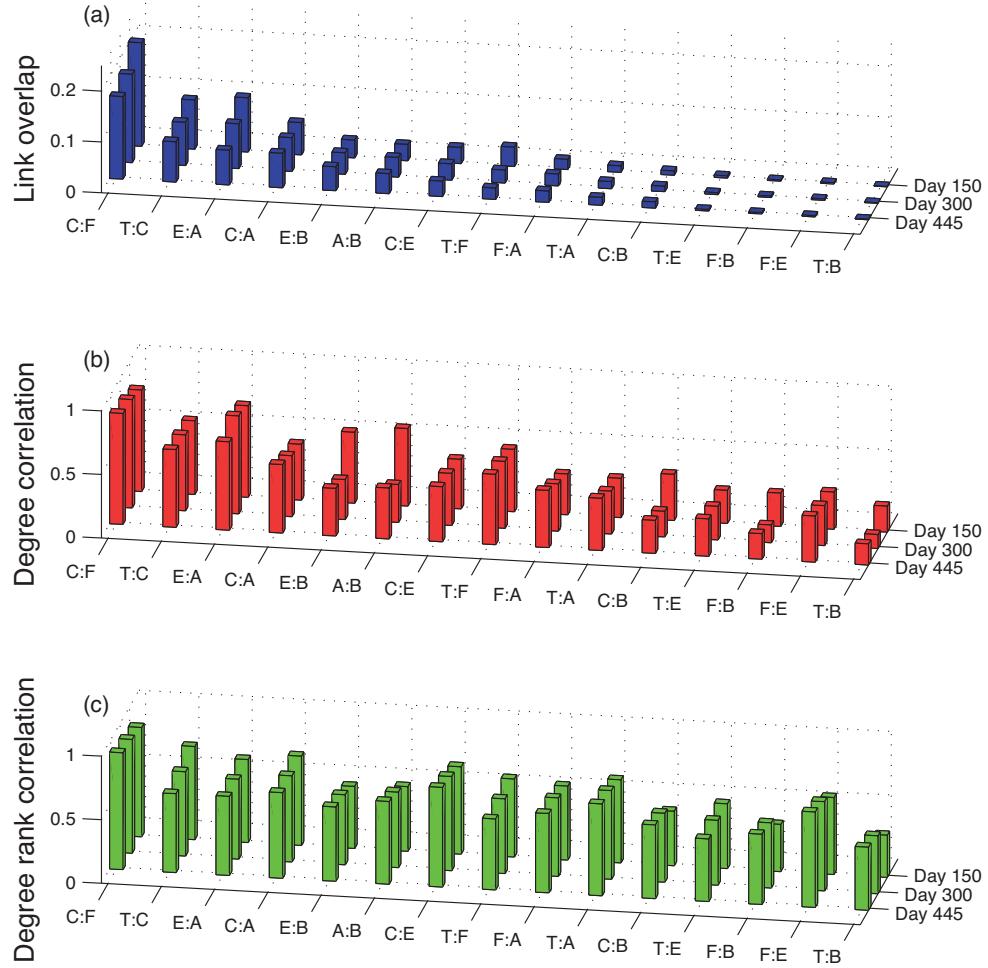


Figure 7.5: Evolving network–network properties (a) link overlap, (b) degree correlation, (c) degree rank correlation, on the days 150, 300, and 445, for all pairs of networks, with the notations E for Enmity, F for Friendship, A for Attack, T for Trade, C for Communication and B for Bounty. Changes are relatively small, except for degree correlations in the E:B, A:B, C:B, and F:B interactions.

$O(100)$ , for all network types  $\alpha$ .

In the following we assign  $+(-)$  to a positively (negatively) connotated link. All friendship links have a value of  $+$ , all enemy links  $-$ . Social balance focuses on signed triads where the sign of a triad is the product of the signs of its three links. Social balance theory – in its strong form [42] – claims that positive triads are ‘balanced’ while negative triads are ‘unbalanced’, see Fig. 7.6. Unbalanced triads are sources of social stress and therefore tend to be avoided by agents when they update their personal relationships. From a physics point of view, the resulting dynamics can be viewed as an energy minimization process which may lead to jammed states [10] due to a rugged energy landscape [150]. There is a ‘weak formulation’ of structural balance [61] which postulates that triads with exactly two positive links are under-

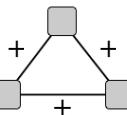
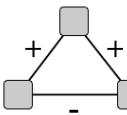
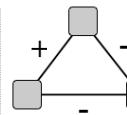
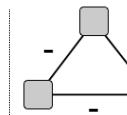
				
Strong formulation of balance	B	U	B	U
Weak formulation of balance	B	U	B	B
$N_\Delta$	26,329	4,428	39,519	8,032
$N_\Delta^{\text{rand}}$	10,608	30,145	28,545	9,009
$z$	71	-112	47	-5

Figure 7.6: Different types of signed triads, balanced (B) or unbalanced (U) according to the strong or weak formulation of structural balance. We show the number of each type of triad  $N_\Delta$  in the friendship-enmity multiplex network, the expected number  $N_\Delta^{\text{rand}}$  of such triads when averaged over 1000 sign-randomizations and the corresponding  $z$ -score (see Section 2.3.5). Triads  $+++$  and  $--$  are over-represented,  $++-$  triads are underrepresented with extraordinary significance.

represented in real networks, while the three other kinds of triads should be much more abundant. In the weak formulation only situations where “the friend of my friend is my enemy” are unstable, whereas in the strong form of structural balance, “the enemy of my enemy is my enemy” is also unstable, see Fig. 7.6.

To test social balance we focus on the multiplex network of friendship and enmity interactions. The number of different types of triads are labeled  $N_\Delta$ . They are compared to the expected number of such triads in the following null model (reshuffled signs of links,  $N_\Delta^{\text{rand}}$ ).

The aspect of a null model becomes important when assessing the relevance of topological structures in a network. A standard procedure consists in comparing this observation against similar observations in null models, i.e. randomized versions of the original network under adequate constraints [151]. In order to test predictions of structural balance theory, we focus on friendship and enmity relations, and leave aside other types of interactions. In a first step, we remove the negligible number of ambiguous links (links between players where one marks the other as friend but is marked back as enemy). Our strategy is now to compare the numbers  $N_{\Delta_i}$  of triads with  $i$  positive links with the expected numbers  $N_{\Delta_i}^{\text{rand}}$  of triads in a null model. A standard choice for a null model consists of random graphs with fixed degree sequences. It has been applied for each network separately in Table 7.1, where we observe that friendship and enmity networks are both more clustered than a random graph. However, this choice is not appropriate to test the arrangement of positive and negative links on the set of existing relations between agents – a reshuffling of topology by keeping degrees fixed would for example considerably change the number of triads which we want to keep fixed. For this reason we define a null model by keeping the topology fixed and by randomly assigning the  $L_+$  plus-signs

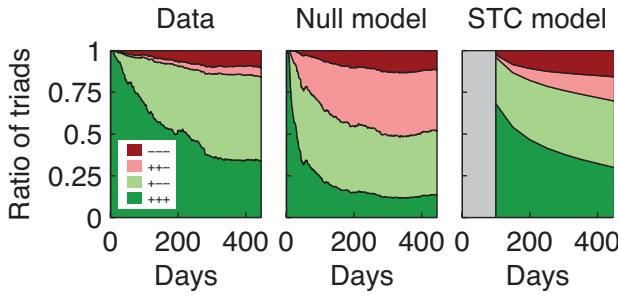


Figure 7.7: Ratio of signed triad types over time. Left: Measured in the data. Center: Expected in the random null model. Right: Simulation of signed triadic closure (STC) with a model based on wedge transition rates. Initial condition: Measured network of day 100. All ratios measured in the data deviate significantly from ratios in the null model, except for the  $--$  triads. The STC model reproduces the observed ratios considerably better.

and  $L_-$  minus-signs on the existing links, where  $L_+$  ( $L_-$ ) are the original numbers of friendship (enmity) links respectively.  $N_{\Delta_i}^{\text{rand}}$  is measured by averaging over 1000 realizations of the null model. Moreover, the deviation of the data from the reshuffled data is evaluated by the so-called  $z$ -score:

$$z_i = \frac{N_{\Delta_i} - N_{\Delta_i}^{\text{rand}}}{\sigma_{\Delta_i}}, \quad (7.1)$$

where  $\sigma_{\Delta_i}$  is the standard deviation of the number of triads  $\Delta_i$  from the 1,000 reshuffled realizations.

Given the ratio  $p := \frac{L_+}{L_++L_-}$  of positive to all links in a signed network, the expected ratio of triad types in the sign-shuffled null model is, following straightforward combinatorial arguments,  $p^3$ ,  $3(1-p)p^2$ ,  $3(1-p)^2p$ , and  $(1-p)^3$  for triad types +++, ++-, +-+, and ---, respectively.

In Fig. 7.6, a standard measure of statistical deviation, the  $z$ -score (see Section 2.3.5), shows that +++ and +-+ triads are heavily over-represented, while ++- triads are heavily under-represented with respect to the unstructured case. Triads of type --- are under-represented to a lesser degree than the three other types, favoring the weak formulation of structural balance over Heider's original formulation of balance theory. It is obvious that triads are characterized by different levels of stability. The robustness of these results is further confirmed by examining the time evolution of the number of triads in friendship/enmity networks over all 445 days, Fig. 7.7.

A detailed dynamical analysis of our data further reveals that a vast majority of changes in the network are due to the creation of new positive and negative links, not due to switching of existing links from plus to minus or vice versa. We illustrate this dominance of link destruction and creation over sign switching on the dynamics of the following triadic structures. Let us define a *wedge* as a signed undirected triad with two links, i.e. a triad with one link missing (a 'hole'). There are three possible wedge types: ++, +-, --. We measure day-to-day transitions from wedges to other

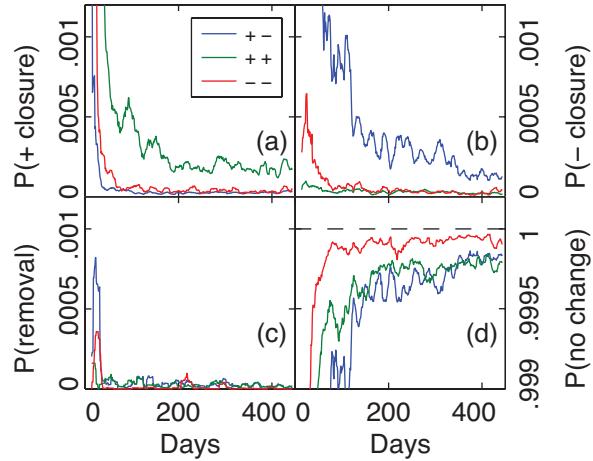


Figure 7.8: Measured day-to-day transition probabilities from wedges to staying the same or to other triadic structures, in the multiplex network consisting of friend and enmity relations. For visual clarity a moving average filter with a time window of 14 days was applied. (a) Probabilities of wedges being closed with a positive link. (b) Probability of a wedge being closed with a negative link. (c) Probability of a wedge having one or both links removed. (d) Probability of a wedge staying unchanged. Note how  $++$  wedges have a clear preference of being closed by a positive link, while  $+-$  wedges have a clear preference of being closed by a negative link. Wedges of type  $--$  have no sign preference for closure and are more likely to remain unchanged than other wedge types.

possible triadic structures. In the vast majority of all cases ( $> 99.9\%$ ), a wedge stays unchanged, see Fig. 7.8 (d). In case of change, most often a hole is closed by either a positive or a negative link, see Fig. 7.8 (a) and (b). The removal of a link is less frequent, see Fig. 7.8 (c); sign switches almost never occur. This result is in marked contrast with many dynamical models of structural balance [10] which assume that a given social network is fully connected from the start and that only the link-signs are the relevant dynamical parameters, which evolve to reduce stress in the system. Our observation underpins that network sparsity and growth are fundamental properties and they need to be incorporated in any reasonable model of dynamics of positive and antagonistic forces in social systems. In full agreement with the results shown in Fig. 7.6, wedges of type  $++$  close preferentially (about 7 times more likely) with a positive link, wedges of type  $+-$  close preferentially (about 11 times more likely) with a negative link. There is no clear sign preference in the closure of type  $--$  wedges.

Indeed, by measuring all day-to-day transitions from wedges (triads with two links, with the possible forms  $++$ ,  $+-$ , and  $--$ ) to the other triadic structures ( $+++$ ,  $++-$ ,  $+--$ ,  $--+$ ,  $++$ ,  $+-$ ,  $--$ ,  $+$ ,  $-$ , and the empty triad) we clarify some of the mechanisms which lead to the social balance observed in Pardus. We measure the following possible transition types: A wedge stays the same, closes with a positive/negative link (with the original links unchanged), has one or both links removed, or has the sign of one or both links switched. Figure 7.8 (a), (b), (c) and

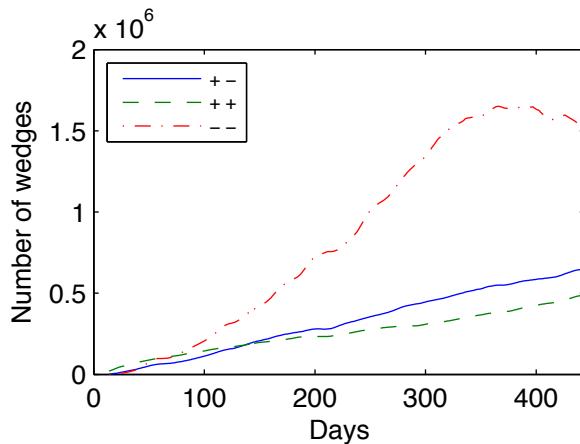


Figure 7.9: Evolution of number of wedges of each type, in the multiplex network consisting of friend and enmity relations. For visual clarity a moving average filter with a time window of 14 days was applied.

(d) show the daily transition probabilities of + closures, - closures, link removals, and of no change, respectively, normalized by the total number of wedges of the corresponding type on that day. Due to lack of notability no panel for the transition type of switching links (with a probability less than a tenth of that of link removal, on average) was included in Fig. 7.8.

Wedges of type ++ close preferentially with a positive link, see green line in Fig. 7.8 (a), wedges of type +- with a negative link, see blue line in Fig. 7.8 (b). These probabilities are decreasing over time and seem to eventually level out. There is no clear sign preference in the closure of type -- wedges (red lines). These observations consistently explain the social balance results shown in Fig. 7.6 and Fig. 7.7. Further, note that -- wedges are much more likely to remain unchanged than other types of wedges, see Fig. 7.8 (d). We conclude that the mechanism of triadic closure [180] has a much weaker influence as a driving force in purely negative tie networks than in positive tie or signed networks.

Figure 7.9 depicts the total number of wedges of each type, for every day. Note how the majority of wedges is of type --, although there are more positive than negative links, see Table 7.1. Also the growth rate for -- wedges is higher than for the other two types (until about day 350, where the number of -- wedges starts to equilibrate). This seemingly paradoxical circumstance is consistent with the marked differences in clustering coefficients, see Table 7.1. It is further consistent with the observation that a number of aggressive players frequently offend many others and consequently get marked as enemy by unconnected players [205]. Since the clustering coefficient measures the ‘closedness’ of all possible triads, a high clustering coefficient in friendship networks implies a relatively small number of ++ wedges, whereas a low clustering coefficient in enmity networks implies a relatively high number of -- wedges.

For assessing to what extent network growth is driven by the closure of triads, we

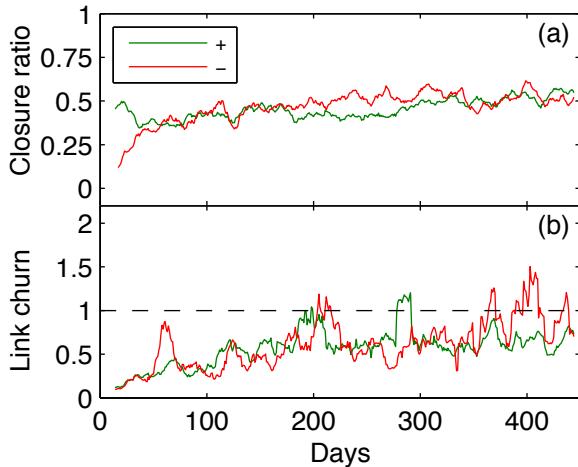


Figure 7.10: (a) Closure ratio, defined as the number of new links which closed at least one wedge divided by all new links, here for each day. For visual clarity a moving average filter with a time window of 14 days was applied. (b) Link churn, defined as the number of links removed divided by the number of links added, here for each time window of 14 days. Evolutions of these ratios are depicted for positive and negative links, within the multiplex network consisting of friend and enmity relations.

define the *closure ratio* as the number of newly added links which close at least one wedge, divided by the number of all new links, over a certain time-window during the evolution of the network. The closure ratio lies between 0 and 1; the higher it is the more new links close a wedge. In practice, the closure ratio is strictly smaller than 1, since a number of cases unavoidably do not allow for the possibility of new links closing a wedge (for example the first and second links which are added into an empty network because no wedges exist at that stage). The measured time-evolution of daily closure ratios in the friend-enmity multiplex network is depicted in Fig. 7.10 (a). Over time the ratio slightly increases and seems to level out at around 0.5 for both positive and negative links. We conclude that half of all links added close at least one wedge, while the other half does not. Thus, a model for network growth using only wedge transition rates shown in Fig. 7.8 could only account for the dynamics of about half of the added links.

Another quantity important for modeling social network dynamics is the number of removed links per time. We define the *link churn* as the number of removed links divided by the number of new links, over a given time-window. The churn is non-negative; there are 3 possible cases: i) Growth ( $0 \leq \text{churn} < 1$ ): More new than removed links, ii) Equilibrium ( $\text{churn} = 1$ ): The same number of new as removed links, iii) Shrinkage ( $\text{churn} > 1$ ): More removed than new links. The higher the *churn*, the more links are removed relative to the number of added links. Note that in the majority of classic network growth models, such as preferential attachment [19], no removal of links is assumed ( $\text{churn} = 0$ ) and the effect of churn is ignored. The measured time-evolution of link churns over time windows of 14 days in the friend/enmity network is depicted in Fig. 7.10 (b). Over time *ch* increases

and fluctuates around  $churn = 0.7$  for both positive and negative links (taken over days 200 to 445,  $churn = 0.66$  for friend links and  $churn = 0.72$  for enemy links). Therefore, at the end for every three new links about two links are removed. Since the number of links removed from wedges is much smaller than links added to close wedges, we conclude that many links are removed from triadic structures other than wedges.

## 7.4 A network evolution model of signed triadic closure

We collect empirical transition rates in a transition matrix  $M_{\text{STC}}$ , which we use in a simple dynamical model for *Signed Triadic Closure* (STC). The measured daily transition rates define a transition matrix

$$M_{\text{STC}} = P \begin{pmatrix} ++ \rightarrow + + + & ++ \rightarrow + + - & ++ \rightarrow + + \\ +- \rightarrow + - + & +- \rightarrow + - - & +- \rightarrow + - \\ -- \rightarrow - - + & -- \rightarrow - - - & -- \rightarrow - - \end{pmatrix}$$

$$= \begin{pmatrix} 0.000212 & 0.000029 & 0.999759 \\ 0.000025 & 0.000279 & 0.999696 \\ 0.000040 & 0.000036 & 0.999924 \end{pmatrix},$$

where the entries are the probabilities of a wedge of given type to another triadic structure. Rows 1, 2, and 3 of  $M_{\text{STC}}$  distinguish between wedge types  $++$ ,  $+-$ , and  $--$ , respectively; columns 1, 2 and 3 distinguish between probabilities for closure with either a positive or a negative link, or the probability of no change, respectively. The constant probabilities of columns 1 and 2 are determined by averaging the corresponding evolving probabilities over the days 100 to 445 (this time window was chosen due to a relatively decreased level of fluctuations in transition probabilities, see Fig. 7.8). The third column is one minus the sum of values in column 1 and 2, since we neglect link-removals and sign-switches. With these parameters we design the following network evolution model, to understand STC:

- At time  $t$  pick wedge  $i$  at random (random sequential update)
- Determine the type of wedge  $i$  and close (or do not close) it according to the relevant entry in  $M_{\text{STC}}$
- Pick next wedge until all wedges are updated
- Continue with time step  $t + 1$

As the initial condition we take the observed friendship and enmity multiplex-network at day 100. Simulating this process leads to the results shown in Fig. 7.7 (right), reproducing the ratio of triads in the data considerably better than the null model.

The STC model ignores three possibly important aspects:

- It does not take into account links added by means other than triadic closure. As we have shown above, the closure ratio  $\approx 0.5$ , i.e. only half of all new links are added in the process of triadic closure. For the other half of new links, one could for example take into account the mechanism of preferential attachment [19], for which exponents were reported in Section 6.2.1 and [205].
- The model does not take into account the removal of links. Because removal happens in a relatively high frequency ( $churn \approx 0.7$ ), a possible extension of the model could involve the measurement and implementation of decay rates, i.e. of transition rates from complete signed triads to wedges (or other triadic structures with a smaller number of links). We suspect that balanced triads are more stable than unbalanced triads.
- The model could be expanded to incorporate directed links and in/out-degree distributions, clustering coefficients, assortativity, etc.

Despite these potential shortcomings, with this model we are able to reproduce the empirical observations to a reasonable extent, see Fig. 7.7 (right).

## 7.5 Discussion of the multi-relational organization of societies

Most empirical studies of large-scale social networks focus on node properties [80], for instance to uncover the topological centrality of social agents or patterns of homophily between agents [51], while being blind to the multiple nature of the links connecting agents. In many social systems, however, a proper description of multiplexity is essential to capture the stress caused by different forces acting on social agents and therefore to uncover the principles shaping the large-scale organization of social interactions. For instance, the interaction and co-existence of multiple relations are crucial to describe the emergence of conflict in social systems [35, 52, 146] or the development of trust in commercial networks [62].

Our work begins to quantitatively measure the multi-dimensionality of human relationships. Its results shed light on macroscopic implications of interaction types: Relations driven by aggression lead to markedly different systemic characteristics than relations of non-aggressive nature. Network-network interactions reveal a non-trivial structure of this multi-dimensionality, and how humans play very different roles in different relational networks. The richness of the data set allows to explore the effect of multiple relations on the structure and stability of a large-scale social network, thereby providing a first empirical basis for the modeling of multiplex complex networks. Future research perspectives include different generalizations of structural balance theory, e.g. to a larger set of social relations, to the case of weighted and/or directed networks or to larger motifs, an extension of the concept of modularity for multiplex [161] or signed [211] networks but also dynamical aspects, for instance the dynamics of non-cooperative organizations [100].

# Chapter 8

## Summary and outlook

In this thesis we have presented a new methodological approach, the extraction and detailed study of the data set of an entire human society from a self-developed massive multiplayer online game. To the growing number of studies on large-scale social systems, in particular on societies in online worlds, we have added a unique, data-driven collection of works on exploring social systems with statistical physics. We have demonstrated the efficiency and fruitfulness of studying these systems from a network perspective using recent statistical physics approaches. In particular, we first studied the “building blocks” of the society, the individuals, their subdiffusive motions and actions, then the evolving network topology spanned by their interactions, and finally the interplay between the different types of interaction networks.

In the large-scale social system constituted by the players of our MMOG, we have first analyzed actions and reactions as isolated events, independent of the topology spanned by these (inter)actions. We have shown that between different types of human behavioral sequences, communication is by far the most dominant activity, followed by aggression and trade. Communication events are about an order of magnitude more frequent than attacks and trading events, and feature typical “back-and-forth” patterns, showing the importance of information exchange between humans. It is possible to understand the collective timeseries of human actions of particular types with a mean-reverting log-normal model. Considering positive and negative actions, we uncovered qualitative differences on the action-reaction level, where negative behavior is likely to be *repeated* and positive behavior is likely to *induce a positive reaction*. In a statistical analysis of the actions as world lines in good-bad space we measured high scaling exponents in the mean square displacements, quantifying the high level of persistency in both positive and negative actions. We also related the lifetimes of players to their relative number of negative actions and found a tendency in the system to favor good behavior. Also, bad players are typically dominant, i.e. they perform significantly more actions than they receive. When splitting the behavioral sequences into words of finite lengths, we found a Zipf law, typical for complex systems, as well as the Shannon  $n$ -tuple redundancy growing with word length, reflecting again the non-trivial statistical structure of behavioral sequences.

We next focused on the mobility of human-controlled avatars within their virtual

game universe. The motion of individuals is characterized by an anomalous diffusion process, again typical for a variety of complex systems including real human motions, where the mean square displacement grows as a power law with an exponent of  $\nu = 0.26$ . We uncovered the presence of memory effects and of long-range correlations in time – caused by players who tend to return to locations which were visited *more recently* – which turned out to be the essential ingredient for modeling the motions of players. Using four measured parameters and a non-Markovian approach we formulated a Time-Order Memory (TOM) model which is able to recover the statistical properties of the observed movements. Further, we found that socio-economic regions have a strong influence on the movement patterns of players, leading to a preference to move within such regions than between them. Again, making use of the comprehensive nature of our data set, we were able to quantify this effect with high precision. Moreover, using only the raw movement data and ignoring the existence of socio-economic regions, we were able to recover – almost perfectly – the underlying socio-economic structures by applying methods of community detection. These results could be relevant to some degree for example to understand the role of borders in the spreading of epidemics.

Shifting our focus to the interactions between players, we measured the properties of the underlying evolving network topologies. We found “aging effects” and a densification of our society, where over time individuals come closer and closer to each other in terms of graph distance. We showed for the first time marked differences between properties of friend and enemy networks (most likely related to aforementioned differences in the *dynamics* of positive and negative actions), and subjected a series of long-standing social-dynamics hypotheses to empirical tests with extraordinary precision for social science standards. We proposed two social laws in communication networks, the first expressing betweenness centrality as the inverse square of the overlap, the second relating communication strength to the cube of the overlap. These laws not only provide strong quantitative evidence for the validity of the classical Weak ties hypothesis of Granovetter, they are moreover fully falsifiable. Our study of the “microscopic building blocks”, the triad significance profiles, empirically confirmed several well-established assertions from social balance theory. We found over-representation (under-representation) of complete (incomplete) triads in networks of positive ties, and vice versa for networks of negative ties. We measured empirical transition probabilities between triad classes and found explicit evidence for the hypothesis of triadic closure, again with unprecedented precision.

Extending the number of considered network types to six, including three positive and three negative interactions, we defined and studied the multiplex network underlying the society. In this multi-relational view, we quantified microscopic interactions (couplings) of the constituents on different interaction networks and we assessed the interactions *between* the different network layers, uncovering their non-trivial organizational structures. This work represents first attempts to quantitatively measure the multi-dimensionality of human relationships. Its results provide information on macroscopic implications of interaction types: Relations driven by aggression lead to quite different systemic characteristics than relations of non-

---

aggressive nature. Network-network interactions reveal a non-trivial structure of this multi-dimensionality, and how humans play different roles in different relational networks. The richness of the data set allows to explore the effect of multiple relations on the structure and stability of a large-scale social network, thereby providing a first empirical basis for the modeling of multiplex complex networks.

First steps to “connect the pieces” were provided in the last part dealing with multiplex networks [203]. Clearly much more work is needed, to explore several aspects *in combination*. For example, how does social performance of individuals relate to their economic performance? Which qualities distinguish players who make many friends (or a lot of money) from other players? How are other features (mobility, negative behavior, etc.) related to social and economic factors? Can one quantify the extent of long-range correlations and the tendency to self-organization in this complex system? Since we have the full data of a closed socio-economic unit available and since we “raised” this society and have all contextual information, we believe it may now for the first time become possible to answer such questions with a precision usually only known from the natural sciences.

A large number of other questions may be approached. As a few examples, the data could be useful for network-theoretical questions such as multiplex community detection [161], time-varying graphs [33], or the dynamics of non-cooperative organizations. Future research perspectives may also include different generalizations of structural balance theory, e.g. to a larger set of social relations, to the case of weighted or directed networks or to larger motifs. Further possible issues include an evolutionary game theory approach (tragedy of the commons happening in the game in the form of e.g. resource exploitation), studying systems dynamics questions (e.g. reaction of the society to shocks), or urban planning (distribution of buildings) in the game, etc. It is also straightforward to test epidemics spreading [55, 120] of mobile individuals who are connected in evolving social and economic networks. Of course Pardus also provides the means for comprehensive economic studies, including the study of price formation and cartelization, distribution of wealth and emergence of inequality, game-theoretic considerations from a behavioral economics perspective, a detailed quantification of systematic “irrationality” in human (economic) behavior [128, 129], the emergence of conflict or crime in social systems [35, 52, 146] and of trust in commercial networks [62], clustering of industries, development of cities, or group dynamics and community formation.

## **8. Summary and outlook**

---

# Bibliography

- [1] www.crowdflow.net.
- [2] www.pardus.at.
- [3] [http://www.pardus.at/index.php?section=manual\\_ref050](http://www.pardus.at/index.php?section=manual_ref050), retrieved October 23rd, 2006.
- [4] [http://www.pardus.at/index.php?section=manual\\_intro020](http://www.pardus.at/index.php?section=manual_intro020), retrieved March 3rd, 2008.
- [5] <http://www.pardus.at/index.php?section=about>, retrieved May 5th, 2010.
- [6] <http://www.bbc.co.uk/news/technology-15672416>, retrieved November 10th, 2011.
- [7] Y.Y. Ahn, J.P. Bagrow, and S. Lehmann. Link communities reveal multi-scale complexity in networks. *Nature*, 466:761, 2010.
- [8] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.
- [9] 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, 2000.
- [10] T. Antal, P.L. Krapivsky, and S. Redner. Social balance on networks: The dynamics of friendship and enmity. *Physica D: Nonlinear Phenomena*, 224(1-2):130–136, 2006.
- [11] A. Arenas, A. Fernández, and S. Gómez. Analysis of the structure of complex networks at different resolution levels. *New Journal of Physics*, 10(5):053039, 2008.
- [12] R. Axelrod. The dissemination of culture. *Journal of conflict resolution*, 41(2):203–226, 1997.
- [13] W.S. Bainbridge. The scientific research potential of virtual worlds. *Science*, 317(5837):472, 2007.

## Bibliography

---

- [14] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality: An explanation of the  $1/f$  noise. *Physical Review Letters*, 59(4):381–384, 1987.
- [15] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J.J. Ramasco, and A. Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.
- [16] P. Ball. The physical modelling of human social systems. *Complexus*, 1(4):190–206, 2003.
- [17] A.-L. Barabási. The origin of bursts and heavy tails in humans dynamics. *Nature*, 435:207, 2005.
- [18] A.-L. Barabási. Scale-free networks: A decade and beyond. *Science*, 325(5939):412, 2009.
- [19] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509, 1999.
- [20] A.-L. Barabási, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A*, 272(1-2):173–187, 1999.
- [21] M. Barthélemy. Spatial networks. *Physics Reports*, 499:1–101, 2010.
- [22] R. Bartle. *Designing virtual worlds*. New Riders Games, 2004.
- [23] V. Batagelj and A. Mrvar. A subquadratic triad census algorithm for large sparse networks with small maximum degree. *Social Networks*, 23(3):237–243, 2001.
- [24] N.K. Baym, Y.B. Zhang, and M.C. Lin. Social interactions across media. *New Media & Society*, 6(3):299, 2004.
- [25] A. Bazzani, B. Giorgini, S. Rambaldi, R. Gallotti, and L. Giovannini. Statistical laws in urban mobility from microscopic GPS data in the area of florence. *Journal of Statistical Mechanics: Theory and Experiment*, 05:P05001, 2010.
- [26] V. Belik, T. Geisel, and D. Brockmann. Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X*, 1:011001, 2011.
- [27] J. Berg and M. Lässig. Correlated random networks. *Physical Review Letters*, 89(22):228701, 2002.
- [28] L. Bettencourt, D.I. Kaiser, J. Kaur, C. Castillo-Chávez, and D.E. Wojick. Population modeling of the emergence and development of scientific fields. *Scientometrics*, 75(3):495–518, 2008.
- [29] L. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G.B. West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104(17):7301, 2007.

- [30] C. Biely, K. Dragosits, and S. Thurner. The prisoners dilemma on co-evolving networks under perfect rationality. *Physica D*, 228(1):40–48, 2007.
- [31] C. Biely, R. Hanel, and S. Thurner. Socio-economical dynamics as a solvable spin system on co-evolving networks. *The European Physical Journal B*, 67(3):285–289, 2009.
- [32] C. Biely and S. Thurner. Statistical mechanics of scale-free networks at a critical point: Complexity without irreversibility? *Physical Review E*, 74(6):66116, 2006.
- [33] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, 2006.
- [34] N. Boccara. *Modeling complex systems*. Springer Verlag, 2010.
- [35] J.C. Bohorquez, S. Gourley, A.R. Dixon, M. Spagat, and N.F. Johnson. Common ecology quantifies human insurgency. *Nature*, 462(7275):911–914, 2009.
- [36] B. Bollobás. *Random graphs*, volume 73. Cambridge University Press, 2001.
- [37] M. Boss, H. Elsinger, M. Summer, and S. Thurner. Network topology of the interbank market. *Quantitative Finance*, 4:677–684, 2004.
- [38] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Z. Nikoloski, and D. Wagner. On finding graph clusterings with maximum modularity. In *Graph-Theoretic Concepts in Computer Science*, pages 121–132. Springer, 2007.
- [39] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.
- [40] M. Buchanan. *The social atom: Why the rich get richer, cheaters get caught, and your neighbor usually looks like you*. Bloomsbury USA, 2007.
- [41] P.J. Carrington, J. Scott, and S. Wasserman. *Models and methods in social network analysis*. Cambridge University Press, 2005.
- [42] D. Cartwright and F. Harary. Structural balance: a generalization of Heider's theory. *Psychological Review*, 63(5):277–93, 1956.
- [43] M. Casey. Real economist learns from virtual world. Wall Street Journal Blog, October 2010.
- [44] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, 2009.
- [45] E. Castranova. *Synthetic Worlds: The Business and Culture of Online Games*. University of Chicago Press, Chicago, 2005.

- [46] E. Castranova. On the research value of large games. *Games and Culture*, 1:163–186, 2006.
- [47] E. Castranova. A test of the law of demand in a virtual world: Exploring the petri dish approach to social science. *CESifo Working Paper Series No. 2355*, 2008.
- [48] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.F. Pinton, and A. Vespignani. Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS One*, 5(7):e11596, 2010.
- [49] A. Chatterjee, S. Sinha, B.K. Chakrabarti, A. Šurda, B.M. Borkent, S.M. Dammer, H. Schoenherr, G.J. Vancso, D. Lohse, N. Yannopoulou, et al. Economic inequality: Is it natural? *Current Science*, 92(10):1383, 2007.
- [50] T. Chesney, S.H. Chuah, and R. Hoffmann. Virtual world experimentation: An exploratory study. *Journal of Economic Behavior & Organization*, 72(1):618–635, 2009.
- [51] N.A. Christakis and J.H. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379, 2007.
- [52] A. Clauset and K.S. Gleditsch. The developmental dynamics of terrorist organizations. *Arxiv preprint arXiv:0906.3287*, 2009.
- [53] A. Clauset, C.R. Shalizi, and M.E.J. Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
- [54] J.S. Coleman. Social capital in the creation of human capital. *American Journal of Sociology*, 94:95–120, 1988.
- [55] V. Colizza, A. Barrat, M. Barthélémy, and A. Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences*, 103(7):2015, 2006.
- [56] V. Colizza and A. Vespignani. Invasion threshold in heterogeneous metapopulation networks. *Physical Review Letters*, 99(14):148701, 2007.
- [57] R. Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.
- [58] T.M. Cover and J.A. Thomas. *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2nd edition, 2006.
- [59] G. Csányi and B. Szendrői. Structure of a large social network. *Physical Review E*, 69(3):36131, 2004.

- 
- [60] G. Csárdi, K.J. Strandburg, L. Zalányi, J. Tobochnik, and P. Érdi. Modeling innovation by a kinetic description of the patent citation system. *Physica A*, 374(2):783–793, 2007.
  - [61] J.A. Davis. Clustering and structural balance in graphs. *Human Relations*, 20(2):181, 1967.
  - [62] C. de Kerchove and P. Van Dooren. The pagetrust algorithm: How to rank web pages when negative links are allowed. In *Proceedings of the SIAM International Conference on Data Mining*, pages 346–352. Citeseer, 2008.
  - [63] W. de Nooy, A. Mrvar, and V. Batagelj. *Exploratory social network analysis with Pajek*. Cambridge University Press New York, NY, USA, 2005.
  - [64] P. Doreian. Evolution of human signed networks. *Metodoloski zvezki*, 1(2):277–293, 2004.
  - [65] P. Doreian and A. Mrvar. A partitioning approach to structural balance. *Social Networks*, 18(2):149–168, 1996.
  - [66] P. Doreian and A. Mrvar. Partitioning signed social networks. *Social Networks*, 31:1–11, 2009.
  - [67] S.N. Dorogovtsev and J.F.F. Mendes. Effect of the accelerating growth of communications networks on their structure. *Physical Review E*, 63(2):25101, 2001.
  - [68] S.N. Dorogovtsev and J.F.F. Mendes. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, 2003.
  - [69] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin. Structure of growing networks: Exact solution of the barabasi-albert’s model. *Physical Review Letters*, 85:4633, 2000.
  - [70] A. Dragulescu and V.M. Yakovenko. Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A*, 299(1-2):213–221, 2001.
  - [71] N. Ducheneaut, N. Yee, E. Nickell, and R.J. Moore. “Alone together?”: Exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 407–416. ACM New York, NY, USA, 2006.
  - [72] N. Ducheneaut, N. Yee, E. Nickell, and R.J. Moore. The life and death of online gaming communities: A look at guilds in World of Warcraft. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 839–848. ACM Press New York, NY, USA, 2007.

## Bibliography

---

- [73] R.I.M. Dunbar. Co-evolution of neocortex size, group size and language in humans. *Behavioral and Brain Sciences*, 16(4):681–735, 1993.
- [74] N. Eagle, A.S. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274, 2009.
- [75] D. Easley and J. Kleinberg. *Networks, crowds, and markets*. Cambridge University Press, 2010.
- [76] B. Entwistle, K. Faust, R.R. Rindfuss, and T. Kaneda. Networks and contexts: Variation in the structure of social ties. *American Journal of Sociology*, 112(5):1495–1533, 2007.
- [77] J.M. Epstein. Agent-based computational models and generative social science. *Complexity*, 4(5):41–60, 1999.
- [78] P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.
- [79] T.S. Evans. Exact solutions for network rewiring models. *The European Physical Journal B*, 56(1):65–69, 2007.
- [80] T.S. Evans and R. Lambiotte. Line graphs, link partitions, and overlapping communities. *Physical Review E*, 80(1):016105, 2009.
- [81] C. Fan, J.L. Guo, and Y.L. Zha. Fractal analysis on human behaviors dynamics. *Arxiv preprint arXiv:1012.4088*, 2010.
- [82] S. Fortunato. Community detection in graphs. *Physics Report*, 486:75–174, 2010.
- [83] S. Fortunato and C. Castellano. Scaling and universality in proportional elections. *Physical Review Letters*, 99(13):138701, 2007.
- [84] J.G. Foster, D.V. Foster, P. Grassberger, and M. Paczuski. Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences*, 107(24):10815, 2010.
- [85] E.B. Fowlkes and C.L. Mallows. *A method for comparing two hierarchical algorithms*, volume 78. Journal of the American Statistical Association, 1983.
- [86] L.C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.
- [87] N. Friedkin. A test of structural features of Granovetter’s strength of weak ties theory. *Social Networks*, 2(4):411–422, 1980.

- 
- [88] S. Gächter and E. Fehr. Collective action as a social exchange. *Journal of Economic Behavior and Organization*, 39(4):341–369, 1999.
  - [89] S. Galam. Minority opinion spreading in random geometry. *The European Physical Journal B*, 25(4):403–406, 2002.
  - [90] D. Garlaschelli and M.I. Loffredo. Patterns of link reciprocity in directed networks. *Physical Review Letters*, 93(26):268701, 2004.
  - [91] R. Gehorsam. The coming revolution in massively multiuser persistent worlds. *Computer*, 36(4):93–95, 2003.
  - [92] J. Genini, L.P.C. Morellato, P.R. Guimarães, and J.M. Olesen. Cheaters in mutualism networks. *Biology Letters*, 6(4):494, 2010.
  - [93] S.A. Golder, D.M. Wilkinson, and B.A. Huberman. Rhythms of social interaction: messaging within a massive online network. *Communities and Technologies 2007*, pages 41–66, 2007.
  - [94] R.L. Goldstone and M.A. Janssen. Computational models of collective behavior. *Trends in Cognitive Sciences*, 9(9):424–430, 2005.
  - [95] M.C. González, C.A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453:779–782, 2008.
  - [96] A. Grabowski and N. Kruszewska. Experimental study of the structure of a social network and human dynamics in a virtual society. *International Journal of Modern Physics C*, 18(10):1527–1536, 2007.
  - [97] M.S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.
  - [98] R. Guimerà, S. Mossa, A. Turtschi, and L.A.N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities’ global roles. *Proceedings of the National Academy of Sciences*, 102:7794–7799, 2005.
  - [99] M. Hamasaki, H. Takeda, T. Hope, and T. Nishimura. Network analysis of an emergent massively collaborative creation community. *Proceedings of the Third International ICWSM Conference*, pages 222–225, 2009.
  - [100] J.T. Hamill. Analysis of layered social networks. Technical report, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, 2006.
  - [101] X.P. Han, Q. Hao, B.H. Wang, and T. Zhou. Origin of the scaling law in human mobility: Hierarchy of traffic systems. *Physical Review E*, 83(3):036117, 2011.

- [102] R. Hanel and S. Thurner. A comprehensive classification of complex statistical systems and an axiomatic derivation of their entropy and distribution functions. *Europhysics Letters*, 93:20006, 2011.
- [103] R. Hanel, S. Thurner, and M. Gell-Mann. Generalized entropies and the transformation group of superstatistics. *Proceedings of the National Academy of Sciences*, 108(16):6390–6394, 2011.
- [104] F. Harary. On the notion of balance of a signed graph. *The Michigan Mathematical Journal*, 2(2):143–146, 1953.
- [105] F. Harary, R.Z. Norman, and D. Cartwright. *Structural Models: An Introduction to the Theory of Directed Graphs*. John Wiley & Sons Inc, 1965.
- [106] C. Haythornthwaite. Exploring multiplexity: Social network structures in a computer-supported distance learning class. *The Information Society*, 17(3):211–226, 2001.
- [107] F. Heider. Attitudes and cognitive organization. *Journal of Psychology*, 21(2):107–112, 1946.
- [108] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, 73(4):1067, 2001.
- [109] D. Helbing. Managing complexity in socio-economic systems. *European Review*, 17(02):423–438, 2009.
- [110] D. Helbing. *Quantitative sociodynamics: stochastic methods and models of social interaction processes*. Springer Verlag, 2011.
- [111] D. Helbing and S. Babiotti. Fundamental and real-world challenges in economics. *Science and Culture*, 2010.
- [112] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, Vol. 407, pp. 487-490, 2000, 2000.
- [113] D. Helbing, W. Yu, and H. Rauhut. Self-organization and emergence in social systems: Modeling the coevolution of social environments and cooperative behavior. *Journal of Mathematical Sociology*, 35(1-3):177–208, 2011.
- [114] J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, M. Alvard, A. Barr, J. Ensminger, et al. “Economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(06):795–815, 2005.
- [115] C.A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausmann. The product space conditions the development of nations. *Science*, 317(5837):482, 2007.

- [116] P. Holme, C.R. Edling, and F. Liljeros. Structure and time evolution of an internet dating community. *Social Networks*, 26(2):155–174, 2004.
- [117] H.-B. Hu and X.-F. Wang. Disassortative mixing in online social networks. *Arxiv preprint 0909.0450v1*, 2009.
- [118] K. Huang. *Statistical Mechanics*. Wiley, New York, 2nd edition edition, 1987.
- [119] Y. Huang, C. Shen, D. Williams, and N. Contractor. Virtually there: Exploring proximity and homophily in a virtual world. In *International conference on computational science and engineering 2009*, volume 4, pages 354–359. IEEE, 2009.
- [120] L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences*, 101:15124–15129, 2004.
- [121] N.P. Hummon and P. Doreian. Some dynamics of social balance processes: Bringing Heider back into balance theory. *Social Networks*, 25(1):17–49, 2003.
- [122] P. Jensen. Network-based predictions of retail store commercial categories and optimal locations. *Physical Review E*, 74(3):035101, 2006.
- [123] H. Jeong, Z. Neda, and A.-L. Barabasi. Measuring preferential attachment in evolving networks. *Europhysics Letters*, 61(4):567–572, 2003.
- [124] Z.Q. Jiang, W.X. Zhou, and Q.Z. Tan. Online-offline activities and game-playing behaviors of avatars in a massive multiplayer online role-playing game. *Arxiv preprint arXiv:0907.5043*, 2009.
- [125] E.M. Jin, M. Girvan, and M.E.J. Newman. Structure of growing social networks. *Physical Review E*, 64(4):46132, 2001.
- [126] N.F. Johnson, C. Xu, Z. Zhao, N. Ducheneaut, N. Yee, G. Tita, and P.M. Hui. Human group formation in online guilds and offline gangs driven by a common team dynamic. *Physical Review E*, 79(6):66117, 2009.
- [127] F. Jordán. Children in time: community organization in social and ecological systems. *Current Science (India)*, 97(11):1579–1585, 2009.
- [128] D. Kahneman. Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review*, 93(5):1449–1475, 2003.
- [129] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.
- [130] B.W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal*, pages 291–307, 1970.

## Bibliography

---

- [131] B.J. Kim, A. Trusina, P. Minnhagen, and K. Sneppen. Self organized scale-free networks from merging and regeneration. *The European Physical Journal B*, 43(3):369–372, 2005.
- [132] P. Klimek, R. Lambiotte, and S. Thurner. Opinion formation in laggard societies. *Europhysics Letters*, 82:28008, 2008.
- [133] R. Kölbl and D. Helbing. Energy laws in human travel behaviour. *New Journal of Physics*, 5:48, 2003.
- [134] G. Kossinets and D.J. Watts. Empirical analysis of an evolving social network. *Science*, 311(5757):88–90, 2006.
- [135] P.L. Krapivsky and S. Redner. Dynamics of majority rule in two-state interacting spin systems. *Physical Review Letters*, 90(23):238701, 2003.
- [136] F. Kyriakopoulos, S. Thurner, C. Puhr, and S.W. Schmitz. Network and eigenvalue analysis of financial transaction networks. *The European Physical Journal B*, 71(4):523–531, 2009.
- [137] G. Labianca and D.J. Brass. Exploring the social ledger: Negative relationships and negative asymmetry in social networks in organizations. *The Academy of Management Review*, 31(3):596–614, 2006.
- [138] R. Lambiotte, V.D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Van Dooren. Geographical dispersal of mobile communication networks. *Physica A*, 387(21):5317–5325, 2008.
- [139] V. Latora and M. Marchiori. Efficient behavior of small-world networks. *Physical Review Letters*, 87(19):198701, 2001.
- [140] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, et al. Computational social science. *Science*, 323(5915):721, 2009.
- [141] E.A. Leicht and R.M. D’Souza. Percolation on interacting networks. *Arxiv preprint arXiv:0907.0894*, 2009.
- [142] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins. Microscopic evolution of social networks. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 462–470. ACM, 2008.
- [143] J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *Proceeding of the 17th international conference on World Wide Web*, pages 915–924. ACM, 2008.

- 
- [144] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, 1(1), 2007.
  - [145] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis. Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30(4):330–342, 2008.
  - [146] M. Lim, R. Metzler, and Y. Bar-Yam. Global pattern formation and ethnic/cultural violence. *Science*, 317(5844):1540, 2007.
  - [147] R.E. Lucas. Econometric policy evaluation: a critique. In *Carnegie-Rochester conference series on public policy*, volume 1, pages 19–46, 1976.
  - [148] E. Majorana. Il valore delle leggi statistiche nella fisica e nelle scienze sociali. *Scientia*, 36(58):58–66, 1942.
  - [149] H.A. Makse, S. Havlin, and H.E. Stanley. Modelling urban growth patterns. *Nature*, 377:608–612, 1995.
  - [150] S.A. Marvel, S.H. Strogatz, and J.M. Kleinberg. Energy landscape of social balance. *Physical Review Letters*, 103(19):198701, 2009.
  - [151] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913, 2002.
  - [152] M. McPherson, L. Smith-Lovin, and J.M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, pages 415–444, 2001.
  - [153] M. Meila. Comparing clusterings—an information based distance. *Journal of Multivariate Analysis*, 98(5):873 – 895, 2007.
  - [154] P.R. Messinger, E. Stroulia, and K. Lyons. A typology of virtual worlds: Historical overview and future directions. *Journal of Virtual Worlds Research*, 1(1):1–18, 2008.
  - [155] R. Metzler and J. Klafter. The random walk’s guide to anomalous diffusion: A fractional dynamics approach. *Physics Reports*, 339:1–77, 2000.
  - [156] S. Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.
  - [157] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenststat, M. Sheffer, and U. Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538, 2004.
  - [158] R. Milo, N. Kashtan, S. Itzkovitz, M.E.J. Newman, and U. Alon. Uniform generation of random graphs with arbitrary degree sequences. *Arxiv preprint arXiv:cond-mat/0312028*, 2003.

## Bibliography

---

- [159] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.
- [160] G. Miritello, E. Moro, and R. Lara. Dynamical strength of social ties in information spreading. *Physical Review E*, 83(4):045102, 2011.
- [161] P.J. Mucha, T. Richardson, K. Macon, M.A. Porter, and J.P. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876, 2010.
- [162] T.M. Newcomb. *The acquaintance process*. Holt, Rinehart and Winston New York, 1961.
- [163] D. Newman. The lines that continue to separate us: Borders in our borderless world. *Progress in Human Geography*, 30(2):143, 2006.
- [164] M.E.J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404, 2001.
- [165] M.E.J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002.
- [166] M.E.J. Newman. Analysis of weighted networks. *Physical Review E*, 70(5):056131, 2004.
- [167] M.E.J. Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101:5200, 2004.
- [168] M.E.J. Newman. Power laws, pareto distributions and zipf’s law. *Contemporary physics*, 46(5):323–351, 2005.
- [169] M.E.J. Newman, S. Forrest, and J. Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, 66(3):035101, 2002.
- [170] M.E.J. Newman and J. Park. Why social networks are different from other types of networks. *Arxiv preprint arXiv:cond-mat/0305612*, 2003.
- [171] M.E.J. Newman, S.H. Strogatz, and D.J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2):026118, 2001.
- [172] J.P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M.A. de Menezes, K. Kaski, A.-L. Barabási, and J. Kertész. Analysis of a large-scale weighted network of one-to-one human communication. *New Journal of Physics*, 9(6):179, 2007.
- [173] J.P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332, 2007.

- 
- [174] J.F. Padgett and C.K. Ansell. Robust action and the rise of the Medici, 1400–1434. *American Journal of Sociology*, pages 1259–1319, 1993.
  - [175] V. Pareto. *Cours d'Economie Politique*. F. Rouge & Cie., Lausanne, Switzerland, 1897.
  - [176] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200–3203, 2001.
  - [177] C.K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H.E. Stanley. Long-range correlations in nucleotide sequences. *Nature*, 356(6365):168–170, 1992.
  - [178] D. Quercia, N. Lathia, F. Calabrese, G. Di Lorenzo, and J. Crowcroft. Recommending social events from mobile phone location data. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 971 –976, 2010.
  - [179] W.M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):pp. 846–850, 1971.
  - [180] A. Rapoport. Spread of information through a population with socio-structural bias: I. assumption of transitivity. *Bulletin of Mathematical Biology*, 15(4):523–533, 1953.
  - [181] C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, and S.H. Strogatz. Redrawing the map of Great Britain from a network of human interactions. *PLoS One*, 5(12):e14248, 2010.
  - [182] E. Ravasz. *Evolution, hierarchy and modular organization in complex networks*. PhD thesis, University of Notre Dame, Indiana, 2004.
  - [183] S. Redner. *A guide to first-passage processes*. Cambridge University Press, 2001.
  - [184] T. Richardson, P. Mucha, and M. Porter. Spectral tripartitioning of networks. *Physical Review E*, 80(3), 2009.
  - [185] J.M. Roberts. Simple methods for simulating sociomatrices with given marginal totals. *Social Networks*, 22(3):273–283, 2000.
  - [186] C.P. Roca, J.A. Cuesta, and A. Sánchez. Effect of spatial structure on the evolution of cooperation. *Physical Review E*, 80(4):046106, 2009.
  - [187] M.A. Rodriguez and J. Shinavier. Exposing multi-relational networks to single-relational network analysis algorithms. *Journal of Informetrics*, 4(1):29–41, 2010.

- [188] T.L. Ross. Constructing a virtual world as a research tool: Lessons learned from the first iteration in the development of Greenland. In *2009 International Conference on Computational Science and Engineering*, pages 1163–1168. IEEE, 2009.
- [189] T.L. Ross and R.D. Cornell. Towards an experimental methodology of virtual world research. In *2010 Second International Conference on Games and Virtual Worlds for Serious Applications*, pages 143–150. IEEE, 2010.
- [190] C. Roth, S.M. Kang, M. Batty, and M. Barthélemy. Structure of urban movements: Polycentric activity and entangled hierarchical flows. *PLoS One*, 6(1):e15923, 01 2011.
- [191] S.F. Sampson. *A novitiate in a period of change: An experimental and case study of social relationships*. PhD thesis, Cornell University, 1968.
- [192] N. Scafetta, V. Latora, and P. Grigolini. Lévy statistics in coding and non-coding nucleotide sequences. *Physics Letters A*, 299(5-6):565–570, 2002.
- [193] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. Campbell. Nextplace: A spatio-temporal prediction framework for pervasive systems. *Pervasive Computing*, pages 152–169, 2011.
- [194] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo. Socio-spatial properties of online location-based social networks. *Proceedings of ICWSM*, 11, 2011.
- [195] T.C. Schelling. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186, 1971.
- [196] T.M. Selee, T.G. Kolda, W.P. Kegelmeyer, and J.D. Griffin. Extracting clusters from large datasets with multiple similarity measures using IMSCAND. In *CSRI Summer Proceedings 2007*, page 87. Citeseer, 2007.
- [197] K. Sigmund. *The calculus of selfishness*. Princeton University Press, 2010.
- [198] R. Sinatra, D. Condorelli, and V. Latora. Networks of motifs from sequences of symbols. *Physical Review Letters*, 105(17):178702, 2010.
- [199] C. Song, T. Koren, P. Wang, and A.-L. Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6:818–823, 10 2010.
- [200] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018, 2010.
- [201] 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, et al. Anomalous fluctuations in the dynamics of complex systems: From DNA and physiology to econophysics. *Physica A*, 224(1-2):302–321, 1996.

- 
- [202] G. Szabó and G. Fáth. Evolutionary games on graphs. *Physics Reports*, 446(4–6):97–216, 2007.
  - [203] M. Szell, R. Lambiotte, and S. Thurner. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31):13636–13641, 2010.
  - [204] M. Szell, R. Sinatra, G. Petri, S. Thurner, and V. Latora. Understanding mobility in a social petri dish. *in review*, 2011.
  - [205] M. Szell and S. Thurner. Measuring social dynamics in a massive multiplayer online game. *Social Networks*, 32:313–329, 2010.
  - [206] C. Thiemann, F. Theis, D. Grady, R. Brune, and D. Dirk Brockmann. The structure of borders in a small world. *PLoS One*, 5:e15422, 2010.
  - [207] S. Thurner, F. Kyriakopoulos, and C. Tsallis. Unified model for network dynamics exhibiting nonextensive statistics. *Physical Review E*, 76:036111, 2007.
  - [208] S. Thurner, M. Szell, and R. Sinatra. Emergence of good conduct, scaling and Zipf laws in human behavioral sequences in an online world. *PLoS One (in print)*, 2011.
  - [209] S. Thurner and C. Tsallis. Nonextensive aspects of self-organized scale-free gas-like networks. *Arxiv preprint arXiv:cond-mat/0506140*, 2005.
  - [210] M.T. Torfason. My brother’s keeper: Patterns of norm violations in a virtual world. *unpublished*, 2011.
  - [211] V.A. Traag and Jeroen Bruggeman. Community detection in networks with positive and negative links. *Arxiv preprint arXiv:0811.2329*, 2008.
  - [212] A.L. Traud, E.D. Kelsic, P.J. Mucha, and M.A. Porter. Community structure in online collegiate social networks. *Arxiv preprint arXiv:0809.0690*, 2008.
  - [213] A. Vazquez. Knowing a network by walking on it: emergence of scaling. *Arxiv preprint arXiv:cond-mat/0006132*, 2000.
  - [214] G.M. Viswanathan, S.V. Buldyrev, S. Havlin, M.G.E. Da Luz, E.P. Raposo, and H.E. Stanley. Optimizing the success of random searches. *Nature*, 401(6756):911–914, 1999.
  - [215] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.
  - [216] D.J. Watts. A twenty-first century science. *Nature*, 445(7127):489, 2007.
  - [217] D.J. Watts and S.H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

## Bibliography

---

- [218] B.J. West, P. Grigolini, R. Metzler, and T.F. Nonnenmacher. Fractional diffusion and levy stable processes. *Physical Review E*, 55(1):99, 1997.
- [219] D. Williams, N. Ducheneaut, L. Xiong, Y. Zhang, N. Yee, and E. Nickell. From tree house to barracks: The social life of guilds in World of Warcraft. *Games and Culture*, 1(4):338, 2006.
- [220] D. Williams, N. Martins, M. Consalvo, and J.D. Ivory. The virtual census: Representations of gender, race and age in video games. *New Media & Society*, 11(5):815, 2009.
- [221] R.T.A. Wood, M.D. Griffiths, and V. Eatough. Online data collection from video game players: Methodological issues. *CyberPsychology & Behavior*, 7(5):511–518, 2004.
- [222] V.M. Yakovenko. Econophysics, statistical mechanics approach to. *Encyclopedia of Complexity and System Science*, 2009.
- [223] N. Yee. The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence: Teleoperators and Virtual Environments*, 15(3):309–329, 2006.
- [224] N. Yee. Motivations for play in online games. *CyberPsychology & Behavior*, 9(6):772–775, 2006.
- [225] D.H. Zanette and P.A. Alemany. Thermodynamics of anomalous diffusion. *Physical Review Letters*, 75(3):366–369, 1995.
- [226] G.K. Zipf. *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley Press, 1949.

# Michael Szell

Section for Science of Complex Systems  
Medical University of Vienna  
Spitalgasse 23  
1090 Vienna, Austria

Phone: (+43) 1 40160 36253  
Fax: (+43) 1 40160 950510  
Office: Bauteil 86, Room 86.02.011  
Email: michael.szell@meduniwien.ac.at  
<http://www.complex-systems.meduniwien.ac.at/people/mszell/>

Birth data: Feb 1983, Budapest  
Nationality: Austria  
Languages: English, German, Hungarian, (Italian)

## Education

Ph.D. candidate in Physics, University of Vienna, graduation planned for Feb 2012

*Dissertation:* Statistical physics approaches to large-scale socio-economic networks

*Advisors:* Stefan Thurner, Christoph Dellago

M.Sc. Technical Mathematics (focus on computer science), Vienna University of Technology, Oct 2007

## Employment

Research assistant, Section for Science of Complex Systems, Medical University of Vienna, Jan 2009 – Dec 2011

Founder and manager of Bayer & Szell OG, Sep 2006 – present

Developer and administrator of the Massive Multiplayer Online Game *Pardus*, [www.pardus.at](http://www.pardus.at), Jul 2004 – present

## Research Interests

Analysis and modeling of structure, evolution, and processes on large-scale (socio-economic) networks

Data mining and statistical physics of human societies, in particular Massive Multiplayer Online Games

Behavioral economics, Evolutionary game theory

## Publications

### *Working Papers*

M. Szell, R. Sinatra, G. Petri, S. Thurner, V. Latora  
Understanding mobility in a social petri dish, *in review*

M. Szell, S. Thurner  
Gender-specific structure and dynamics of large-scale social networks, *in preparation*

### *Peer-reviewed*

S. Thurner, M. Szell, R. Sinatra  
Emergence of good conduct, scaling and Zipf laws in human behavioral sequences in an online world, PLoS One (2011), *in print*

- M. Szell, S. Thurner  
 Measuring social dynamics in a massive multiplayer online game, *Social Networks* 32, 313–329 (2010)
- M. Szell, R. Lambiotte, S. Thurner  
 Multirelational organization of large-scale social networks in an online world, *Proceedings of the National Academy of Sciences* 107, 13636–13641 (2010)

### *Other*

- S. Thurner, M. Szell (Eds.)  
 Book of abstracts ECCS'11 Vienna, *European Conference on Complex Systems 2011*, Löcker (2011)
- M. Szell, G. Petri, K. Zhao, D. Levin  
 Human mobility in an online world, *Santa Fe Summer School 2010 Project Paper* (2010)
- M. Szell, S. Thurner  
 Social dynamics in a large-scale online game, bookchapter in ‘Successes in complex systems science’ (2011), *in print*
- M. Szell  
 Finite Differenzen Verfahren zur numerischen Lawinensimulation, Institute of Analysis and Scientific Computing, Vienna University of Technology (2007), *Diploma thesis*
- M. Szell, F. Judex  
 Petri Net Modelling of Different Strategies in ARGESIM Comparison C4 “Dining Philosophers” with HPSim, *Simulation News Europe* 17, 1 (2007)

### Talks

- Macro Connections* Seminar at Harvard University, Cambridge, USA, Oct 2011  
 Understanding collective socio-economic behavior in an online world
- ECCS 2011 Conference, Vienna, Austria, Sep 2011  
 Human mobility on a network
- NetSci 2011 Conference, Budapest, Hungary, Jun 2011  
 Multi-relational Socio-economic Networks in a Large-scale Online Community
- Net2011 Conference at University Bicocca Milano, Italy, May 2011  
 Multi-relational socio-economic networks in a massive multiplayer online game
- INSITE Kickoff-Meeting, Venice, Italy, Apr 2011  
 The artificial world of the Pardus game: First lessons
- Games Technology Workshop at ETH Zürich, Switzerland, Mar 2011  
 Browser-based MMOGs: Technology and use as research tools
- Interdisciplinary Seminar *Modeling Complex Socio-Economic Systems and Crises* at ETH Zürich, Switzerland, Feb 2011  
 Social Networks in Multiplayer Online Games
- Invited talk at ETH Zürich, Switzerland, Feb 2011  
 Massive Multiplayer Online Games as Socio-economic Labs
- 7th Vienna Central European Seminar in Particle Physics and Quantum Field Theory, Austria, Nov 2010  
 Complex human mobility dynamics on a network
- ECCS 2010 Conference, Lisbon, Portugal, Sep 2010  
 Multi-relational social dynamics in a large-scale online game society
- Trade, Conflict and Sentiments: Multi-relational organization of large-scale social networks

- IIASA Xevents Workshop*, Laxenburg, Austria, Mar 2010  
 Extreme events in a large-scale, virtual human society
- COST workshop *Processes on Networks* at Medical University of Vienna, Austria, Mar 2010  
 Establishment of a large-scale “laboratory” for socio-economic behavior
- Invited talk at University of Catania, Italy, Oct 2009  
 Measuring the evolution of socio-economical structure in an online game
- Workshop *Modelling Science* at Rathenau Institute, Amsterdam, Netherlands, Oct 2009  
 Measuring the evolution of socio-economical structure in an online game
- DPG Spring Meeting*, Dresden, Germany, Mar 2009  
 Measuring the evolution of socio-economical structure in an online game
- Invited lecture at the Department of Education and Human Development, University of Vienna, Austria, Jun 2008  
 Pardus – Funktion und Aufbau eines Browser-Spiels
- Workshop for avalanche dynamics at the Austrian service for Torrent and Avalanche Control*, Schwaz, Austria, Oct 2007  
 Finite Differenzen Verfahren zur numerischen Lawinensimulation
- ARGESIM Simulation Workshop*, Siena, Italy, Sep 2007  
 Finite Differenzen Verfahren zur numerischen Lawinensimulation  
 Pardus – Funktion und Aufbau eines Browser-Spiels

## Poster Presentations

- Evolution of Cooperation - TECT 09* Conference at IIASA, Laxenburg, Austria, Sep 2009  
 Evolution of cooperation- and defection-networks in a multiplayer online game

## Scientific Activities

- Academic visit and scientific collaboration at the Laboratory for Complex Systems, University of Catania, Italy, May – Jun 2011
- Academic visit and scientific collaboration at the Chair of Sociology, in particular of Modeling and Simulation, ETH Zürich, Switzerland, Feb – Mar 2011
- Organizer and Program Committee member of the ECCS’11 Satellite XNet – Complexity & Networks at the Medical University of Vienna, Austria, Jan – Sep 2011
- Organizer and Program Committee member of the *European Conference on Complex Systems 2011 (ECCS’11)* at the Medical University of Vienna, Austria, Oct 2010 – Sep 2011
- Academic visit and scientific collaboration at the Laboratory for Complex Systems, University of Catania, Italy, Sep 2010
- COST scientific mission at the Institute of Mathematical Sciences, Imperial College London, UK, Aug – Sep 2010
- Participant in the *SFI Complex Systems Summer School 2010* at the Santa Fe Institute, USA, Jun 2010
- Organizer of COST MPo801 workshop *Processes on Networks* at the Medical University of Vienna, Austria, Mar 2010
- Participant in summer school *Complex Systems and Social Simulation* at the Central European University, Budapest, Hungary, Jul 2009
- Participant in summer school *Steps in Evolution* at the Jacobs University Bremen, Germany, Jun 2009

Reviewer for PLoS One, Social Networks

## Skills

### Programming

Advanced: MatLab, L<sup>A</sup>T<sub>E</sub>X, PHP, MySQL, HTML/CSS

Basic: C/C++, Java, R, Javascript, Processing

### Applications

MatLab, Pajek, Gephi, Inkscape, Illustrator, Photoshop, InDesign

### Software development and data handling

Computer/Online game development, Web development, Database management, Data mining

## In the Media (Selection)

Wien live: *80-dimensionale Ameisen*, Mar 2011

Radio TV México: *Las relaciones virtuales son más estables que las habituales*, Aug 2010

ORF Futurezone: *Beziehungsanalysen im Onlinegame*, Jul 2010

ORF Science: *Positive Beziehungen halten uns zusammen*, Jul 2010

Der Standard: *Soziale Gesetze von Freund und Feind*, Jul 2010

Science Daily: *The friend of my enemy is my enemy*, Jul 2010

Slashdot: *Study of MMOG Proves Human Interaction Theory*, Jul 2010

Numerama: *Un jeu massivement multijoueurs analysé pour démontrer les liens sociaux*, Jul 2010

Physorg: *The friend of my enemy is my enemy*, Jul 2010

Santa Fe Institute: *The friend of my enemy is my enemy*, Jul 2010

Imperial College London: *The friend of my enemy is my enemy*, Jul 2010

Austrian National Radio Ö1: *Dimensionen – Die Welt der Wissenschaft*, Jan 2010

Die Presse: *Netzwerkanalysen: Ansteckung im Computer*, Nov 2009

T-Online: *Pardus - Strategiegame aus Österreich*, Feb 2008

Radio Niederösterreich: *Interview mit Michael Szell*, Feb 2008

PC PowerPlay: *Pardus: Der Newcomer aus Österreich*, JuJ 2007

Telekommunikations & IT Report: *Auf Erfolgskurs*, May 2007

wiener: *Lost in Space*, Apr 2007

Austrian National Radio FM4: *Interview mit Michael Szell*, Jan 2007

Telekommunikations & IT Report: *Pardus: Erfolg für Community-Plattform*, May 2006

GameOgre: *Pardus - Best Free Online RPG of the Year 2005*, Dec 2005

Telekom Austria: *Online Universum Pardus*, Nov 2005