Insight into social physics: Uncovering the structure and dynamics of social relationships

by

Diego Escribano Gómez

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in

Mathematical Engineering

Universidad Carlos III Madrid

Advisor:

José Antonio Cuesta Ruiz

Tutor:

José Antonio Cuesta Ruiz

June 2023

A mis padres,
simplemente gracias.

Agradecimientos

Bueno, quién me iba a decir que la parte más complicada de escribir de todo este trabajo iba a ser esta sección de agradecimientos. Sentarme delante de un folio en blanco, echar la vista atrás a todo este viaje y volver a recordar una gran cantidad de momentos que tengo grabados en la memoria... Los que más me conocéis ya sabéis todo lo que me cuesta abrirme y expresar mis sentimientos, pero voy a intentar hacerlo lo mejor posible. Allá voy.

Tengo muy claro por dónde empezar. Muchas gracias, Jose, por acompañarme durante todos estos años. Sé que esta no es la primera tesis que diriges, pero sí la primera bajo estas condiciones. Por ejemplo, una pandemia mundial, que nos tuvo un tiempo encerrados en casa, o tu puesto como director del departamento. Y sólo puedo decir que, fuese cual fuese el momento, cuando te lo he pedido siempre has encontrado un hueco para mí. El talento que tienes es increíble. La de veces que he llegado a tu despacho con un problema, de esos que llevan días dando vueltas a tu cabeza, y tú has sido capaz de resolverlo en un momento. Te admiro, y te aseguro que eso es algo que puedo decir de muy pocas personas.

Anxo, aunque como siempre decías cuando me pedías algo (*"pero yo no soy tu jefe, eh"*), para mí ha sido una verdadera suerte poder aprender a tu lado. Muchas gracias por contar siempre conmigo para cualquiera de los proyectos que surgían, he hecho cosas que ni se me habrían pasado por la cabeza y lo he disfrutado muchísimo. Por ti he conocido otra forma de pensar y de hacer ciencia que me guardo para siempre. Sinceramente, todo lo que he aprendido de trabajar contigo me va a valer más en la vida que cualquier título de doctorado, lo tengo clarísimo.

Ignacio, estuvimos juntos sólo al principio de esta tesis, pero, sin duda, una gran parte del trabajo también lleva tu marca. Dejaste unos cimientos muy sólidos, a partir de los cuáles ha sido mucho más sencillo seguir avanzando. Me queda la espina de saber hasta donde podríamos haber llegado. Si alguna vez llegas a leer esto, quiero que sepas que estoy súper agradecido por lo que me ayudaste.

Por supuesto, recordar también a todas las personas que durante este tiempo me han hecho sentir como en casa cada vez que venía a Leganés. A mis compañeras de despacho (Vicky, Francy y Sara), sois pura amabilidad y simpatía. No me he podido sentir más a gusto a vuestro lado. A los que ya habéis abandonado el departamento (Pablo Lozano, Beatriz, Miguel, Juan, Pau...), no sabéis todo lo que se os echa de menos. Y a los que aún seguís (Pilar, Nello, Alberto, Jose Camacho, Pablo Catalán, Jorge...), habéis sido un pilar fundamental en esta tesis con vuestros consejos y apoyo diario. Mil gracias a todos (y todas) por cada conversación, café (o plátano), comida o cerveza. Me voy, pero todos esos momentos se vienen en la mochila conmigo. Me habéis demostrado que lo mejor de la ciencia es la gente que la hace.

Quiero agradecer también a toda la gente del IES Blas de Otero. En especial, a Francis y a Silvia. Sin vuestra ayuda habría sido imposible realizar esta investigación. Nos habéis abierto las puertas de vuestro centro y siempre habéis estado dispuestos a ayudar con cada una de nuestras propuestas. Espero que seáis capaces de identificar la gran cantidad de aportaciones vuestras que hay en la tesis.

A Marta y a Gema, que no podían faltar en estas líneas. Con el tiempo cada vez valoro más todo lo que me ayudasteis y apostasteis por mí en su momento. Gracias por enseñarme que lo fácil es seguir a los demás y olvidar el pensamiento crítico, pero que lo valiente y beneficioso a largo plazo es defender en lo que tú crees. Lo tengo siempre presente.

A mi gente, los de siempre, vosotros sabéis quiénes sois. Tengo tantas cosas por las que daros las gracias después de tantos años que no sé ni por dónde empezar. Da igual cómo nos vaya todo, dónde estemos cada uno o el tiempo que llevemos sin vernos. Cuando nos juntamos las cosas fluyen y todos nos entendemos incluso sin hablar. Habéis estado interesándoos todo el tiempo por la evolución de este trabajo, aunque en muchas ocasiones tenía clarísimo que no os enterábais de lo que os estaba contando. Vuestra amistad es de las cosas que no cambiaría por nada.

A mis padres y a mi hermana. Quiero que sepáis que de la cosa que más afortunado me siento en el mundo es de teneros detrás y sentir que siempre me habéis apoyado en cada cosa que he decidido. Sólo por tener estas líneas para poder dejarlo por escrito, para mí ya ha merecido la pena el esfuerzo de estos años. Sé que no os lo digo muy a menudo, pero os quiero. Os quiero mucho. Voy a intentar decíroslo más, aunque no prometo nada. Espero que os sintáis orgullosos de mí, con eso ya lo tengo todo.

Emma, que llegaste a mi vida cuando esta tesis estaba empezando y lo único que quiero es que ya no te vayas nunca. Hiciste cambiar mi forma de pensar: todo lo que antes era yo, ahora es nosotros. Eres la única persona que conoce realmente el esfuerzo que hay detrás de todo esto, y quien en los momentos más complicados ha estado tirando de mí. Somos un equipaso. Llega ahora ahora una etapa nueva a nuestras vidas, no sabemos qué se viene, pero tengo clarísimo que, sea lo que sea, quiero vivirlo a tu lado. Gracias, de verdad.

Y por último, a los que ya no están. Os echo de menos.

Diego

PUBLISHED AND SUBMITTED CONTENT

Chapter 2 (including Appendix A) is composed entirely of the results published in:

- **Diego Escribano,** Francisco J. Lapuente, José A. Cuesta, Robin I. M. Dunbar, and Angel Sánchez. Longitudinal study of friendships and enmities in middle school. *Submitted for publication.* URL: arXiv:2303.03786 [physics.soc-ph]

The author of this thesis is the first author of this publication. The material from this source included in this thesis is not singled out with typographic means and references.

Chapter 3 (including Appendix B) is composed entirely of the results published in:

- **Diego Escribano,** Victoria Doldán Martelli, Francisco J. Lapuente, José A. Cuesta, and Angel Sánchez. Evolution of social relationships between first-year students at middle school: From cliques to circles. Scientific Reports, 11, 11694 (2021). DOI: 10.1038/s41598-021-90984-z

The author of this thesis is the first author of this publication. The material from this source included in this thesis is not singled out with typographic means and references.

Chapter 5 (including Appendix D) is composed partially of the results published in:

- **Diego Escribano**. Statistical mechanics of social relationships. Master's thesis. Universidad Carlos III de Madrid (2019).

The author of this thesis is the author of this master's thesis. The material from this source included in this thesis is not singled out with typographic means and references.

Chapter 6 (including Appendix E) is composed entirely of the results published in:

- **Diego Escribano,** and José A. Cuesta. Free-energy density functional for Strauss's model of transitive networks. Physical Review E, 106, 054305 (2022). DOI: 10.1103/PhysRevE.106.054305

The author of this thesis is the first author of this publication. The material from this source included in this thesis is not singled out with typographic means and references.

Chapter 8 (including Appendix G) is composed entirely of the results published in:

- **Diego Escribano,** Victoria Doldán-Martelli, Katherine A. Cronin, Daniel B. M. Haun, Edwin J. C. van Leeuwen, José A. Cuesta, and Angel Sánchez. Chimpanzees organize their social relationships like humans. Scientific Reports, 12, 16641 (2022). DOI: 10.1038/s41598-022-20672-z

The author of this thesis is the first author of this publication. The material from this source included in this thesis is not singled out with typographic means and references.

SUMMARY

This thesis investigates the emerging interdisciplinary field of social physics, which applies concepts and methods from physics, mathematics and anthropology to understand human behaviour in social systems. Our research seeks to elucidate how humans organise their social relationships and how they evolve over time by examining the universal principles underpinning these phenomena. The basis of our investigation is the concept of "social atom", which serves as a foundation for studying ego-networks at the micro-level and exploring the collective behaviour of social systems at the macro-level.

We embark on two complementary research approaches to address this complex problem. Our first approach involves conducting field research by surveying high school students about their friendships and enmities over two academic years. This empirical data enables us to analyse the organisation and evolution of social relationships, providing valuable insights that can be shared with school principals to foster a more positive social atmosphere and prevent important issues such as bullying.

Our second approach aligns with the conventional scientific method. It involves the formulation of hypotheses, the development of network models and their testing. To do that, we employ exponential random graph models and density functional theory, a technique originating from statistical mechanics for analysing lattice gases. This approach demonstrates that social networks can exhibit phenomena comparable to those observed in fluids or gases, such as phase transitions. These findings contribute to a more profound understanding of the behaviour exhibited by social systems.

Moreover, we expand the applicability of these models to include other species, such as primates, demonstrating their relevance beyond human social relationships. We establish a formalism that can be employed to address social physics problems more effectively by synthesising the insights derived from both research approaches. This integrative method advances our understanding of the discipline and paves the way for more accurate and effective solutions.

Through the combination of field research, network modelling and the extension of these models to other species, this thesis makes a substantial contribution to the field of social physics. Our research provides a solid foundation for future studies and applications aimed at improving the understanding and management of complex social systems by uncovering the fundamental mechanisms governing human social behaviour.

CONTENTS

IV Final remarks   113

9 Conclusions and future work   115

Appendices   125

References   193

1

Introduction

Physics is a science that has been exploring the mysteries of nature for centuries. It is one of the oldest and most fundamental sciences, with a history that dates back to ancient civilisations. Scientists from diverse historical periods and backgrounds (such as Archimedes, Galileo, Newton or Einstein) have been trying to uncover the laws that govern the universe.

The diversity of problems physics addresses is particularly noteworthy, as it studies the behaviour of matter and energy from the level of atoms and molecules to the entire universe through the development of complex models and mathematical theories. However, it has only been recently that physics has focused on handling a question that has been present from the beginning of time: how humans organise their social relationships and interact with each other in society.

The first documented reference to this school of thought might be traced back to the 18th century, when the British philosopher Hume (1739) suggested creating a new science to study human behaviours rooted in the principles of mathematics and physics. In spite of this dissertation, he presented these ideas from a much more philosophical than purely numerical perspective.

It was not until a few decades later, during the 19th century, that the development of two new physical theories altered this framework. Firstly, thermodynamics changed perspectives by introducing the abstract concept of "system", which was previously unexplored. Secondly, electromagnetism revealed how two distinct phenomena might be explained from a unified perspective. These changes led to a new viewpoint that promotes the idea that human

1. INTRODUCTION

behaviour may be governed by universal laws akin to those in physics. The French philosopher Comte and the Belgian statistician Quetelet were the two main representatives of this school of thought. The first proposed ideas that constituted a revolution in understanding society as a system governed by general rules, while the second was a pioneer in applying probability theory to human data. More specifically, Quetelet (1835) published his "Essays in Social Physics",[1] where he used a Gaussian distribution to derive statistical laws that describe the average human behaviour. Quetelet named the statistical approach he had developed "social physics", creating a new discipline which is still a subject of study nowadays.

The early 20th century experienced a boom in scientific production: two major theories, relativity and quantum mechanics, completely revolutionised the perception of nature. This caused researchers to focus solely on traditional physics, while interdisciplinary studies were poorly received and regarded as of little importance by the scientific community. Therefore, the investigation that merged physics and the analysis of social systems was minimal during this period. One such example can be found in the article "Concerning 'social physics'"[2] written by Stewart (1948). In this article, the author formulated the existence of an invisible force, similar to gravity, that organised the population of the US in rural and urban regions. Stewart conducted this analysis using telephone call data, becoming a pioneer in one of the foundations of the discipline today: the use of experimental data to model the system.

The discussion around social physics was not reopened until the second half of the 20th century, since applying physics to social sciences was considered demeaning physics. Bernard and Killworth (1979) published an essay titled "Why are there no social physics?", where they discuss the two reasons they observe for the failure to formulate a complete theory for this topic. The authors refer to the first reason as the "forests and trees" dilemma, which consists on involving an excessive focus on individual details to the detriment of the overall perspective. They argue that this is precisely what happens with field work, where only specific information can be observed, not average behaviours, making it challenging to formulate the underlying principles. The authors' second point is that social processes continually evolve, not enduring in a static state, and it is hard to infer a dynamic process from a mere observation of it. Despite this, Bernard and Killworth strongly promoted the creation of a social physics theory that employs quantitative models to make predictions and assess their accuracy through empirical evidence. That is, developing a model and comparing its predictions with data rather than attempting to interpret the data directly.

It seems that the aspirations of these two authors are becoming a reality, as the discipline of social physics has experienced significant growth in the late 20th and early 21st century, primarily motivated by two reasons. On the one hand, experts and specialists from various fields have joined forces to investigate complex systems, including social ones, creating an interdisciplinary process of collaboration that transcends traditional science. Incorporating ideas and methods from statistical physics yields insights into previously unexplored problems

---

[1]The foundation of these essays lies in notebooks written in French that are actually located in the Royal Academy of Sciences, Letters and Arts of Belgium, in Brussels. These notebooks are unpublished, but their contents have been transcribed and translated into English by Aubin (2014).

[2]In the article, the author states that the use of quotation marks in the title indicates that social physics is not an accepted science, although it could become one in the future. This highlights the limited recognition of interdisciplinary work within the scientific community at that time.

from this viewpoint (Albert & Barabási, 2002; Ball, 2003; Castellano et al., 2009; Desmarais & Cranmer, 2012; Dorogovtsev et al., 2008; Strogatz, 2001). On the other hand, computers have experienced remarkable development in the last years, leading to a rise in computational power and causing a revolution in the field. In the past, computers were used mainly for data storage and processing, but now they can perform a wide range of tasks. This, combined with the growing popularity of Big Data, has motivated the emergence of new exploration techniques, increased access to more plentiful data and led to significant progress in data-driven approaches to social phenomena (Borgatti et al., 2009; Galam, 2014; Lazer et al., 2009; Pentland, 2014; Sawyer, 2005). Although some authors, such as Schweitzer (2018), claim that this recent direction in the field is more related to computational social sciences than to the foundations of social physics established by Quetelet in the 19th century, the discipline is clearly thriving. The increasing attention it is receiving from the scientific community suggests a promising short and long-term future.

This is the framework to which this thesis belongs. Our objective is to shed light on how humans organise their social relationships and how these relationships evolve over time. We rely on well-established anthropological theories, outlined in detail in the next section, to understand the mechanisms underlying these phenomena. Our starting point is the concept of "social atom", introduced by Tamarit (2019) in his PhD thesis. Tamarit combines these anthropological theories with statistical physics techniques to explain the structure of ego-networks[3] (micro-level) in humans. We focus on explaining the rules governing the interaction between the constituents to clarify the collective behaviour of social systems (macro-level), considered as ensembles of "social atoms" interacting with one another.

We tackle this problem from two different but complementary perspectives. Firstly, we conducted field research by surveying high school students on their friendships and enmities over time. By analysing the data, we hope to gain insight into the organisation of social relationships and their evolution. Furthermore, we use this data to provide feedback to the school directors in order to help them to improve the social atmosphere in the school and prevent serious issues such as bullying. The second approach is closer to the conventional scientific method (hypothesise, model, test). We build models from first principles using a family of network models, the exponential random graphs, and some techniques developed in traditional physics to analyse lattice gases, the density functional theory. We show that social networks can exhibit phenomena similar to those observed in gases, such as phase transitions. The use of these techniques enables us to tackle the problems and gain an understanding of the system's behaviour. Finally, we extend these models to other species, such as primates, to show that they can be applied more generally and are not limited to human relationships.

Our motivation is to better understand social physics problems by combining the results obtained through these two different perspectives and creating a formalism that can be used to address them. We believe this comprehensive approach can advance our understanding of the discipline and result in more accurate and effective solutions in the future.

---

[3]The definition of the term ego-network could be ambiguous or unclear to the reader as there is no uniform definition across different disciplines. To avoid confusion, the concept of ego-network used in this work refers uniquely to the connections between a single individual (the ego) and others, excluding relationships between others. This definition will be consistently used throughout the whole text whenever the term ego-network is mentioned.

1.1 The Social Brain Hypothesis

Primates stand out among other animal species for having larger brains, considering their body size (Jerison, 1973). Moreover, it has always been widely accepted that there is a strong correlation between this characteristic and their more extraordinary cognitive abilities. This has caused natural scientists to be interested in understanding the connection between brain size and social behaviour in primates. The appearance in the 19th century of Darwin's theory of evolution by natural selection highly augmented this interest.

The debate reached its peak in the second half of the 20th century. In this period, three predominant theories emerged to explain how primates had evolved large brains. The first theory emphasised the evolution of cognitive abilities from an ecological viewpoint (Clutton-Brock & Harvey, 1980; Gibson, 1986). This theory exposed that primates reside in many diverse habitats, and these cognitive abilities have been developed to enhance their adaptation to different environments, giving them an advantage over other species to survive. The second theory approached this issue from a social perspective (Humphrey, 1976; Kummer, 1982). These authors underlined that primates' social lives are highly complex; consequently, they had been required to develop exceptional abilities to handle them. Finally, the third theory was a purely biological explanation (Martin, 1981). It argued that maternal nutrition is crucial in explaining the larger brain size of primates. However, the first two were the most widely accepted theories within the scientific community, as some research that questioned the latter's validity had also been published (Harvey & Pagel, 1991; Pagel & Harvey, 1988). These authors reasoned that the cost of maintaining a large brain is enormous and that it is inherently impossible for brains to evolve in size simply due to these biological reasons.

Under all these premises, Whiten and Byrne (1988) presented their book *Machiavellian Intelligence Hypothesis*, so named in honour of the Italian political and philosopher of the Renaissance Niccolò Machiavelli, author of the famous and most influential book *The Prince*, which consists on some pieces of advice to rulers that are as valid nowadays as they were in the 15th century. This hypothesis suggested that some species have developed specific cognitive adaptations for mastering the complexities of their social interactions and the environment. The authors referred to these adaptations as "Machiavellian intelligence". Whiten and Byrne defended that these abilities had provided evolutionary advantages to primates and could be used to explain the development of their intelligence.

A few years later, in the early 1990s, Dunbar (1991) published his initial research on understanding the reason behind primate grooming behaviour. This article challenged the traditional view about grooming, which held that it was exclusively carried out for hygiene and health purposes. Dunbar argued that these two are not the only purposes of grooming, but it also serves a social and pleasure-related function.[4] Furthermore, he studied the relationship between grooming time, body size and group size for various primates species and found that grooming time correlated much better with group size than with body size. This result was surprising and provided strong support for his hypothesis.

---

[4]To emphasise this argument, Dunbar (1991) points out that some primates allocate approximately 20% of their entire day to both giving and receiving grooming. For comparison, other animals of similar body sizes do not devote even 2% of their time to this activity.

With this in mind, Dunbar (1992a) decided to participate in the ongoing debate between ecological and social theories. In his paper "Neocortex size as a constraint on group size in primates", he questioned the validity of the "Machiavellian intelligence" as a factor in explaining the development of cognitive skills in primates due to the fact that this concept is somewhat vague and difficult to quantify. As an alternative, Dunbar investigated whether a correlation existed between the average group size of different primate species ($\mathcal{N}$) and their neocortex ratio ($\mathcal{C}_R$).[5] He found that the following log-log relationship ($R^2 = 0.764$) holds:

$$\log_{10}(\mathcal{N}) = 0.093 + 3.389 \log_{10}(\mathcal{C}_R). \tag{1.1}$$

Dunbar (1992a) also investigated any relationship between the average group size and different variables related to the ecological theory, such as the percentage of fruit consumed in the diet, range area size and total length of the day's journey. However, he didn't find any significant results. These findings suggested that group size may be determined by neocortical volume, and the brain capacity limits the organism's ability to process information. Furthermore, primates that live in large groups have more potential friendships than those in small groups, so they need larger brains to manage them. Therefore, when a group size surpasses its limit, it becomes unstable and breaks into smaller subgroups. All these findings together provided clear evidence supporting the social theory and against the ecological one. The foundations of a new social theory, the Social Brain Hypothesis, had just been established.

1.1.1 Dunbar's number

Shortly after discovering the relationship between average group size and neocortex size of different primate species, Dunbar (1993) asked himself the obvious question: does this relationship have any implications on the natural size of human groups? To answer it, Dunbar directly applied Equation 1.1 to the case of humans using the data of their neocortex ratio, whose value is $\mathcal{C}_R = 4.1$.[6] The prediction, when the human neocortex size is plugged into the equation, is 147.8. The results are presented in Figure 1.1. However, this value is frequently rounded to 150 and is widely referred to as Dunbar's number.[7]

This prediction was revolutionary, as 150 appears to be a surprisingly small number for the average natural size of human groups. For this reason, Dunbar decided to include in the same study the analysis of experimental data that supported his results. He used a dataset of 21 hunter-gatherer societies worldwide. Each of these societies is formed by communities that consist of a group of people who live in a common territory and hunt together. Dunbar reported that the sizes of these communities ranged from 100 to 200 individuals, with an average of 148.4, which closely matches the theoretical prediction of 147.8.

---

[5]The neocortex is the region of the brain with the greatest cognitive capacity. The neocortex ratio refers to the ratio of neocortex volume to the volume of the rest of the brain. Dunbar (1992a) studies the relationship between the average group size of different primate species and many other variables, such as neocortex total volume or neocortex ratio (vs. hindbrain). Finally, he uses the neocortex ratio (vs. the rest of the brain) as a predictor because it presents the highest correlation with the average group size. Considering the success his theory has achieved, it seems he made the right decision.

[6]Dunbar (1993) points out that this value is roughly 30% larger than the highest value for any other species, which could explain the superior development exhibited in comparison to other primates.

[7]Dunbar (2021) confesses in his book *Friends: Understanding the power of our most important relationships* that Gonçalves et al. (2011) were the first people that referred to this concept as Dunbar's number in a scientific publication.
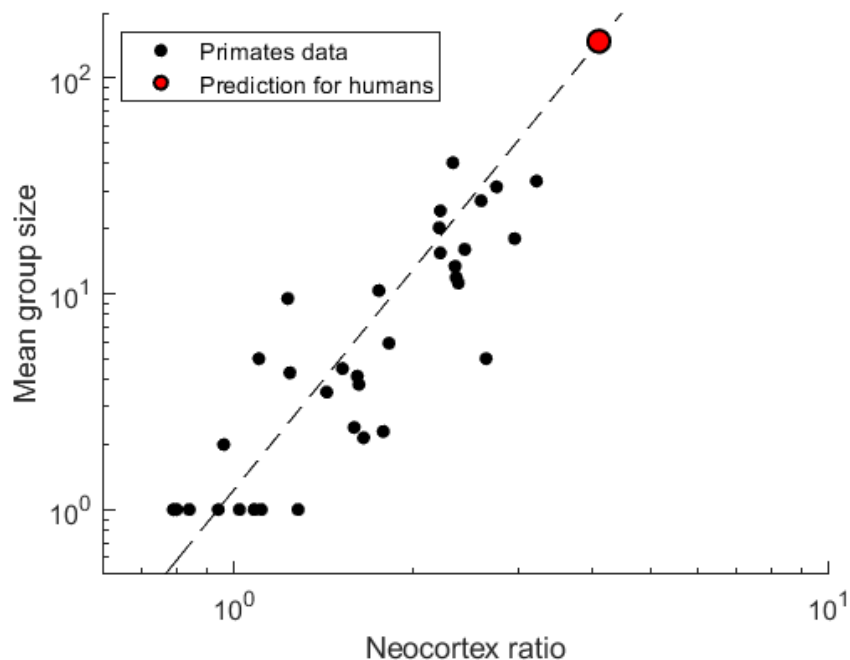
Figure 1.1: The Social Brain Hypothesis - The figure represents the mean group size of different primate species against their neocortex ratio. The black dots are the experimental data observed for 33 primate species, while the red dot shows the prediction for humans. The dotted line corresponds with the log-linear fit governed by the Equation 1.1. This figure is a replica of Figure 1 in Dunbar (1992a), using the data published in the article. For an alternative representation[8], see Figure 1 (page 51) in Dunbar (2021), where the primates species are divided based on four different grades in the social brain relationship. Additionally, an independent fit is included for each of these groups.

In the following years, various studies were published challenging the validity of Dunbar's number, and its robustness has been nothing short of amazing. It has been observed that this pattern holds in many different situations: postal mail exchanges (Killworth et al., 1984), face-to-face interactions (Roberts et al., 2009), mobile phone calls (MacCarron et al., 2016), the online world (Dunbar et al., 2015; Gonçalves et al., 2011; Haerter et al., 2012) and even in the number of people to whom you would send a Christmas card (Hill & Dunbar, 2003).

Furthermore, in recent times the development of technology has led to the emergence of more detailed studies to determine the size of the human brain accurately (Kanai et al., 2012; Kwak et al., 2018; Lewis et al., 2011; Noonan et al., 2018). The results of these studies have been compared with the ego-network size of the participants, and all have confirmed the same hypothesis: individuals with smaller brain sizes can handle fewer relationships simultaneously. These findings have emphasised the dominance of social factors over ecological ones and have led to the popularity of the Social Brain Hypothesis and Dunbar's number.

---

[8]I would like to take this opportunity to thank Robin Dunbar for allowing me to use the figures from his book *Friends: Understanding the power of our most important relationships* in this text. However, I finally decided not to include them and create alternative representations to ensure no copyright issues.

1.1.2 The hierarchical structure: Dunbar's circles

When Dunbar (1993) examined the data of hunter-gatherer societies worldwide, he quickly realised that all of them exhibited a common pattern in terms of their social structure. Moreover, the complexity of this structure was beyond what he could have expected. Each society was formed by different communities scattered across the same territory. In turn, each of these communities was divided into three or four camp groups. Dunbar observed that bigger formations were also created, so several communities were grouped into mega-bands and several mega-bands were combined to form a tribe. This was the first evidence of societies being divided into layers representing a characteristic set of relationships.

Dunbar (1998b) decided to expand this investigation from a more quantitative perspective. He noticed that hunter-gatherer societies were organised into structural groups clustered around a series of characteristic values: 5, 12, 35, 150, 500 and 2000. Dunbar also argued that these values might represent points of stability in the familiarity of all types of relationships, from the closest ones to the more distant ones. In other words, not all relationships are equal; instead, they differ in closeness or level of intimacy. These factors are responsible for the division into groups, with the more intimate relationships corresponding to smaller groups.

A few years later, Hill and Dunbar (2003) continued examining the hierarchical structure of social relationships. Hence, they designed an experiment where they asked a group of white British people[9] about who they would send a Christmas card to (i.e. individuals whose relationship they value). The participants were also requested to supply some information about the person(s) they would contact, such as distance, type of relationship, social status, last contact or emotional closeness (on a scale from 0 to 10). The analysis of the responses provided some revealing results. The average size of the participants' networks was 153.5, very close to the prediction of 150 obtained for humans. Moreover, the authors identified hierarchical grouping levels in the data very similar to those observed in hunter-gatherer societies. As a result, Hill and Dunbar suggested that varying levels of emotional closeness could determine this hierarchical structure.

The existence of these highly similar patterns in both studies caught the attention of Zhou et al. (2005), who decided to reanalyse[10] the data using more sophisticated mathematical techniques. More specifically, the authors utilised techniques used for fractal analysis and identified a uniform trend in all these societies: relationships are organised into layers (or circles) with characteristic sizes that follow a geometric progression with a scaling ratio around 3.[11] These layers are sequentially inclusive of one another so that the innermost layers, which correspond to the most intense relationships, are also included in the outermost ones. The estimated sizes of each layer are approximately 5, 15, 50, 150, 500 and 1500. This result is illustrated in Figure 1.2 and represents the hierarchical structure observed in human societies, also known as Dunbar's circles.

---

[9]Hill and Dunbar (2003) state that the reason for choosing only white British respondents was to minimise cultural influences.

[10]To be more precise, in this study, the authors reviewed the data from the hunter-gatherer societies and the Christmas card sending experiment, but also analysed other datasets from the social network literature (Adams et al., 2002; Kef, 1997; Marsden, 2003).

[11]Zhou et al. (2005) reported an approximate scaling ratio of 3.2 with a confidence level of 0.993.
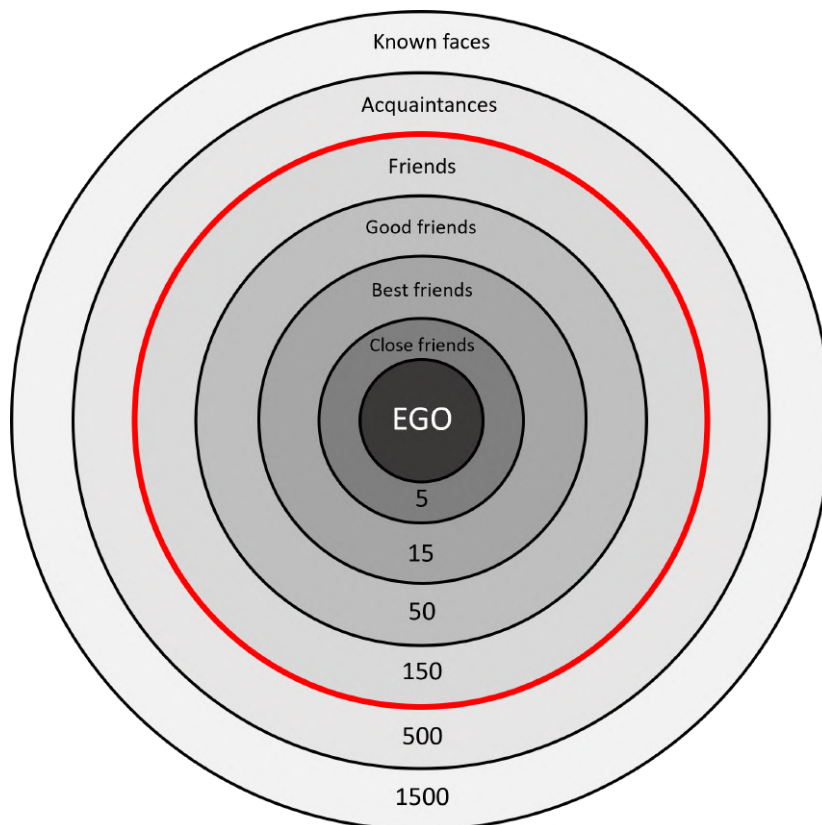
Figure 1.2: The hierarchical structure: Dunbar's circles - The figure represents the characteristic layered structure that forms society. The numbers represent each layer's approximate size, including the layers within it. These sizes follow a geometric progression with a scaling ratio of approximately 3. The red line shows the limit to the number of simultaneous relationships an individual can maintain, the so-called Dunbar's number. This figure is a replica of Figure 3 (page 71) in Dunbar (2021).

Dunbar (2021) provides some insight into the connection between these layers and the emotional closeness of their associated relationships in a more familiar way. The 5-layer (or support clique) corresponds to close friends, the 15-layer (or sympathy group[12]) to best friends, the 50-layer to good friends and the 150-layer to just friends. These 150 relationships constitute each individual's ego-network. Dunbar also points out that the meaning of the two remaining layers is less clear. The 500-layer may correspond to individuals with whom someone is familiar, such as coworkers, and the 1500-layer to the number of people someone can name, even if they have never been met in person.

Recent studies suggest the existence of two additional layers. On the one hand, researches into online social networks (Arnaboldi et al., 2017; Arnaboldi et al., 2015; Gonçalves et al., 2011) and telephone calls (MacCarron et al., 2016) uncover a layer with a size of around 1.5. This layer would correspond to people with whom the relationship is at its maximum, such

---

[12]This term was first introduced by Buys and Larson (1979), who were the first authors that suggested the idea that there is a limit to the number of people for whom one can feel sympathy.

as a romantic partner or a very close relative. On the other hand, another layer with an approximate size of 5000 has also been predicted. These findings are supported by research conducted by Jenkins et al. (2018) in the area of facial recognition. Their results indicate that an individual's ability to recognise faces, even without knowing their names, is estimated to be around this value. It's important to emphasise again that, even including these two new layers, the scaling ratio between successive ones continues to be approximately 3.

In recent years, evidence has also been published showing that this characteristic pattern is not exclusive to humans. Other species with complex social systems exhibit similar groupings. Hill et al. (2008) analysed data for some species of mammals and found the same pattern[13] and scaling ratio as the observed for humans. Dunbar et al. (2018) revealed similar trends in their analysis of average group sizes among various primate species. These results suggest the hypothesis that the social world is organised into a series of characteristic layers, and this structure is highly consistent across individuals, societies, cultures and even species.

Additionally, the robust nature of these results has drawn the attention of the scientific community for years. As a result, a large number of researchers have tried to find the reasons behind this characteristic hierarchical structure. The most widely accepted hypothesis is that the intensity of a relationship depends on the time invested in it. As time is a limited resource, each person allocates it to others based on how highly they value each relationship.

The first result in this direction was proposed by Max Burton, who showed the expectation of mutual help is directly proportional to the amount of time invested in a relationship. Sutcliffe et al. (2012) put numbers to the question. By analysing the contact frequencies of each person among the different circles of their social network, they showed that 40% of the time was devoted to the closest layer, the support clique, an additional 20% to the rest of the people in the sympathy group and only the remaining 40% to all other relationships, devoting less time to those more distant. Miritello et al. (2013) demonstrated that these patterns are not limited to face-to-face interactions. They analysed a database of phone calls and found that individuals who called more people did not spend more time on the phone, but instead spent less time calling each of their contacts. Moreover, the optimal number of contacts is 150.

In light of all these results, it can be concluded that time is limited and constrains how social relationships are organised. There exists a trade-off between the time invested in the different relationships and the benefit gained from them, leading to the emergence of the hierarchical layered structure of social networks.

1.1.3 The "social atom"

Making predictions in social sciences is challenging due to human behaviour's complex and unpredictable nature. In general, these predictions are rather obvious or too simplistic. This makes the robustness of the results proposed by Dunbar even more remarkable. However, the problem remained partially unsolved because some questions still needed to be answered: Why is the scaling ratio between the different layers constant? And more importantly, why

---

[13]Hill et al. (2008) found the same pattern up to the layer of 50 individuals, as the mammal groups they studied did not go beyond this value.

1. INTRODUCTION

is the value of this scaling ratio approximately 3?[14]

Once again, these questions attracted the attention of the scientific community. Sutcliffe et al. (2016) were the first who tried to answer them. These authors designed an agent-based model to tackle the problem, reproducing the observed layered structure using a set of behavioural rules and specific parameter values. They found that this structure only arises when there are substantial benefits for individuals and extensive levels of social interaction, which might mirror what happens in the natural world. However, this model was difficult to understand and computationally expensive, making it difficult to gain more insights from it.

Thanks to interdisciplinary collaboration in science, Tamarit et al. (2018) provided a new and very different perspective on the problem. These authors interpreted the organisation of social networks as a resource allocation problem and used a technique from statistical physics, the maximum entropy principle, to explain the layered structure. They assumed two basic and easy-to-accept constraints: the number of relationships that can be maintained simultaneously is limited, and the cost of maintaining each varies based on its emotional intensity.[15] From them, the authors showed that a hierarchical structure emerges, which is governed by a single parameter μ. Depending on the value of this parameter, two different regimes can arise: the standard one ($\mu > 0$) and the inverse one ($\mu < 0$).

On the one hand, the standard regime corresponds to the traditional layered structure observed, where the number of individuals increases in successive circles. Tamarit et al. (2018) highlight that this is the regime observed when studying a community with a large population, where individuals have more opportunities to select their social relationships. On the other hand, the model predicted the inverse regime, which had never been noticed before. This would be the expected regime in small communities, where the individual has fewer opportunities to create weak relationships, so unused time is employed to create more close relationships. The authors confirmed this hypothesis by using data from four immigrant communities, where the options for choosing friends are limited to a small number of members. In consequence, the relationships that emerge are more intense.

The most remarkable point here is that the model predicted an unexpected and surprising regime, which prompted the authors to search for data to verify it. This approach is contrary to the traditional way of proceeding in the social sciences, where, first, data is analysed, and later, a model is developed to explain it. In other words, the traditional process moves from data to the model instead of from the model to the data, as Tamarit et al. (2018) did. This perfectly exemplifies the need for interdisciplinary science, where techniques from very different fields are combined to obtain astonishing results.

Moreover, Tamarit et al. (2018) provide an explanation for the layered structure of social networks: the maximum entropy principle dictates it is the most probable organisation of these relationships, considering the constraints. This leads to the concept of "social atom" (Tamarit, 2019), as it replicates a system similar to that observed in physical particles, and represents how each individual (the "atom") manages their social relationships.

---

[14]Zhou et al. (2005) claim that the fundamental question is to determine the origin of the hierarchical layered structure, and there is no clear explanation for why the scaling ratio must be approximately 3.

[15]As explained in detail in the previous subsection, this emotional intensity highly correlates with the time invested or the benefit gained from each of these relationships.

Tamarit et al. (2022) extended this discussion. The traditional structure of ego-networks had always been considered a discrete set of layers with characteristic sizes, so, in their first model, they assumed a discrete set of categories to allocate resources. However, they proposed an expanded version of the model in which the layers were omitted and the social relationships were classified on a continuous scale. The authors also observed the two regimes identified in the discrete version of the model in the continuous one. Furthermore, this new model allowed them to study continuous[16] data measuring the intensity of social relationships, without having to categorise it. Consequently, Tamarit et al. (2022) introduced a universal parameter η, that does not depend on an arbitrary choice of the number of layers and explains how an individual's social relationships are organised.

1.1.4 Personal effects and homophily in social networks

One of the most complex challenges humans face in their daily lives is dealing with the social world. Individuals react to this situation in many different ways depending on their abilities. For this reason, all patterns observed in human social behaviour are averages that can vary significantly from one person to another. Of course, not everyone has 150 friends; some people have more, and others have fewer. This is a manifestation of the natural variability present in the real world. Determining the factors contributing to this variability has been a subject of study in the last few years.

Stiller and Dunbar (2007) laid the foundations in this direction, arguing that differences in the number of relationships an individual maintains might be related to specific mental abilities, such as memory. Pollet et al. (2011) proceeded in this direction by demonstrating that personality also has a significant influence. The authors proposed the existence of a "social capital" that extroverts distribute along more people. Thus, their ego-networks are larger than those of introverts. A scientifically-backed answer to this question has also been proposed. Advancements in neuroimaging techniques have allowed for a high level of precision in determining the size of the different brain regions. Several studies have confirmed that the greater the number of friends an individual has, the larger the brain areas[17] involved in social skills (Kwak et al., 2018; Lewis et al., 2011; Noonan et al., 2018). Again, these findings are not limited to face-to-face interactions but extend to online relationships (Kanai et al., 2012). These results suggest differences in social networks based on an individual's cognitive abilities.

Moreover, other biological factors have been identified as necessary in accounting for variations in network sizes. One such factor is age. Wrzus et al. (2013) found that networks increase in size initially with age, reaching a peak around the thirties and then decreasing as one becomes older. These authors also show that it is the friendship component of the network that changes in size over time, while the family one remains relatively stable. Another biological factor that has a significant influence is gender. Some studies have exposed that

---

[16]Tamarit et al. (2022) study three different datasets that measure the intensity of social relationships on a continuous scale: duration of phone calls (Saramäki et al., 2014), face-to-face interaction time (Isella et al., 2011) and number of likes in Facebook (Arnaboldi et al., 2012).

[17]The prefrontal cortex, the temporal lobe and the temporo-parietal junction (TPJ) are the brain regions associated with social abilities.

men and women have different[18] approaches to social relationships (Barrett et al., 2000; Dunbar, 2016). Men prioritise a sense of belonging to a group, while women place greater importance on intimate dyadic relationships. A recent study by Kiesow et al. (2020) revealed differences in brain structure between both sexes, which might contribute to the differences observed in social behaviour.

Another key aspect in explaining the structure of social networks is homophily. People tend to build friendships with those with similar characteristics, such as gender, age or ethnicity. Dunbar (2021) incorporates a list of seven cultural factors into this classification, based on the findings of various studies (Curry & Dunbar, 2013a, 2013b; Launay & Dunbar, 2015). He names these factors "The seven pillars of friendship", which are the following: language, growing location, educational and career experiences, hobbies, worldview, sense of humour and musical tastes.

Hence, very diverse factors influence how an individual socially relates to others. These factors are specific to each person, embody their personality and tend to remain stable over time. Dunbar (2021) refers to them as the "social fingerprint". He argues that this must be considered when understanding the social structure, as it complicates making generalisations about overall behaviour. Each person has a unique "social fingerprint" that sets them apart from others. Therefore, the conclusions reached are usually averages that provide an overview of the system, but may not apply to every individual case.

1.2 Exponential random graphs

After introducing the anthropological theory on which our research is based, the Social Brain Hypothesis, we explore some techniques for modelling social systems. Over the years, different methods have been developed to understand these systems, but networks have undoubtedly been one of the most impactful tools, allowing the field to advance significantly. Networks provide a powerful way to visualise and analyse the complex relationships that shape social systems. For example, they can be used to identify influential individuals within a society, understand the flow of information through it or predict how changes in its structure might affect the formation of communities.

In recent decades, the interdisciplinary nature of science has become increasingly evident due to the significant effort invested by the scientific community in developing realistic models of social networks. As a result of this interdisciplinary approach, a variety of methods have been employed to study these networks, including mathematical modelling, statistical analysis and computational simulations. Researchers from different disciplines, including economics (Jackson et al., 2012), computer science (Wellman, 2001), biology (Alm & Arkin, 2003; Mason & Verwoerd, 2007) or statistics (Snijders, 2011), have devoted considerable effort to formulating them. The field of statistical physics has also significantly contributed to the study of social networks (Albert & Barabási, 2002; Girvan & Newman, 2002; Newman, 2018; Toivonen et al., 2009). Even, some of the most popular techniques used today for studying social systems, ensemble models of networks, are closely related to this discipline.

---

[18]The authors emphasise that it is essential to remember that all of these conclusions refer to average behaviours.

Ensemble models of networks (or random graph models) are an important theoretical tool used to investigate properties that emerge from random processes, providing insights into the fundamental mechanisms of network formation. Among them, exponential random graphs (ERGs) have proven to be helpful in analysing social networks. These models allow us to create graphs that exhibit similar characteristics to networks observed in the real world, offering a valuable understanding of their structure and dynamics.

1.2.1 A brief historical introduction

Moreno and Jennings (1938) conducted a study that is considered the foundational work in the analysis of social networks. These authors were pioneers in using statistical methods to compare the social interactions among the participants in their research with those expected by a null model. While the statistical methods used were relatively simple, they suggested that the relationships that appeared could not be purely random, but instead, there had to exist underlying structural effects that introduced some bias.

A few years later, Erdös and Rényi (1959) presented the first work in the field of random graph theory, the well-known Erdös-Rényi model of random graphs. In this model, a graph with n vertices is constructed by including each possible edge with probability p, independent of all the other links. This revolutionary idea profoundly impacted the study of network models at that time.

Holland and Leinhardt (1981) developed a statistical model based on dyadic independence. They named it the "p1 model", and it was the first attempt to go beyond simple random graph distributions. This model provides a characterisation of a graph by counting the number of edges that connect vertices in it. Furthermore, its probability distribution can be used to estimate the likelihood of a specific graph within the ensemble. Even though the authors only conducted dyad analysis, this work laid the foundation for exponential random graph models (ERGMs), so-called because the logarithm of their probability distribution is linear.

Another key idea in the development of this family of models was proposed by Frank and Strauss (1986). They concluded that adopting an approach that accounted for dependence was essential, rather than looking for new methods based on network independence. For this reason, they introduced Markov random graph models intending to adapt spatial statistical techniques to network analysis.

The article published by Wasserman and Pattison (1996) marked the turning point that led to the overall adoption of ERGs and their surge in popularity. They emphasised the form of the log-linear probability distribution that defines this family of models, accentuating their simplicity and utility. Wasserman and Pattison also highlighted the estimation of newly proposed parameters, which allowed the model to be adapted to various social networks. Furthermore, the authors proved that conditional independence assumptions could be used to derive this family of models, subject to the conditions dictated by the Hammersley-Clifford theorem.

Since then, the ERGs have become one of the most widely used tools by the social network community (Robins et al., 2001; Snijders, 2002, 2011). Moreover, the increase in computing power and the ease of fitting real data have allowed researchers to analyse larger and more

complex networks leading to new insights into their structure, including the discovery of new patterns and properties that were previously unknown (Fronczak, 2014; Koskinen et al., 2013; Lusher et al., 2013; Robins et al., 2007; Snijders et al., 2010; Stivala et al., 2016).

1.2.2 Traditional statistics approach

In the field of complex systems, network modelling analysis has become an essential tool for understanding the relationships and interactions between different components (Snijders, 2011). When formulating a model for network analysis, the objective is to create a mechanism that reproduces the characteristics of real networks.

Ensemble models of networks address this problem using purely probabilistic methods; the observed networks emerge as a result of applying a stochastic process over a set of them that share some features (at least, the same number of nodes). Usually, the stochastic process that generated the observed networks is unknown, so the objective is to formulate a reasonable and theoretically justified hypothesis for it (Robins et al., 2007). This process is governed by a probability distribution $P(G)$, which determines the likelihood of a network appearing in the ensemble so that networks that are more similar to those observed are more likely to be present. Therefore, an ensemble model is entirely determined by two elements: the set $\mathscr{G}$ of all possible realisations of the network and its probability distribution $P(G)$.

To determine the probability distribution $P(G)$, the starting point is a set of graphs with data available on a total of p distinct properties. In other words, the values of $\{x_i\}$ where $i = \{1, ..., p\}$, have been obtained through empirical observations of the networks. Furthermore, it is possible to estimate its average value $\langle x_i \rangle$.[19] In consequence, the probability distribution satisfies

$$\langle x_i \rangle = \sum_{G \in \mathscr{G}} P(G) x_i(G), \tag{1.2}$$

where $x_i(G)$ is the value of the observable x_i on the graph G.

The observables x_i are called sufficient statistics because they comprise all the information in the data. In the probability distribution $P(G)$, each has an associated parameter θ_i. The value of these parameters determines the probability of a specific graph G appearing within the complete distribution.

Wasserman and Pattison (1996) generalised the probability distribution of the exponential random graphs (ERGs) family. This probability distribution provides a direct and elegant way to link each observable to its respective parameter, thanks to its log-linear form. It is defined by

$$P(G) = \frac{\exp\left(-\boldsymbol{\theta} \cdot \boldsymbol{x}(G)\right)}{\sum\limits_{G' \in \mathscr{G}} \exp\left(-\boldsymbol{\theta} \cdot \boldsymbol{x}(G')\right)} = \frac{\exp\left(-\boldsymbol{\theta} \cdot \boldsymbol{x}(G)\right)}{K(\boldsymbol{\theta})}, \tag{1.3}$$

where $\boldsymbol{x}(G) = (x_1(G), x_2(G), ..., x_p(G))$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)$ are two vectors formed by the observables of the network and their associated parameters, respectively. $K(\boldsymbol{\theta})$ is a normalisation constant to ensure that the probabilities sum 1.

---

[19]In practice, it is common to have only a single network realisation available. In such cases, the value of each observable is treated as its average value.

The probability distribution determined by Equation 1.3 can represent any configuration on the graph space that assigns a positive probability to each possible graph. This issue can be accomplished by computing the maximum likelihood values of the parameters θ_i for specific values of the observables x_i. However, calculating the normalisation constant becomes intractable analytically when these observables involve more complex structures, such as triangles. In these cases, it becomes necessary to appeal to Monte Carlo techniques, which are computationally expensive.

1.2.3 Statistical physics approach

Park and Newman (2004b) tackled this problem from a different perspective, leveraging the interdisciplinary nature of science. These authors realised that the probability distribution given by Equation 1.3 arises from applying the maximum entropy principle subject to the constraints outlined in Equation 1.2. With this in mind, they integrated ERGs into the language of statistical physics and related them to other types of models commonly used in this discipline (fluids, ferromagnetic materials, gases...). Well-established techniques to deal with these problems allowed for obtaining approximate analytical results without requiring simulations (Grandy Jr, 2012; Newman, 2018). Thanks to this approach, these techniques can be extended to network analysis.

The derivation proposed by Park and Newman maximises the Gibbs entropy

$$S = - \sum_{G \in \mathscr{G}} P(G) \log P(G), \tag{1.4}$$

subject to the constraints outlined in Equation 1.2 and the normalisation condition

$$\sum_{G \in \mathscr{G}} P(G) = 1. \tag{1.5}$$

By utilising the Lagrange multipliers α and β_i, the probability distribution $P(G)$ must satisfy the equation

$$\frac{\partial}{\partial P(G)} \left[S + \alpha \left(1 - \sum_{G'} P(G') \right) - \sum_i \beta_i \left(\langle x_i \rangle - \sum_{G'} P(G')\, x_i(G') \right) \right] = 0. \tag{1.6}$$

This expression is equivalent to

$$\log P(G) + 1 + \alpha - \sum_i \beta_i x_i(G) = 0, \tag{1.7}$$

from where it is deduced that the probability distribution $P(G)$ takes the form

$$P(G) = \frac{e^{-H(G)}}{\Xi}. \tag{1.8}$$

Therefore, the Hamiltonian[20] of the network is

$$-H(G) = \sum_i \beta_i x_i(G). \tag{1.9}$$

---

[20]Notice that the sign conventions used here are different than those in Park and Newman (2004b), to agree with those commonly adopted in statistical mechanics of lattice gases. This is the convention we will use throughout all this work.

The grand partition function associated with the Hamiltonian is obtained as

$$\Xi = e^{\alpha+1} = \sum_{G \in \mathscr{G}} e^{-H(G)}. \tag{1.10}$$

One can observe that the functional form of Equation 1.8 is identical to the probability distribution in Equation 1.3. Thus, this derivation justifies the choice of the exponential form which defines the ERGs: it is the least biased distribution compatible with the constraints in Equation 1.2. From the Bayesian point of view, this is the optimal choice since $P(G)$ includes exclusively the information provided by the constraints and nothing else. Any additional detail about the distribution would require modifying its form.

Consequently, an ERG is entirely defined by its probability distribution, Hamiltonian and grand partition function. In statistical physics, $P(G)$ is known as the Boltzmann-Gibbs distribution over all the possible microstates of the system, which in this case are all the graphs G that belong to the ensemble $\mathscr{G}$.

The expected value of any property of the graph can be obtained as

$$\langle x_i \rangle = \sum_{G \in \mathscr{G}} P(G)\, x_i(G) = \frac{1}{\Xi} \sum_{G \in \mathscr{G}} e^{-H(G)}\, x_i(G). \tag{1.11}$$

However, this average can be computed more easily by just defining the grand potential[21] of the system

$$\Omega = -\log \Xi. \tag{1.12}$$

Then, the mean values of the observables x_i are obtained throughout

$$\langle x_i \rangle = -\frac{\partial \Omega}{\partial \beta_i}, \tag{1.13}$$

where β_i is the Lagrange multiplier associated with the observable x_i in the Hamiltonian of the graph.

1.3 Summary and objectives

In the preceding sections, we introduced the theoretical foundations of our research as a starting point for understanding it. Our work falls under the discipline of social physics, an interdisciplinary and emerging field that applies concepts and methods from physics and mathematics to study social phenomena. The ultimate goal of social physics is to uncover universal principles that underlie human behaviour in social systems.

Our research is based on the Social Brain Hypothesis, a theory proposed by the British anthropologist Dunbar (1998b). This theory posits that the number of social relationships an individual can maintain simultaneously is limited to around 150, which is often referred

---

[21]Park and Newman (2004b) refer to this term as free energy because it is the nomenclature adopted in the physics of spin models. However, we are adopting the nomenclature commonly used in statistical mechanics of lattice gases. We will use it throughout all this work.

to as Dunbar's number. This limit is believed to be a result of the cognitive capacity of the human brain. According to the Social Brain Hypothesis, the evolution of the large human brain was driven by the demands of living in complex social groups. Therefore, the size of the neocortex, the part of the brain responsible for complex cognitive processes, is related to the size of an individual's ego-network. Moreover, all these relationships are organised in a characteristic layered structure known as Dunbar's circles. The size of each inclusive layer (5, 15, 50 and 150) is inversely proportional to the intensity of its associated links and follows approximately a geometric progression with a scale parameter of 3.

For a long time, there has been no explanation for this characteristic structure of the ego-networks. However, Tamarit et al. (2018) have recently proposed a model that interprets their organisation as a resource allocation problem. The model justifies this structure using the maximum entropy principle and shows that a maximum likelihood analysis, subject to only two basic constraints, yields a one-parameter probability distribution that describes the likelihood of a given relationship occupying a specific layer. This has led to the development of the concept of "social atoms", which explains the micro-level social structure of humans.

Our research aims to apply this formalism to the macro-level, where social systems can be understood as ensembles of interacting "social atoms", with the objective of comprehending the phenomena that emerge from the interactions. To achieve this, we use exponential random graphs (ERGs), a family of network models whose probability distribution can be derived using the maximum entropy principle. These models play a similar role in network analysis as the Boltzmann-Gibbs distribution does in statistical physics by providing the maximum likelihood prediction of parameters that considers the constraints imposed by observations.

Therefore, our motivation for this work is to gain a deeper understanding of the social structure. We hope to uncover some universal principles by applying the formalism of social physics to the study of complete social networks, which are macro-level systems. We believe that this approach has the potential to revolutionise our understanding of human behaviour and provide new perspectives on the fundamental mechanisms that govern social interactions by bringing together concepts and methods from physics and other related disciplines.

With these objectives in mind, we have divided this thesis into four parts. The first part utilises a data-driven approach to tackle the problem. We conducted multiple surveys on high school students, gathering data on their friendships and enmities throughout different academic years. We will explore this data to identify universal patterns in human social behaviour and highlight the significance of previously unexplored factors that play a key role in determining the structure and dynamics of social relationships, such as reciprocity or the impact of negative relationships. Additionally, we will introduce an online application that we have developed to enable school directors to leverage our findings and improve the social atmosphere in their schools.

In the second part, we will build models from first principles using exponential random graph models that allow us to explain the structure of social relationships. In particular, we will demonstrate that, for certain parameter values, these models exhibit a phase transition similar to that observed by Park and Newman (2005). We will apply density functional theory to obtain an energy functional that allows us to detail this phase transition. Lastly, we will

use the data on the relationships between high school students to validate these models. We will estimate the parameters that govern the formation of social relationships among the students by fitting the models to the data. These parameters will provide us insights into the underlying mechanisms that shape the social structure of the school and the emergence of social phenomena.

In the third part, we will demonstrate that all these models can be applied more generally than solely to human relationships. We will use grooming data from four different groups of chimpanzees living in the Chimfunshi Orphanage in Zambia to show that they exhibit the same behaviour patterns as those observed in humans by Tamarit et al. (2022). Our results will not only provide further evidence of the universality of these models but also shed light on the similarities between human and chimpanzee social behaviour.

Finally, in the fourth part, we will present our conclusions and discuss the implications of our findings. We will highlight the contributions of our study and propose potential avenues for future research.

PART I

DATA-DRIVEN PERSPECTIVE

2

UNDERSTANDING THE DYNAMICS OF THE SOCIAL WORLD

Social structure and dynamics constitute the foundations of human social life, shaping the interactions, behaviours and experiences of individuals within the social world. Therefore, analysing them is essential to understanding the functioning and evolution of societies, as well as the ways in which humans navigate and interact within them.

In this chapter, we explore the temporal evolution of social relationships among high school students over two consecutive academic years. Our study is based on survey data collected at five different points in time (waves), providing a comprehensive picture of the social landscape. This constitutes a unique opportunity to investigate the resilience of social structures and their evolution, which may prove particularly useful for understanding the social environment in high schools.

Our results show a very high degree of consistency among the survey waves, despite the complex and sometimes difficult-to-decipher dynamics of teenage social relationships. This consistency suggests a high degree of persistence of social structures in high schools. We investigate the role of various factors, such as being in the same class, gender homophily or reciprocity. Furthermore, we also examine enmities, which are reported to a much lesser extent and are highly volatile. Our findings contribute to the growing body of evidence supporting the layered structure of human social networks and offer valuable insights into the social dynamics of high school students.

2. UNDERSTANDING THE DYNAMICS OF THE SOCIAL WORLD

2.1 Data description

2.1.1 Data collection

We conducted a study on a group of students from IES Blas de Otero, a public high school in Madrid, Spain. The study was carried out during two complete academic years, 2020-2021 and 2021-2022, and consisted of 5 data collection sessions in which we surveyed the students about their social relationships and their skills. Two of these sessions were carried out during the academic year 2020-2021, in December 2020 (wave 1) and May 2021 (wave 2), and the remaining three during the academic year 2021-2022, in October 2021 (wave 3), February 2022 (wave 4) and May 2022 (wave 5). Academic years in Spanish middle schools begin around mid-September and end around mid-June. The participants in our study were in 1st, 2nd and 3rd year of middle school (in Spanish, Enseñanza Secundaria Obligatoria, i.e. ESO) during the academic year 2020-2021 and were between 12 and 16 years old.

The study was approved by the Ethics Committee of Carlos III University of Madrid and carried out in accordance with the approved guidelines. Consent was obtained from the school, which adopted this as a research project of its own and, in turn, got informed consent from the parents of the participants. Students always participated voluntarily and signed informed consent prior to beginning the study.

Data were obtained from surveys carried out in the computer lab of the high school or on the own mobile phones of the participants, always under the supervision of a teacher. These surveys were conducted using SAND, a software developed by our research group in collaboration with other universities. This application has allowed us to obtain very diverse information about the students, such as their friendship and enmity relationships and their associated intensities, their social and cognitive abilities or the classmates with whom they prefer to collaborate. Despite the richness of this data, we will focus only on studying the structure and dynamics of their social relationships. However, the school principals have access to all this information and are using it to improve the social atmosphere in the centre.

The first question of the survey related to social relationships was the following:

- Who are your friends within the school?

Then, a list with the names and surnames of all their schoolmates (from all courses) appeared on the screen, and they could select all those whom they considered their friends without having a minimum or maximum limit. After finishing answering the question and pressing the accept button to move to the next screen, they were asked about their best friends:

- Of those previously selected, who are your best friends within the school?

Then, a list appeared on the screen that only showed the schoolmates they had selected in the first question, allowing the user to choose their best friends from among them.

For negative relationships, an analogous process was followed. In this way, for each of the participants, we were able to know their "best friends", "friends", "enemies" and "worst enemies". The information is gathered as a list of student IDs labelled as $\{+2, +1, -1, -2\}$, respectively. The combination of all these relationships forms the links of our high school

networks. Note that each student provides one such list, so we are extracting a directed network. This process is repeated independently in each of the five data collection sessions conducted throughout the study.

As for the differences between waves, in the academic year 2020-2021, due to COVID prevention measures, the structure of the school was different from the standard one. Students who were in 1st and 2nd year of ESO during this course were divided into eight classes with some 15 students each. In the same year, students in the 3rd year of ESO were divided into five classes of some 25–30 students each, and within each class, they were split into two subgroups which, because of COVID, attended school physically on alternate days. In the academic year 2021-2022, these students advanced to the 2nd, 3rd and 4th year of ESO as already stated, and the school returned to a pre-COVID structure, i.e. 5, 5 and 4 classes with some 25 students in each year, respectively, attending school physically on a daily basis. In addition, there are two teaching itineraries in this school, one that is taught mostly in English, except Spanish and Mathematics, and another that is taught mostly in Spanish, except Arts and Physical Education, which are taught in English. Approximately 40% of the students take the English itinerary. For a complete description of the school's demographics, see subsection A.1.1.

Finally, it is also important to highlight that the older participants, i.e. those from the 3rd year of ESO during the period 2020-2021, had already participated in a similar study[1] that we conducted previously in the same school during the academic year 2018-2019 (Escribano et al., 2021). Unfortunately, we cannot establish a correlation between the results of both studies as they were carried out using two different applications, and due to privacy concerns, we can only identify each student through an ID which is different in both applications, and we have no way to connect them. Despite this limitation, our current study provides sufficient information to investigate the social dynamics of high school students.

2.1.2 Data curation

In this second study, we made an adaptation of our survey application as a consequence of what we learned from the first study (see section 3.1 for a detailed discussion). This greatly simplified the data curation process. In this case, we have only focused on ensuring that the responses were consistent throughout the five waves, so we have filtered out participants who did not meet some requirements.

The first requirement we have imposed is that the student had participated in all five waves. Among them, outliers were removed according to the following criterion: those who reported more than 100 relationships or more than a 200% change (upwards or downwards) in the number of selected relationships from one wave to the next, were discarded from the sample. From the 285 students that answered in all five waves, we discarded 64 outliers: 22 reported more than a hundred relationships, while 42 had too much variation between the number of answers in consecutive samples. This leaves us with a final sample of answers from 221 students for the five waves.

---

[1]In spite of the chronological sequence, we have decided to present the results in a different order to give more coherence to the discussions addressed in each chapter. However, the ideas presented in section 3.1 may help to understand the development of the questionnaires and the reasons behind their current form.

2.2 The evolution of social relationships

To start our longitudinal study about the structure and dynamics of social relationships, we decided to explore the friendship relationships reported by the participants. The number of friendships and best friendships among students remained remarkably consistent across the five waves, despite the vastly different situations in the two academic years under study, as can be observed in Figure 2.1. This finding is consistent with the results presented by Kucharski et al. (2018), who concluded that the reported number of relationships over time showed great consistency. In our study, even with drastic changes in the composition of the groups and the school situation in waves 3, 4 and 5, the results remained almost constant across the five waves.

Furthermore, the results were consistent across all data subsets, including class, gender, itinerary and whether the student was a "repetidor" (i.e., taking the course for the first or second time, with the latter being referred to as "repetidor"). This finding applied to both friendships and best friendships. When a regression line was plotted through the number of friends in each of the five waves for each student, the median of the slope was exactly 0 (see subsection A.1.4).

To further examine the structure of ego-networks, we utilised the parameter μ as a tool for analysing the layered structure of an individual. This parameter was first introduced by Tamarit et al. (2018) and discussed in detail in our previous study (Escribano et al., 2021). To obtain its value for each individual, we fitted the following analytical expression:

$$\mu = \log\left(\frac{C_2 - C_1}{C_1}\right), \tag{2.1}$$

where C_1 represents the reported number of friendships in the innermost circle and C_2 in the circle associated with the sympathy group.

When μ is greater than 0, the circles exhibit a characteristic structure where the number of friendships increases rapidly as we move away from circle 1, the innermost one. A value of μ close to 0.7 is typically observed when the scaling ratio between the sizes of the circles is around 3, which is often the case. Conversely, when μ is less than 0, most friendships are concentrated in circle 1, and the rest of the circles have very few additional people. Negative values of μ are observed in situations where the number of possible links is limited (e.g., sailors on a boat, communities of migrants, etc.) or for introverted individuals. The distribution of the μ parameters is relatively constant, and in fact, most individual values of μ change very little across the five waves (see subsection A.1.3), despite the organisational changes and the composition of the groups when they transition from the first to the second academic year.

To study the evolution of relationships, we compared the nature of links between pairs of subjects in consecutive waves, w_n and w_{n+1}, using labels of +2 for "best friend", +1 for "friend", 0 for "no link", −1 for "enemy" and −2 for "worst enemy", and computed the corresponding conditional probabilities. Therefore, Figure 2.2(a) shows $P(x, w_5 | +2, w_4)$ and Figure 2.2(b) $P(x, w_5 | +1, w_4)$, while Figure 2.2(c) shows $P(x, w_5 | -2, w_4)$ and Figure 2.2(d) $P(x, w_5 | -1, w_4)$. All the other conditional probabilities are shown in subsection A.1.5.
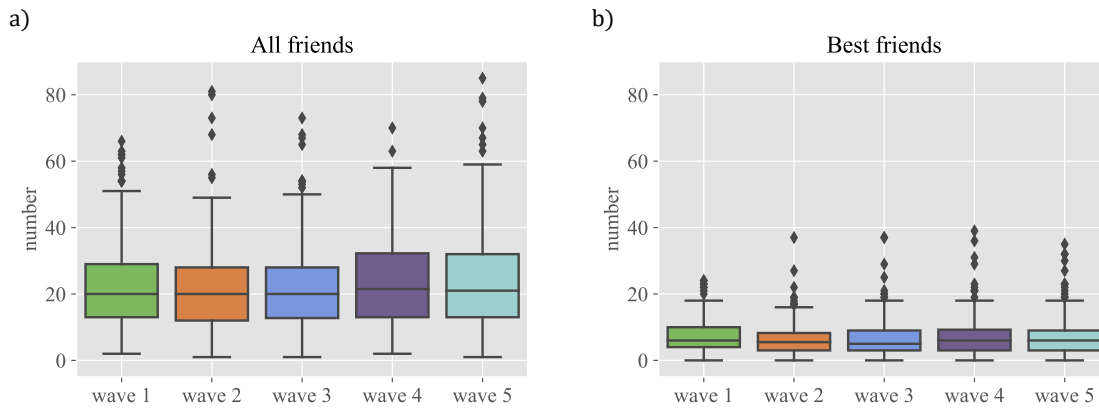
a)

b)

Figure 2.1: The structure of friendships - The figure represents boxplots corresponding to: a) the number of friendships and b) the number of best friendships reported by the students in each of the five waves of the survey.

The first two circles related to friendships (best friends and just friends, respectively) evolve in very different ways: best friends are quite stable, and when they stop being best friends, they usually end up as just friends. In contrast, just friends are more dynamic and may disappear from the radar or, in some cases, even become best friends.

We have also looked at the opposite evolution, namely where students that appear for the first time as best friends were in previous waves, i.e. $P(x, w_{n-1} | + 2, w_n)$. We have observed that best friends were often already best friends in the previous wave, and new ones come primarily from being just friends. Therefore, the first two circles show apparent differences in the stability of their relationships, arising from the different intensities in best-friend vs. just-friend relationships.

Enmities are very few and highly volatile. The total number of negative relationships is an order of magnitude smaller than that of positive ones. As can be observed in Figure 2.2, most relationships marked as worst enemies or just enemies in one wave are not kept in the next wave. Interestingly, worst enemies are retained with higher frequency than plain enemies. These results point to friendships and enmities having a different nature, with friendships being more long-lasting and enmities reflecting, in general, more the heat of a specific conflict. However, very bad relationships may last longer. All this has to be taken with a grain of salt though, because of the poor statistics of these relationships. More compelling data is needed before any conclusion can be drawn reliably.

The results described so far, the stability of the number of relationships across waves and the higher turnover of the outer friendship layer $(+1)$ compared to the inner one $(+2)$, suggest picturing individuals as "social atoms". In this metaphor, layers play the role of atomic orbitals, whereas individuals act like electrons. Inner orbitals attract electrons more strongly than outer ones, so there is less turnover. Also, electrons may leave their orbitals for good, leaving a "hole" quickly replaced by a new electron. Likewise, friends who leave the ego-network get replaced by new friends, so their average number remains constant.

On the other hand, we observe a larger degree of turnover than that reported by Roy
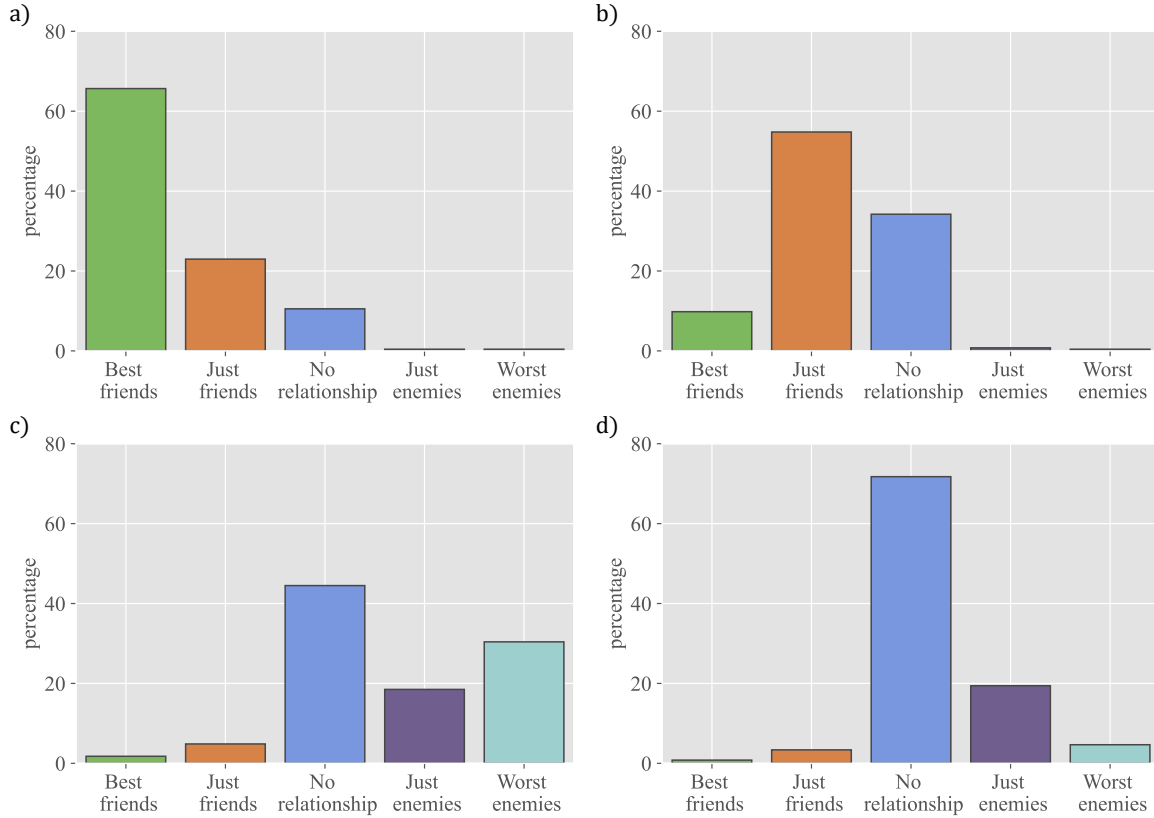
Figure 2.2: The evolution of social relationships over time - The figure represents the percentage of individuals that ended up in a given category in wave 5, when they were marked in wave 4 as: a) best friend, conditional probability $P(x, w_5| + 2, w_4)$, b) just friend, conditional probability $P(x, w_5| + 1, w_4)$, c) worst enemy, conditional probability $P(x, w_5| - 2, w_4)$, or d) just enemy, conditional probability $P(x, w_5| - 1, w_4)$.

et al. (2022). This may be because the latter study deals with data obtained from phone calls between adults. Interestingly, even in that study participants aged 17–21 showed a larger turnover than those older than 21 years old. Nonetheless, even Roy et al. (2022) reported differences between layers similar to what we find here. This points to the role of developmental issues in the evolution of the structure of personal friendship networks. Care has to be taken, though, because the phone data should capture the general structure of people's friendships, whether they are family, workmates, friends, etc., and familiar ties, for instance, are particularly resilient.

In contrast, here we restrict the students to tell us only about their relationships within the school. In this respect, we see that Dunbar's circle structure reproduces itself in each domain of relationships: a fraction of each student's cognitive capabilities are devoted only to school. Then, from that limited capability, the structure emerges as predicted by Tamarit et al. (2018). The more rapid turnover could be related to the smaller cognitive capacity devoted to the specific niche of school relationships and to the mean-less of these ties compared to family or lifelong friendships.
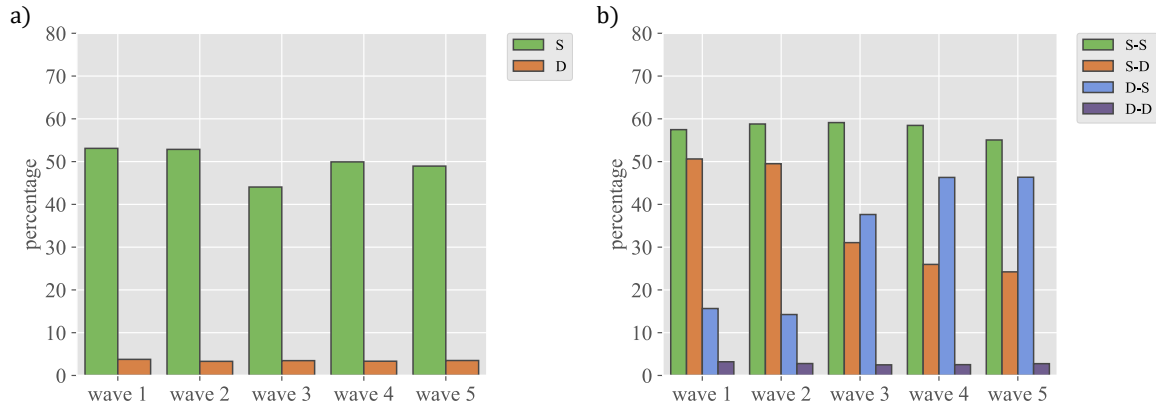
Figure 2.3: The importance of the group for the existence of relationships - The figure represents the following: a) percentage of relationships formed between individuals in the same group (S) versus those in different groups (D) relative to the total number of relations that might potentially form in each of the two cases and b) percentage of relationships between individuals that are in the same group both academic years (S-S), in the same group the first year but different the second (S-D), in a different group the first year and the same group the second (D-S) and different group both years (D-D), referred in each case to the total number of possible relationships.

One confound that may influence the evolution of relationships is the distribution of students in the different classes. Generally speaking, approximately 70% of the relationships among the students are with other students in their same year and, of those, a majority are with students in their same class. In fact, among all potential relationships within the same class, approximately 50% of them are actually reported, whereas less than 5% of all potential relationships with students in different classes exist, as can be observed in Figure 2.3(a).

The fact that this is an important factor can also be observed in Figure 2.3(b), where pairs of students are divided into groups according to whether they were in the same class in the two academic years included in our data, 2020-2021 and 2021-2022 (hereafter referred to as S-S), in different classes both years (hereafter D-D), in the same class in the first year and in a different class in the second year (hereafter S-D) and vice versa (hereafter D-S). Then, Figure 2.3(b) shows the percentage of relationships in each of these groups that were actually reported in each wave. Importantly, the change in the academic year between wave 2 and wave 3 students is associated with a reshuffling of the classes. This is reflected in a decrease of relationships S-D, going from values close to 50% to 25%, (orange bars in Figure 2.3(b)), i.e., the separation led to the disappearance of half of the existing relationships. On the contrary, the plot shows an increase in the percentage of relationships D-S (blue bars in Figure 2.3(b)). In this case, the percentage rises from 15% to almost 45%, comparable to the starting point of the other group. This observation should be compared to the almost constant percentages of S-S and D-D relationships.

This clearly shows that being in the same class is a very relevant driver for relationships to decay or start. It also speaks of a certain weakness of the ties formed in middle school, compared to those arising in different contexts. Moreover, it is interesting to note that when
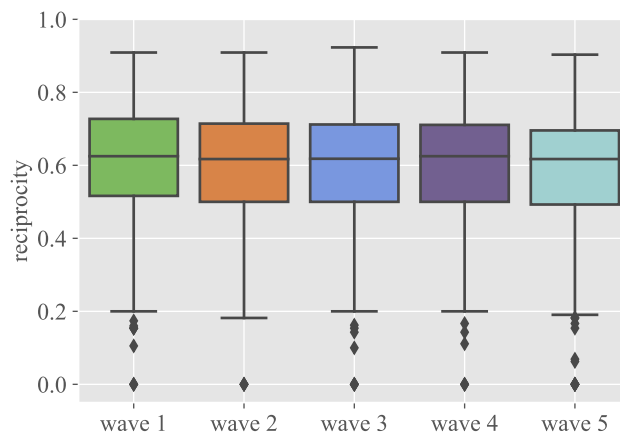
Figure 2.4: The evolution of reciprocity over time - The figure represents boxplots that correspond to the distributions of individual reciprocal relationships in each wave. It is surprising to observe how the median value approximately remains constant, around 60% in the five waves, since there are no external mechanisms that induce it.

students are separated, the number of relationships that still remain in the second year is almost twice as large as relationships D-S in the first year, meaning that it takes longer for relationships to disappear due to separation than to form upon becoming together.

2.3 The importance of reciprocity

Our longitudinal study enables us to examine the reciprocity of relationships, which is a critical issue in social science research. As illustrated in Figure 2.4, we found that the overall percentage of reciprocal relationships remained consistently high at around 60% across all five waves of our study. While most individuals showed a similar degree of reciprocity in their relationships, there were some participants for whom reciprocity was significantly lower.

Moreover, we have observed that these findings were consistent across different groups, genders and itineraries of the participants. These results can be observed in subsection A.1.7. The fact that reciprocity levels did not significantly differ across different groups, genders, and itineraries indicates that this tendency is a fundamental aspect of human social behaviour rather than specific to certain groups.

As reciprocity is also a property of relationships, it is worth considering their dependence on the personal characteristics of both people involved. Figure 2.5(a) shows the percentage of reciprocal links between individuals of the same gender and also the percentages of the four types of temporal evolution discussed above. Regarding gender, the plot shows that homophilic links are generally more reciprocal, while mixed-gender links are less reciprocal. Interestingly, when mixed-gender links are not reciprocal, it is not due to a gender bias (see subsection A.1.7).

We can also observe in Figure 2.5(b) that reciprocity is quite high in relationships between students that remain together the two academic years (S-S), is lowest for students that
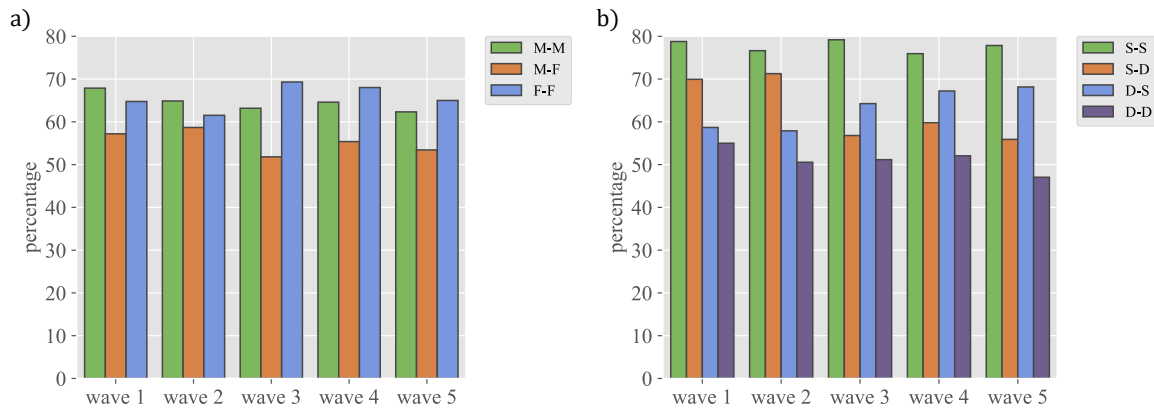
a)

b)

Figure 2.5: The importance of the group for the reciprocity of relationships - The figure represents the following: a) percentage of reciprocal links according to the individual's gender, M-M (male-male), M-F (male-female), F-F (female-female), and b) percentage of reciprocal links according to whether the pair of individuals are in the same or different class in consecutive years.

are always separate (D-D), while S-D and D-S links decrease or increase, respectively, in reciprocity in later waves. Similar results arise when looking at relationships in the same class or itinerary. Reciprocated friendships are also more stable, as are triangles formed only by positive relationships. In both cases, they are much more stable than any other combination.

Finally, our study suggests that the reciprocity of relationships is not only a fundamental aspect of human social behaviour, but it also plays a critical role in forming and maintaining social networks. As individuals tend to reciprocate social interactions, a lack of reciprocity or the appearance of a negative link in a relationship may indicate an imbalance of power on it. We will explore these phenomena and their consequences in detail in the following chapter.

2.4 Discussion

In this chapter, we have studied the temporal evolution of relationships among 12-16 years old students that attend the same high school. The study consisted of five waves of surveys during two consecutive academic years and included positive and negative relationships and their intensities. The number of students answering all five waves of the survey was 224.

In spite of what one could expect, we do not observe any signs of fatigue among the students and their responses are remarkably similar in every wave, thus confirming the consistency of the data collection reported by Kucharski et al. (2018). Furthermore, the number of reported friends and best friends is quite constant in the five waves, irrespective of groups and ages, genders, itineraries or being a "repetidor". Therefore, we have a very rich longitudinal dataset that can be used to address several important issues.

The survey results show that friendships in the innermost circle (best friends) are more stable than the rest of the friendships in the second circle (just friends). This observation provides further evidence of the key role of Dunbar's circles in the organisation of relationships

but also supports the idea that the intensity of a relationship in the first circle is higher than in the second one, apparently making them more stable. On the contrary, enmities are few, much less frequent than friendships and highly volatile, with many simply disappearing from one survey wave to the next. Note, however, that this does not mean that learning about enmities is irrelevant, as we will explore in the following chapter. This fact is, therefore, important for the daily dynamics of the class and, as such, it is highly valuable information for the school principals.

Our study also points to the importance of being in the same class for forming and stabilising friendships. As discussed above, the strongest friendships arise among students in the same class during the two academic years we have studied. The change from being in the same class one year and in a different one the following year leads to the loss of a sizable fraction of friendships, which are then refocused on new classmates. Friendships among students that never shared class are much rarer in comparison. These observations suggest that we tend to have our relationship structure occupied at all times, as the friends lost because of the separation are replaced with new classmates. In addition, it also highlights the importance of frequent interaction in keeping or weakening relationships.

Another interesting observation concerns the topic of reciprocal friendships, an important issue given its connection to performance (Candia et al., 2022) or to the success of behavioural interventions (Almaatouq et al., 2016). We have observed that reciprocity is remarkably constant. For most individuals, a percentage between 50% and 70% of their relationships are reciprocal. In general, gender homophilic relationships are slightly more reciprocal, and vice-versa, male-female relationships tend to be less reciprocal, with both genders being equally responsible for this effect. The reciprocity of a relationship also shows the effect of group reshuffling and evolves in a manner similar to the friendships themselves. Furthermore, there are a few individuals whose reciprocity is very low, which could be an indicator of possible socialisation problems for those particular subjects, providing yet another valuable hint for the school principals.

We want to end this discussion by summarising the big picture that can be inferred from this study. As already mentioned, everybody seems to have a predefined structure of their relationships, despite their frequent turnover. The structure is akin to that of an atom with its electrons, with less turnover in the inner layers than in the outer ones, hence the "social atom" metaphor. This suggests the possibility of studying the formation of social networks as a statistical-mechanical system in equilibrium with every relationship having an associated "binding energy", which is the cost to remove the link. Consequently, the question to investigate is if one could map this system into one of the models available in the statistical mechanics of networks. For example, those models that are based on exponential random graphs (Escribano & Cuesta, 2022; Park & Newman, 2004b; Strauss, 1986).

From a philosophical point of view, one such mapping would imply that the "total energy", or the Hamiltonian in the statistical-mechanical jargon, describes a social system better than a graph. Graphs are volatile and constantly changing, whereas the energy uniquely determines a network ensemble, of which any observed social network would be but a specific instance. This perspective would open a big avenue to re-think social systems from a new viewpoint and will be discussed in detail in the second part of this thesis.

3

THE ROLE OF NEGATIVE RELATIONSHIPS IN SOCIAL STRUCTURE

In the realm of human social relationships, it is often the positive relationships that take centre stage. Friendships are commonly celebrated for their potential to enrich our lives, foster personal growth and enhance the social fabric. In chapter 2, we examined the structure and dynamics of positive relationships, which helped us develop the "social atom" concept further. We also explored negative relationships, which are characterised by conflict or rivalry. Our findings demonstrate that these relationships exhibit a different structure than positive ones, as they are less common and less enduring over time. However, the impact of these negative relationships should not be underestimated because they play a critical role in shaping social structures and community dynamics.

In this chapter, we explore the complex and multifaceted world of negative relationships in the context of high school social environments. Here, we explore how these challenging relationships influence community structure. Our focus on understanding the relationships among high school students provides valuable insights into the interplay between negative relationships, social cohesion and community dynamics. Therefore, this chapter lays the groundwork for understanding the complexities of enmities and their impact on communities. Furthermore, our analysis also considers the role of social balance, a fundamental concept in social network theory that suggests individuals are more likely to form and maintain relationships that minimise tension and conflict within the network.

3. THE ROLE OF NEGATIVE RELATIONSHIPS IN SOCIAL STRUCTURE

3.1 Data description

3.1.1 Data collection

We conducted a study on a group of students from IES Blas de Otero, a public high school in Madrid, Spain. The students were in their first year of secondary school, and their ages ranged from 12 to 13 years old. Most students were new to the high school, having completed their primary education elsewhere in the preceding year. However, a few students were repeating the first year after failing the previous academic year. Therefore, most of the participants did not know each other or had a minimal relationship before the start of the academic course, except for those who came from the same primary school.

Our research involved two data collection sessions held in December 2018 (wave 1) and May 2019 (wave 2), respectively. A total of 151 students, consisting of 73 boys and 78 girls (for more detailed information, refer to subsection B.1.1), along with their families, agreed to participate in the study. The participation rate was very high, with 97% of the students answering the surveys in the first wave and 90% in the second wave.

The students surveyed for this study were born in 2006 (75%), 2005 (23%), and 2004 (2%). They were organised into five groups (classes) and were all enrolled in the first year of ESO, a mandatory secondary education program similar to middle school. The groups were labelled with letters, ranging from A to E. Groups A and B received most of their instruction in English, with the exception of Maths, Spanish, and the optional course in either Religion or Ethical Values. On the other hand, groups C through E were taught primarily in Spanish, except for Arts and Physical Education classes, which were conducted in English.

The main goal of this study was to understand the structure of the social relationships among the students in terms of both their friendships and enmities. For this reason, all the respondents were given a list containing all their schoolmates' names and were asked to indicate their relationships with them by marking their choices. More specifically, they had to respond to the following questions (in Spanish in the original survey):

- Questions regarding positive relationships (friendships):

  1. Who are your friends inside the school?

  2. Considering your friends: Who do you have the closest relationship with?

  3. Finally, among your closest friends: who would you say are your best friends? (We are referring to those people with whom you are "flesh and bone").

- Questions regarding negative relationships (enmities):

  1. Which partners do you not like at all or do not have a good relationship with?

  2. Considering the people you don't like at all: who do you dislike or have problems with?

  3. Finally, considering the people you dislike: are there any people with whom you have a particularly bad or troublesome relationship?

The study was approved by the Ethics Committee of Universidad Carlos III de Madrid, the institution responsible for funding the project and conducted following its guidelines. The privacy and confidentiality of the participants have been protected at all times, and all data was kept securely and confidentially.

3.1.2 Data curation

At this point, it is essential to highlight that, despite the logical order we have chosen to show the results in this thesis, this study was conducted before the one presented in chapter 2. The software used in our first study was more rudimentary, and many things we learned from analysing this data helped us improve the new software. Even though some students overlap between the respondents of both studies, we can only identify them by their unique IDs (not by their names) due to data protection issues. Therefore, we can not compare the results between the two studies for those students who participated in both because the software programs are different and the IDs are unrelated.

As mentioned in the previous section, a total of 151 students participated in our study. Each participant was presented with a list of the remaining 150 schoolmates, and they were asked to select their relationships as either friendship or enmity using three different levels of intensity. While we were supervising the study, we started to suspect that, for different reasons, especially their young age, the participants could not correctly discern between the three different intensity levels of each possible relationship. We confirmed this intuition when we analysed the data and noticed the high variability in the two most intense categories in the responses of different students. For this reason, we decided to merge the two levels of higher intensity, for both friendships and enmities, into a single one.[1] Therefore, positive and negative relationships are now classified into two different intensity levels, which we will henceforth refer to as "friend" and "best friend" for positive relationships and "enemy" and "worst enemy" for negative ones.

After transforming the data, we found that the average number of "best friends" selected by each student in the first wave was 14.48, and the average number of "friends" was 27.01, while in the second wave, the average number of "best friends" was 11.27 and the average number of "friends" was 28.47. These values are slightly higher in both waves than those presented in section 2.1, where the average values also remain approximately constant across the five waves. The reason for this difference is the filtering of outliers we made in chapter 2 to eliminate students who decided to "troll" the survey. However, we believe that these differences do not influence the results presented in this chapter, which aim to understand the importance of negative relationships.

As for reported negative relationships, their absolute number is much smaller than that of positive ones, with average values of 4.82 for "enemies" and 2.07 for "worst enemies" in wave 1, while 7.83 and 2.11, respectively, in wave 2. Even if there are a few negative links, we will show their importance in explaining the social structure in the following sections.

---

[1]Here, we should point out that the students not distinguishing correctly between the three categories is one of the issues we learned from the data analysis, which we have corrected in the new survey software. We have reduced the number of categories in the intensity of the relationships, both positive and negative, to two instead of the three we had in our first study before the merge (see section 2.1 for details).

3.2 The formation of communities

One of the most extensively researched areas in the field of complex systems is the problem of identifying communities in networks since it has relevance in many disciplines such as biology, physics, computer science or social network analysis (Newman, 2018). This problem amounts to identifying groups of nodes in a network that have a high degree of interconnectivity among them and are relatively more loosely connected to the rest of the network. Therefore, communities can be seen as clusters of nodes with similar structural properties or functional roles, and they can provide insights into the organisation, dynamics and functionality of complex systems.

Historically, the problem of finding communities in networks has been approached from very diverse perspectives (Albert & Barabási, 2002; Girvan & Newman, 2002). However, the introduction of the concept of "modularity" by Newman and Girvan (2004) revolutionised the field. This concept quantifies the degree to which a network can be partitioned into densely connected groups of nodes that are only sparsely connected to each other and can be regarded as a measure of the quality of a particular community partitioning of a network: the higher the modularity value, the better the partitioning. In other words, modularity quantifies how far is the studied network structure from a random distribution of links.

More specifically, to incorporate all the information gathered from the student surveys in our analysis, we utilise the most general definition of modularity Q proposed by Granell et al. (2011). This definition considers the presence of directed, weighted and signed links in the network, that is

$$Q = \frac{w^+}{w^+ + |w^-|} Q^+ - \frac{|w^-|}{w^+ + |w^-|} Q^-, \tag{3.1}$$

where w^+ and w^- represent the total strength (sum of weights with signs) of positive and negative relationships, respectively. The construction of Q using both the positive (Q^+) and negative (Q^-) modularities implicitly involve the possibility of establishing signed links independently. Therefore, total modularity provides a balance between the inclination of positive relationships to build communities and the tendency of negative ones to break them up. For our case of study, links in the network can only take values $\{-2, -1, 0, 1, 2\}$ depending on the relationship sign (enmity or friendship) and intensity.

The positive modularity Q^+ is defined as

$$Q^+ = \frac{1}{2w^+} \sum_i \sum_j \left(w_{ij}^+ - \frac{w_i^{+,\text{out}} w_j^{+,\text{in}}}{2w^+} \right) \delta\left(C_i, C_j\right), \tag{3.2}$$

where w_{ij}^+ is the (i, j) element of the positive weighted adjacency matrix and $w_i^{+,\text{out}}$ and $w_j^{+,\text{in}}$ are the total strength of the positive links getting out of node i and of those coming into node j, respectively. In other words, $w_i^{+,\text{out}} = \sum_k w_{ik}^+$ and $w_j^{+,\text{in}} = \sum_k w_{kj}^+$. C_i represents the community to which individual i belongs, so the Kronecker delta $\delta(C_i, C_j)$ takes the value 1 if both nodes share community and 0 otherwise. The negative modularity Q^- is defined by an analogous expression to Equation 3.2 but using the absolute values of the negative weights.

The optimisation of modularity is a computationally challenging problem that belongs to
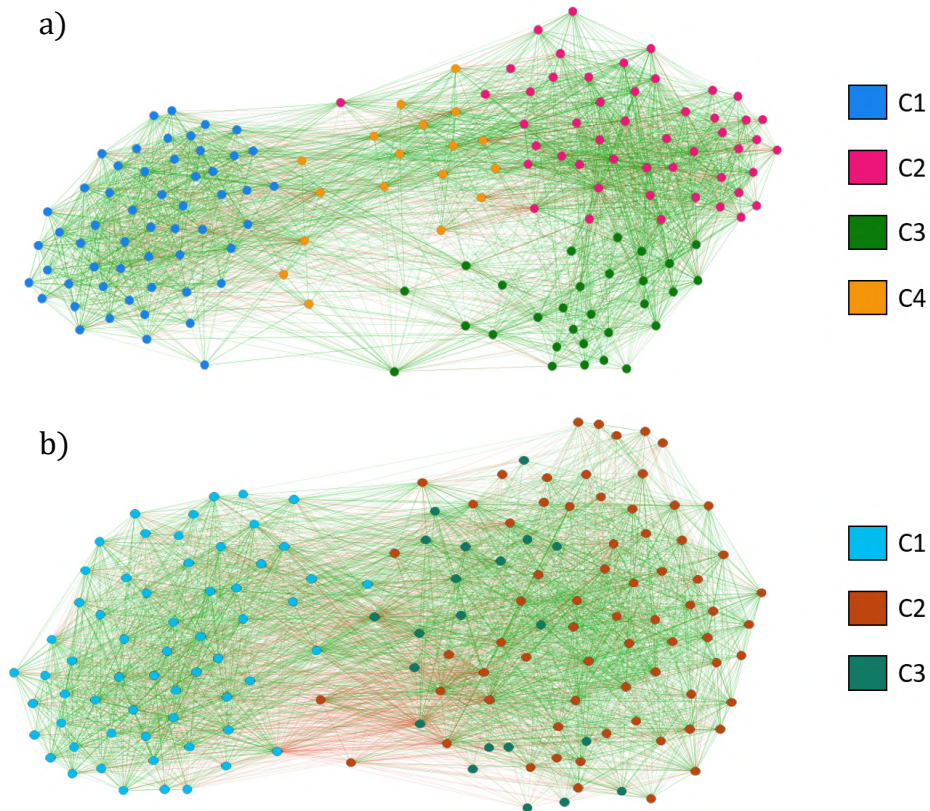
Figure 3.1: Distribution of social relationships and communities in the complete school - The figure represents the network characterising the social relationships and communities in the complete school divided by different time stamps: a) wave 1 and b) wave 2. Nodes belonging to each community are marked by colour as indicated in the figure. Friendship relationships are drawn in green, whereas enmity relationships are drawn in red.

the family NP-hard, as the number of possible combinations grows exponentially with the size of the network. Therefore, we need to employ a combination of heuristic algorithms to find the best partition (or a close one) of the network into communities (i.e., the one that maximises Q as defined in Equation 3.1). These algorithms aim to find a suitable solution within a reasonable amount of time, rather than guaranteeing the optimal solution.

In our study, we start by discussing the corresponding results and exploring the school at a global level, which includes all the students and their associated relationships in a single network. Figure 3.1 represents the global network obtained from the two waves, with the nodes separated into communities. We used the plotting algorithm ForceAtlas2 (Jacomy et al., 2014), which by itself yields a community analysis. The high correlation between colours and positions in the plot is an outcome of this algorithm and has not been externally imposed. This, in turn, indicates that the communities found are quite robust.

The data summarised in Table B.2 shows that in the network of wave 1, community C1 is formed by the students in the two English-speaking groups, whereas community C3 contains

groups C and D, and community C4 mainly contains group E. Community C2 is formed by a minority of students from almost every group, but mainly from group D. The separation observed in the plot between community C1 and the rest indicates a profound separation between the two English-speaking groups and the three Spanish-speaking ones. This division most likely originates from students who just entered high school two months before wave 1 and still rely more on colleagues from the same primary school. Unfortunately, we do not have data on their school of origin, which could help us understand this phenomenon.

It is worth noting that the network plot in wave 2 appears more spatially extended than that of wave 1. This is due to almost doubling the number of negative relationships mentioned earlier, while the number of positive links remains more or less constant. However, at the whole group level, negative links do not appear to influence the community structure strongly.

As shown in the figures in section B.2, if the community analysis is conducted without the negative links, the only change is in wave 1, where the community arising from group E is merged with community C2, which is formed by an assortment of students from all Spanish-speaking groups. In wave 2, there is no significant change in the number or composition of the clusters. Therefore, we conclude that identity factors (i.e., English- vs. Spanish-speaking) are the most relevant at this level, and negative links lead to minor changes if any.

In wave 2, Figure 3.1 shows that the separation between groups that speak different languages still persists despite a whole academic course passing. However, the increase in the number of friends is noticeable in the fact that the two communities are now somewhat closer. Interestingly, our algorithm detects one fewer community because all Spanish-speaking groups are now part of a single community, C3. C1 remains the English-speaking part of the students, now including all but one of them, while C2 is a smaller community formed by people from groups C and D. It is worth noting that group D remains split into two communities in the two waves, making it the only separated group. Something is likely acting as a dichotomising criterion here, such as the school or country of origin, but unfortunately, we do not have the data to assess the mechanism behind this splitting. In the new survey software, we have included some questions related to the school of origin of the students that may provide us with some insight into this topic. However, other information, such as the country of origin, has been impossible to obtain due to student privacy issues.

The table of the number of people that change between the communities in section B.1 confirms that the movements of individuals between communities took place, as discussed earlier. Additionally, the numerical data shows that, while communities C1 and C3 inherit the gender distribution from their constituent groups, with a 2:1 girl-to-boy ratio in C1 and an almost 1:1 ratio in C3, C2 is practically a male-only community, with 19 boys and 2 girls.

After discussing the network as a whole, we will now turn our attention to each group separately. For this analysis, we only consider the links within the five groups and discard the reported friendships with students in other groups. At this level, the primary variable observed is a strong division by gender. For example, let's consider group C (see Figure 3.2 and section B.1). The group consists of 15 boys and 15 girls, and in wave 1, our community algorithm returns two communities, one with 12 boys and 5 girls, and the other with 10 girls and 3 boys. This division persists in wave 2, with only two students exchanging communities.
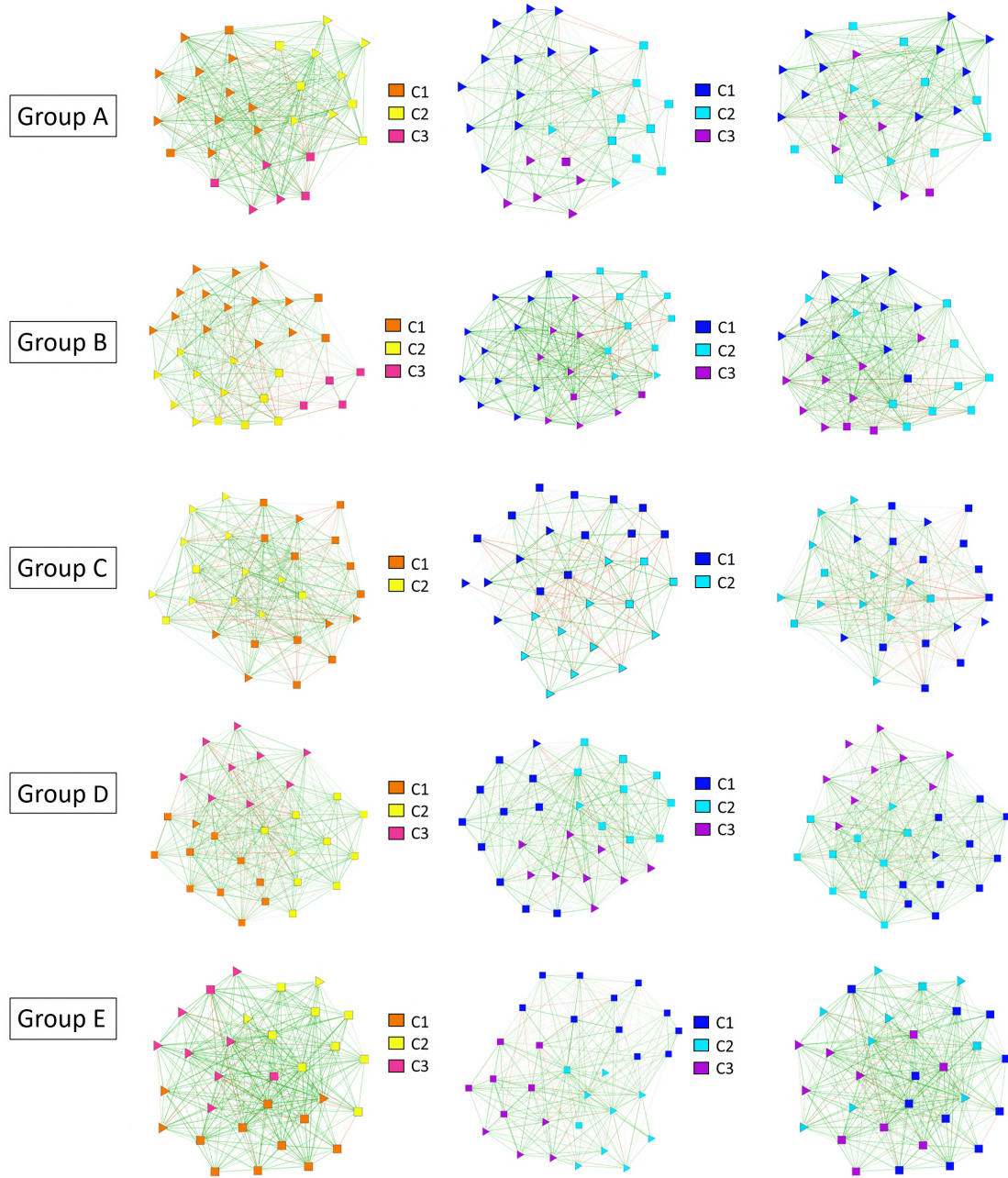
Figure 3.2: Distribution of social relationships and communities within each group - The figure represents the network characterising all the social relationships and communities within each group in wave 1 (left) and wave 2 (centre) along with the community structure of the network. The networks on the right also correspond to wave 2 (centre) but, in order to facilitate comparison, nodes are laid out in the same position as in wave 1, whereas links and colour codes for communities correspond to wave 2. The shape of each node represents its gender: triangles for females and squares for males.

However, the situation becomes a bit more complicated when the groups are less gender balanced. Let's observe one of the English-speaking groups, e.g., group A, with 20 girls and 9 boys. In wave 1, we find a community of 11 girls and only 2 boys, whereas the other communities are better (or exactly) gender-balanced. In wave 2, we again find three communities, but now all of them are gender dominated: 12 girls and 0 boys, 3 girls and 8 boys, and 5 girls and 1 boy.

Figure 3.2 shows that the evolution of the communities has been quite complex in the period between the two waves, with many students exchanging communities. For instance, if we look at community C1 in group A and in wave 1, only 5 of its 13 original members remain, with two groups of 4 students moving to the other two communities. This reinforces our interpretation of an evolution strongly dominated by gender homophily. However, it also suggests that the limits imposed by Dunbar's circles are destabilising the process of forming a unique, large cluster composed only of same-gender children, making it difficult to accommodate. The same is observed in the other groups, except for group D, which turns out to be remarkably stable in comparison. Its 10 boy-1 girl, 9 boy-1 girl, and 9 girls-only clusters almost do not change between waves (again, with the exception of 3 students who exchange communities).

We want to point out that our algorithm is agnostic regarding the number of communities it should produce, finding simply the best partition in terms of modularity. Therefore, the fact that we never observe a cluster with more than 13 persons of the same gender, despite the gender homophily bias and the intense dynamics of the communities, supports the idea that other mechanisms, such as those behind the formation of Dunbar's circles, are at work here.

At this point, it is important to note that Dunbar's circles refer to ego-networks, which are the set of relationships of one person. In contrast, communities pertain to the realm of social networks, which are the set of all relationships among a group of people. Nonetheless, we believe that the connection we have suggested makes sense given that, as mentioned in chapter 2, these networks are highly reciprocal, turning the clusters into something closer to cliques. Each ego-network would contain everybody else within the limit of a perfect clique, and, with the high percentage of reciprocity, ego-networks in a cluster still include most of the rest of the group. Thus, Dunbar's limits could have a say in their evolution.

Therefore, our analysis of the school's social network reveals several interesting patterns. At the global level, we observe a strong division between English- and Spanish-speaking groups, with the latter dividing into two communities in wave 2. At the group level, the gender plays a significant role in community formation, with larger groups tending to split into multiple communities. Moreover, the intense dynamics of the communities and the limitations imposed by Dunbar's circles suggest that other mechanisms are at work beyond gender homophily in the formation and evolution of these communities.

Our analysis also reveals the key role played by the negative relationships, despite their relatively small number, at the scale of the group. As shown in Figure 3.3, the network's community structure corresponding to group B in wave 1 differs significantly with and without the negative links. We can see that communities C1 and C2 would merge almost perfectly if
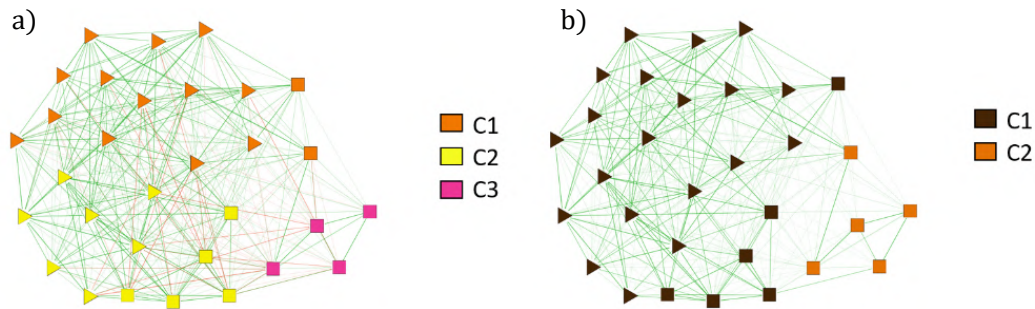
Figure 3.3: Analysis of Group B communities in wave 1 - The figure represents the analysis of Group B communities in wave 1 considering the following: a) both positive and negative links and b) only positive links. Nodes belonging to each community are marked by colour, as indicated in the figure. Friendship links are drawn in green, whereas enmity links are drawn in red. The shape of each node represents its gender: triangles correspond to females and squares to males. The distinction between communities C1 and C2 is clearly due to the presence of negative links.

the negative links did not exist. Given that the resulting community would contain 26 nodes compared to the 15 and 12 of the actual communities, one wonders if negative relationships are tied to the natural size of the sympathy group, making it more likely for them to appear when the community exceeds its typical size. It's worth noting that, contrary to what happens at the level of the whole network, by looking at relationships within single groups, the results cannot be understood in terms of "us vs. them", in-group vs. out-group or any other identity label, leading to a more relevant role for the negative links.

Thus, it is interesting to observe that the negative links, despite their small number, play a crucial role in determining the community structure at the scale of the group. This finding sheds light on the importance of negative relationships in social dynamics and highlights the limitations of relying solely on positive links in understanding social networks. Furthermore, our analysis suggests that negative links may have a more significant impact in smaller groups where the size of the sympathy network is limited. In such situations, the negative links may act as a mechanism to prevent the formation of a large, unwieldy community and help to maintain the typically observed size of the group's sympathy network.

3.3 Exploring social balance

A natural question about a network with positive and negative links is whether it satisfies social balance. The social balance theory was proposed by Heider (1946), who explored the connection between the structure of a network and the formation and stability of social relationships. According to social balance theory, people tend to associate with others who share similar attitudes, values and beliefs, forming cohesive groups and social networks. These patterns of association can lead to a phenomenon known as triadic closure, captured by the aphorism "my friend's friend is my friend", which means that two people who share a mutual friend are more likely to form a friendship with each other.

Moreover, social balance theory also suggests that individuals in a network strive for
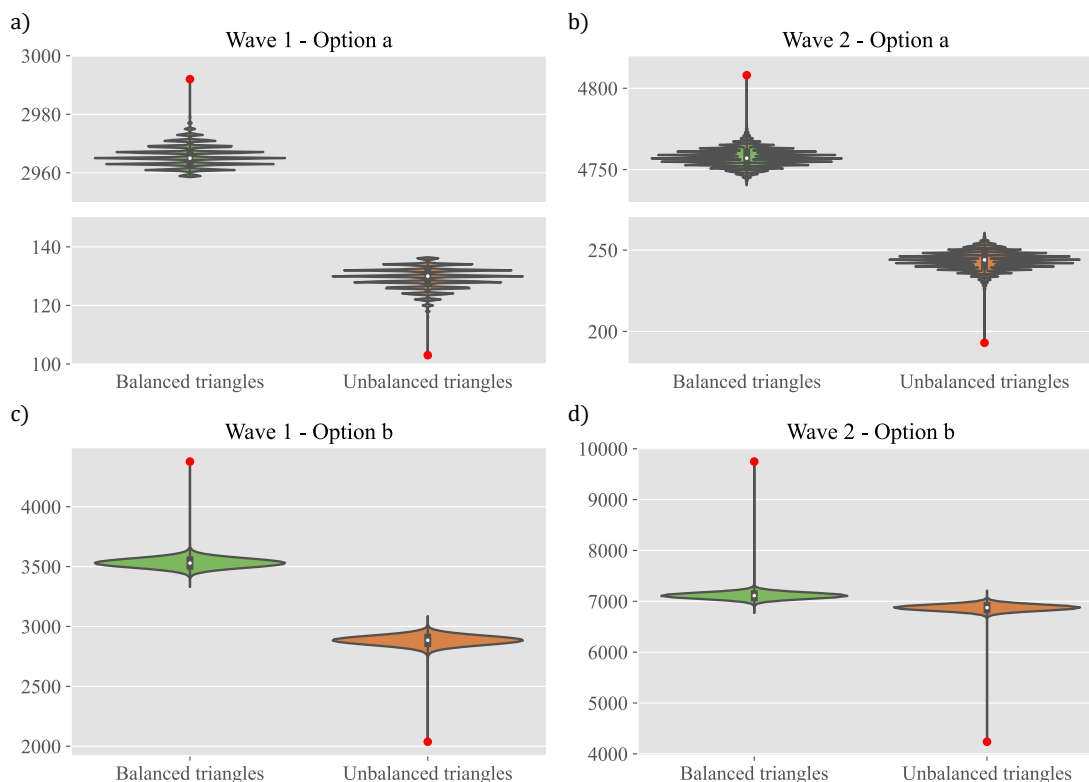
Figure 3.4: Comparison between the observed number of balanced and unbalanced triangles and our simulations - The figure represents the comparison between the observed number of balanced and unbalanced triangles and the results obtained in our simulations. Each panel shows a different wave/method: a) Wave 1 - option a, b) wave 2 - option a, c) wave 1 - option b and d) wave 2 - option b. In option a, a link is considered negative only if the interaction between a pair of nodes is negative in both directions. In contrast, in option b, a link is considered negative even when there is a single negative relationship between a pair of nodes. The red point in the plots corresponds with the observed number of triangles in the network and the violin distribution with the results of the simulations.

balance and avoid imbalanced relationships, where one person is friends with two people who dislike each other. In such cases, the network can become unstable and lead to the dissolution of relationships or the formation of new social groups. Consequently, triangles with an odd number of negative links create a cognitive dissonance sin some (or all) of its members that makes them unstable and prone to be resolved into a more balanced configuration by changing the sign or removing one link.

Cartwright and Harary (1956) generalised the theory mathematically. They proved that a perfectly balanced network could be decomposed into two positive subnetworks joined by purely negative links. This result is known as the Cartwright-Harary theorem and has important implications for understanding the structure of balanced social networks. Under this strict definition of balance, no real network with positive and negative links is ever balanced. For example, ours contain many unbalanced triangles. However, one can relax this condition by introducing a null model against which to test whether the number of

unbalanced triangles can be considered relevant.

The usual null model to test social balance in real networks is obtained by randomly permuting the links belonging to some triangle without changing their signs. However, this test has been criticised because positive and negative networks have very different properties (Feng et al., 2022). Negative networks are essentially random, whereas positive ones present very characteristic structures. Therefore, permuting both kinds of links can alter the nature of the network and overestimate imbalance. Feng et al. (2022) propose that a less biased null model can be obtained if links are classified according to their embedding (i.e., the number of triangles to which they belong), and only links with the same embedding are permuted.

The networks obtained in our study contain links of two intensities, which calls for an extension of the social balance theory. However, negative links are too few to maintain such a distinction and still have statistical significance. Accordingly, we have ignored the intensity of links to test social balance. Furthermore, our links are directional and not always reciprocated. With this in mind, we have adopted two criteria to project our network onto a bidirectional network: a) consider a negative link only if the interaction between a pair of nodes is negative in both directions, or b) assume that the presence of a single negative interaction between a pair of nodes is already evidence of a conflict and also consider those cases as negative links.

The number of triangles with 0 through 3 negative links is very different in both cases (see section B.1), but regardless of this, embedding-preserving permutations never produce configurations with fewer unbalanced triangles than the empirical network. As a matter of fact, Figure 3.4 shows that this number is actually more significant except for very few cases. We then conclude that, even though negative links create many unbalanced triangles, the configuration in which they appear is compatible with an extreme bias toward balance. The presence of negative links in social networks can indeed contribute to the formation of unbalanced triangles, but they are not the sole reason for the observed prevalence of social balance in real-world networks. Therefore, our results force us to conclude that there seems to be a strong tendency to reduce unbalance in middle school classrooms.

3.4   Discussion

Social networks consist of both positive (friendships) and negative (enmities) relationships. These two types of relationships have vastly different structures, with positive ones being more numerous and enduring over time. Nonetheless, negative relationships have a profound influence on the overall structure of the network. Consequently, understanding their role in social networks is crucial for gaining a insight on how these networks are structured and their evolution over time.

With this in mind, we have looked at the community structure of social networks both at the level of the five groups considered together and at the group level. The communities found from the analysis of the first wave are somewhat mixed, made up of both boys and girls, but they evolve so that in the second wave they appear to be largely segregated by gender. At the group level, the size of each community was stabilised around 12 people, which agrees with the size of the Dunbar's second circle, known as the sympathy group in

social psychology. Consequently, in classes with around 20 students of the same gender, their group split into two separate communities of about 10 each to stay below the size of the sympathy group. Interestingly, we have also observed that for this separation to arise in the community analysis it is necessary to consider the importance of negative relationships. Otherwise, the analysis yields a unique community formed by almost all students of the same gender. Another relevant finding is that, when the community structure is more stable, relationships with best friends intensify, and there is an evolution opposite to the general one of cliques relaxing to circles: more stable communities result in an intensification of the focus on best friends.

The existence of negative links also puts the question of balance into play. The social balance theory predicts that triangles with an odd number of negative links create a cognitive dissonance that renders them unstable and prone to disappear over time. Ideally, social networks should be free of unbalanced triangles. However, negative links are constantly being formed, so the question is not so much whether the network is free from those triangles as to whether they appear in a number that is significantly below what a null model would yield for that particular number of negative ties. We have run a recently proposed test whose null model takes into account the different nature of the positive and negative subnetworks, and have found that in one million realisations of the null model a network with fewer unbalanced triangles than the original network never shows up. This result forces us to conclude that there seems to be a strong tendency to reduce unbalance in school classrooms.

Aside from their relevance towards understanding the role of negative relationships, which is particularly important in such a crucial age in development as early adolescence, we believe our results have practical implications for the school social network and atmosphere. Thus, we have observed that the co-existence of two working languages in the same high school leads to the splitting of the social network of the school into two groups with very little communication or very few positive connections. To avoid this, it would be important for the school directors to design joint activities so that the students of both groups would get to know each other.

Another very relevant finding is the instability associated with gender imbalance in the groups. In such a situation, the gender homophily typical of this age leads to tensions between all students of the majority gender wanting to be in the same community and the fact that there are limits to the number of friendships of a given intensity. In this case, we have seen that negative relationships appear that split the majority gender community in two, which could lead to the increased polarisation of the group atmosphere. Our results suggest that this is a problem that can be easily avoided by making gender balance a priority or else by reducing the group size. Another related measure would be to consider this when forming groups in subsequent academic years by separating students from such gender-segregated communities into different classes to avoid perpetuating intra-group divisions.

All these insights point to the fact that, beyond its appealing interest from the scientific viewpoint, this kind of social network analysis is an efficient and easy-to-implement tool that can be used to foster a friendlier school environment which could also have connections with the performance of the students. With both objectives in mind, we have developed the application that we have used in our study and is presented in the next chapter.

4

A TOOL TO MONITOR THE SOCIAL CLIMATE OF A SCHOOL

Scientific research can provide valuable insights and solutions to a wide range of issues that affect our daily lives. It is true that its main objective is to advance the understanding of the natural world and its fundamental principles, but it can nonetheless have significant practical applications and its contribution to solving real-world problems cannot be underestimated. One of the most significant problems currently faced within schools is bullying. To eradicate it, it is essential to create an inclusive atmosphere that promotes the adaptation of students.

For this reason, in this chapter, we move away from the scientific perspective and show an application that we have developed during this thesis to improve the social environment of educational institutions. Initially, the purpose of the application was to analyse data more quickly for research purposes, but we realised that it could have a significant impact on the information managed by those responsible for improving coexistence in schools. In this way, instead of providing static reports, they can have access to an interactive application that allows them to filter the information they need or the students they want to analyse based on the problem, resulting in a more well-founded response to the situations they face daily.

We provide a trial username and password to test the application and explore its various functionalities. Additionally, we briefly review some of these features and their usefulness in addressing the problem of bullying. Finally, we present a summary of the feedback received and its contribution to promoting a positive learning environment.

4. A TOOL TO MONITOR THE SOCIAL CLIMATE OF A SCHOOL

4.1   The application

The application consists of an interactive dashboard developed using the Shiny package for R software. It was initially designed for research purposes to enable faster data analysis on questions that might arise during the development of the project. However, once developed, we realised that its functionalities could extend far beyond this and be useful to the school principals for detecting or analysing conflictive situations among the students. Moreover, the application is hosted on a server, which is owned by our research group, and is accessible online at the URL https://harpo.uc3m.es/ies_blas_otero_uc3m/. In order to provide an interactive view of the dashboard, we have created a demo user with the following credentials:

- <u>User</u>: uc3m_thesis_user

- <u>Password</u>: RjAzbgP8psANmBP9

Once you access the application, a menu appears on the left side, allowing you to select the panel to be displayed in the central part of the screen. This menu consists of a total of 10 interactive options, which are the following:

- Summary table.

- Networks.

- Ego-networks.

- Personal networks.

- Friendship vs. enmity dispersion.

- In-out friendship dispersion.

- Reciprocity dispersion.

- Relationship count boxplot.

- μ parameter boxplot.

- Relationship count histogram.

By default, the panel that appears selected is the summary table. This table compiles the personal information associated with each student, such as their group, gender or itinerary. It also includes quantitative information about their personal relationships, both friendships and enmities. We can observe the number of "best friends", "just friends", "just enemies" and "worst enemies" that each user has selected and how many have selected them. All of this is visible at a single glance on the screen. As it is an interactive table, we can quickly search for any user we want or sort the users based on the number of relationships simply by clicking the associated column. The other panels offer different ways to visualise and analyse the relationships among students. Next, we will provide a brief analysis of some of the most prominent and useful panels, based on the feedback received. To view and understand all the remaining ones, we recommend accessing the website with the provided user credentials and exploring them all.
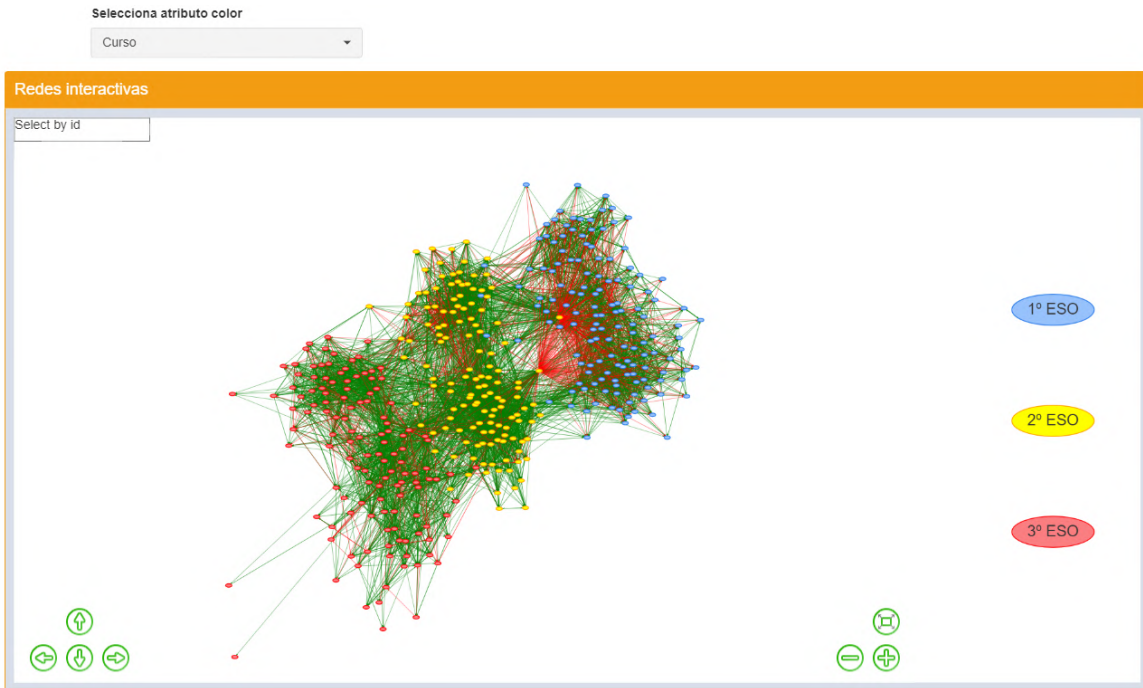
Figure 4.1: Panel for the interactive network - The figure represents a screenshot of the app that captures the structure of the complete high school in wave 1. The green arrows represent friendships and the red ones enmities. The width of each arrow indicates the intensity of the associated relationship. The filter at the top of the panel allows to change the colour of the nodes based on the selected attribute. The attribute "course" is selected in the figure's screenshot. The distribution of the nodes in the network is determined using an algorithm that positions those with an existing relationship closer to each other than those without one. Moreover, it also considers the intensity of the relationships. In this way, it captures information related to the community structure of the school. The figure clearly shows clusters corresponding to the different classrooms.

We begin by introducing the interactive networks panel. In Figure 4.1, we present a screenshot of a panel that shows the structure of the complete high school in wave 1. Green arrows represent friendships and red arrows represent enmities. Additionally, the width of the arrows characterises the intensity of the associated relationship. All the relationships related to the survey responses appear in the application without performing any prior statistical treatment to filter outliers. The filter appearing at the top of the panel allows changing the colour of the nodes based on the selected attribute. One can zoom in on the image and click particular nodes to highlight its features. Nodes can also be dragged and put anywhere on the screen.

The key point to understanding the structure is the distribution of the nodes in the network. This distribution is achieved through an algorithm that positions individuals who maintain a relationship closer together and those who do not further apart, taking into account the weights of the links. Therefore, it captures information related to the community structure. In this way, this panel can help identify isolated students by finding those nodes
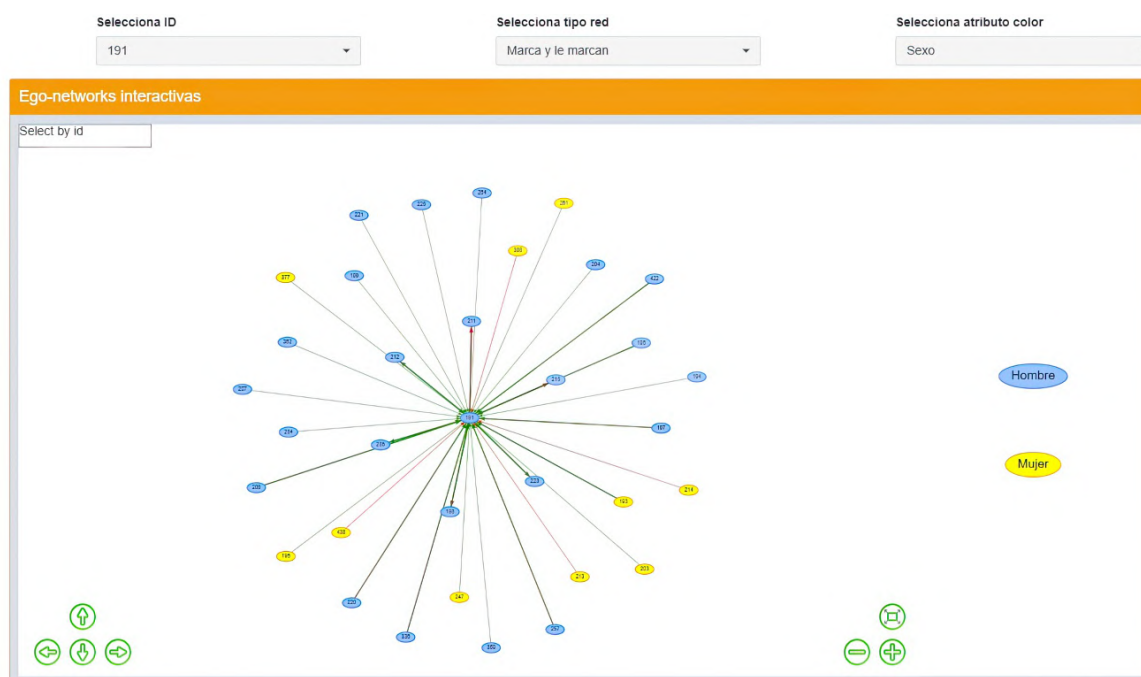
Figure 4.2: Panel for the interactive ego-network - The figure represents a screenshot of the app that captures the structure of the ego-network corresponding to ID 191 in wave 1. This network captures exclusively the relationships this individual maintains with each of the remaining students. The green arrows represent friendships and the red ones enmities. The width of each arrow indicates the intensity of the associated relationship. The left filter allows changing the selected ID, the central one lets us choose between the relationships displayed in the image (all, only those marked by the individual or only those in which the individual has been marked), and the right filter colours the nodes according to the selected attribute. In the screenshot of the figure, the attribute "sex" is selected.

which appear distant in the network. This would indicate a low density of relationships and serve as an alert for the school principals. Furthermore, individuals who have many enmity relationships may also require monitoring to investigate the reasons behind it.

In the second panel (interactive ego-networks), an individual can be selected to visualise all the relationships in which he or she is involved (in and out arrows). Furthermore, the colour of each node in the figure represents a characteristic attribute. All these variables can be controlled using the filters appearing at the top of the panel. The filter on the left allows selecting the ID of the focal individual, the central one determines which relationships to consider (all, only those marked by the individual or only those in which the individual has been marked) and the right filter sets the attribute that determines the colour of the nodes.

In Figure 4.2, we present an example of these ego-networks, the one of ID 191 in wave 1. This kind of visualisation is particularly useful for spotting potential problems. When there is a suspicion of bullying or related issues concerning a student, the school principals can access this panel of the application and quickly identify the individuals with whom the
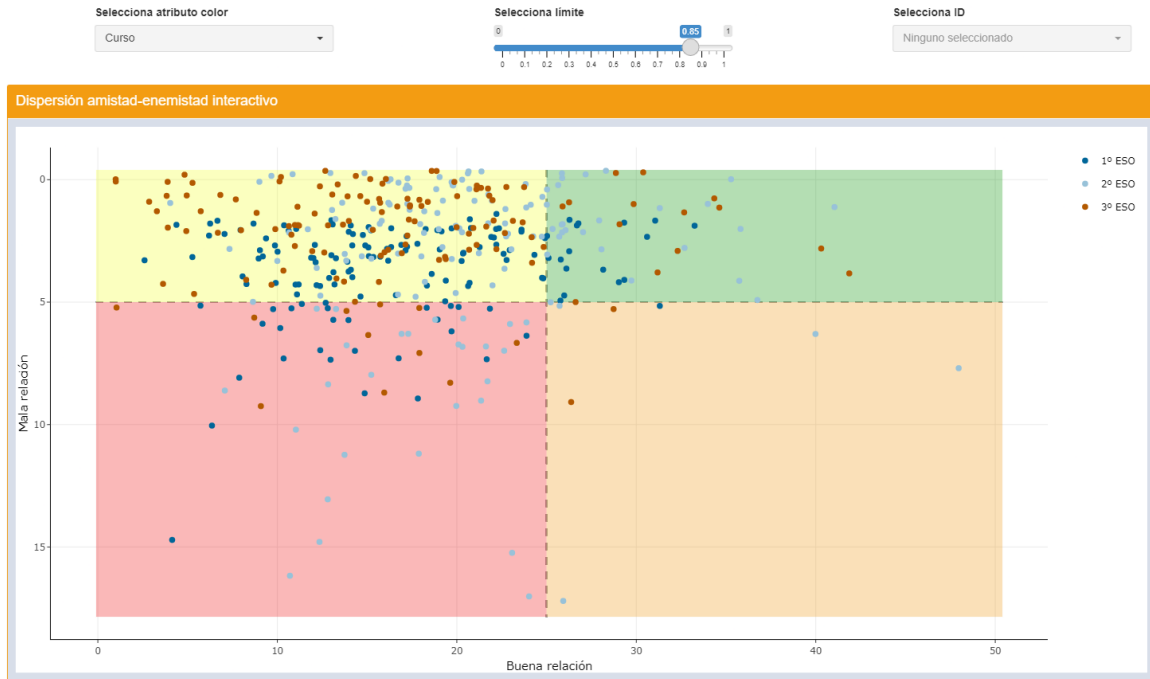
Figure 4.3: Panel for the interactive friendship vs. enmity dispersion - The figure represents a screenshot of the app that captures the friendship vs. enmity dispersion in the complete high school in wave 1. The x-axis represents the number of times the individual has been marked as a friend (the origin is on the left) and the y-axis represents the number of times they have been marked as an enemy (the origin is at the top). The background colours indicate each of the zones: green represents being marked as a friend many times and as an enemy only a few times, yellow represents being marked as a friend and as an enemy only a few times each, orange represents being marked as a friend many times and as an enemy many times and red represents being marked as a friend only a few times and as an enemy many times. This chart can provide information about a student's situation at the high school. The left filter allows for changing the colour of the points based on the selected attribute. The attribute "course" is selected in the figure's screenshot. The central one allows the user to move the position of the background colour panels based on the percentile of the total number of relationships marked for the division. Finally, the right filter allows the user to select and highlight a specific ID within the figure.

student has friendship or enmity relationships. With this information, interventions can be much more effective in improving coexistence and fostering a positive school environment.

The last panel we discuss is the interactive friendship vs. enmity dispersion. This panel consists of a scatter plot in which the x-axis represents the number of times the individual has been selected as a friend by other students and the (inverted) y-axis represents the same variable for enmities. In Figure 4.3, we show the results obtained for the complete high school in wave 1. We divide the students into four groups based on the number of friendships and enmities they have been marked: many friends and few enemies (green zone), many friends and many enemies (orange zone), few friends and few enemies (yellow zone) and few friends and many enemies (red zone). The percentile determines the limits of these zones. The

default value is 0.85, but it can be adjusted by ganging the central filter. Additionally, the filter on the left allows changing the colour of the points according to the selected attribute and the filter on the right allows highlighting a specific student.

This panel is particularly useful for detecting students at potential risk of some sort. In general, those individuals who are in the red zone of the figure need to be taken care of. Understanding the reasons behind this being in the red zone allows for targeted interventions to address their specific needs and challenges. Moreover, those who are in the orange zone should also be carefully examined, as it is unusual to be marked both as friend and enemy simultaneously by many other students. Usually, these students tend to be popular but controversial, so they generate strong positive and negative reactions. In these cases, it is also important to detect such situations and respond to them with tailored interventions. Although these interventions may differ from those used for students in the red zone, the final goal is the same. Finally, it may also be interesting to identify individuals in the green zone, those who have a favourable social standing within the high school. These students tend to have many friends and few enemies, reflecting their positive relationships and effective social skills. They can serve as role models or mentors for other students to create a positive school culture or to learn from their experiences and strategies. Therefore, regularly updating and analysing the data from this panel can help track the progress and effectiveness of the interventions, allowing for adjustments and improvements as needed. Ultimately, the goal is to move students out of the dangerous zones and foster a positive and safe environment.

Once we have discussed the most relevant panels and their potential applications, we now move on to introducing the most interesting part of the application: the filters. A filter panel unfolds by clicking on the icon in the top right corner, allowing the user to analyse a personalised situation. The filter is divided into three sections: relationship type, source nodes info and target nodes info. In the relationship type section, the user can select the scholar year and the wave, filter by friendships or enmities or determine the intensity of the relationships. In the other two sections, we can set the characteristic attributes of the source and target nodes for which the relationships are considered. We can filter by all available attributes, such as course, group, sex or itinerary. These filters provide users with a powerful tool to examine specific aspects of the social network, allowing for more in-depth analysis and understanding of the structure and dynamics within the high school. We encourage the reader to access the application with the provided credentials, test the filters and experience the full potential it has to offer.

4.2   Feedback received

One of the aspects that fills us with pride in our research is the widespread acceptance that our application is receiving among the school principals, the guidance team and the teaching staff of the high school we are collaborating with. Although this was not the main goal when we started our study, this acceptance encourages us to continue developing the project and identify areas where we can further improve and refine it. In this section, we present the feedback received from the guidance team responsible, Silvia Ibáñez Morcillo. Its content consists of two emails, whose original text (in Spanish) can be found in subsection C.1.1. These emails allow us to understand first-hand the potential of our application, and hearing

directly from those who are using the application in their daily work provides valuable insights that will help us shape the future direction of the project.

The first email is dated October 12th, 2021, just a few days after we granted them access to the application for the first time. This email highlights the ability of the application to offer insights into student interactions within and across different groups. The guidance team responsible found the tool particularly helpful as she does not teach classes and relies on such information to support the students better. Moreover, she exposes that the application has been instrumental in monitoring students requiring special attention. Thanks to the application, she has been able to identify potential motivators for absentee students and confirm the social acceptance of others she had doubts about. Therefore, our tool has enabled her to recognise students with severe emotional issues who have more support than initially thought, facilitating the process of finding accompanying peers for them.

The second email is more recent, dated January 29th, 2023. It shows how Silvia has become more familiar with the tool after over a year of use. For this reason, she focuses more on analysing individual cases and emphasises the value of the application in identifying specific students of concern and tailoring intervention efforts accordingly. She even mentions that she spent more than two hours analysing two cases that concerned her, creating a table for each one based on the insights provided by our application. The insights gleaned allowed her to confirm the level of concern for each student and to identify key relationships that could be reinforced. The engagement and dedication to using the application demonstrate the positive impact it has had on the school guidance team to address the unique needs of each student.

Both emails underscore the usefulness of the application in subtly strengthening student relationships without explicitly revealing survey data. The teachers can use the insights to make informed decisions about group projects or seating arrangements in classes. This fosters a more inclusive and supportive learning environment, as it encourages positive connections among students who may not have realised their shared appreciation. Furthermore, the feedback received from the two emails emphasises the considerable utility of the application in an educational context. Its ability to provide valuable insights into student relationships and dynamics has proven to be an essential resource for the school guidance team. Using the application they can try to prevent bullying, more effectively support their students and create an environment conducive to personal growth and success.

4.3 Discussion

In this chapter, we have moved away from scientific research and into the transference of our research to society by introducing an online application that we have developed during the course of this thesis. The initial purpose of this application was to streamline data analysis within our research, but once it was developed, we realised its potential as an educational tool. We noticed that it could be of great value for the school staff to anticipate problems within the school, respond to them and improve the social atmosphere. For this reason, we stopped generating static reports based on the collected data, created specific user accounts and granted them full access. Furthermore, to allow readers to become more familiar with the

functionalities of the application, we have created a demo user account that has full access to all the details. This demo account provides an opportunity to explore our software, observe how it can be effectively utilised in an educational setting and gain insight into the potential benefits it offers.

Regarding the application, it consists of a total of 10 panels, each of which provides different information that can be useful for understanding the structure and dynamics of social relationships in high school students. Throughout the chapter, we have discussed in detail the three panels that have received the most positive feedback in terms of use found by the guidance team (interactive networks, interactive ego-networks and interactive friendship vs. enmity dispersion). These three panels have proven to be valuable tools for identifying potential issues and designing targeted interventions. The interactive networks panel allows users to visualise the overall structure of social connections among students, revealing patterns and potential areas of concern. The interactive ego-networks panel offers a closer look at individual students' social circles, providing insights into their relationships and helping to identify those who may need additional support. Lastly, the interactive friendship vs. enmity dispersion panel enables users to distinguish between positive and negative relationships, highlighting potential sources of conflict or tension within the student population. In addition to these three key panels, the application offers a range of other features designed to support a comprehensive understanding of the social landscape within the school. Again, we encourage the reader to use the provided demo credentials and explore the diverse array of tools and visualisations available.

Finally, we have presented the feedback received from the guidance team responsible, Silvia Ibáñez Morcillo. Silvia's feedback has been invaluable in demonstrating the practical benefits and real-world impact of our application on the educational environment. In this way, her positive responses and engagement with the tool show that it has been effective in addressing the unique needs of the students and providing targeted support. Moreover, her insights have also allowed us to identify areas for further improvement and refinement of the software. As a result, we have focused on adapting the application to serve their needs better, incorporating their feedback to make it more user-friendly and effective. This collaborative effort has allowed us to create a tool that not only assists in our research but also empowers informed decisions and proactive measures to address social issues within the school community.

PART II

MODEL-DRIVEN PERSPECTIVE

5

MODELLING SOCIAL STRUCTURE AND COLLECTIVE BEHAVIOUR

In the first part of this thesis, we analysed the structure and dynamics of social relationships from a data-driven perspective. This approach has enabled us to discover that, despite frequent changes, it seems that everyone possesses a predetermined structure for their social relationships. Moreover, this structure exhibits similarities with the behaviour of an atom and its electrons and remains relatively stable over time. Therefore, this evidence suggests the possibility of examining the formation of social networks as a statistical-mechanical system. In consequence, this chapter aims to formulate a mathematical model for understanding the collective behaviour of individuals within a community. We strive to develop a theory that allows us to comprehend the relationships formed between the members of a system as they interact, similar to those we use to describe physical systems of particles.

Our starting point is the concept of "social atom". From this, we aim to establish the foundations of the model and seek to understand its interactions to explain the structure of the whole system. Exponential random graphs provide an appropriate tool to address this issue. All we need is to obtain a Hamiltonian and compute the grand partition function associated with its probability distribution. The explicit calculation of this probability distribution over the states of the system allows us to identify the maximum likelihood values for the parameters with those obtained from a specific dataset. These parameters enable us to understand the equilibrium behaviour, calculate the expected values of macroscopic variables and investigate the response to external perturbations.

5.1   Pairwise approximation

With this in mind, we construct a hypothetical society consisting of N individuals, each possessing r layers (or levels of emotional intensity) for allocating their social relationships. The relationships associated with each layer entail different cognitive costs and equilibrium within the system is achieved when all constituents have their layers filled. Once this happens, all changes in the system are due to "thermal" fluctuations, just as the ensemble of atoms in a gas lose and recover their electrons as a consequence of thermal agitation.

In our hypothetical society, each member can have a relationship with their neighbours at each level i, where $i \in \{0, 1, 2, ..., r\}$. If $i = 0$, there is no relationship, and the other person does not occupy any space in the layers. All the information regarding the relationships and their intensities in the system can be represented using a matrix σ. This matrix is of size $N \times N$, asymmetric, and its entries satisfy $\sigma_{ab} = i$ if b is in layer i of a. The diagonal elements of the matrix are zero because self-relationships are meaningless in this model.

We begin by assuming pairwise interactions between individuals, as this approximation offers a straightforward approach to modelling the system. Specifically, we consider each pair of connections independently and assume that the overall behaviour of the entire network can be fully explained by combining all these pairs. The composition of each pair is determined solely by the two levels (i, j), regardless of the order. In other words, $(i, j) = (j, i)$ applies.

The essential ingredients of the model are the parameters associated to each pair. These parameters depend only on the levels that define the pair and not on the specific individuals forming it. In order to introduce them, hence to define the Hamiltonian, we are going to make a mapping of our system to a classical system of statistical mechanics. We will assume that individuals are the nodes of a complete graph and that each link can be either empty (no link between the individuals occupying the end nodes) or filled with one "particle" (and only one) out of a set of different "species". Species of particles are labelled according to the two possible relationships existing between the two connected individuals. Thus, the link connecting nodes a and b will be of type (i, j) if $\sigma_{ab} = i$ and $\sigma_{ba} = j$. This maps our model to Potts's model. In the pairwise approximation, we are assuming that different links do not interact (otherwise, interactions would involve more than two nodes), so the model is ideal. The only parameters of an ideal Potts's model are the chemical potentials of each type of particle E_{ij}.[1] Despite this interpretation, in this context, we will refer to these parameters as "efficacies", and we need to keep in mind that the higher the efficacy the higher the probability that the corresponding link is formed.

5.2   Formulation of the pairwise model

As we have just mentioned, the simplest model for studying social relationships is the pairwise model. In it, each potential pair of individuals has a corresponding efficacy E_{ij}, which solely depends on the levels (i, j) that define the pair. These parameters govern the stability of

---

[1]The notation E_{ij} is reminiscent of the initial formulation of the model, where these parameters were interpreted as "energies" and had a sign convention opposite to the one we are using in this thesis. We adopt the interpretation as chemical potentials for consistency with further chapters but keep the original notation.

Figure 5.1: The pairwise approximation - The figure represents an schema of the pairwise approximation. In this approximation, the levels (i, j) completely define the pair. Additionally, the symmetry property is satisfied so that the direction of the links is irrelevant.

social relationships, as the lower their value, the harder it becomes to find a pair of this nature in the system.

The symmetry we are imposing on the pairs of the system forces us to assume $E_{ij} = E_{ji}$. The meaning of this symmetry is the identity of the individuals does not matter, only how they relate to each other. Furthermore, when there is no link in either direction of the pair $E_{00} = 0$. This fixes an arbitrary reference level for the efficacies. Consequently, the total number of parameters required to characterise the model fully is

$$N_{\text{par}} = \frac{(r+1)(r+2)}{2} - 1 = \frac{r(r+3)}{2}. \tag{5.1}$$

The Hamiltonian $H(G)$, where G denotes the configuration of links, is then defined as

$$-H(G) = \sum_{a=1}^{N} \sum_{b>a}^{N} E_{\sigma_{ab} \sigma_{ba}} = \sum_{i=0}^{r} \sum_{j=0}^{i} E_{ij} R_{ij}, \tag{5.2}$$

where R_{ij} denotes the total number of pairs in the system with associated levels (i, j). Both forms of the model are equivalent and can be used interchangeably. In the first expression, there is no need to incorporate the sums across the r layers of each individual, as this information is in the σ matrix. Conversely, the second expression implicitly includes sums over the N members of the system when calculating the R_{ij} values.

With this Hamiltonian, the probability distribution over G is given by

$$P(G) = \frac{1}{\Xi} e^{-H(G)}, \tag{5.3}$$

where Ξ, the grand partition function of the system or the normalization constant of the probability distribution, is calculated by summing over all possible configurations as follows:

$$\Xi = \sum_{G} e^{-H(G)} = \sum_{\{\sigma_{ab}\}} e^{-H(G)}. \tag{5.4}$$

Considering the Hamiltonian defined by Equation 5.2, the grand partition function is obtained as

$$\Xi = \prod_{a=1}^{N} \prod_{b>a}^{N} \left[\sum_{\sigma_{ab}=0}^{r} \sum_{\sigma_{ba}=0}^{r} e^{E_{\sigma_{ab} \sigma_{ba}}} \right] = \prod_{a=1}^{N} \prod_{b>a}^{N} \Xi_1, \tag{5.5}$$

where Ξ_1 is the grand partition function for a single pair,

$$\Xi_1 = \sum_{i=0}^{r} \sum_{j=0}^{r} e^{E_{ij}}. \tag{5.6}$$

The result shown in Equation 5.5 holds considerable importance because it reveals that we can treat each potential pair in the network as identical and independent systems. As a consequence, the grand partition function Ξ can be obtained as

$$\Xi = \Xi_1^{\binom{N}{2}}, \tag{5.7}$$

with $\binom{N}{2}$ corresponding to the total number of pairs in the system.

Considering the symmetry $E_{ij} = E_{ji}$, the overall sum of terms in Equation 5.6 results in

$$\Xi_1 = \sum_{i=0}^{r} \sum_{j=0}^{i} \left(2 - \delta_{i,j}\right) e^{E_{ij}}, \tag{5.8}$$

where $\delta_{i,j}$ is the Kronecker delta ($= 1$ if $i = j$ and $= 0$ otherwise).

Finally, the grand potential of the system Ω is obtained as

$$\Omega = -\log \Xi = -\binom{N}{2} \log \Xi_1. \tag{5.9}$$

Calculating the mean values $\langle R_{ij} \rangle$ is straightforward. This expression can be obtained as partial derivatives of the grand potential Ω with respect to the different efficacies. Thus,

$$\langle R_{ij} \rangle = -\frac{\partial \Omega}{\partial E_{ij}} = \left(2 - \delta_{i,j}\right) \binom{N}{2} \frac{e^{E_{ij}}}{\Xi_1}. \tag{5.10}$$

In particular, considering that $E_{00} = 0$,

$$\langle R_{00} \rangle = \binom{N}{2} \frac{1}{\Xi_1}, \tag{5.11}$$

so we can rewrite Equation 5.10 as

$$\langle R_{ij} \rangle = \left(2 - \delta_{i,j}\right) \langle R_{00} \rangle \, e^{E_{ij}}. \tag{5.12}$$

Up to this point, we have derived the formulas for the average values of each pair of relationships, depending on the model parameters. Given a specific dataset, the empirical values for the efficacies E_{ij} can be calculated using Equation 5.12, so

$$E_{ij} = \log \left(\frac{\langle R_{ij} \rangle}{\left(2 - \delta_{i,j}\right) \langle R_{00} \rangle} \right). \tag{5.13}$$

Estimating the errors associated with these parameters is feasible as well. This can be accomplished using a Gaussian approximation by constructing the likelihood $\ell(G)$, which is defined as

$$\ell(G) = -\log P(G) = -H(G) + \Omega(G), \tag{5.14}$$

and expanding up to the second order around the values determined by Equation 5.13. We compute the second derivatives of $\ell(G)$ with respect to the parameters explicitly and evaluate them at the efficacy values to construct the Hessian matrix $\mathcal{H}(G)$:

$$\mathcal{H}(G) = \left(\frac{\partial^2 \ell(G)}{\partial E_{ij}\, \partial E_{kl}} \bigg|_{E=E_m} \right). \tag{5.15}$$

A detailed calculation of these derivatives can be found in subsection D.1.1. Moreover, the variance-covariance matrix $V(G)$ associated with these parameters is derived as the inverse of the Hessian matrix,

$$V(G) = \left[-\mathcal{H}(G) \right]^{-1}. \tag{5.16}$$

The diagonal elements of this matrix represent the variances associated with each parameter of the model and indicate their uncertainty level.

5.2.1 Results

Once we have formulated the pairwise model, we need to validate it using data obtained from real-world networks. To achieve this, we will use the data collected in our longitudinal study on the structure and dynamics of social relationships among high school students, presented in detail chapter 2. Specifically, we will concentrate on conducting an independent analysis of the structure for each class of the school, obtaining the parameters of the model for each group independently and comparing the results. Furthermore, when we analysed this data in chapter 2, we also observed that the structure of friendship and enmity relationships is very different, with friendships being much more abundant and stable over time. Although in chapter 3 we demonstrated that enmities play a significant role in explaining the structure of these networks, we will focus on explaining the friendship networks to keep the model (and future calculations) as simple as possible. Therefore, we have relationships with three different levels of intensity: $+2$ ("best friends"), $+1$ ("just friends") and 0 ("no relationship" or "enmity").

In such a case, we need to particularise the pairwise model for $r = 2$ layers, so the number of parameters that define this model is 5. These parameters are the efficacies associated with all possible pairs of levels (i, j), with $i, j = \{0, +1, +2\}$, that can exist in the system. The Hamiltonian defined by Equation 5.2 is reduced to

$$-H(G) = R_{10}E_{10} + R_{11}E_{11} + R_{20}E_{20} + R_{21}E_{21} + R_{22}E_{22}, \tag{5.17}$$

where the different E_{ij} are the parameters of the model and the R_{ij} are the total number of pairs of links of each type.

The grand partition function can be factorized according to Equation 5.7, where Ξ_1 is

$$\Xi_1 = 1 + 2\, e^{E_{10}} + e^{E_{11}} + 2\, e^{E_{20}} + 2\, e^{E_{21}} + e^{E_{22}}. \tag{5.18}$$
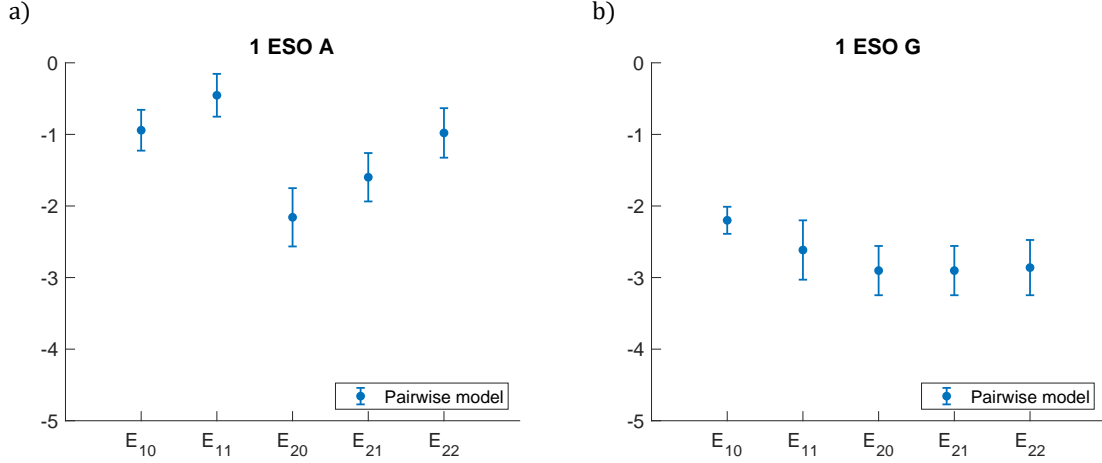
Figure 5.2: Parameters for the pairwise model - The figure represents the values and 95% confidence intervals for the maximum likelihood parameters in the pairwise model for the following groups in wave 1: a) 1 ESO A and b) 1 ESO G. These two groups have been selected because they exhibit different densities of relationships: 1 ESO A has a higher density, while 1 ESO G has a lower density. This is evidenced by the values of efficacies E_{ij}, which are larger in the former case. This indicates that these relationships are more likely to appear. The figures representing the values of the parameters for each of the remaining groups in wave 1 can be found in subsection D.3.1.

From these results, obtaining the model parameters for our dataset is immediate. The efficacy values are given by Equation 5.13. Their values and 95% confidence intervals are shown in Figure 5.2 for two groups in wave 1, 1 ESO A and 1 ESO G. The results for all the remaining groups in this wave are presented in subsection D.3.1.

The reason why only these two groups have been selected to be included in the main text is they exhibit different densities of relationships, i.e. the proportion of the number of observed relationships to the total possible number. 1 ESO A exhibits a higher density of relationships, resulting in higher values for the associated parameters, as these values are directly proportional to the probability of their respective relationships appearing in the system.

Moreover, the structure of these two groups also varies. In 1 ESO A, the reciprocal relationships are the most prevalent, whereas, in 1 ESO G, those with a weak intensity link in a single direction are more common. The likelihood of observing the remaining relationships in this latter group is similar because the values of their associated efficacies are practically equal. If we compare these results with the figures for the remaining groups in subsection D.3.1, the values of the parameters are more similar to those associated with 1 ESO A. Based on this, we can conclude that reciprocal and equal relationships are the most probable, and pairs of links in which a strong emotional relationship is not reciprocal are the least common. Finally, it should be noted that the structure observed in Figure 5.2(a) is the most common among the groups studied, while the organisation in Figure 5.2(b) is only observed in three of them, which have a low density of relationships. Therefore, we consider the first as the "typical" structure, while the second may be interpreted as "anomalies".

5.3   Linear model with reciprocity

The pairwise model presented in section 5.2 provides a simple approach to investigating the structure of human social relationships. The parameters of the model E_{ij} characterise the behaviour of each pair of links and, consequently, the overall behaviour of the system. However, this model does not offer any insight into the factors that these parameters rely on, something that is essential for identifying the elements influencing the structure of social relationships.

The foundations of our study are the evidence derived from the Social Brain Hypothesis and the concept of "social atom", so it is natural to inquire whether the factors explaining the individual behaviour also account for the macroscopic structure of the system. Consequently, we start by considering two factors to explain the form of the parameters E_{ij}. First, the number of personal relationships an individual can maintain simultaneously is limited, and second, the cost of maintaining each relationship increases as its emotional intensity grows.

Denoting l_i as the total number of relationships a person has within a layer i and s_i as the cost of each relationship, we can mathematically write the conditions that the entire system satisfies on average:

$$\langle L \rangle = N \sum_{i=1}^{r} \langle l_i \rangle, \qquad \langle S \rangle = N \sum_{i=1}^{r} s_i \langle l_i \rangle, \tag{5.19}$$

where N is the number of nodes in the system, L is the total number of relationships, and S is their associated cost.

In my Master's thesis (Escribano, 2019), we attempted to fit the parameters E_{ij} using only these two factors but ultimately concluded that it did not work. However, looking at figures like Figure 5.2(a), which is typical, we realised that reciprocity plays a role because the more reciprocal a relationship, the higher its efficacy. Thus, we concluded that it was also necessary to include a term associated with the reciprocity of the relationships. The simplest way to consider both the conditions imposed by Equation 5.19 and reciprocity is to express the efficacies as

$$E_{ij} = \lambda \left(2 - \delta_{i,0} - \delta_{j,0} \right) + \mu \left(i + j \right) + \beta \left(1 - \delta_{i,0} \right) \left(1 - \delta_{j,0} \right), \tag{5.20}$$

where λ and μ are the parameters associated with the total number of relationships in the system and the total cost of these relationships, respectively, and a positive β enhances reciprocal links. Notice that this term ignores the intensities of the reciprocal pair because this is already accounted for by the other two terms.

The Hamiltonian of the system, defined by Equation 5.2, can be rewritten as

$$-H(G) = \lambda L + \mu S + \beta R, \tag{5.21}$$

where L is the total number of links, S is the cost of all links, and R is the number of pairs of reciprocal relationships in the system.

It is important to note that this model can be applied to an arbitrary number r of layers. However, since the data we are working with is the one presented in chapter 2 and corresponds

to a case with $r = 2$ layers, we will focus on the results for this specific scenario. Therefore, the grand partition function for a single pair is

$$\Xi_1 = 1 + 2\,e^{(\lambda+\mu)} + e^{(2\lambda+2\mu+\beta)} + 2\,e^{(\lambda+2\mu)} + 2\,e^{(2\lambda+3\mu+\beta)} + e^{(2\lambda+4\mu+\beta)}. \tag{5.22}$$

From this, we can derive the grand potential Ω and the averages of the different variables,

$$\langle L \rangle = -\frac{\partial\Omega}{\partial\lambda} = 2\,\binom{N}{2}\,\frac{\left(e^{-\mu}+1\right)\left(e^{-(\lambda+2\mu+\beta)}+e^{-\mu}+1\right)}{(1+e^{-\mu})^2 + e^{-(\lambda+2\mu+\beta)}\left(2+2e^{-\mu}+e^{-(\lambda+2\mu)}\right)}, \tag{5.23}$$

$$\langle S \rangle = -\frac{\partial\Omega}{\partial\mu} = 2\,\binom{N}{2}\,\frac{\left(e^{-\mu}+2\right)\left(e^{-(\lambda+2\mu+\beta)}+e^{-\mu}+1\right)}{(1+e^{-\mu})^2 + e^{-(\lambda+2\mu+\beta)}\left(2+2e^{-\mu}+e^{-(\lambda+2\mu)}\right)}, \tag{5.24}$$

$$\langle R \rangle = -\frac{\partial\Omega}{\partial\beta} = \binom{N}{2}\,\frac{\left(e^{-\mu}+1\right)^2}{(1+e^{-\mu})^2 + e^{-(\lambda+2\mu+\beta)}\left(2+2e^{-\mu}+e^{-(\lambda+2\mu)}\right)}. \tag{5.25}$$

By dividing Equation 5.23 by Equation 5.24, we can express the parameter μ as

$$\mu = \log\left(\frac{\langle S \rangle - \langle L \rangle}{2\langle L \rangle - \langle S \rangle}\right). \tag{5.26}$$

Furthermore, dividing Equation 5.23 by Equation 5.25 yields

$$\lambda + \beta = \log\left(\frac{2\langle R \rangle\,(2\langle L \rangle - \langle S \rangle)^2}{\langle L \rangle\,(\langle S \rangle - \langle L \rangle)\,(\langle L \rangle - 2\langle R \rangle)}\right). \tag{5.27}$$

By substituting the last two expressions into Equation 5.25, we obtain the expression for β as

$$\beta = \log\left(\frac{2\langle R \rangle\,[N(N-1) - 2\langle L \rangle + 2\langle R \rangle]}{(\langle L \rangle - 2\langle R \rangle)^2}\right). \tag{5.28}$$

Finally, by using this expression in Equation 5.27, we can determine the parameter λ:

$$\lambda = \log\left(\frac{(2\langle L \rangle - \langle S \rangle)^2\,(\langle L \rangle - 2\langle R \rangle)}{\langle L \rangle\,(\langle S \rangle - \langle L \rangle)\,[N(N-1) - 2\langle L \rangle + 2\langle R \rangle]}\right). \tag{5.29}$$

Therefore, the values of these three parameters fully characterise the model. Additionally, as we did with the pairwise model, we can analytically calculate confidence intervals for the values of these parameters. All of these calculations are detailed in subsection D.1.1.

5.3.1 Results

To verify the performance of the linear model with reciprocity, we must compare its results with those obtained using the pairwise model. In Figure 5.3, we observe that the parameters obtained using these two models closely approximate each other for both groups studied: 1 ESO A and 1 ESO G, even despite the difference in the density of relationships between groups. The same happens when we study all the remaining classes in wave 1. The figures associated with each of them can be found in subsection D.3.2.
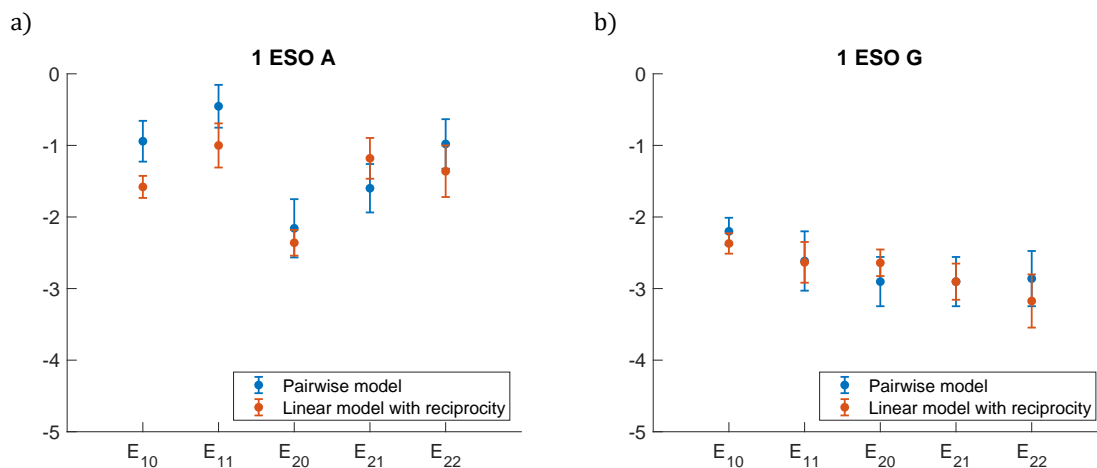
Figure 5.3: Parameters for the linear model with reciprocity - The figure represents the values and 95% confidence intervals for the maximum likelihood parameters in the linear model with reciprocity and its comparison with the ones obtained in the pairwise model for the following groups in wave 1: a) 1 ESO A and b) 1 ESO G. The figures representing the values of the parameters for each of the remaining groups in wave 1 can be found in subsection D.3.2.

The reasonably good alignment of the values obtained for the parameters using the linear model with reciprocity with those of the former model implies that the three factors considered largely explain the shapes of the parameters when utilising the pairwise approximation and, consequently, describe the probability distribution $P(G)$ that governs the ensemble. If this finding holds true, it may be possible to characterise the macroscopic structure of social relationships using only three individual properties of the constituents of the system and their interactions, as there is a strong correlation between the results obtained using the linear model with reciprocity and the ones with the pairwise model. However, before doing so, we need to validate the pairwise approximation to determine if it can reproduce macroscopic properties that characterise real-world social networks.

5.4 Validation of the pairwise approximation

The main objective of formulating and developing a model for network analysis is to generate a realistic and comprehensive structure that can reproduce the observed characteristics found in real-world systems. In particular, social networks exhibit two differential features that are present in most systems within this category. Firstly, the number of reciprocal relationships is substantially higher compared to a null model in which the links were randomly distributed. Secondly, the observed transitivity is quite large, implying that the probability of a pair of individuals having a relationship increases if they have mutual friends within the network. Thus, we need to verify if these two features are present when employing the proposed pairwise approximation in order to validate its ability to replicate real-world systems.

Reciprocity is already taken care of in the Hamiltonian, so it only remains to analyse if the transitivity produced by our model reproduces that of the real network. To do it, we compare the total number of triangles of each type in the system with those predicted by the
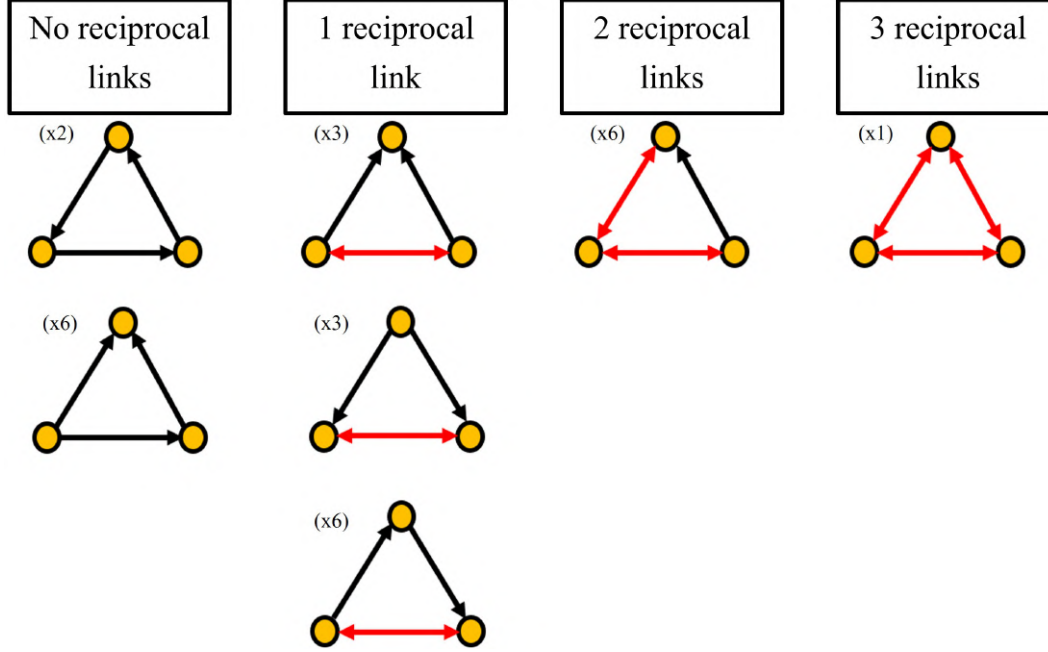
Figure 5.4: Set of triangles that can be formed in the system - The figure represents the total set of triangles that can be formed in the system, ignoring symmetries, grouped by the number of reciprocal links. The number of unique triangles of each type, including all symmetries, is shown in parentheses.

model. Due to the limited size of the groups, there is not enough data to estimate the average number of triangles of all possible types. To circumvent this problem, we ignore the direction of links as well as emotional intensity. We only consider whether links are reciprocal or not. The different types of triangles resulting from this simplification, ignoring symmetries, are listed in Figure 5.4.

To estimate the number of expected triangles in the system, the following considerations are made:

• The probability p_{ij} for the existence of a pair of type (i, j) is equal to the fraction of the number of pairs of that type over the total number of possible pairs:

$$p_{ij} = \frac{R_{ij}}{\binom{N}{2}} = \frac{2\,R_{ij}}{N(N-1)}. \tag{5.30}$$

• As the model is ideal, the probability of forming a triangle, where each edge has associated a pair of links of type (i, j), is the product of the probabilities for the three individual pairs:

$$p_{ij-kl-mn} = p_{ij}\,p_{kl}\,p_{mn} = \frac{8\,R_{ij}\,R_{kl}\,R_{mn}}{\left(N(N-1)\right)^3}. \tag{5.31}$$

• The total number of triangles predicted by the model is calculated by multiplying the probability of a specific type of triangle by the total number of possible triangles in the

system:

$$T_{ij-kl-mn} = p_{ij-kl-mn} \binom{N}{3} = \frac{4\,R_{ij}\,R_{kl}\,R_{mn}(N-2)}{3\,(N(N-1))^2}. \tag{5.32}$$

- When the link pairs forming the edges of the triangle are heterogeneous, that is two equal and one different, the number of triangles $T_{ij-kl-mn}$ predicted by the model is multiplied by 3. The reason is that there are three possible positions for the different pair, then:

$$T_{ij-kl-mn} = 3\,p_{ij-kl-mn} \binom{N}{3} = \frac{4\,R_{ij}\,R_{kl}\,R_{mn}(N-2)}{(N(N-1))^2}. \tag{5.33}$$

We can also obtain confidence intervals for the number of different types of triangles. These calculations are detailed in subsection D.1.2.

The comparison between the observed number of triangles and the expected one, together with the 95% confidence intervals, is shown for each group in wave 1 in Table D.1. For all groups, the observed values for triangles with zero, one, and two reciprocal links are within the bounds of the confidence interval. However, the expected values for triangles with three reciprocal links are much lower than the observed ones in the majority of the groups and, in most cases, the real value is not even within the confidence interval.

This finding indicates that our models are unable to characterise the transitivity of the system because if we were to simulate a network by randomly distributing pairs of links, the number of triangles that would appear would be lower than the actual number. Therefore, if we want to simulate real-world networks, we need to consider a mechanism that incentives the appearance of triangles with all three reciprocal links. The most natural way to address this issue is to include a term in the Hamiltonian of these models associated with this type of triangle.

With this in mind, we have expanded the Hamiltonian of the pairwise model, defined by Equation 5.2, to incorporate an additional term[2] that addresses this issue:

$$-H(G) = \sum_{\langle ab \rangle} E_{\sigma_{ab}\sigma_{ba}} + \frac{\gamma}{N} \sum_{\langle abc \rangle} T_{abc}, \tag{5.34}$$

where $\sigma_{ab} \in \{0, 1, ..., r\}$, $\langle ab \rangle$ denotes pairs of links and $\langle abc \rangle$ refers to triangles. Furthermore, we introduce the variable T_{abc}, which takes value $T_{abc} = 1$ if there is a reciprocal triangle spanning the three nodes a, b and c of the triangle and $T_{abc} = 0$ otherwise. In other words,

$$T_{abc} \equiv (1 - \delta_{\sigma_{ab},0})(1 - \delta_{\sigma_{ba},0})(1 - \delta_{\sigma_{ac},0})(1 - \delta_{\sigma_{ca},0})(1 - \delta_{\sigma_{bc},0})(1 - \delta_{\sigma_{cb},0}). \tag{5.35}$$

The problem we face when working with the Hamiltonian defined by Equation 5.34 is that it has terms that involve more than two nodes, so we cannot perform the factorisation

---

[2]The reason why the parameter associated with this term is written as γ/n will be discussed more in-depth in chapter 6. The idea behind it is to consider that there are $O(N^3)$ triangles in the network as opposed to $O(N^2)$ links. By incorporating this factor, both terms become comparable for large N if the constants are of the order of $O(1)$.

of the grand partition function as in Equation 5.7. This makes it impossible to obtain exact analytical expressions for the parameters since the calculation of the normalisation constant becomes intractable for large values of N.

5.5 Mapping into a Hamiltonian without levels

In section 5.4, we have shown that the previous models do not provide a complete description of the system because they are not capable of capturing transitivity, a characteristic feature of social networks. The number of triangles with all three reciprocal links that appear when randomly distributing pairs of links is lower than what is observed in real networks.

For this reason, we have added a term to the Hamiltonian defined by Equation 5.34 that enhances the presence of reciprocal triangles. However, this new Hamiltonian includes terms that involve more than two nodes, making it impossible to factorise the grand partition function and obtain exact analytical expressions for the parameters. Additionally, the graph we are working on is directed and weighted, with a total of r different possible weights. In this section, we will try to simplify the Hamiltonian so that we can compare it with classical models of the literature and gain insight into the system.

To start, we define the following variables on every pair ab:

$$\tau_{ab} \equiv (1 - \delta_{\sigma_{ab},0})(1 - \delta_{\sigma_{ba},0}). \tag{5.36}$$

These variables can only take two values, $\tau_{ab} = 0$ and $\tau_{ab} = 1$, depending on whether $\sigma_{ab}\sigma_{ba} = 0$ or $\sigma_{ab}\sigma_{ba} \neq 0$. Then, we consider the sum over the pair ab in the grand partition function defined by Equation 5.5,

$$S_{ab}(\boldsymbol{\tau}) \equiv \sum_{\sigma_{ab}=0}^{r} \sum_{\sigma_{ba}=0}^{r} \exp\left[E_{\sigma_{ab}\sigma_{ba}} + \tau_{ab}J_{ab}(\boldsymbol{\tau})\right], \tag{5.37}$$

where, for simplicity, we have introduced the short-hand notation

$$J_{ab}(\boldsymbol{\tau}) \equiv \frac{\gamma}{N} \sum_{c \neq a,b} \tau_{ac}\tau_{bc}. \tag{5.38}$$

We can split this sum as

$$S_{ab}(\boldsymbol{\tau}) = \begin{cases} \displaystyle\sum_{\sigma_{ab},\sigma_{ba}\,:\,\tau_{ab}=0} \exp(E_{\sigma_{ab}\sigma_{ba}}), & \text{if } \tau_{ab} = 0, \\ \displaystyle\sum_{\sigma_{ab},\sigma_{ba}\,:\,\tau_{ab}=1} \exp\left[E_{\sigma_{ab}\sigma_{ba}} + J_{ab}(\boldsymbol{\tau})\right], & \text{if } \tau_{ab} = 1, \end{cases} \tag{5.39}$$

which becomes

$$S_{ab}(\boldsymbol{\tau}) = \begin{cases} 1 + 2\displaystyle\sum_{i=1}^{r} e^{E_{i0}}, & \text{if } \tau_{ab} = 0, \\ e^{J_{ab}(\boldsymbol{\tau})}\displaystyle\sum_{i=1}^{r}\sum_{j=1}^{i}(2 - \delta_{i,j})e^{E_{ij}}, & \text{if } \tau_{ab} = 1, \end{cases} \tag{5.40}$$

Thus, if we introduce the variable

$$\phi = \log\left(\sum_{i=1}^{r}\sum_{j=1}^{i}(2-\delta_{i,j})e^{E_{ij}}\right) - \log\left(1 + 2\sum_{i=1}^{r}e^{E_{i0}}\right), \tag{5.41}$$

we can rewrite

$$S_{ab}(\boldsymbol{\tau}) = \left(1 + 2\sum_{i=1}^{r}e^{E_{i0}}\right)\sum_{\tau_{ab}=0,1}\exp\left\{\tau_{ab}\big[\phi + J_{ab}(\boldsymbol{\tau})\big]\right\}. \tag{5.42}$$

Therefore, we have an alternative expression for the grand partition function:

$$\Xi = \left(1 + 2\sum_{i=1}^{r}e^{E_{i0}}\right)^{\binom{N}{2}}\sum_{\boldsymbol{\tau}}e^{-\tilde{H}(G)}, \tag{5.43}$$

where

$$-\tilde{H}(G) \equiv \phi\sum_{\langle ab\rangle}\tau_{ab} + \frac{\gamma}{N}\sum_{\langle abc\rangle}\tau_{ab}\tau_{ac}\tau_{bc} = \phi L(G) + \frac{\gamma}{N}T(G). \tag{5.44}$$

In this expression, $L(G)$ is the number of reciprocal links in the graph, regardless of the level, and $T(G)$ is the number of reciprocal triangles.

Through this process, we have simplified the Hamiltonian defined by Equation 5.34 to reduce the complexity of the model so that the new one corresponds to an undirected and unweighted graph. Obviously, this is not exclusive to the pairwise model. For the linear model with reciprocity, we can use an analogous derivation (detailed in subsection D.1.3) using a transformation defined by

$$\phi = \beta + 2\lambda + 2\mu - \log\left(1 + 2e^{\lambda+\mu}\frac{1-e^{r\mu}}{1-e^{\mu}}\right) + 2\log\left(\frac{1-e^{r\mu}}{1-e^{\mu}}\right). \tag{5.45}$$

The Hamiltonian defined by Equation 5.44 corresponds to Strauss's model of transitive networks (Strauss, 1986), which has been analysed using mean-field techniques by Park and Newman (2005). These authors showed that this model exhibits a first-order phase transition when the interaction is strong enough, from a low-density to a high-density phase. For this reason, Strauss's model has been deemed unsuitable for producing networks with intermediate fractions of links, raising certain doubts about its usefulness. To solve this issue, in the following chapter, we study this phase transition using the language of lattice gases and characterise it in detail.

5.6 Discussion

In this chapter, we have constructed a model to explain the structure of social networks based on analysing the interactions between their members. The starting point is the concept of "social atom", introduced by Tamarit (2019) in his PhD thesis and supported by the results obtained in the first part of this thesis through an analysis of the data on social relationships

among high school students. The mathematical tools we have used are exponential random graphs, as their formulation in terms of statistical physics allows us to characterise the graph ensemble only using the Hamiltonian of the system and its associated grand partition function. We have explored a pairwise approximation, in which we assume that each pair of links is entirely defined by the intensity levels of their associated relationships, regardless of the identity of the nodes (hence with symmetric interactions).

The first model we have studied is the pairwise model. The parameters of this model, referred to as efficacies, characterise the stability of links and define the probability of their appearance in the system. To prove this, we compared the results of two groups with different densities of links, one high and one low, and observed that a higher value of the parameter indicates a greater likelihood of the associated relationship appearing when the system is simulated using the model. By analysing the values of the parameters for all the groups, we have found that reciprocal relationships in which the two intensity levels within the pair are equal are the most likely ones, while non-reciprocal relationships in which the present link has maximum intensity are the least probable ones.

The main drawback of the pairwise model is that it does not provide any information about the factors shaping the form of the parameters. For this reason, we introduced the linear model with reciprocity, in which we considered that the number of relationships that an individual can maintain simultaneously is limited, and their cognitive cost varies depending on the intensity of their associated links, in agreement with the Social Brain Hypothesis. We also added a term that enhances the reciprocity of the links. By doing so, we were able to fit the maximum likelihood values obtained through the pairwise model. These results suggest that we can explain the structure of social networks only considering these three variables.

The next step was validating the pairwise approximation by checking if it satisfies some of the characteristic macroscopic features observed in real-world social networks, especially reciprocity and transitivity. The latter is related to the number of triangles in the system, and we have demonstrated that the total number of triangles with three pairs of reciprocal links is larger than that which emerges when randomly distributing the pairs of relationships. To solve this issue, we introduced a term in the Hamiltonian that enhances the presence of this kind of triangles. However, this adds terms to the Hamiltonian that involve more than two nodes simultaneously, causing the grand partition function not to be factorisable. As a consequence, for systems with a large number of individuals, it becomes intractable.

This highlights the need for alternative methods to obtain results. Therefore, we have developed a transformation of the parameters that maps the Hamiltonian to that of Strauss's model of transitive networks, a model for undirected and unweighted graphs. This model has been previously explored by many researchers and exhibits a phase transition, suggesting that the model would be unsuitable for producing networks with intermediate fractions of links. Furthermore, small changes in the control parameter associated with the number of triangles cause very abrupt changes in the system. For this reason, the validity of this model, in particular, and exponential random graphs, in general, to describe real systems has been questioned. In the following chapter, we use the language of lattice gases to deal with this issue.

6

A COMPREHENSIVE ANALYSIS OF STRAUSS'S MODEL

Strauss's model of transitive networks is an exponential random graph model that enhances clustering by introducing in the Hamiltonian of the system an interaction term associated with the number of triangles. This model exhibits a first-order phase transition from a low-density to a high-density phase when the interaction is strong enough, which suggests that there are values of some observables that no graph in the ensemble can attain for a certain set of parameters. Furthermore, in this regime, the typical graphs of the ensemble abruptly change from sparse to dense upon a slight variation of the control parameter. These facts have cast serious doubts in the past on the usefulness of the model.

In this chapter, we will introduce the language of lattice gases to show that there is no qualitative difference between the phase transition exhibited by Strauss's model and the condensation transition of an Ising lattice gas. In this way, thermodynamics will provide a description of the sort of networks that we must expect for those "forbidden" values of the observables in Strauss's model. Moreover, we will address this problem using a density-functional formalism especially tailored for lattice gases. This formalism provides a method to construct a mean-field-like free energy of the system, from which everything else can be derived. It also has the advantage that the non-homogeneous counterpart of Strauss's model can be solved with no extra effort. Networks in which nodes of different types interact in different ways are of this kind, and using them we can study, for example, the effect of homophily in social networks.

6.1 Strauss's model and its lattice-gas interpretation

Strauss's model is an ERG ensemble of undirected graphs with N nodes defined by the Hamiltonian

$$-H(G) = \phi L(G) + \frac{\gamma}{N} T(G), \tag{6.1}$$

where $L(G)$ is the number of links (edges) in the graph and $T(G)$ is the number of triangles (clustering). In this model, a positive parameter ϕ enhances the creation of links and a positive parameter γ enhances the formation of triangles. The factor N^{-1} in front of $T(G)$ accounts for the fact that there are $O(N^3)$ triangles in the network compared to $O(N^2)$ links. With this factor, both terms are comparable for large N if the constants are $O(1)$. The model was introduced by Strauss (1986) to describe graphs with a clustering higher than that of a typical Erdös-Rényi graph.

In this chapter, we will deal with a non-homogeneous version of this model. To write down the Hamiltonian, we need to introduce some notation. Let $\mathscr{N}$ denote the set of all nodes of an undirected graph G, $\mathscr{N}_2$ the set of all subsets of 2 elements of $\mathscr{N}$, hence the potential links, which will be denoted by their indexes $\{i_1, i_2\}$, and $\mathscr{N}_3$ the set of all subsets of 3 elements of $\mathscr{N}$, hence the potential triangles, which will be denoted $\{i_1, i_2, i_3\}$. If $|\mathscr{N}| = N$, then $|\mathscr{N}_2| = \binom{N}{2}$, $|\mathscr{N}_3| = \binom{N}{3}$. We also denote $\{\tau_{ij}\}$ the adjacency matrix of the graph G, where $\tau_{ij} = 1$ if the link $\{i, j\}$ exists in G and $\tau_{ij} = 0$ otherwise. Then, in terms of this variables a triangle $\{i, j, k\}$ exists if and only if $\tau_{ij}\tau_{jk}\tau_{ki} = 1$. Thus, the Hamiltonian of the non-homogeneous version of Strauss's model can be written as

$$-H(G) = \sum_{\{ij\} \in \mathscr{N}_2} \phi_{ij} \tau_{ij} + \sum_{\{ijk\} \in \mathscr{N}_3} \frac{\gamma_{ijk}}{N} \tau_{ij} \tau_{jk} \tau_{ki}. \tag{6.2}$$

The parameters ϕ_{ij} and γ_{ijk} are local versions of the ones associated with the Hamiltonian defined by Equation 6.1. Therefore, a positive parameter ϕ_{ij} enhances the creation of the link $\{i, j\}$ and, similarly, a positive parameter γ_{ijk} enhances the formation of the triangle $\{i, j, k\}$.

The variables τ_{ij} play the role of "particles" sitting on the links of the complete graph over the set of nodes $\mathscr{N}$. If $\tau_{ij} = 1$, it means that the link $\{ij\}$ is occupied by a particle, whereas if $\tau_{ij} = 0$, the link is empty. Thus, G can also be interpreted as the configuration of $\binom{N}{2}$ such particles in the complete graph. Under this interpretation, ϕ_{ij} can be regarded as a (local) chemical potential, and the grand partition function of the system can be defined as

$$\Xi = \sum_{G \in \mathscr{G}} e^{-H(G)}, \tag{6.3}$$

where $\mathscr{G}$ is the set of all possible realisations of the graphs (Lafuente & Cuesta, 2004, 2005; Lavis & Bell, 1999).

As particles occupy the links, rather than the nodes, of a complete graph, the "space" where these particles live is weird. As a matter of fact, in this dual network, every two particles are either neighbours or second neighbours to each other. The reason is that if links $\{ij\}$ and $\{kl\}$ are not neighbours, i.e. have no common nodes, then they are both neighbours
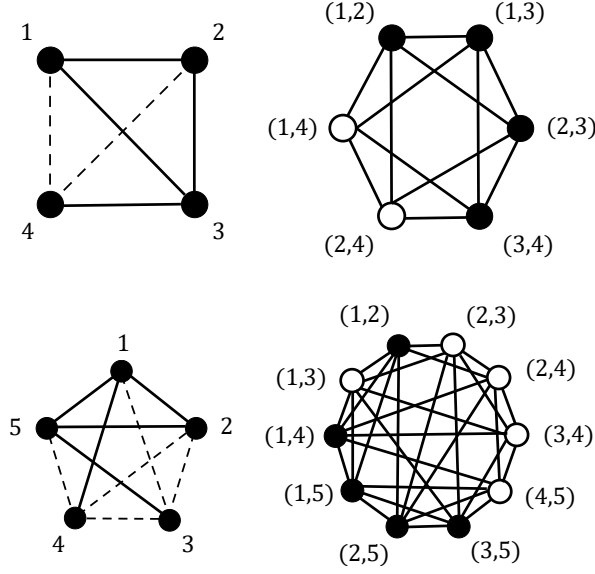
Figure 6.1: Dual networks for the complete graphs - The figure represents the dual networks for the complete graphs of four (up) and five nodes (down). In dual networks, links are represented as nodes, and two nodes are neighbours if they share a node in the original graph. Furthermore, in these dual networks, every two nodes are either neighbours or second neighbours to each other. For an alternative representation, see Figure 1 in Palla et al. (2004).

to a common link (e.g. $\{ik\}$). Figure 6.1 illustrates these dual networks for the complete graphs of 4 and 5 nodes. Each link is neighbour to $2(N-2)$ other links, and second neighbour to the remaining $\binom{N-2}{2}$.

The grand potential of this system $\Omega = -\log \Xi$ is a function of all conjugate fields $\boldsymbol{\phi} = \{\phi_{ij}\}$ and $\boldsymbol{\gamma} = \{\gamma_{ijk}\}$, from which the probability that link $\{ij\}$ is occupied, henceforth "density", can be obtained as

$$-\frac{\partial \Omega}{\partial \phi_{ij}} = \langle \tau_{ij} \rangle = \rho_{ij}. \tag{6.4}$$

A Legendre transform on the grand potential yields the free energy

$$F(\boldsymbol{\rho}, \boldsymbol{\gamma}) = \sum_{\{ij\}} \phi_{ij}(\boldsymbol{\rho})\rho_{ij} + \Omega(\boldsymbol{\phi}, \boldsymbol{\gamma}), \tag{6.5}$$

where $\phi_{ij}(\boldsymbol{\rho})$ is obtained by solving Equation 6.4 for fixed $\boldsymbol{\rho}$. Differentiating the free energy F with respect to the densities,

$$\frac{\partial F}{\partial \rho_{ij}} = \phi_{ij} + \sum_{\{kl\}} \frac{\partial \phi_{kl}}{\partial \rho_{ij}}\rho_{kl} + \sum_{\{kl\}} \frac{\partial \Omega}{\partial \phi_{kl}} \frac{\partial \phi_{kl}}{\partial \rho_{ij}}, \tag{6.6}$$

so, if we take Equation 6.4 into account, we finally find that

$$\frac{\partial F}{\partial \rho_{ij}} = \phi_{ij}, \tag{6.7}$$

which is the usual equation for the chemical potential.

It can be proven that, given the free-energy density functional $F(\boldsymbol{\rho})$ of a system, its equilibrium density is the unique density profile that minimises the functional defined by $\Omega(\boldsymbol{\rho}) \equiv F(\boldsymbol{\rho}) - \boldsymbol{\phi} \cdot \boldsymbol{\rho}$ (Hansen & McDonald, 2013, Appendix B). Alternatively, it minimises $F(\boldsymbol{\rho})$ at constant mean density. Thus, Equation 6.7, which is dual to Equation 6.4, is the expression of this variational principle. Consequently, its solution provides the values of the densities for a given set of chemical potentials $\boldsymbol{\phi}$.

6.2 Fundamental-measure approximation

The technique we will use to find an approximation to the free energy of Strauss's model is known in the field of lattice gases as *fundamental-measure theory* (Lafuente & Cuesta, 2004, 2005). It is a reformulation of the well-known cluster variation method (Lavis & Bell, 1999). The idea behind this technique is to decompose the system in overlapping clusters and express the free energy as a sum of the free energies of those clusters, controlling for overcounting.

In the case of Strauss's model, the geometry of the Hamiltonian suggests that the simplest possible clusters are triangles. Thus, as a first approximation, the free energy is obtained as the sum of the contributions to the free energy of all the triangles within the complete graph. However, in doing so, every link participates in $N - 2$ triangles, so we need to subtract $N - 3$ times the contribution to the free energy of all links. In other words, the fundamental-measure approximation to the free energy will be

$$F(\boldsymbol{\rho}, \boldsymbol{\gamma}) = \sum_{\{ijk\}} \Phi_3(\rho_{ij}, \rho_{jk}, \rho_{ki}, \gamma_{ijk}) - (N - 3) \sum_{\{ij\}} \Phi_2(\rho_{ij}), \tag{6.8}$$

where Φ_2 and Φ_3 are the free energies of a single link and a single triangle, respectively.

The expression for Φ_2 is easy to obtain. Denoting $\Xi_{ij} \equiv e^{\phi_{ij}}$, the grand partition function for a single link $\{ij\}$ is simply $\Xi_2 = 1 + z_{ij}$. Thus,

$$\rho_{ij} = z_{ij} \frac{\partial}{\partial z_{ij}} \log \Xi_2 = \frac{z_{ij}}{1 + z_{ij}},$$

from which

$$z_{ij} = \frac{\rho_{ij}}{1 - \rho_{ij}}, \qquad \Xi_2 = \frac{1}{1 - \rho_{ij}}.$$

Substituting these expressions in the Legendre transform $\Phi_2 = \rho_{ij} \log z_{ij} - \log \Xi_2$, we end up with

$$\Phi_2(\rho_{ij}) = \rho_{ij} \log \rho_{ij} + (1 - \rho_{ij}) \log(1 - \rho_{ij}), \tag{6.9}$$

which is simply the free energy of an ideal lattice gas.

The calculation of Φ_3 is rather more involved (see subsection E.1.1). Introducing the shorthand

$$\zeta_{ijk} \equiv \exp(\gamma_{ijk}/N) - 1 = \frac{\gamma_{ijk}}{N} + O\left(\frac{1}{N^2}\right), \tag{6.10}$$

its expression turns out to be

$$\Phi_3 = \Phi_2(\rho_{ij}) + \Phi_2(\rho_{jk}) + \Phi_2(\rho_{ki}) + \rho_{ij} \log \left(1 - \frac{\rho_{ijk}}{\rho_{ij}} \right)$$
$$+ \rho_{jk} \log \left(1 - \frac{\rho_{ijk}}{\rho_{jk}} \right) + \rho_{ki} \log \left(1 - \frac{\rho_{ijk}}{\rho_{ki}} \right) - 2 \log(1 - \rho_{ijk}), \tag{6.11}$$

where ρ_{ijk} is one of the real solutions of the cubic equation

$$\zeta_{ijk}(\rho_{ij} - \rho_{ijk})(\rho_{jk} - \rho_{ijk})(\rho_{ki} - \rho_{ijk}) = \rho_{ijk} \left(1 - \rho_{ijk} \right)^2. \tag{6.12}$$

This "triangle density" ρ_{ijk} is related to $T_{ijk} = \langle \tau_{ij}\tau_{jk}\tau_{ki} \rangle$, the probability that the nodes i, j, k form a triangle, as (see subsection E.1.2 for full detailed calculations)

$$T_{ijk} = \frac{1 + \zeta_{ijk}}{\zeta_{ijk}} \rho_{ijk} = \frac{1}{1 - e^{-\gamma_{ijk}/N}} \rho_{ijk}. \tag{6.13}$$

If we now substitute Equation 6.9 and Equation 6.11 into Equation 6.8, and take into account that

$$\sum_{\{ijk\}} (A_{ij} + A_{jk} + A_{ki}) = (N - 2) \sum_{\{ij\}} A_{ij} \tag{6.14}$$

for any link-dependent magnitude A_{ij}, we finally conclude that the free energy F satisfies

$$F = \sum_{\{ij\}} \left[\rho_{ij} \log \rho_{ij} + (1 - \rho_{ij}) \log(1 - \rho_{ij}) \right]$$
$$+ \sum_{\{ijk\}} \left[\rho_{ij} \log \left(1 - \frac{\rho_{ijk}}{\rho_{ij}} \right) + \rho_{jk} \log \left(1 - \frac{\rho_{ijk}}{\rho_{jk}} \right) \right.$$
$$\left. + \rho_{ki} \log \left(1 - \frac{\rho_{ijk}}{\rho_{ki}} \right) - 2 \log(1 - \rho_{ijk}) \right]. \tag{6.15}$$

6.3 Homogeneous networks

6.3.1 Free energy

We can recover Strauss's original model by assuming $\rho_{ij} = \rho$ for every link $\{ij\}$ and $\gamma_{ijk} = \gamma$ for every triangle $\{ijk\}$. Then, the free energy *per link* $f(\rho, \gamma) = \binom{N}{2}^{-1} F(\boldsymbol{\rho}, \boldsymbol{\gamma})$ will be

$$f = \rho \log \rho + (1 - \rho) \log(1 - \rho) + (N - 2) \left[\rho \log \left(1 - \frac{\rho_{\mathrm{T}}}{\rho} \right) - \frac{2}{3} \log(1 - \rho_{\mathrm{T}}) \right], \tag{6.16}$$

where ρ_{T} is the only real root of

$$\zeta(\rho - \rho_{\mathrm{T}})^3 = \rho_{\mathrm{T}}(1 - \rho_{\mathrm{T}})^2. \tag{6.17}$$

With the change of variable

$$t = \frac{\rho - \rho_T}{1 - \rho_T}, \qquad \rho_T = \frac{\rho - t}{1 - t}, \tag{6.18}$$
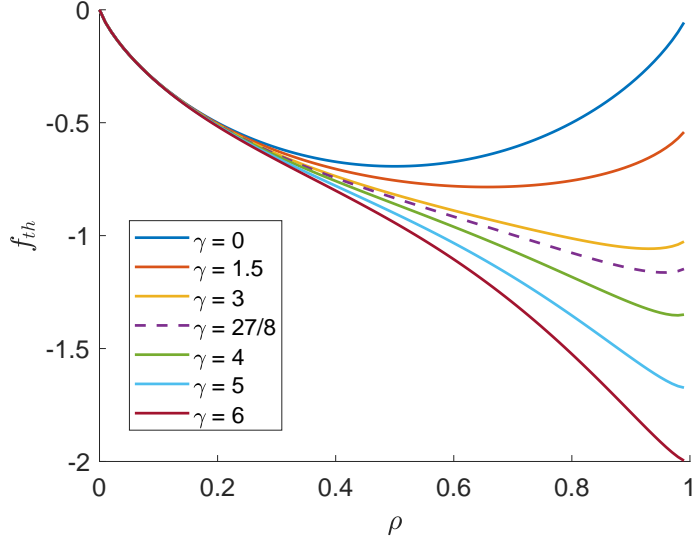
Figure 6.2: Thermodynamic free energy - The figure represents the thermodynamic free energy of the model for different values of γ below and above the critical value γ_c (γ increases from top to bottom). The curves illustrate the onset of the concavity as γ grows past $\gamma_c = 27/8$, represented by the purple dashed line.

Equation 6.17 can be rewritten as

$$t^3 + \frac{t}{\zeta(1-\rho)} - \frac{\rho}{\zeta(1-\rho)} = 0. \tag{6.19}$$

This equation has only one real root, which is given by the formula (Birkhoff & Mac Lane, 1997, pp. 102–103)

$$t = \frac{2}{\sqrt{3\zeta(1-\rho)}} \sinh\left[\frac{1}{3}\sinh^{-1}\left(\frac{3}{2}\rho\sqrt{3\zeta(1-\rho)}\right)\right]. \tag{6.20}$$

6.3.2 Thermodynamic limit

In the thermodynamic limit $N \to \infty$ one can see, either from Equation 6.17 or directly from Equation 6.20, that the thermodynamic free energy becomes simply

$$f_{\text{th}} = \rho \log \rho + (1-\rho)\log(1-\rho) - \frac{\gamma\rho^3}{3}. \tag{6.21}$$

This free energy is convex as long as

$$\frac{\partial^2 f_{\text{th}}}{\partial \rho^2} = \frac{1}{\rho(1-\rho)} - 2\gamma\rho > 0, \tag{6.22}$$

which is equivalent to $2\gamma\rho^2(1-\rho) < 1$. As the maximum value of $\rho^2(1-\rho)$ is $4/27$, reached at $\rho_c = 2/3$, the condition above implies $\gamma < 27/8$. So, the values $\gamma_c = 27/8$ and $\rho_c = 2/3$ mark a critical point above which the system exhibits a first-order phase transition.
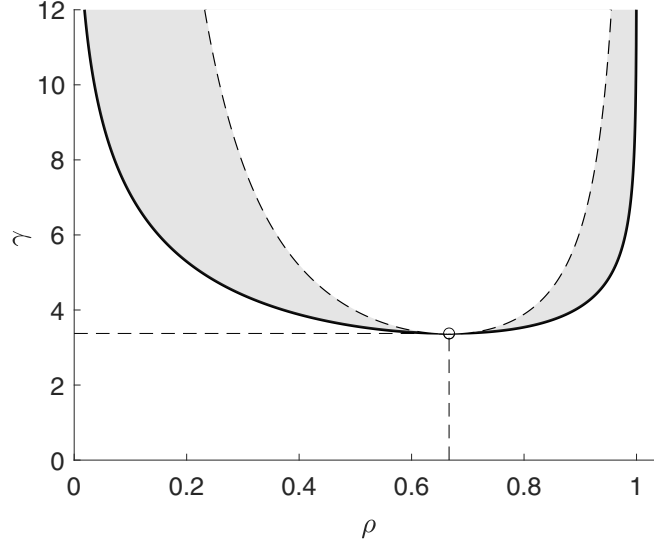
Figure 6.3: Coexistence curve, spinodal curve and critical point - In the upper region delimited by the solid curve, the system is not homogeneous but separated into two coexisting phases whose respective densities are given by the values of the curve at the corresponding γ. The dashed line represents the spinodal (defined by Equation 6.24), i.e. the curve at which the compressibility vanishes. Hence, the thermodynamic free energy changes from convex to concave. The circle, where both curves meet, marks the critical point. Within the shaded region, the system may be trapped in a metastable, homogeneous state.

Figure 6.2 illustrates the concavity that f_{th} develops as γ increases past γ_c. It is the fingerprint of a condensation transition in lattice gases (Lavis & Bell, 1999) because concave free energy implies thermodynamic instability; that is, the compressibility is negative. The homogeneous "fluid" separates into two phases, each of a different density, in thermodynamic equilibrium. The fraction occupied by each phase must be such that the overall density matches the prescribed one.

The thermodynamic equilibrium is determined by "chemical" equilibrium (equality of chemical potentials) and "mechanical" equilibrium (equality of pressures). In this way, the first condition implies $f_\rho(\rho_1, \gamma) = f_\rho(\rho_2, \gamma)$, and the second $\rho_1 f_\rho(\rho_1, \gamma) - f(\rho_1, \gamma) = \rho_2 f_\rho(\rho_2, \gamma) - f(\rho_2, \gamma)$. Both conditions are summarised in the equation

$$f_\rho(\rho_1, \gamma) = f_\rho(\rho_2, \gamma) = \frac{f(\rho_2, \gamma) - f(\rho_1, \gamma)}{\rho_2 - \rho_1}, \tag{6.23}$$

which represent Maxwell's double tangent construction (Huang, 1987). For f_{th}, the solution of these equations is represented in Figure 6.3. Given any $\rho_1 < \rho < \rho_2$, there will be a fraction x of the graph of density ρ_1 and a fraction $1 - x$ of density ρ_2 such that $\rho = x\rho_1 + (1 - x)\rho_2$.

On the other hand, the condition $f_{\rho\rho}(\rho, \gamma) = 0$ marks the points where the compressibility vanishes, that is, the point where the system is no longer mechanically stable. This curve is known as the *spinodal*, represented in Figure 6.3. According to Equation 6.22, this curve is
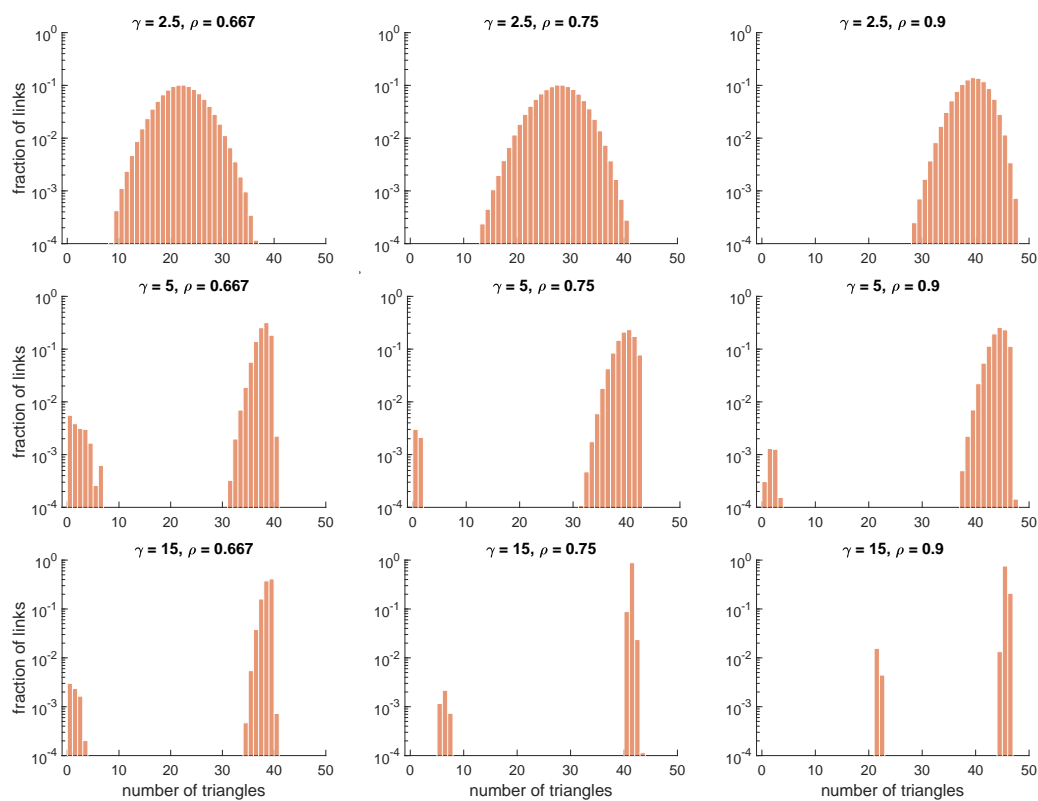
Figure 6.4: Monte Carlo simulations using Kawasaki dynamics - The figure represents nine panels that depict the average fraction of links belonging to a given number of triangles, as obtained from Monte Carlo simulations, using Kawasaki dynamics, of a Strauss network with $N = 50$ nodes, for three values of the interaction parameter γ (below, just above, and well above the critical point) and three different densities.

defined by

$$\gamma = \frac{1}{2\rho^2(1-\rho)}. \tag{6.24}$$

Within the region between the coexistence curve and the spinodal, which corresponds to the shaded area of Figure 6.3, the system can still be prepared in a homogeneous but metastable state. This explains the origin of the hysteresis observed in first-order phase transitions. This particular one is presented in Park and Newman (2005).

It is difficult to guess the nature of the phase transition that this system undergoes above the critical point. Recall that no more than one intermediate neighbour separates any two links. The very notion of "space" breaks down in such a system, so the picture of the usual condensation transition, where gas and liquid occupy different portions of the volume, has no reasonable counterpart in a complete graph. Nonetheless, the transition may be illustrated by computing a histogram of the number of links belonging to a given number of triangles (Tamm et al., 2014). As a way of illustration, we have obtained such histograms by performing Monte Carlo simulations using the dynamics of Kawasaki (1972), which preserves the number of links, and hence the density ρ. In this dynamics, a Monte Carlo step amounts to first removing a link at random and then creating a link also at random. The results, obtained for three different values of the interaction γ (below, just above, and well above the critical point) and three different densities, are depicted in Figure 6.4. In each of these simulations, we perform 5×10^5 Monte Carlo steps. When $\gamma < \gamma_c$, the histograms show a single peak that shifts to the right and shrinks as the density increases, whereas if $\gamma > \gamma_c$, the distribution exhibits two very neat peaks, one at high values and the other one at lower values of the number of triangles. Obviously, the links forming each of the two peaks belong to each of the low and high-density phases. Figure 6.4 reveals that networks within the coexisting region do exist, but they have different structural properties than those outside this region.

6.3.3   Finite networks

We can use an asymptotic expansion in N to obtain ρ_T from Equation 6.20. The first two terms are

$$\rho_T = \frac{\gamma\rho^3}{N} \left[1 + \frac{\gamma(1 - 6\rho^2 + 4\rho^3)}{2N} + O\left(\frac{1}{N^2}\right) \right], \tag{6.25}$$

and, consequently, the free energy can be expanded as

$$f = \rho\log\rho + (1-\rho)\log(1-\rho) - \frac{\gamma\rho^3}{3}\left[1 - \frac{2}{N} + \frac{\gamma(1 - 3\rho^2 + 2\rho^3)}{2N} + O\left(\frac{1}{N^2}\right)\right]. \tag{6.26}$$

The critical point will then be a solution of the equations $f_{\rho\rho} = f_{\rho\rho\rho} = 0$, which yields

$$\gamma_c(N) = \frac{27}{8}\left[1 + \frac{45}{16N} + O\left(\frac{1}{N^2}\right)\right],$$
$$\rho_c(N) = \frac{2}{3} + O\left(\frac{1}{N^2}\right). \tag{6.27}$$

The numerical solution for $\gamma_c(N)$ is depicted in Figure 6.5 along with the asymptotic expansion above. As for $\rho_c(N)$, within the numerical resolution, we find that its value is
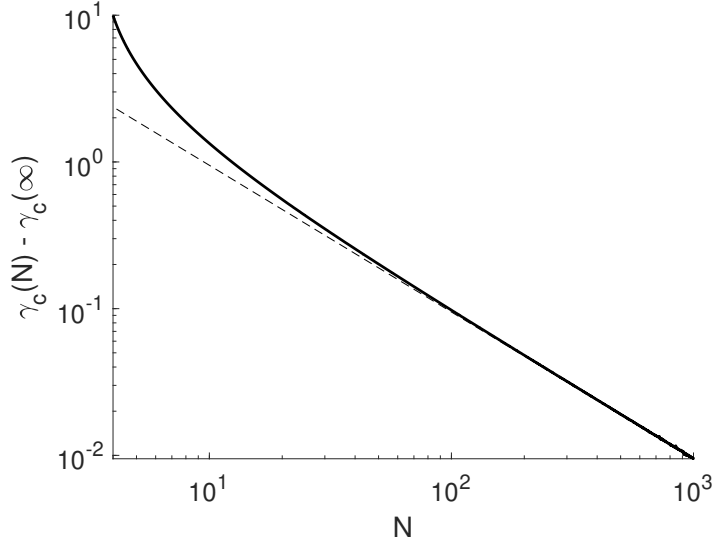
Figure 6.5: Critical value in finite networks - The figure represents the difference between the critical value $\gamma_c(N)$ for a network with N nodes and its limit for $N \to \infty$, as a function of N. The solid line is obtained by numerically solving the equations for the critical point and the dashed line arises from the asymptotic expression derived in Equation 6.27.

always 2/3. It is noteworthy that the curve of $\gamma_c(N)$ diverges somewhere between $N = 4$ and $N = 3$.

In spite that the uniform free energy defined by Equation 6.26 predicts a critical point and a first-order phase transition for arbitrary N (as low as $N = 4$, see Figure 6.5), we know that this is not possible. In other words, this phase transition is not real. The free energy exhibits a concavity for some values of ρ when $\gamma > \gamma_c(N)$ because the equilibrium solution is not truly uniform for any density value, even though for some densities it is indistinguishable from a uniform one. The transition from the regions where the solution is almost uniform to those in which the structure is like those shown in Figure 6.4 (third row) is continuous, albeit probably abrupt. Thus, the approximation we are using here shows up as a phase transition (i.e., the convex envelope of the free energy as a function of ρ represents a good approximation of the true free energy of the system). To illustrate this point, we compare our approximate free energy for $N = 4$ with the exact one (see subsection E.1.3) in Figure 6.6.

6.3.4 Comparison with Park and Newman's mean-field calculations

A fair question is how the present theory compares with the mean-field calculations of Park and Newman (2005). In spirit, this theory is also mean-field-like, but clearly, its construction follows a very different approach. Because of the high dimensionality of this system, one expects that in the thermodynamic limit, it becomes exact (Park & Newman, 2005), so it would be desirable that, if not for all N, at least in this limit, both theories coincide. Figure 3(b) of Park and Newman (2005) shows the expected number of triangles T (among other things) as a function of the interaction parameter γ/N, for $\phi = -0.53$ and $N = 500$. We can obtain T and ϕ as a function of ρ and γ through Equation 6.7 and Equation 6.13,
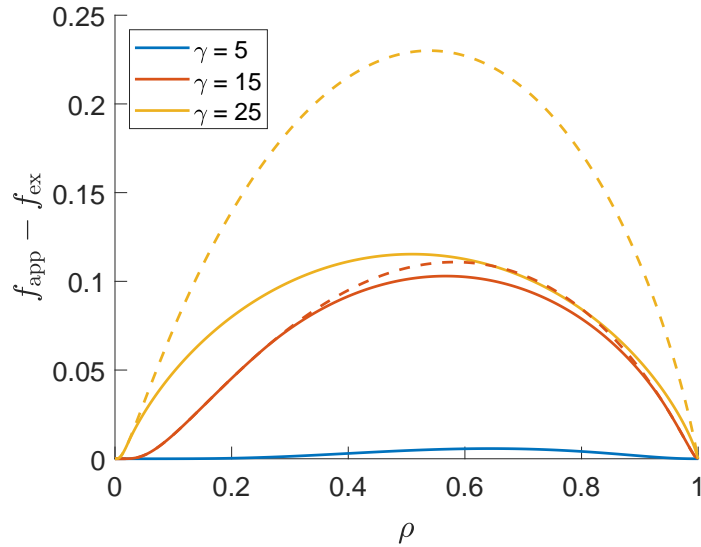
Figure 6.6: An approximation to the real free energy - In the figure, the dashed lines represent the difference between the approximate free energy (f_{app}), defined by Equation 6.26, and the exact free energy (f_{ex}), defined by Equation E.15, for the complete graph with $N = 4$ nodes and three values of the interaction parameter γ (below, just above, and well above the pseudo-critical point). As the free energy has an unphysical concave region above the critical point, we also represent in solid lines the difference between the *convex envelope* of the approximate free energy and the exact one since this convex envelope is a better approximation to the real free energy.

respectively. From these two calculations, we can obtain parametrically the curve $T(\gamma)$ for fixed ϕ and different values of N. The discrepancy between our results and those of Park and Newman is shown in Figure 6.7. Figure 6.7(a) illustrates that the difference between the predictions of both theories decreases with system size—so that both coincide in the thermodynamic limit. However, for very small networks, their predictions differ significantly (e.g., for $N = 10$, the discrepancy may be as high as $\sim 20\%$). Figure 6.7(b) compares the predictions of both theories for $N = 10$ along with Monte Carlo simulations performed using a Metropolis-Hastings algorithm (Snijders, 2002). Observable magnitudes are averaged over 10^6 configurations of the Markov chain. This figure highlights the higher accuracy of the current theory in calculating results for small networks.

Given that most real networks are large, the discussion of this section may seem like an academic issue of little practical relevance. Nevertheless, small networks of about 20-30 nodes are common, for example, in social science, anthropology, or biology. Thus, one can find instances of these small networks in studies of different social organisations (Escribano et al., 2021; Everett & Borgatti, 2014; Huitsing & Veenstra, 2012; Stadtfeld et al., 2020), in bands of hunter-gatherers (Migliano et al., 2020; Page et al., 2017), or in groups of social animals (Escribano et al., 2022; Ilany et al., 2013; Kasper & Voelkl, 2009). For these studies, the improved accuracy provided by the current approach might not be negligible. As a result, researchers working with smaller networks may find it beneficial to employ the models and techniques discussed in this section.
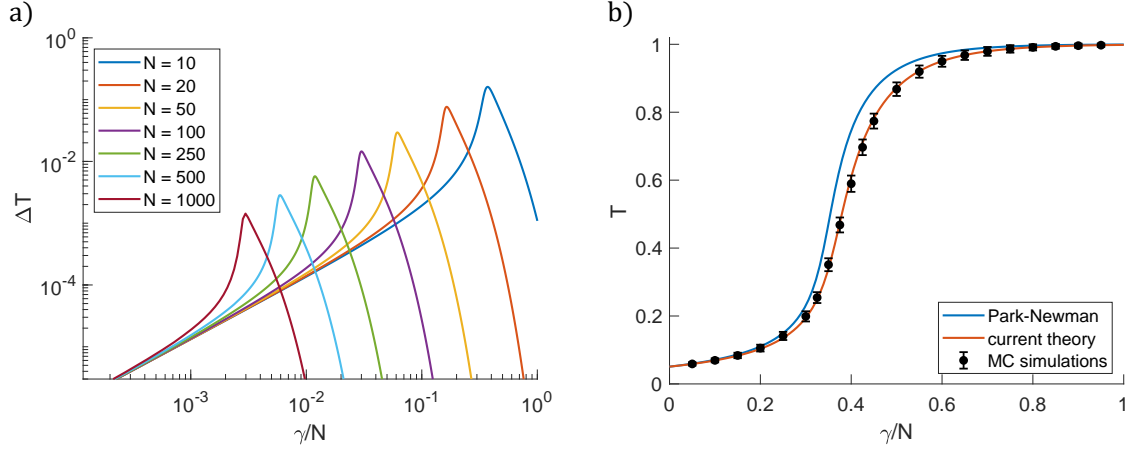
Figure 6.7: Comparison with Park and Newman's mean-field calculations - Each of the panels in the figure represents the following: a) Difference between the expected number of triangles T as obtained from Park and Newman's mean-field calculations and from the current theory. b) Expected number of triangles for $N = 10$ according to both theories, along with Monte Carlo simulations. Error bars represent the standard deviation of the number of triangles along the simulations. In both panels $\phi = -0.53$.

6.4 Non-homogeneous networks: Homophily

Having an expression for the free energy of the non-homogeneous Strauss's model allows us to tackle other interesting cases. Particularly important is the case where there are different types of nodes in the network with different interaction parameters. This case can model, e.g., homophily in a social network, where similar nodes are more prone to form links or triangles than different nodes are (McPherson et al., 2001). A particular version of this model has already been used to study segregation on Strauss networks where triangles are both favoured or disfavoured (Avetisov et al., 2018).

Suppose we have two types of nodes in the network, A and B. Since the underlying graph is complete, the actual location of these nodes is irrelevant; only how many of each type there are matters. So let us assume that there are N_A of type A and $N_B = N - N_A$ of type B. Accordingly, $\binom{N_A}{2}$ links are homophilic of type AA, $\binom{N_B}{2}$ of type BB, and $N_A N_B$ are of mixed type. Likewise, there will be $\binom{N_A}{3}$ homophilic triangles of type AAA, $\binom{N_B}{3}$ of type BBB, $\binom{N_A}{2} N_B$ mixed triangles of type AAB, and $\binom{N_B}{2} N_A$ of type ABB. Hence, the free energy of the system can be obtained as

$$
\begin{aligned}
F = &\sum_{X=A,B} \binom{N_X}{2} \left[\rho_{XX} \log \rho_{XX} + (1 - \rho_{XX}) \log(1 - \rho_{XX}) \right] \\
&+ N_A N_B \left[\rho_{AB} \log \rho_{AB} + (1 - \rho_{AB}) \log(1 - \rho_{AB}) \right] \\
&+ \sum_{X=A,B} \binom{N_X}{3} \left[3\rho_{XX} \log \left(1 - \frac{\rho_{XXX}}{\rho_{XX}} \right) - 2 \log(1 - \rho_{XXX}) \right] \\
&+ \sum_{X=A,B} \sum_{Y \neq X} \binom{N_X}{2} N_Y \left[\rho_{XX} \log \left(1 - \frac{\rho_{XXY}}{\rho_{XX}} \right) + 2\rho_{XY} \log \left(1 - \frac{\rho_{XXY}}{\rho_{XY}} \right) - 2 \log(1 - \rho_{XXY}) \right],
\end{aligned}
$$

where $\rho_{XY} = \rho_{YX}$ is the density of links of type XY, and the densities associated with the
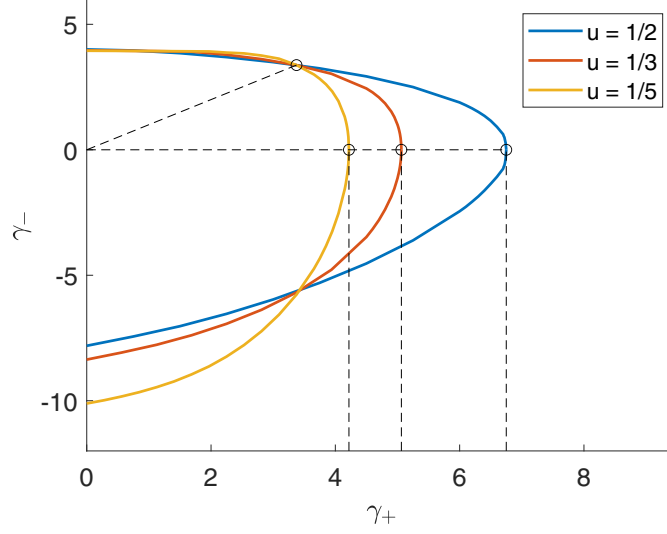
Figure 6.8: Critical curves for the interaction parameters - The figure represents the critical curves γ_- vs. γ_+ for different values of $0 \leq u \leq 1/2$. The free energy of the non-homogeneous model is convex only for the points on the left of the curve. The curves reach their rightmost values of γ_+ (marked with circles and vertical dashed lines) for $\gamma_+ = 27/8(1 - u)$, $\gamma_- = 0$. The oblique dashed line is $\gamma_+ = \gamma_-$. It meets all critical curves at one point (marked with a circle): $\gamma_+ = \gamma_- = 27/8$, the critical point of the homogeneous system.

triangles are the solutions of

$$\zeta_{\text{XXY}}(\rho_{\text{XX}} - \rho_{\text{XXY}})(\rho_{\text{XY}} - \rho_{\text{XXY}})^2 = \rho_{\text{XXY}}(1 - \rho_{\text{XXY}})^2, \tag{6.28}$$

To reduce the number of parameters of the model, we will henceforth assume that it is only homophily, and not the nature of the nodes, that determines interactions. This means that there are only two values of the interaction parameter instead of four, namely $\gamma_{\text{AAA}} = \gamma_{\text{BBB}} \equiv \gamma_+$, $\gamma_{\text{AAB}} = \gamma_{\text{BBA}} \equiv \gamma_-$. Furthermore, in the thermodynamic limit, the solution to Equation 6.28 is

$$\rho_{\text{XXY}} = \rho_{\text{XX}}\rho_{\text{XY}}^2 \frac{\gamma_\pm}{N} + O\left(\frac{1}{N^2}\right), \tag{6.29}$$

where the subindex of $\gamma_\pm$ depends on whether X=Y ($+$) or X$\neq$Y ($-$). In this same limit, and setting $N_{\text{A}} = uN$, $N_{\text{B}} = (1-u)N$, the free energy per link $f \equiv \binom{N}{2}^{-1}F$ turns out to be

$$
\begin{aligned}
f = {} & u^2\big[\rho_{\text{AA}} \log \rho_{\text{AA}} + (1 - \rho_{\text{AA}}) \log(1 - \rho_{\text{AA}})\big] \\
& + (1-u)^2\big[\rho_{\text{BB}} \log \rho_{\text{BB}} + (1 - \rho_{\text{BB}}) \log(1 - \rho_{\text{BB}})\big] \\
& + 2u(1-u)\big[\rho_{\text{AB}} \log \rho_{\text{AB}} + (1 - \rho_{\text{AB}}) \log(1 - \rho_{\text{AB}})\big] - \frac{\gamma_+}{3}\big[u^3\rho_{\text{AA}}^3 + (1-u)^3\rho_{\text{BB}}^3\big] \\
& - \gamma_- u(1-u)\rho_{\text{AB}}^2\big[u\rho_{\text{AA}} + (1-u)\rho_{\text{BB}}\big].
\end{aligned}
\tag{6.30}
$$

The convexity of this function is linked to the positive definiteness of its Hessian matrix, where entries are ordered as AA, BB, and AB. This matrix is defined by

$$H=\begin{pmatrix} \dfrac{u^2}{\rho_{\mathrm{AA}}(1-\rho_{\mathrm{AA}})} - 2\gamma_+ u^3 \rho_{\mathrm{AA}} & 0 & -2\gamma_- u^2(1-u)\rho_{\mathrm{AB}} \\[2ex] 0 & \dfrac{(1-u)^2}{\rho_{\mathrm{BB}}(1-\rho_{\mathrm{BB}})} - 2\gamma_+(1-u)^3 \rho_{\mathrm{BB}} & -2\gamma_- u(1-u)^2 \rho_{\mathrm{AB}} \\[2ex] -2\gamma_- u^2(1-u)\rho_{\mathrm{AB}} & -2\gamma_- u(1-u)^2 \rho_{\mathrm{AB}} & \dfrac{2u(1-u)}{\rho_{\mathrm{AB}}(1-\rho_{\mathrm{AB}})} - 2\gamma_- u(1-u)\bar{\rho}(u) \end{pmatrix},$$

where we have introduced the shorthand notation $\bar{\rho}(u) \equiv u\rho_{\mathrm{AA}} + (1-u)\rho_{\mathrm{BB}}$. This translates into the positiveness of the first two diagonal elements plus $\det H > 0$. In other words,

$$\chi_{\mathrm{AA}} \equiv \frac{1}{\rho_{\mathrm{AA}}(1-\rho_{\mathrm{AA}})} - 2\gamma_+ u\rho_{\mathrm{AA}} > 0,$$
$$\chi_{\mathrm{BB}} \equiv \frac{1}{\rho_{\mathrm{BB}}(1-\rho_{\mathrm{BB}})} - 2\gamma_+(1-u)\rho_{\mathrm{BB}} > 0, \tag{6.31}$$

and removing trivial positive factors reduces to

$$\chi_{\mathrm{AA}}\chi_{\mathrm{BB}}\chi_{\mathrm{AB}} - 2u(1-u)\gamma_-^2 \rho_{\mathrm{AB}}^2 (\chi_{\mathrm{AA}} + \chi_{\mathrm{BB}}) > 0, \tag{6.32}$$

with

$$\chi_{\mathrm{AB}} \equiv \frac{1}{\rho_{\mathrm{AB}}(1-\rho_{\mathrm{AB}})} - \gamma_- \bar{\rho}(u). \tag{6.33}$$

Without loss of generality, we may assume $0 \leq u \leq 1/2$. With this assumption, inequalities defined by Equation 6.31 hold for any set of densities provided, so that

$$\gamma_+ < \frac{27}{8(1-u)}. \tag{6.34}$$

For any γ_+ satisfying this constraint, the critical value of γ_- is obtained as the smallest value for which inequality defined by Equation 6.32 breaks down for some set of densities. This curve is represented in Figure 6.8 for several values of u.

To validate the expression of the free energy defined by Equation 6.30, we have performed Monte Carlo simulations, using the same method as before, to calculate the fractions of the different kinds of triangles, as functions of γ_+, for a system with $N = 50$ nodes, for two values of $u = N_{\mathrm{A}}/N$ (1/2 and 2/5), and fixed values of the other parameters ($\gamma_-/N = 0.04$ and $\phi_{\mathrm{AA}} = \phi_{\mathrm{BB}} = \phi_{\mathrm{AB}} = -0.25$). Analytic expressions for those fractions of triangles are obtained from Equation 6.13. As in the case of uniform nodes, the agreement between theory and simulations, presented in Figure 6.9, suggests that the free energy expression might be exact in the thermodynamic limit.

6.5   Discussion

In this chapter, we have solved approximately Strauss's model of transitive networks using a technique specific to the statistical physics of lattice gases, the density-functional theory. The solution we found is more accurate than a standard mean-field approximation for small systems but coincides with it (and probably with the exact solution) in the thermodynamic
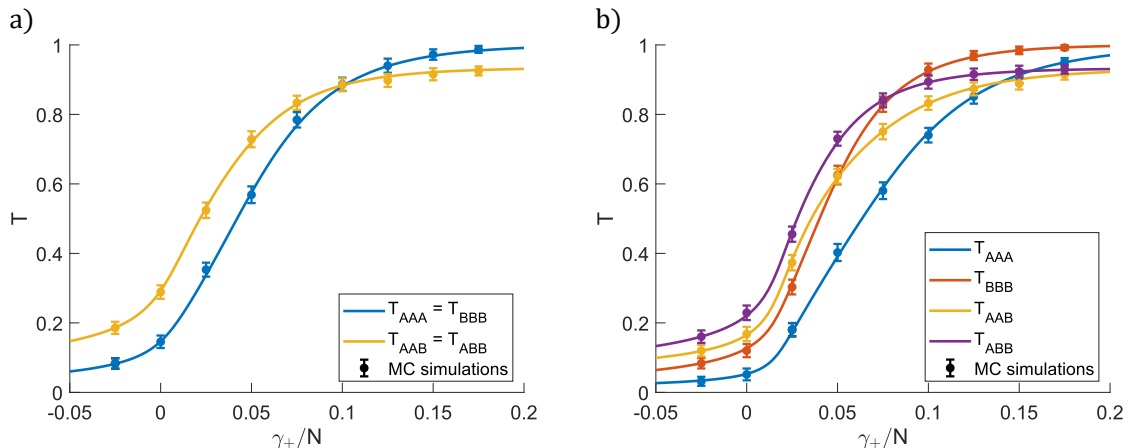
Figure 6.9: Estimation of triangles in a system with homophily - The figure represents the fractions of the different kinds of triangles vs. γ_+/N for a system with $N = 50$ nodes of two different types, A and B. The fraction of A nodes is a) $u = 1/2$ and b) $u = 2/5$. Solid lines are the curves obtained from Equation 6.13, and bullet points are the Monte Carlo results. Error bars represent the standard deviation of the number of triangles along the simulations. In both panels $\gamma_-/N = 0.04$ and $\phi_{AA} = \phi_{BB} = \phi_{AB} = -0.25$.

limit of infinitely many nodes. The model exhibits a first-order phase transition for triangle interactions above a critical threshold γ_c. For $\gamma > \gamma_c$, when the probability of link creation is increased, the system crosses a region where two solutions are possible; one with a low and one with a high fraction of links (density). Because of this fact, this model had been deemed unsuitable for producing networks with intermediate fractions of links.

The density-functional formalism we have employed reveals that the canonical ensemble (constant density and "temperature", i.e., triangle interaction) is the natural description for this system if we want to access these "forbidden" intermediate states. In this ensemble, the system behaves as a fluid undergoing a condensation transition. The two (low-density and high-density) phases are akin to a gas and a liquid, and at those intermediate densities, both phases coexist in chemical and mechanical equilibrium. A histogram of links belonging to a given number of triangles shows that links in the graph form two separate groups, each associated with one of these two phases. Hence, graphs within this coexisting region have a different structure than those outside it.

Under this interpretation, the problem of generating graphs in the studied "inaccessible" coexisting region with this model amounts to performing Monte Carlo simulations using Kawasaki dynamics, which keeps the number of links constant. The idea of accessing these intermediate states by controlling an extensive parameter had already been suggested (Miller, 2009; Newman, 2009), although the preferred control variable has always been the number of triangles.

Whether the graphs produced by Strauss's model are a suitable model for some real networks is still an open question that we will try to answer in the following chapter. It is true that the peculiar structure of these graphs has never been observed so far, but it is also true that the existence of the phase transition in Strauss's model (and its extensions) is an

unavoidable consequence of the specific interaction among its links.

The density-functional formalism we have developed can further be applied to systems where the interaction constants are link- or triangle-dependent. This way, we can study systems in which nodes have different types and interactions depending on the type of nodes involved. Homophily is one of the situations that can be so described. The analysis of the simplest example of homophilic interactions shows that homophily favours the stability of uniform networks (networks with uniform density) by increasing the value of the critical point. The predictions for this case have been validated with Monte Carlo simulations, which, as in the case of uniform networks, suggest that the free energy obtained here is exact in the thermodynamic limit. Further studies are needed to fully characterise the complex phase behaviour of a system like this.

The main contribution of this chapter is to provide a formalism that can be extended to tackle other ERG models. Its advantage with respect to more standard mean-field approaches is that it provides a systematic procedure to deal with them. Our method can be extended to other complex systems beyond network science and can provide a novel perspective and potentially useful tools for the study of many other systems. Therefore, our work provides a new way of studying network models and sheds light on their rich and complex behaviour.

7

RE-EVALUATING THE MODELS OF SOCIAL ORGANISATION

In the previous chapter, we approximately solved Strauss's model of transitive networks using the density-functional formalism of lattice gases, characterised the phase transition it exhibits, and determined how to simulate graphs in the "inaccessible" region by performing Monte Carlo simulations using Kawasaki dynamics. This is particularly interesting for our work because Strauss's model is at the heart of the models introduced in chapter 5 to analyse the social structure of school classrooms. We can relate the Hamiltonian that defines these models to that of Strauss's model using a transformation of the parameters (see section 5.5). This allows us to describe our systems entirely.

In this chapter, we re-evaluate these models considering the inclusion of the clustering term. This motivates the appearance of higher-order interactions in the Hamiltonian, which precludes the pairwise factorisation of the grand partition function, as in the pairwise models, and complicates obtaining analytical results. For this reason, we determine the maximum likelihood values of the parameters using numerical methods. Moreover, we compare these results with those obtained through Monte Carlo simulations of the system and find very similar results. This confirms the validity of our method and highlights it as an effective alternative to Monte Carlo simulations, as it is much more computationally efficient. Finally, we study the distributions of the parameters obtained for the different groups. Their form resembles a physical system in equilibrium, so we discuss the "social fluid" concept as an alternative to a network description of social systems.

7.1 Re-estimating the parameters

7.1.1 Pairwise model with clustering

In chapter 5, we introduced the pairwise model with clustering. This model allows us to fix the transitivity of the system, a property underestimated by the pairwise approximation, by fixing the number of triangles in the network with the three reciprocal links. Its Hamiltonian is defined by

$$-H(G) = \sum_{\langle ab \rangle} E_{\sigma_{ab}\sigma_{ba}} + \frac{\gamma}{N} \sum_{\langle abc \rangle} T_{abc}, \tag{7.1}$$

where T_{abc} is a term whose value is 1 if the three nodes a, b and c form a reciprocal triangle and 0 otherwise.

Furthermore, in section 5.5, we derived a transformation of the parameters that enables us to obtain a simplified expression for the grand partition function

$$\Xi = \left(1 + 2 \sum_{i=1}^{r} e^{E_{i0}} \right)^{\binom{N}{2}} \sum_{\tau} e^{-\tilde{H}(G)}, \tag{7.2}$$

where $\tilde{H}(G)$ corresponds to the Hamiltonian of Strauss's model.

Consequently, the grand potential Ω is determined by

$$\Omega = -\log \Xi = -\binom{N}{2} \log \left(1 + 2 \sum_{i=1}^{r} e^{E_{i0}} \right) + \Omega_S, \tag{7.3}$$

with Ω_S being the grand potential associated to Strauss's model.

From Equation 7.3, we can derive the mean values of the different observables

$$\langle R_{i0} \rangle = -\frac{\partial \Omega}{\partial E_{i0}} = \left[\binom{N}{2} + \frac{\partial \Omega_S}{\partial \phi} \right] \frac{2e^{E_{i0}}}{1 + 2 \sum\limits_{k=1}^{r} e^{E_{k0}}}, \tag{7.4}$$

and

$$\langle R_{ij} \rangle = -\frac{\partial \Omega}{\partial E_{ij}} = -\frac{\partial \Omega_S}{\partial \phi} \frac{(2 - \delta_{i,j})e^{E_{ij}}}{\sum\limits_{k=1}^{r} \sum\limits_{l=1}^{k} (2 - \delta_{k,l})e^{E_{kl}}}. \tag{7.5}$$

In particular, we can observe that

$$\sum_{i=1}^{r} \sum_{j=1}^{i} \langle R_{ij} \rangle = -\frac{\partial \Omega_S}{\partial \phi}. \tag{7.6}$$

It is noteworthy that the sum on the left-hand side is the total number of reciprocal links. So, combining this result with Equation 7.4, we obtain

$$\sum_{i=1}^{r} \langle R_{i0} \rangle = \left[\binom{N}{2} - \sum_{i=1}^{r} \sum_{j=1}^{i} \langle R_{ij} \rangle \right] \frac{2 \sum\limits_{k=1}^{r} e^{E_{k0}}}{1 + 2 \sum\limits_{k=1}^{r} e^{E_{k0}}}, \tag{7.7}$$

which yields

$$1 + 2 \sum_{k=1}^{r} e^{E_{k0}} = \frac{\binom{N}{2} - \sum_{i=1}^{r} \sum_{j=1}^{i} \langle R_{ij} \rangle}{\binom{N}{2} - \sum_{i=1}^{r} \sum_{j=0}^{i} \langle R_{ij} \rangle}. \tag{7.8}$$

Notice that the total number of pairs with no links can be rewritten as

$$\binom{N}{2} - \sum_{i=1}^{r} \sum_{j=0}^{i} \langle R_{ij} \rangle = \langle R_{00} \rangle. \tag{7.9}$$

Thus, substituting back in Equation 7.4, we find that

$$e^{E_{i0}} = \frac{\langle R_{i0} \rangle}{2 \langle R_{00} \rangle}, \tag{7.10}$$

an expression that immediately yields the parameters E_{i0} from empirical data as

$$E_{i0} = \log \left(\frac{\langle R_{i0} \rangle}{2 \langle R_{00} \rangle} \right). \tag{7.11}$$

Finally, using Equation 7.5 and by performing a similar derivation, we obtain

$$e^{E_{ij} - \phi} = \frac{\langle R_{ij} \rangle}{(2 - \delta_{i,j}) \langle R_{00} \rangle} \left(\frac{\langle R_{00} \rangle + \sum_{k=1}^{r} \langle R_{k0} \rangle}{\sum_{l=1}^{r} \sum_{m=1}^{l} \langle R_{lm} \rangle} \right), \tag{7.12}$$

which allows us to determine the parameters E_{ij} as

$$E_{ij} = \log \left(\frac{\langle R_{ij} \rangle}{(2 - \delta_{i,j}) \langle R_{00} \rangle} \right) + \log \left(\frac{\langle R_{00} \rangle + \sum_{k=1}^{r} \langle R_{k0} \rangle}{\sum_{l=1}^{r} \sum_{m=1}^{l} \langle R_{lm} \rangle} \right) + \phi. \tag{7.13}$$

Therefore, to find the values of the efficacies E_{ij}, we first need to use Equation 7.6 to determine the parameter ϕ. However, the value of ϕ also depends on the value of γ, which is the parameter associated with the clustering term. The issue we have is that the value of γ cannot be obtained either analytically or numerically from the expressions we have derived previously because all empirical networks are in the region of phase coexistence. Therefore, we need to find an empirical procedure to determine its value.[1] This procedure is exemplified in Figure 7.1 for the group 1 ESO A. Firstly, we need to study the observed network and find the total number of reciprocal triangles in the system and the distribution of the total

---

[1] In fact, in the simulations, what is analysed is the value of the complete clustering parameter γ/N. However, as the value of N is fixed and known for each of the groups, we can easily determine the value of γ.
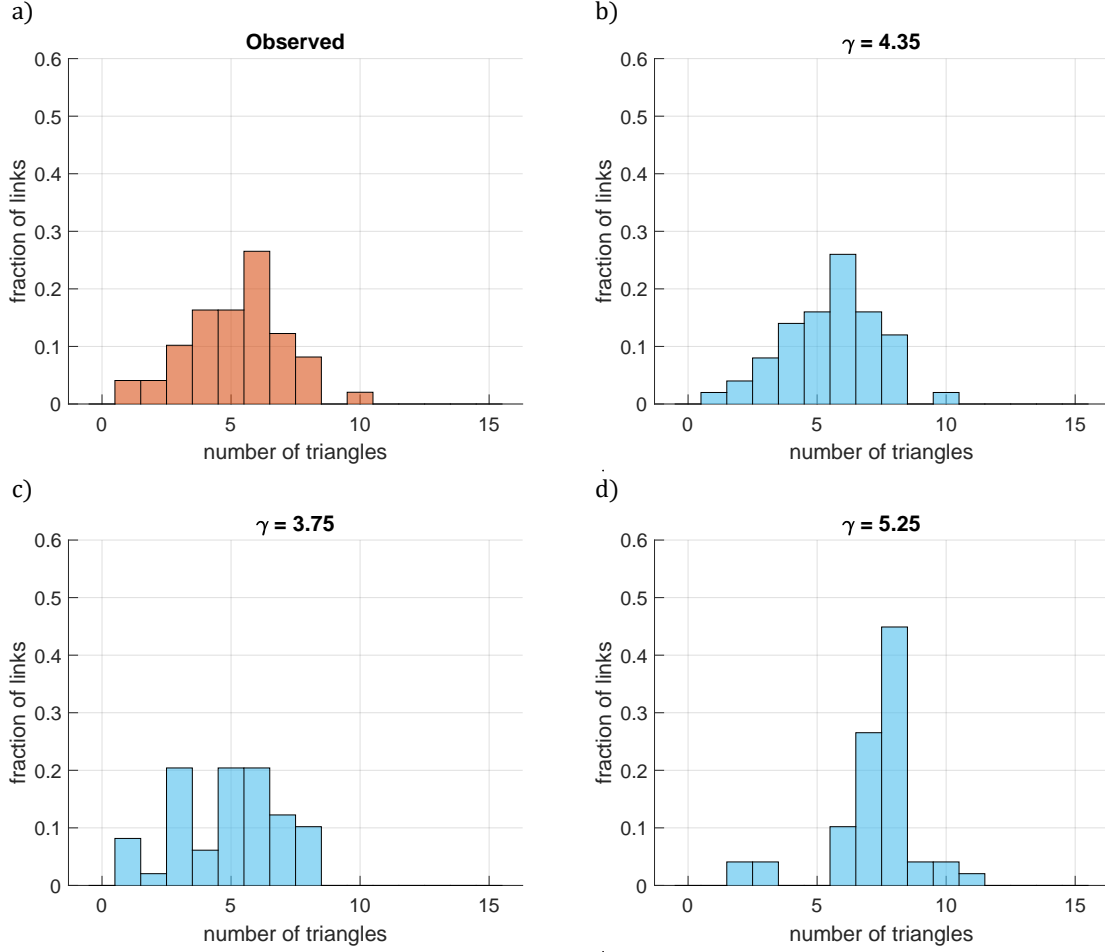
Figure 7.1: Procedure for estimating the value of the γ parameter - The figure represents the procedure for estimating the value of the γ parameter for the group 1 ESO A. Since we have neither an analytical nor a numerical procedure to estimate this parameter, we have to do it empirically. To do so, we compare the observed distribution of the number of reciprocal triangles in which each link is involved (panel a) with those obtained through Montecarlo simulations using Kawasaki dynamics. The selected value for the parameter corresponds to the one where the simulated system agrees with the observed one (panel b) both in the total number of triangles and their distribution related to the pairs of links. We can observe that the distributions vary for lower values (panel c) or higher values (panel d) of the parameter γ.

number of triangles in which each link is involved, as illustrated in Figure 7.1(a). After that, we iterate over the parameter γ and carry out Monte Carlo simulations of the system using Kawasaki dynamics (i.e. at a constant density of links ρ) until the distribution associated with the simulated system resembles the observed one, as in Figure 7.1(b). Thus, we select this value for the parameter γ. In the remaining two panels of Figure 7.1, we can check how the distributions of the total number of triangles in which each link is involved for slightly lower and higher values of the parameter γ (respectively) are different and, in consequence, so are the systems. We repeat this procedure to determine the clustering parameter γ for each of the groups in our study.

This method may seem inefficient due to the cost of iterating over different values of γ and the need to empirically check the relationship between the observed and the simulated system. But we have no other way we have of determining γ because we do not have an analytical expression relating it with the triangle distribution of Figure 7.1. Notwithstanding, the results we present in section 7.2 confirm the validity of the method. Once the value of γ is fixed, we can determine the parameters E_{ij} through Equation 7.11 and Equation 7.13.

The first analysis of the obtained efficacies E_{ij} is to compare them with those obtained using the pairwise approximation in chapter 5. The parameters associated with non-reciprocal pairs of links coincide in both models. As for those associated with reciprocal pairs of links, the first term on the right-hand side of Equation 7.13 is the result obtained for the pairwise model, and the other two terms add a correction to it.

The fact that corrections only affect the parameters associated with reciprocal pairs of links was expected since the term we have added to the Hamiltonian of the system only involves these pairs. Additionally, as we have not differentiated between the different types of reciprocal pairs of links in the calculation of the number of triangles, the correction term is the same for all of them and does not depend on the levels (i, j) that define the pair. Whether this is a sufficient correction of the model or not remains to be determined. With the statistics we have so far, we cannot decide on this use, so Occam's razor suggests keeping the model as simple as possible.

In Figure 7.2, we compare the results for the pairwise model and for the pairwise model with clustering for two groups in wave 1: 1 ESO A and 1 ESO G. These two groups have already been explored in chapter 5 and have been selected because they exhibit different densities of relationships (1 ESO A has a higher density than 1 ESO G). This is reflected in the values of the parameters, as they increase with the probability that the associated pair of links appears in the system. In the results for both groups, we can observe what we mentioned in the previous paragraph: the parameters associated with non-reciprocal pairs of links do not vary between the two models. Regarding the relevance of the density of links in the parameters of the pairwise model with clustering compared to the original model, we can see that the corrections are smaller the lower the density of links, in agreement with the fact that a lower number of links induces fewer triangles. The figures representing the comparison of the parameters for each of the remaining groups in wave 1 can be found in subsection F.1.1.

Moreover, the values of the parameters for the pairwise model with clustering associated with reciprocal pairs of links are lower for both groups. The reason behind this phenomenon is that the term associated with triangles with the three reciprocal links enhances the presence of these links to satisfy transitivity. Thus, smaller values of these parameters are enough to induce the same number of reciprocal links.

Therefore, the pairwise model with clustering provides a more accurate representation of the system, as it accounts for transitivity by taking into account the presence of triangles with three reciprocal pairs of links. This model better captures the structural properties of real-world networks. However, once more (as in the pairwise model), this model does not account for the factors that shape the efficacies, only for the behaviour of the system.
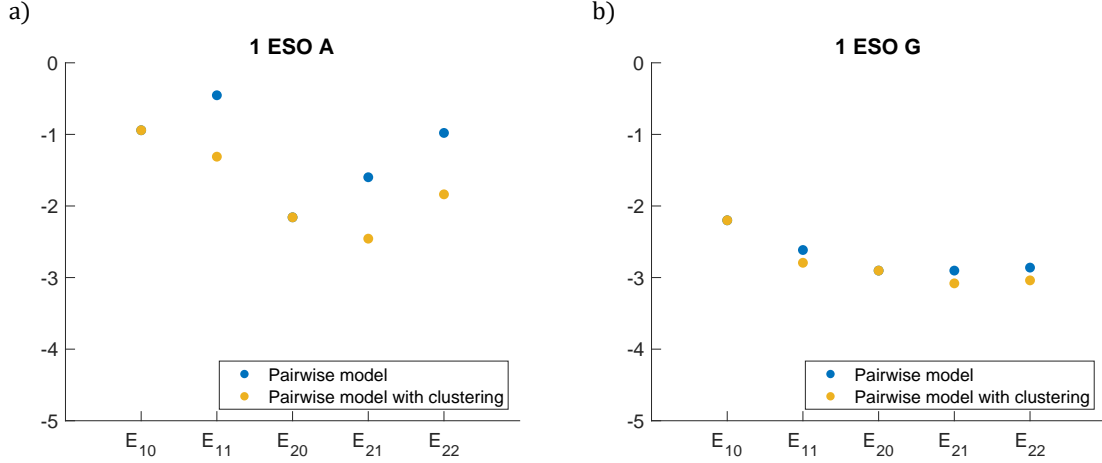
Figure 7.2: Comparison between the parameters for the pairwise model and the pairwise model with clustering - The figure represents the comparison of the values for the maximum likelihood parameters in the pairwise model and the pairwise model with clustering for the following groups in wave 1: a) 1 ESO A and b) 1 ESO G. These two groups have been selected because they exhibit different densities of relationships: 1 ESO A has a higher density, while 1 ESO G has a lower density. This is evidenced by the values of efficacies E_{ij}, which are larger in the former case. It can be observed in the figure that the values of the efficacies associated with non-reciprocal pairs of links do not vary between both models, while those associated with reciprocal pairs of links do change. This variation is directly proportional to the link density. The figures representing the comparison of the parameters for each of the remaining groups in wave 1 can be found in subsection F.1.1.

Therefore, we need to analyse the linear model with reciprocity and clustering to check if it fits the parameters in a way that allows us to explain these factors, just as with the pairwise model.

7.1.2 Linear model with reciprocity and clustering

We can perform an analogous derivation to estimate the parameters of the linear model with reciprocity and clustering. This model is defined by the Hamiltonian

$$-H(G) = \sum_{\langle ab \rangle} \left[\lambda(2 - \delta_{\sigma_{ab},0} - \delta_{\sigma_{ba},0}) + \mu(\sigma_{ab} + \sigma_{ba}) + \beta(1 - \delta_{\sigma_{ab},0})(1 - \delta_{\sigma_{ba},0}) \right] + \frac{\gamma}{N} \sum_{\langle abc \rangle} T_{abc}.$$

In this expression, the first term corresponds to the Hamiltonian associated with the linear model with reciprocity, and the second term enhances transitivity by fixing the average number of triangles with the three reciprocal pairs of links.

In subsection D.1.3, we introduced a transformation of the parameters that allows us to rewrite the grand partition function of the system as

$$\Xi = \left(1 + 2e^{\lambda + \mu} \frac{1 - e^{r\mu}}{1 - e^{\mu}} \right)^{\binom{N}{2}} \sum_{\tau} e^{-\tilde{H}(G)}, \tag{7.14}$$

where $\tilde{H}(G)$ is the Hamiltonian of Strauss's model.

Therefore, the grand potential Ω is determined by

$$\Omega = -\log \Xi = -\binom{N}{2}\log Q + \Omega_S, \tag{7.15}$$

where Ω_S is the grand potential of Strauss's model, and we have introduced the short-hand Q, which will be used throughout this section to refer to the term

$$Q \equiv 1 + 2e^{\lambda+\mu}\frac{1-e^{r\mu}}{1-e^{\mu}}. \tag{7.16}$$

From Equation 7.15, we can derive the average values of the different observables by calculating the partial derivatives with respect to the parameters. First, we start with the number of reciprocal pairs of links, so

$$\langle R \rangle = -\frac{\partial \Omega}{\partial \beta} = \sum_{\langle ab \rangle}\langle(1-\delta_{\sigma_{ab},0})(1-\delta_{\sigma_{ba},0})\rangle = -\frac{\partial \Omega_s}{\partial \phi}. \tag{7.17}$$

Similarly, we can calculate the mean value associated with the total number of relationships in the network. That is

$$\langle L \rangle = -\frac{\partial \Omega}{\partial \lambda} = \sum_{\langle ab \rangle}\langle 2 - \delta_{\sigma_{ab},0} - \delta_{\sigma_{ba},0}\rangle = \binom{N}{2}\frac{\partial \log Q}{\partial \lambda} - \frac{\partial \Omega_s}{\partial \phi}\frac{\partial \phi}{\partial \lambda}, \tag{7.18}$$

where

$$\frac{\partial \log Q}{\partial \lambda} = 1 - \frac{1}{Q}, \tag{7.19}$$

and

$$\frac{\partial \phi}{\partial \lambda} = 1 + \frac{1}{Q}. \tag{7.20}$$

Finally, we repeat the process for the term associated with the varying costs of relationships depending on their emotional intensity:

$$\langle S \rangle = -\frac{\partial \Omega}{\partial \mu} = \sum_{\langle ab \rangle}\langle \sigma_{ab} + \sigma_{ba}\rangle = \binom{N}{2}\frac{\partial \log Q}{\partial \mu} - \frac{\partial \Omega_s}{\partial \phi}\frac{\partial \phi}{\partial \mu}, \tag{7.21}$$

where

$$\frac{\partial \log(T)}{\partial \mu} = \frac{e^{\mu}\left[1 + (r-1)e^{r\mu} - re^{(r-1)\mu}\right]}{(1-e^{\mu})^3}\frac{\left(2e^{\lambda+\mu}\right)\left(1-e^{\mu}\right)}{Q}, \tag{7.22}$$

and

$$\frac{\partial \phi}{\partial \mu} = 2 - \frac{\partial \log Q}{\partial \mu} - \frac{2re^{r\mu}}{1-e^{r\mu}} + \frac{2e^{\mu}}{1-e^{\mu}}. \tag{7.23}$$

From the expressions above, we can determine the parameter values for a specific dataset. Although the equations are too complex to obtain analytical results, as in the previous cases, we solve them numerically. In this way, we obtain the values of the three parameters that define the efficacies and compare them with those of previous models. Again, to do so, we
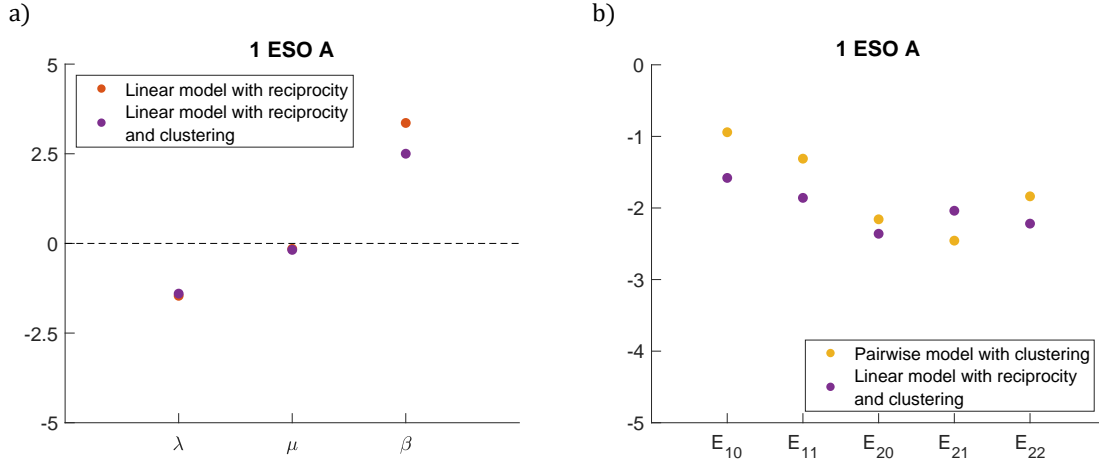
Figure 7.3: Comparison between the parameters associated with the linear model with reciprocity and clustering and other previously studied models - The figure represents the values for the maximum likelihood parameters and the comparison between the linear model with reciprocity and clustering and other previously studied models for the group 1 ESO A in wave 1: a) Linear model with reciprocity and b) Pairwise model with clustering. The figures representing the comparison of the parameters for each of the remaining groups in wave 1 can be found in subsection F.1.2.

need to estimate first the value of the parameter γ using the empirical method introduced in subsection 7.1.1.

In Figure 7.3(a), we compare the values of the parameters obtained by solving the previous non-linear system of equations with the analytical values obtained in chapter 5 for the linear model with reciprocity for the group 1 ESO A in wave 1. The values of the parameters λ, associated with the total number of relationships, and the parameters μ, associated with the variable cost of each of them, nearly coincide. These differences, being so slight, are likely due more to numerical errors than to real differences in the models. However, the value of the parameter β varies considerably. Moreover, if we compare the results for all groups (see subsection F.1.2) we can observe that the higher the number of links, the more significant the difference between the values of the parameters β associated with each of the models. This makes sense because β is the parameter controlling for reciprocity, which, to a larger extent, is already taken care of by the clustering parameter γ.

These results are consistent with those obtained for the pairwise model with clustering in subsection 7.1.1 and agree with what we could expect before evaluating the model. The reason behind this is that, when introducing the term related to triangles with the three pairs of reciprocal links into the Hamiltonian of the system, the total number of relationships nor their associated cost vary. Therefore, the values of their associated parameters do not change. What does change is the way in which each pair of links is distributed within the system. This is reflected in the value of the parameter β. As already explained, this parameter is always lower in the model with clustering than in the former model because the clustering term already accounts for reciprocal links. As a result, the number of reciprocal links is similar in both models, but their structure is different, closer to that of networks observed in

the real world.

Another interesting question concerns the comparison of parameters between the two models with the clustering term included, to verify whether the form of the efficacies can still be explained by the two parameters derived from the concept of "social atom" plus the reciprocity term. The results presented in Figure 7.3(b) and Figure 5.3 for the group 1 ESO A in wave 1 show that the fits are very similar. It can be observed in subsection F.1.2 that the same holds for all the remaining groups analysed. In view of this, the conclusions drawn in chapter 5 for the models without the clustering term are also valid when this term is included. This suggests that including the clustering term does not significantly affect the overall conclusions and relationships between the parameters. Adding the clustering term helps capture more nuanced aspects of the system, allowing for a better understanding of the underlying social structures and behaviours. However, the core relationships and findings from chapter 5 remain valid and provide a solid foundation for analysing and modelling these social networks.

7.2 Bayesian Monte Carlo estimation of the parameters

In the previous section, we discussed how the main drawback of models incorporating the clustering term in the Hamiltonian of the system is that an exact analytical calculation of the grand partition function becomes intractable as the size increases. This is because it no longer factorises in pairs because the clustering term introduces higher-order interactions in the system. Analytical calculations are thus more complex and challenging to perform. For this reason, and unlike what happened in the models presented in chapter 5, we can only derive (approximate) analytical expressions for the maximum likelihood values of the parameters but not for their confidence intervals. Therefore, to get an idea of the validity of our method to reproduce the empirical results, we make use of Bayesian methods in order to estimate the posterior distribution of the parameters, given the empirical networks of relationships.

In particular, we have implemented a Markov Chain Monte Carlo (MCMC) technique, which applies the Metropolis-Hastings sampling method through the Python package `PyMC`. This technique amounts to constructing a Markov chain in the set of parameters. The chain is designed so that, after a sufficiently large number of iterations, the distribution of the variables converges to the target posterior distribution. Once the chain has converged, samples from the chain can be used as random samples of the posterior distribution of the model parameters. This approach enables us to estimate the values of the parameters and their uncertainties. In our simulations, we use a total of 5×10^5 Monte Carlo steps, each consisting of a small random change of one of the parameters, also chosen randomly. To start the method, we use the values of the parameters estimated from analytical expressions as explained in the previous section and "burn-in" the first 3×10^4 iterations to let the Markov chain reach the stationary state. This initial condition accelerates the convergence of the method. We have initiated the chain with a random set of parameters, and the results obtained are the same, but the convergence time is longer.

In Figure 7.4, we show the comparison between the values of the parameters obtained
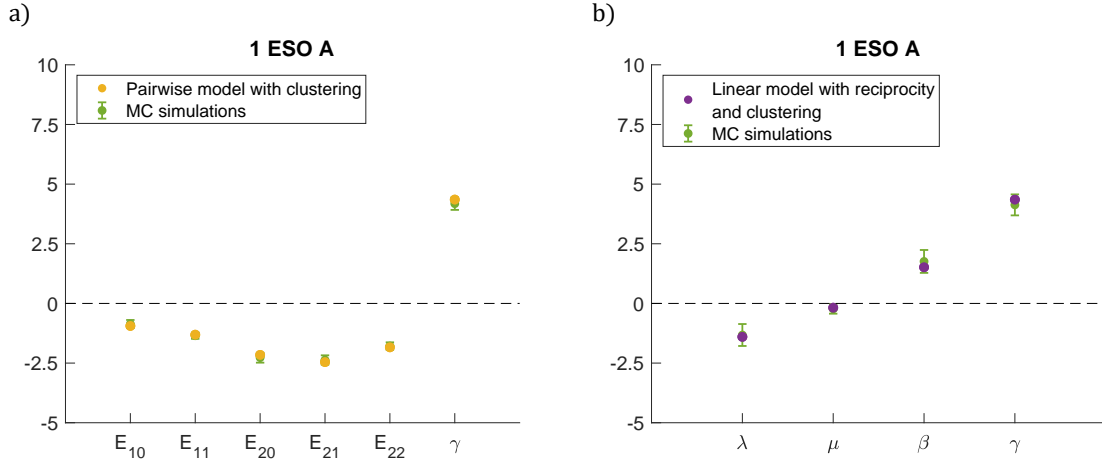
a)

b)

Figure 7.4: Comparison between the parameter values and Monte Carlo simulations for the models considering clustering - The figure represents the values for the maximum likelihood parameters and its comparison with Monte Carlo simulations for the group 1 ESO A in wave 1 for the following models: a) Pairwise model with clustering and b) Linear model with reciprocity and clustering. The figures representing the comparison of the parameters for each of the remaining groups in wave 1 can be found in subsection F.1.3 (for the pairwise model with clustering) and subsection F.1.4 (for the linear model with reciprocity and clustering).

using our method for the pairwise model with clustering and the linear model with reciprocity and clustering and Monte Carlo simulations of the system for the group 1 ESO A in wave 1. For the remaining groups in wave 1, the results can be found in subsection F.1.3 and subsection F.1.4 for each of the models, respectively. We can observe that the values obtained using our method and the MCMC simulations approach each other quite satisfactorily for all parameters in both models. Both methods are capable of capturing important aspects of the system, such as the difference between the various types of links, the variable costs of each relationship based on its intensity, and the importance of reciprocity. Notice, in particular, the good agreement of the parameter γ, which was obtained through an empirical procedure.

This convergence between the approximate analytical results and those obtained with Monte Carlo simulations demonstrates the reliability and robustness of the models, which confirms their ability to represent the underlying social structure effectively. The agreement between the two methods reinforces the validity of the models and their parameters, providing confidence in the insights derived from their analysis. This result has two main implications. The first one is model validation because we can overcome the limitations posed by the intractability of the analytical calculations by using MCMC techniques and obtain a more comprehensive understanding of the performance of the model. The second is methodological versatility, as it is important to note that MCMC sampling can be computationally expensive, especially for complex models and large datasets. The fact that the analytical method and the Monte Carlo simulations yield similar results confirms the versatility of the methodologies employed in studying social networks. Therefore, in many cases, it may be beneficial to use the analytical method proposed in this chapter, as it is much more computationally efficient and returns similar results.

7.3 The "social fluid"

7.3.1 Analysing the distribution of the parameters

The maximum likelihood values of the parameters associated with a system entirely define the form of its Hamiltonian and, consequently, characterise the structure of the system. These values indicate the probability of their associated relationships appearing among individuals. By comparing their values, we can determine similarities and differences in the macroscopic structure of the different groups. This comparison enables us to identify patterns or trends across the groups, which may lead to insights into the general structure of social networks.

We begin the analysis by exploring each of the groups from the five waves in the study to obtain the most complete and comprehensive statistics of the distribution of the parameters. In this way, we obtain results for 80 different groups. All these groups have a similar size, with values ranging from 20 to 30 people. Figure 7.5 shows the distribution of the parameters associated with the pairwise model with clustering across all 80 groups. The distributions associated with each of the parameters vary. They all have a form centred around a value, which is different for each parameter. This indicates that some pairs of relationships are more likely to be observed within the system while others appear less frequently - recall that the probability that a link appears in the system increases with the value of its associated parameter. Our results indicate that the most common relationships, in general, are the reciprocal ones with a low level of intensity in both directions of the pair (parameter E_{11}), while the least likely are non-reciprocal relationships with a very intense link (parameter E_{20}). These results confirm the differences between the different levels of relationship intensity, providing further evidence of the layered structure of social networks.

Figure 7.6 shows the distribution of the parameters associated with the linear model with reciprocity and clustering. Similar to the distributions presented in Figure 7.5, the values of the parameters in this model are distributed around a central value, with the standard deviation in the distribution being even smaller in this case, except for γ. Furthermore, the distribution of the parameter λ, associated with the constraint on the total number of relationships that can be maintained simultaneously, shows a very pronounced peak. This confirms that the limit on the number of relationships is similar across all groups. When studying the cost associated with relationships of different intensities, we observe that the parameter μ always has a negative value, indicating that the cost of maintaining a relationship is larger when the emotional level of the relationship increases. Therefore, these results obtained for the analysis of the system's structure at the macroscopic level are in clear agreement with those presented by Tamarit et al. (2018) when analysing each ego-network individually, that is, at the microscopic level. Finally, the distribution of the parameter β, associated with the reciprocity of the system, clearly correlates with the distribution of the density of links ρ (see subsection F.1.5). The parameter associated with reciprocity is always positive and grows with the density, so the probability of reciprocal link pairs appearing in the system increases with ρ.
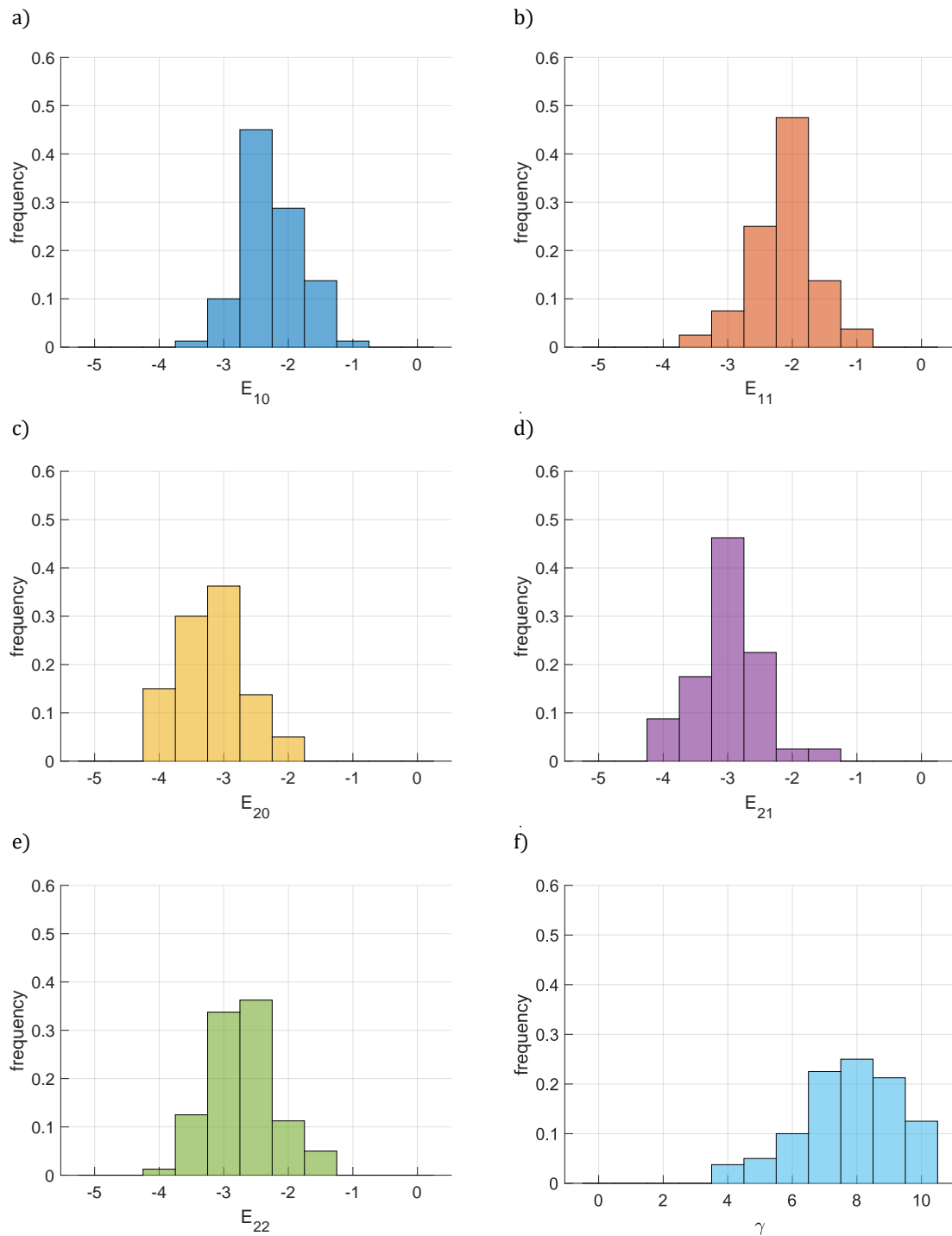
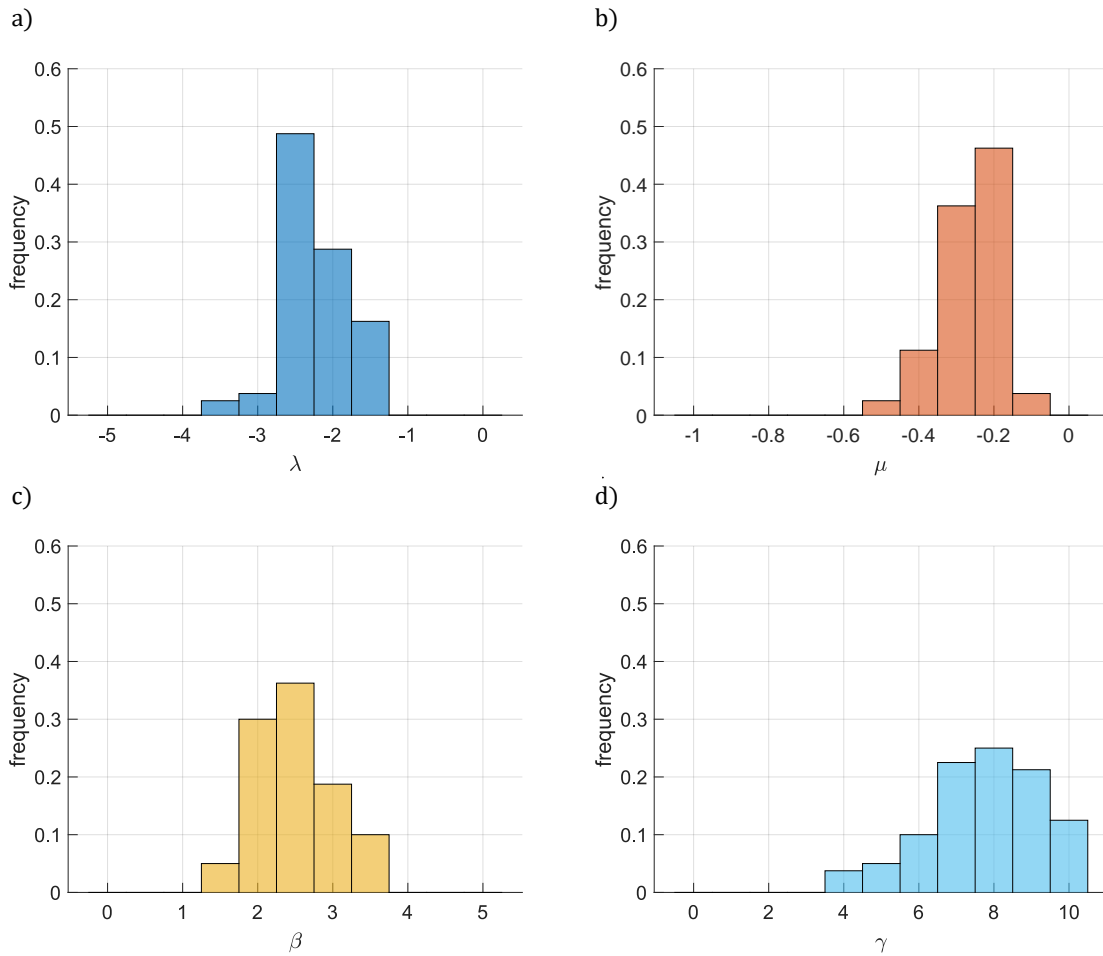**Figure 7.5: Distribution of the parameters associated with the pairwise model with clustering** - The figure represents the distribution of the values associated with the pairwise model with clustering for the maximum likelihood parameters for all the groups in the five waves. Each panel corresponds to the distribution of a different parameter: a) E_{10}, b) E_{11}, c) E_{20}, d) E_{21}, e) E_{22} and f) γ. All the parameters show a distribution with a peak around a different value, which highlights the differences between the various types of pairs of links, their variable intensities and the probability of finding the associated relationship within the system.

Figure 7.6: Distribution of the parameters associated with the linear model with reciprocity and clustering - The figure represents the distribution of the values associated with the linear model with reciprocity and clustering for the maximum likelihood parameters for all the groups in the five waves. Each panel corresponds to the distribution of a different parameter: a) λ, b) μ, c) β and d) γ.

It is also important to note that both models show the same distribution of the parameter γ associated with the clustering term, as it is obtained empirically by comparing the observed and simulated distribution of triangles for each group, which does not vary between models. The values of this parameter are all within a similar range. This parameter presents a positive value in all the groups, which indicates that it appears more significant clustering than what is obtained when randomly distributing all pairs of links in the system. Moreover, this value is higher in those groups where the difference between the observed number of triangles and the expected one is larger.

7.3.2 An alternative representation to networks

Networks have traditionally been the tools used to sketch social relationships because they allow us to describe individuals (nodes) and relationships (edges) that appear between them.

Furthermore, the fact that in networks links can be of different types, such as directed, weighted or signed, enables us to represent all these relationships more accurately (if they are reciprocal or non-reciprocal, friendships or enmities, with varying intensity...). In recent years, techniques have even emerged that allow for studying networks with multiple layers and their temporal evolution. These approaches provide a deeper understanding of the complexities and dynamics of social relationships, as they account for changes over time and the multifaceted nature of relationships between individuals. However, the techniques we have used in this thesis to analyse social structures differ significantly from the traditional over widely used in networks and are closer to those employed to analyse physical systems, such as fluids.

Tamarit (2019) does a preliminary discussion in this direction in his PhD thesis. He argues that liquids or gases can also be studied using network theory. At a specific time stamp, a snapshot of the system can be taken to determine the positions of the different particles within it and identify the edges involving these particles based on the distances between them. This snapshot provides a static view of the relationships and interactions at that particular moment, offering insights into the structure, connectivity and relative positioning of the particles in the system. We can gain insights into the system that help us understand how the particles behave within it by repeating this process multiple times. However, we know that these techniques will always limit our understanding of the system and that statistical mechanics provides a much more comprehensive and practical theoretical foundation for describing such systems. It allows us to analyse the overall behaviour and properties of those systems by considering the collective dynamics of particles rather than focusing on individual snapshots, and takes into account the probability distributions of particle states and interactions, leading to a deeper understanding of the system's behaviour, its equilibrium properties or its response to external influences.

We believe that social systems could be treated as a similar case. Although networks may be a valuable tool for understanding their structure, as evidence to date, they are insufficient to capture the full complex and intricate nature of social systems. As an alternative, we propose what we call the "social fluid", which consists of ensembles of "social atoms" that interact with each other, much like physical particles do in a fluid. The results discussed in the subsection 7.3.1 clearly reinforce this idea. The fact that the distributions of the parameters exhibit that characteristic form around a central value, with a small standard deviation, suggests that all the networks whose data we have compiled are just realisations of an interacting system at "thermal" equilibrium. Despite the differences and variations in social relationships at the microscopic level among different constituents, the general structure is maintained at the macroscopic level. Again, this is reminiscent of what we find when studying the interaction of particles in a gas or a fluid.

This similarity suggests that the "social fluid" approach could be a powerful way to analyse and understand social systems. This approach offers a novel perspective on social relationships by leveraging concepts and methodologies from the domain of physics to gain insights into the underlying structure and dynamics of social systems. In this way, we can better capture the emergent properties and equilibrium states that arise from the complex interplay of individual relationships. This perspective may open up new avenues for research and provide a more comprehensive framework for studying social systems, their properties and their dynamics.

7.4 Discussion

In this chapter, we have re-evaluated models for the analysis of social structure when a term associated with clustering is introduced into the Hamiltonian, allowing us to fix the number of triangles with all reciprocal links that appear in the system. Higher-order interactions that involve more than two individuals emerge when adding this term. This precludes the pairwise factorisation of the grand partition function and complicates the calculation of analytical expressions for the parameters and their confidence intervals.

First, we study the pairwise model with clustering, which better captures the structural properties of real-world networks compared to the model without the clustering term. The parameters associated with non-reciprocal pairs of links remain unaltered between the two models, while a correction term appears for those associated with reciprocal pairs of links. Regarding the differences between groups, the variations in values of the parameters are related to the density of relationships. The groups with a lower density of relationships also exhibit smaller changes in the parameters. Moreover, the values of the parameters for the pairwise model with clustering associated with reciprocal pairs of links are found to be lower for all the groups than in the former model. This is because the term associated with triangles with three reciprocal links enhances the presence of these links to satisfy transitivity. In this way, we manage to maintain the density and increase transitivity, making the probability distribution of the system more accurately represent reality. However, once again, this model does not provide us with any information about the factors that shape the parameters. For this reason, we study the linear model with reciprocity and clustering, in order to fill this gap.

The analysis of the parameters of the linear model with reciprocity and clustering provides us with results in two directions. On the one hand, by comparing the values of the parameters obtained with those associated with the model without the clustering term, we observe that the values of the parameters λ, associated with the total number of relationships maintained, and μ, associated with the variable cost based on intensity, barely change between both models. The only parameter whose value differs is the parameter β, associated with the reciprocity of relationships. The variation is correlated with the density of pairs of links in the system. This result suggests that the addition of the clustering term helps to capture more complete aspects of the system, leading to a more accurate representation of real-world networks. This term encourages the presence of reciprocal pairs of links, so the value of the parameter associated with the reciprocity must be lower in order to preserve the system's density. On the other hand, comparing the results with those of the pairwise model with clustering, we observe that the adjustments appear to be very similar to the ones without the clustering term, further supporting the idea that including this term does not significantly affect the overall conclusions and relationships between the parameters. Therefore, we can still explain the form of the parameters using only three factors: the two aspects derived from the "social atom" concept and reciprocity.

Moreover, to confirm the validity of the insights from our models, we have conducted Monte Carlo simulations to compare their results. We have used the Metropolis-Hastings sampling method because it is a powerful technique for exploring complex parameter spaces and estimating their posterior distributions when analytical solutions are not feasible. This

technique helps us to validate the model, allowing for a more robust comparison of results and a better understanding of the underlying social network structure. The agreement between our method and the simulations reinforces the validity of the models and their associated parameters, providing confidence in the insights derived from their analysis. This has two primary implications. The first implication is model validation. Bayesian methods overcome the limitations posed by the intractability of the analytical calculations and provide robust parameter estimations and uncertainty quantification. The second implication corresponds with methodological versatility, as the analytical method and the Monte Carlo simulations yield similar results, while the former is computationally more efficient. Therefore, the use of our method can be beneficial in studying social networks because it offers a balance between computational efficiency and accuracy in capturing the complexities of the system.

Finally, we have questioned networks as the most appropriate method for schematising and understanding social relationships. As an alternative, we propose the "social fluid". This approach, inspired by the concepts of statistical mechanics, allows for a more comprehensive examination of the underlying structure of social systems. We can better understand the emergent properties of social systems by considering them as ensembles of "social atoms" interacting. This includes their collective behaviour, equilibrium states and responses to external stimuli or changes. This perspective may provide deeper insights and help to develop more effective models for studying social systems. Therefore, we firmly believe that the "social fluid" approach offers a promising direction for future research in social network analysis by borrowing concepts and methodologies from statistical mechanics.

PART III

BEYOND HUMANS: APPLICATION TO PRIMATES

8

THE COMPLEXITIES OF SOCIAL INTERACTIONS IN CHIMPANZEES

Human and non-human primates exhibit complex relational structures that govern their social interactions. Understanding these structures is essential for comprehending the evolutionary underpinnings of social behaviour and cognition in non-human primates. In this chapter, we pretend to gain an understanding of this topic by extrapolating the continuous model that explains the layered structure of the ego-networks in humans to chimpanzees. With this in mind, we explore the world of chimpanzee social networks. Our objective is to uncover the underlying organisational principles that guide the formation and maintenance of social bonds in chimpanzees. We will analyse whether this structure is consistent with that of humans, given the inherently limited resources of cognition and time might apply to both species alike.

This research is based on meticulous observations of grooming behaviour among four different groups of chimpanzees residing in the Chimfunshi Wildlife Orphanage in Zambia. Grooming, a characteristic social activity for chimpanzees, serves as a reliable proxy for social bonding and allows us to gain insight into their relationships. Through this analysis, we will not only illuminate the remarkable similarities between the social organisation of chimpanzees and humans but also deepen our understanding of the factors that have shaped the social relationships among these primates.

8.1 Data description

8.1.1 Environment description

The Chimfunshi Wildlife Orphanage provides a naturalistic habitat for chimpanzees with large forested enclosures ranging from 20 to 77 hectares of grasslands and forests, which are similar to the natural habitat of wild chimpanzees (Ron & McGrew, 1988). The enclosures provide ample space for the chimpanzees to engage in species-typical behaviours, including natural fission-fusion dynamics (van Leeuwen et al., 2019). The chimpanzees are divided into four groups of different sizes and live in separate enclosures with no possibility even of inter-group visual encounters, except for a small section between Groups 3 and 4. Although they cannot see each other, the groups live within hearing distance of each other, covering a distance of about 3km between environments. The four groups comprise wild-born and Chimfunshi-born chimpanzees from various phylogenetic and geographic backgrounds, with multiple subspecies. The chimpanzees engage in natural foraging behaviour, primarily on fruiting trees, but also consume insects and small mammals present in their enclosures. They also receive two daily feedings of fruits and vegetables to supplement their diet. Moreover, they construct their own nests in the woodland enclosures, where they sleep at night (for further details, see van Leeuwen et al. (2018) and van Leeuwen et al. (2022)).
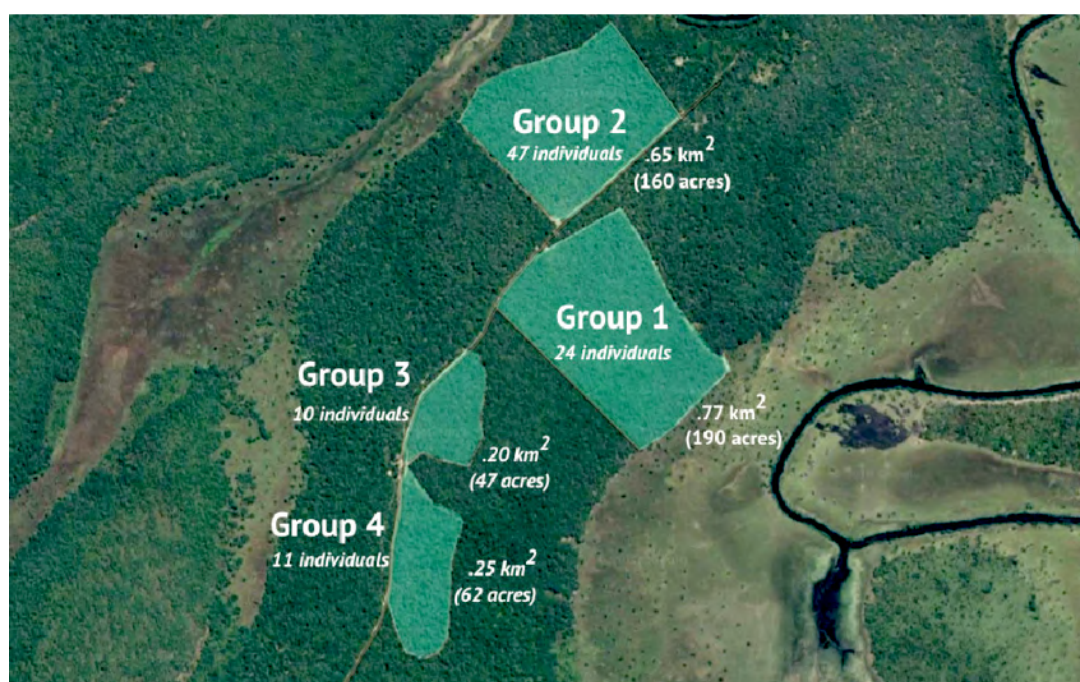

Figure 8.1: Aerial view of the Chimfunshi Wildlife Orphanage - The figure represents the aerial view of the Chimfunshi Wildlife Orphanage in Zambia. The habitat is divided into four different groups with no possibility even of inter-group visual encounters, except for a small section between Groups 3 and 4. The data presented in our study shows slight variations in the number of chimpanzees living in each group (see subsection 8.1.3) compared to the information provided in the figure, which is sourced from van Leeuwen et al. (2022).

8.1.2 Data collection

The grooming data was collected as part of a larger and ongoing data-collection effort at Chimfunshi to assess chimpanzee sociality over time (van Leeuwen et al., 2018). Trained staff members conduct focal follow daily with an every-minute scan sampling technique in the ZooMonitor application (Lincoln Park Zoo, 2020). This is a protocol implemented and maintained by our collaborators Katherine Cronin, Daniel Haun and Edwin van Leeuwen since 2015. The protocol comprises 10min focal follows in which ten scan points are scored. On each scan, all instances of proximity (<1m), grooming, social play, and aggression by the focal individual are scored, including the identities of the interaction partners. Data were semi-randomly collected from the fence line, restricted by visibility. Our collaborators work in a sanctuary where chimpanzees have ample space to retreat into the forest. As per sanctuary stipulations, they do not enter their enclosures ever, which prevents them from following the chimpanzees into the forest. Hence, the next best thing is to divide the fence line into different sections and start the observations randomly from these different sections, also randomising the direction (clockwise vs. counter-clockwise) in which the search for chimpanzees commences (Cronin et al., 2014; van Leeuwen et al., 2012). Upon encountering a chimpanzee within eye-sight, behavioural observations on the respective individual begin, using established focal follow protocols. After finishing the respective focal follow, the nearest chimpanzee is searched to start the subsequent focal follow. Overall, if the focal follow lasted 5 minutes or less (i.e., due to visibility challenges), the focal follow is discarded. The observation efforts start each day at a different location, upon which the first-seen chimpanzee is chosen as the focal. The observation efforts were distributed across the day: typically, for each of the groups, one hour was collected between 7:00 am-11:00 am and one hour between 2:30 pm and 5:00 pm, after which the chimpanzees retreat into the forest to spend their nights there. All individuals were sampled except for dependent offspring clinging to their mothers.

8.1.3 Data curation

Our study focuses on grooming behaviour due to the well-established connection between grooming and the quality of dyadic relationships (Watts, 2002). Furthermore, we cleaned the grooming data, collected over a period of four years (from 2015 to 2019), to be more consistent and reliable. To analyse this data, we considered a grooming interaction to have occurred when an individual was observed grooming another during the 10-minute focal follow, regardless of whether the grooming action occurred throughout the entire period or only a fraction. This approach enabled us to standardise the criteria for identifying a grooming interaction and reduce data uncertainty, as grooming bouts often start or end outside the focal observation period. As an alternative approach, we also assigned a weight to each grooming bout based on the number of minutes it occurred within the observation window. Data analysis using this criterion yields results in excellent agreement with the previous one presented here, so we retained the first procedure as it is standard in the field.

We are studying four different groups of chimpanzees (see van Leeuwen et al. (2018) for details), with varying population sizes (groups 1-4, n=$26, 60, 11, 14$, respectively). The analysis is individual for each chimpanzee. For this reason, before constructing the networks, we filtered chimpanzees that had groomed less than five individuals in the study period, although we still considered grooming actions directed toward them. This condition does not

influence the conclusions and is based on the fact that five is the size of the core of grooming ego networks in primates (Dunbar et al., 2018). Having less than five other individuals would lead to large errors in the fitting procedure, making the analysis results meaningless for that specific chimpanzee. Most of the individuals filtered out due to this criterion have been observed less than 20% of the mean number of observations obtained for the population (~ 300 times). This criterion also excluded a few others with more observations, typically immature individuals who groom very little and are still very dependent on the actions of their mothers.

Using this approach, we selected for the analysis only chimpanzees who were aged nine years or older at the end of the observation period, excluding infants and individuals who died between 2015 and 2019. Consequently, after curating the data, the number of individuals included in the analysis is reduced (groups 1-4, n=21, 32, 10, 10, respectively). Some detailed information regarding each chimpanzee's demographics is provided in subsection G.1.1. This restriction homogenises the population of chimpanzees studied and allows us to extrapolate the results to the case of adults.

8.2 The model

8.2.1 Theoretical background

To provide some context for the analysis of the ego-networks's structure in chimpanzees presented below, it is helpful to summarise the main findings of the theoretical approach to the continuous layered structure presented by Tamarit et al. (2022). This mathematical model is the formalism we use here to study the social structure of chimpanzees.

In the discrete case, L is defined as the total number of relationships in an ego-network and σ is the average cognitive cost of a relationship. Relationships belong to r different categories, each of them bearing a different cost $s_{\max} = s_1 > s_2 > \cdots > s_r = s_{\min}$. As described in detail by Tamarit et al. (2018), using a maximum entropy approach, it is possible to obtain the probability that a given relationship of the ego-network belongs to category k as

$$p_k = Z_r^{-1} e^{-\hat{\mu} s_k}, \qquad Z_r = \sum_{k=1}^{r} e^{-\hat{\mu} s_k}, \tag{8.1}$$

where $\hat{\mu}$ is fixed by letting σ be the expected cost $\sigma = \mathbb{E}(s_k)$. Assuming a linear distribution of costs $s_k = (s_{max} - s_{min})\dfrac{k-1}{r-1}$, with $k = 1, ..., r$, we obtain

$$t \equiv \frac{s_{max} - \sigma}{s_{max} - s_{min}} = e^{\mu} \frac{(r-1)e^{r\mu} - re^{(r-1)\mu} + 1}{(r-1)(e^{r\mu} - 1)(e^{\mu} - 1)}. \tag{8.2}$$

Using this probability distribution, we can calculate χ_k, which is the expected number of relationships with costs larger than or equal to that of category k (i.e., the size of the social circles, with $k = 1$ corresponding to the innermost one), as

$$\chi_k = \frac{e^{k\mu} - 1}{e^{r\mu} - 1}, \tag{8.3}$$

where $\mu \equiv \hat{\mu}(s_{max} - s_{min})/(r-1)$. It can subsequently be shown that, for large values of μ, the scaling ratio, i.e., the size of one circle divided by the previous one, behaves approximately as

$$\frac{\chi_{k+1}}{\chi_k} \sim \begin{cases} e^\mu, & \mu \to \infty, \\ 1, & \mu \to -\infty. \end{cases} \tag{8.4}$$

As discussed by Tamarit et al. (2018), this result predicts the known regime for values of $\mu > 0$, in which the circles satisfy an approximate scaling relation; in particular, for $\mu \approx 1$, the expected value of 3 found on empirical data by Dunbar et al. (2015) is recovered. On the other hand, it also predicts a so-called "inverse" regime when $\mu < 0$, in which most of the relationships are in the closest circle. This second behaviour had not been described before this publication, where it was confirmed using data from small migrant communities in Spain.

In the continuum approach presented by Tamarit et al. (2022), we can take the continuum limit (namely $r \to \infty$, $\mu \to 0$) in Equation 8.2 with $\eta = \mu(r-1)$ constant. In terms of this new parameter η,

$$t \equiv \frac{s_{max} - \sigma}{s_{max} - s_{min}} = \frac{e^\eta}{e^\eta - 1} - \frac{1}{\eta}, \tag{8.5}$$

and thus $\eta = \eta(t)$ is a function of the parameter t defined in the equation above, which represents a normalised measure of the cost of a relationship ($t = 0$ corresponding to the highest cost and $t = 1$ to the lowest one). Once η is determined, the fraction of relationships with a normalised cost not larger than t is given by

$$\chi(t) = \frac{e^{\eta t} - 1}{e^\eta - 1}, \tag{8.6}$$

an expression derived from Equation 8.3 in the continuum limit. This is the curve that should fit the data. Notice that each individual will be characterised by its own value of η.

The scaling ratio of the circles can be obtained from the asymptotic behaviour, for large η, of the logarithmic derivative of $\chi(t)$, the fraction of links whose "distance" to the individual is not larger than t, which turns out to be

$$\frac{\dot{\chi}(t)}{\chi(t)} = \frac{\eta e^{\eta t}}{e^{\eta t} - 1} \sim \begin{cases} \eta, & \eta \to \infty, \\ 0, & \eta \to -\infty. \end{cases} \tag{8.7}$$

In this approach, the separation between the normal and the inverted regimes also takes place at $\eta = 0$. Finally, to connect the two formalisms, we can use the fact that the discrete version of the left-hand side of Equation 8.7 is $(\chi_{k+1} - \chi_k)/\chi_k \Delta t$. Then, a comparison between Equation 8.4 and Equation 8.7 in the ordinary regime leads to $\eta \Delta t \approx e^\mu - 1$. Since $\Delta t \approx (r-1)^{-1}$, we obtain the equivalence

$$\eta \approx (r-1)(e^\mu - 1). \tag{8.8}$$

Interestingly, this result shows that the value of μ in the discrete model depends on the total number of layers r. This fact had not been noticed in previous research because of the implicit assumption of the existence of $r = 4$ layers in the structure of ego-networks. Setting $r = 4$ in Equation 8.8 and assuming, as empirically observed by Dunbar et al. (2015), that $e^\mu \approx 3$, we then find $\eta \approx 6$.

8.2.2 Parameter estimation

With the above approach in mind, given a dataset of relationships with continuous weights, the scaling parameter η can be estimated using the maximum-likelihood method. Tamarit et al. (2022) show that such an analysis leads to an expression equivalent to Equation 8.5 to connect the range of data weights to the theoretical parameters η and σ. Thus, for an empirical dataset, we can find the values of s_{max} and s_{min}, which are the largest/smallest possible costs an individual can invest in a relationship, respectively. Then, the value of σ, which is the total cost per item, is determined by

$$\sigma = \bar{s} = \frac{1}{L} \sum_{i=1}^{L} s_i, \tag{8.9}$$

where s_i are the costs associated with each of the relationships, measured in the same units as s_{max} and s_{min}, and L is the total number of relationships that an individual has. Once these variables are fixed, the parameter η, characterising each individual's ego-network structure, can be estimated numerically by solving Equation 8.5. Moreover, an expression for the $1 - 2\delta$ confidence intervals associated to the parameter η can be found (see Tamarit et al. (2022) for details). We choose a 95% confidence interval for this work using $\delta = 0.025$.

8.3 Results

The amount of grooming between primates is often used to indicate their relationship quality (Massen et al., 2010). We have applied the formalism presented in section 8.2 to the grooming data of chimpanzees residing at the Chimfunshi Wildlife Orphanage in Zambia for four years, between 2015 and 2019. Chimpanzees live in four distinct populations at this sanctuary, with no interaction occurring between individuals from different groups. Each group lives in different environments with identical conditions; thus, they only differ in size. This allowed us to analyse the grooming behaviour within each population and investigate the social structure of chimpanzees within their respective groups.

Considering the theoretical approach, given a dataset of relationships with continuous weights, the scaling parameter η can be estimated using the maximum-likelihood method. The basic idea of the fitting procedure is as follows: for every chimpanzee, we have the list of other individuals groomed and how often the focal chimpanzee was observed grooming each of them. From these data, we can obtain the range of grooming investment allocated across different partners and the total number of observations devoted to these activities. We can then get the corresponding η parameter characterising the chimpanzee's distribution of grooming times, and by inserting the value of η in Equation 8.6, we have the function $\chi(t)$ describing the whole distribution.

It is important to note that this methodology differs significantly from the commonly used approach in primatology, which involves regressing the response variable (e.g. grooming times) onto socio-demographic factors such as age, sex, and kinship. Rather than regressions with additional factors, we employ an analytical formalism to fit the distribution of grooming behaviour observed in the chimpanzees. It is noteworthy to remark that this model does not explicitly depend on any individual socio-demographic factors, which only enter indirectly
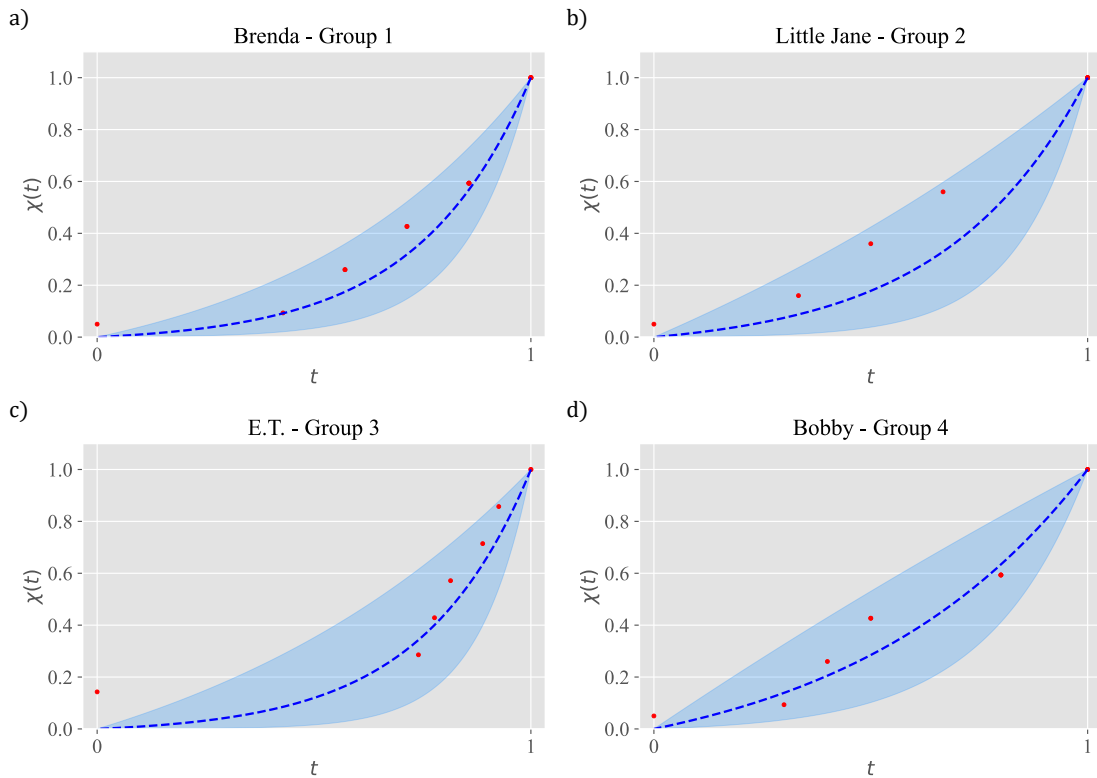
Figure 8.2: Examples of fittings for a chimpanzee of each group - The figure represents selected chimpanzees for which there were more available data points in each group. Selected individuals are: a) Brenda - Group 1, b) Little Jane - Group 2, c) E.T - Group 3 and d) Bobby - Group 4. Each figure represents $\chi(t)$, the fraction of links whose "distance" to the individual is not larger than t. Red dots are actual data, representing the number of individuals who receive no more grooming than a fraction t of the maximum. The blue dashed line is the fitted function $\chi(t)$, and the blue-shadowed region is the confidence interval.

through the capacity to maintain relationships (i.e., the total amount of time devoted to grooming). Consequently, the model provides a different kind of information. Finally, a limitation of our approach is that chimpanzees with very few observed relationships cannot be included in the study, despite reasonable sampling effort, due to the inaccuracies in the corresponding fits to our analytical expressions, as explained in detail in section 8.1.

The function $\chi(t)$ is used to determine the structure of the continuous circles in the grooming networks of individual chimpanzees, as exemplified in Figure 8.2. Plots of the fits for all individuals in the study can be found in subsection G.2.2. While the fits are not perfect, most data points lie within the 95% confidence interval for the fitted distribution, with those not being close to it. Moreover, some fits are better or worse than others, but the ones shown in Figure 8.2 were selected because they had more data for fitting. It is important to remember that the chimpanzee data are relatively noisy because the animals were observed in large, naturalistic enclosures with varying levels of visibility, and multiple observers collected data over a four-year period. Despite these challenges, the fits can be considered very good and are similar to those reported by Tamarit et al. (2022). These
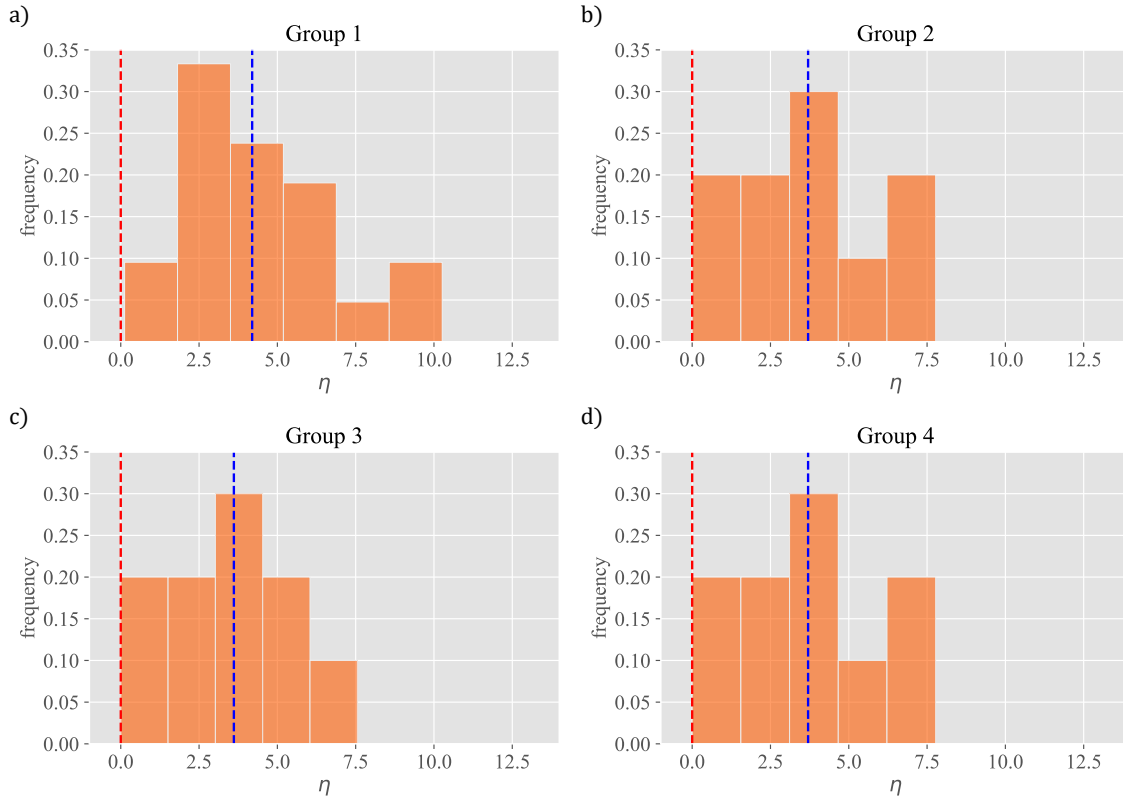
Figure 8.3: Histograms for the parameter distribution - The figure represents histograms for the η parameter in each group: a) Group 1, b) Group 2, c) Group 3 and d) Group 4. The red dashed line represents the regime change, from the standard to the inverse one at $\eta = 0$, and the blue dashed line represents each group's mean value.

results suggest that the continuum theory accurately describes how a chimpanzee distributes its grooming time to others.

These findings have important implications for our understanding of social behaviour in primates. By showing that the continuum theory is a good description of how chimpanzees distribute their grooming time, our study contributes to introducing the layered structure that shapes social networks in human groups to primates. Furthermore, our approach offers a novel way to interpret social behaviour less dependent on individual socio-demographic factors than all other traditional methods.

We summarise the analysis of the parameters characterising all individuals studied in Figure 8.3. The parameter η obtained from the fits has a mode of approximately $\eta = 4$, with a mean value close to the mode for all groups except Group 2, which is closer to 6. The range of values for η falls within the expected range, and the mode suggests that, typically, the scaling ratio of the circles for chimpanzees is somewhat smaller than for humans, except in Group 2. We believe that this difference arises because Group 2 is significantly larger than the other groups before and after filtering the data. This larger size allows group members to develop more complex social relationships involving shorter grooming intervals with more
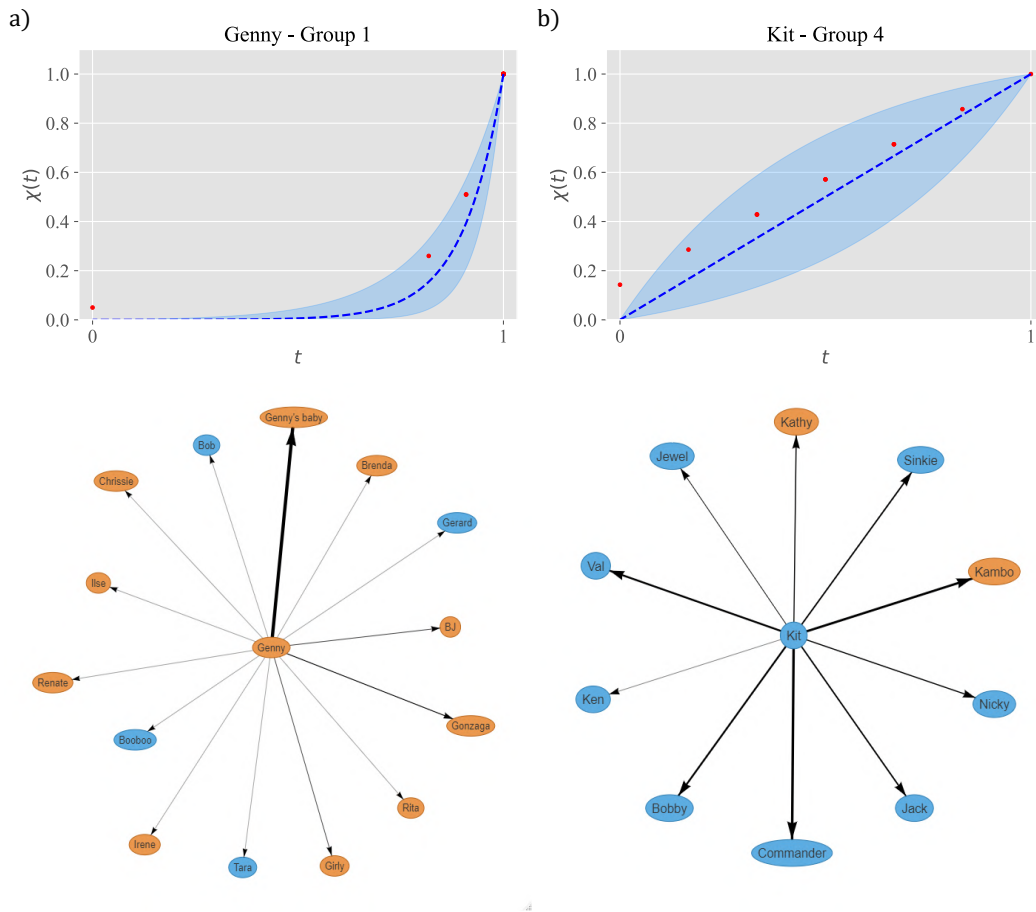
Figure 8.4: Relationship between the fittings and the structure of ego-networks - The figure represents the relationship between the $\chi(t)$ functions and the structure of the ego-networks for different individuals: a) Genny - Group 1 ($\eta = 10.3$) and b) Kit - Group 4 ($\eta = 0.14$). The reason for selecting these two chimpanzees is that they present the highest and lowest η parameter values in the entire dataset. The arrows connect the focal individual with those it grooms, while the width of the arrows represents the total amount of time devoted to grooming a specific chimpanzee. Orange ovals represent females and blue ones males.

individuals, leading to higher values of η and more low-intensity relationships. These results are consistent with previous research indicating that social structure in primate groups can be influenced by group size (Massen et al., 2010). Moreover, they support that the distribution of grooming behaviour in primate groups can be described by a layered continuum structure, as proposed by Tamarit et al. (2022) for humans.

Furthermore, the histograms presented in Figure 8.3 reveal that none of the fits yielded negative values of η, which would indicate an inverted structure of relationships as observed in human communities of migrants by Tamarit et al. (2018). However, some values of η are close to zero, which means they lie on the border between the two regimes and should therefore result in a higher fraction of individuals in the inner part of the distribution $\chi(t)$. As expected, these values primarily emerge in groups where the population size is smaller

(i.e., Group 3 and Group 4). In these groups, chimpanzees have fewer relationships but devote more grooming time to each other, likely contributing to the observed distribution of the parameter values.

These previous hypotheses are also supported by the results presented in Figure 8.4, which illustrates the relationship structure of the ego-networks and the $\chi(t)$ fitted function for two contrasting individuals. One example is Kit, whose $\eta = 0.14$ indicates an intermediate structure between the two regimes. Kit spends a considerable amount of time grooming primarily Kambo, Commander, Bobby and Val, as well as other chimpanzees. This finding is consistent with the fact that Kit is in Group 4, the group where fewer individuals are available to groom. The opposite extreme on the η scale is illustrated by Genny, in Group 1, with $\eta = 10.3$. Genny's grooming behaviour is typical of the normal regime, with a significant amount of grooming directed toward her baby and some toward her other daughter Gonzaga, and minimal grooming for other individuals. This outcome is expected, as Group 1 has many more chimpanzees that Genny can interact with.

Overall, this analysis provides insights into the social structure of chimpanzees living in different groups and contributes to our understanding of the continuum theory. It also highlights the similarities between the social structures of chimpanzees and humans, despite the substantial differences between the two species.

8.4   Discussion

The results of our study indicate that chimpanzees organise their grooming time in a way that aligns with the continuum resource allocation theory applied to humans by Tamarit et al. (2022). Essentially, chimpanzees allocate their grooming time among group members in a manner similar to how humans distribute their attention to individuals when relationship intensity is treated as a continuous variable. These findings are consistent with prior research suggesting that grooming is a resource allocation problem (Barrett et al., 1999; Dunbar, 1992a; Fruteau et al., 2011; Kaburu & Newton-Fisher, 2015; Waal, 1997). Similar to humans, some chimpanzees invest a large amount of grooming time in a few other individuals, mainly when the group is small, while others distribute smaller amounts of grooming time among many other individuals. In other words, chimpanzees exhibit exactly the same two allocation strategies observed in human relationships, and inverted structures are more likely in smaller groups. These findings confirm grooming as a way of expressing friendship in non-human primates (Silk, 2002) and suggest that different social strategies may be at play among chimpanzees, depending on their immediate group structure. Our results demonstrate that chimpanzees in larger groups employ their social capital differently than those in smaller groups, similar to how humans behave (Bernard & Killworth, 1973; Tamarit et al., 2018). This suggests that chimpanzees can adapt flexibly to their social environment, distributing social resources widely among group members when necessary but investing more intensely in a few others when possible. Additionally, any social forces influencing group cohesiveness (e.g., whether a group forms a whole or a modular, sub-group structure) indirectly shape a primate's resource allocation strategy. This connection is currently only known from the field of human sociology (Wellman & Berkowitz, 1988).

The key finding of our study is that chimpanzees organise their affiliative relationships in a manner that mirrors the pattern we have previously described for humans. More precisely, the mean group sizes of all primate species follow the same pattern observed in real human ego-networks (Dunbar et al., 2018), and some primate species' global network structures also exhibit similar internal anatomies (Dunbar & Shultz, 2021; Hill et al., 2008; Kudo & Dunbar, 2001). By analysing grooming data using the continuum approach, we demonstrate that ego-networks in chimpanzees exhibit a specific circle-based organisation. Understanding the similarities in the organisation of social networks across different primate species can provide important insights into the evolution of social behaviour. By comparing the patterns of social organisation in humans and chimpanzees, we can better comprehend the social forces that shape the behaviour of both species. For example, the flexibility in resource allocation strategies observed in chimpanzees suggests that they are able to adapt their social behaviour to different group sizes and social environments, just as humans do. Furthermore, our findings support the idea that grooming serves as a means of expressing friendship in non-human primates, highlighting the importance of social bonding in the lives of these animals.

It is important to note that while our study focused on chimpanzees, other non-human primates also exhibit complex social behaviours and may have similar patterns of social organisation. Therefore, future research should explore whether the continuum approach can be used to analyse social networks in other primate species and provide a more comprehensive understanding of the evolution of social behaviour. Future research could use behavioural data from other primate species to determine whether the continuum analysis of ego-networks reveals a consistent pattern in other species living in large social groups.

Part IV

Final remarks

9

Conclusions and future work

In this chapter, we present a synthesis of our research findings within the interdisciplinary field of social physics, highlighting the key contributions derived from our investigation. By examining the structure and dynamics of human social relationships through the lens of social physics, our comprehensive analysis not only consolidates the outcomes of the various research stages but also emphasises their significance in the broader context of the study. Furthermore, we reflect on the limitations and challenges encountered throughout the research process, underscoring the areas that warrant further exploration.

Moreover, we discuss the future research directions that have emerged as a result of our findings, outlining potential avenues for extending and refining our work. These opportunities encompass various aspects of the research domain, which include different methodologies, applications and interdisciplinary collaborations. With this in mind, we aim to provide a roadmap for future investigations, fostering continued growth and development within the field.

Finally, we discuss the broader implications of our research in social physics, considering its potential impact on practice and society at large. We underscore the relevance and importance of our work beyond academia, demonstrating how our findings can contribute to tangible improvements in various sectors, especially in education by preventing important issues such as bullying. Therefore, this chapter serves as a culmination for this thesis, offering a reflective perspective on the progress made, the challenges faced and the exciting prospects that lie ahead in the realm of understanding the structure and dynamics of social relationships.

9.1 Conclusions

9.1.1 Data-driven perspective

In the first part of this thesis, we have tackled the problem of understanding the structure and dynamics of human social relationships using a data-driven approach. Additionally, we have used different methods and techniques to uncover hidden patterns in social networks, ultimately providing valuable insights into the complex nature of human interactions.

In chapter 2, we have examined the evolution of social relationships among high school students by conducting a longitudinal study consisting of five waves of surveys during two consecutive academic years. Our analysis discovered a consistent structure in the number of reported friends and best friends across the five waves. The innermost circle of friendships (best friends) demonstrated more stability than the outer circle (just friends), which supports the concept of Dunbar's circles in the organisation of social relationships and emphasises the critical role of the different intensities in maintaining stable connections. Moreover, our findings indicate that enmities were much less frequent and more volatile than friendships, with many negative relationships disappearing from one survey wave to the next. All these results do not depend on different factors, such as course, age or gender, which highlights their consistency.

Apart from that, we also observed that sharing the same class had a significant impact on the formation and stability of relationships among students. Our findings showed that when students were separated into different classes between academic years, they tended to lose a substantial proportion of their previous friendships, and these lost connections were often replaced with new ones formed with new classmates. Furthermore, our study revealed that reciprocal relationships remained remarkably consistent across different waves, around 60%, with gender homophilic friendships exhibiting slightly higher levels of reciprocity compared to cross-gender ones. These results not only reinforced the importance of frequent interaction in maintaining or weakening relationships but also led us to continue supporting the "social atom" metaphor. From this, we suggest that the layered structure of social relationships may be studied as statistical-mechanical systems in equilibrium, with each relationship possessing an associated "binding energy", which determines the probability of appearance associated with each link.

In chapter 3, we have investigated the role of negative relationships in social networks and community structure. We analysed the community structure at the complete school level and also when considering each group individually. In this way, we were able to uncover interesting patterns and dynamics. We observed a distinct evolution of gender segregation over time in the communities, with negative relationships playing a critical role in exposing these gender-based divisions within the network. Furthermore, the size of each community eventually stabilised at around 12 individuals, which intriguingly coincides with the size of the second Dunbar's circle. This finding reinforces the idea that human social networks tend to naturally organise around specific numbers, suggesting that there may be innate constraints on the size of stable communities.

Moreover, our research also remarked on the significance of the social balance theory in understanding the stability of triadic relationships within social networks. This theory posits that triangles with an odd number of negative links tend to create a kind of cognitive dissonance, rendering them unstable and likely to disappear over time. By analysing the data and comparing it with a recently proposed null model, we have concluded that there is indeed a strong inclination towards minimising unbalanced configurations.

Finally, in chapter 4, we have shifted our focus from scientific research to the development of an online application designed to streamline data analysis and provide valuable insights into social dynamics within schools. Initially intended for research purposes, we soon realised the application's potential to assist school staff in anticipating and addressing issues, thereby improving the social atmosphere. We granted staff full access to the application and created a demo user account for readers to explore its features and potential benefits.

More in detail, the application comprises ten panels, with three key panels discussed in detail due to their usefulness in facilitating targeted interventions: interactive networks, interactive ego-networks and interactive friendship vs. enmity dispersion. All these panels provide valuable insights into the social structure, individual students' relationships and potential sources of conflict. We also discussed feedback from the guidance team leader, Silvia Ibáñez Morcillo, which has been invaluable in demonstrating the real-world impact of our application and leading to further improvements and refinement. Therefore, our collaborative efforts have resulted in a tool that not only aids our research but also supports informed decision-making and proactive measures to address social issues within the school community.

9.1.2 Model-driven perspective

In the second part of this thesis, we have employed techniques from statistical mechanics to explain the structure of social relationships. Through this alternative approach, we have been able to explore new viewpoints that enable us to uncover new details about human social behaviour.

In chapter 5, we have developed a model to explain the structure of social networks by analysing interactions between their members. We built upon the "social atom" concept and used exponential random graphs, considering the pairwise approximation for the analysis. We started by exploring the pairwise model, which characterises the stability of pairs of links and defines the probability of their appearance in the system. However, the pairwise model lacks information about the factors shaping the parameters. To address this limitation, we introduced the linear model with reciprocity, which takes into account the limited cognitive cost of maintaining relationships, their variable intensities, and the fact that there is a high level of reciprocity in these relationships. This model effectively fits the maximum likelihood values from the pairwise model, suggesting that social network structures can be explained using these three variables.

We then aimed to validate the pairwise approximation by examining reciprocity and transitivity, which are characteristic macroscopic properties observed in real-world social networks. We have demonstrated that the first condition is satisfied, but the second one is not. For this reason, we have added a term related to transitivity in the Hamiltonian. However, this

brought the model to a new level of difficulty because the grand partition function becomes intractable for systems with a large number of individuals. We attempted to address this issue by developing a transformation of the parameters to simplify the Hamiltonian and finally mapped it to Strauss's model of transitive networks. This model exhibits a phase transition and becomes sensitive to small changes in the control parameter. This is the reason why its validity as a model to describe real networks has been questioned.

With this result in mind, in chapter 6, we have tackled Strauss's model of transitive networks using density-functional theory, a technique borrowed from the statistical physics of lattice gases. Our solution demonstrated greater accuracy in describing small systems than more traditional mean-field approaches and revealed that the canonical ensemble is the natural description for accessing intermediate states. The model exhibits a first-order phase transition, with low-density and high-density phases akin to gases or liquids. To conclude, we demonstrated that Monte Carlo simulations using Kawasaki dynamics could help access the intermediate states.

Moreover, we have analysed how our density-functional formalism can also be applied to systems with link- or triangle-dependent interaction constants, for example, allowing for the study of homophilic interactions. This analysis showed that homophily increases the stability of uniform networks by raising the critical point value. While further studies are needed to characterise such systems fully, our method offers a systematic approach to extending other ERG models. In this way, the main contribution of this formalism is that it can be applied to various network models and extended to other complex systems beyond network science, ultimately shedding light on their rich and complex behaviour.

Finally, in chapter 7, we have re-evaluated the models for the analysis of social structure by introducing a clustering term into the Hamiltonian, which allows us to fix the number of triangles with all three reciprocal links in the system. As explained in chapter 6, this addition results in higher-order interactions and complicates the calculation of analytical expressions for the parameters and their confidence intervals. We first examined the pairwise model with clustering, which captures better real-world network properties than the model without clustering. However, it does not provide insights into factors that shape the parameters. Therefore, as we did with the pairwise model, we studied the linear model with reciprocity and clustering to fill this gap. Analysing the parameters of this model suggests that their form can still be explained using only three factors (apart from clustering): the two derived from the "social atom" concept and reciprocity.

We have also conducted Monte Carlo simulations to validate our models. The agreement between our method and simulations supports the validity of the models and their associated parameters, offering confidence in the insights derived from their analysis. Moreover, our method provides a balance between computational efficiency and accuracy in capturing the complexities of the system, making it an advantageous alternative to costly Monte Carlo simulations for studying social systems.

To conclude, we have proposed the "social fluid" approach as an alternative paradigm to that of networks for understanding social relationships. Inspired by statistical mechanics, this approach allows for a more comprehensive analysis of the underlying structure of social

systems by considering them as ensembles of interacting "social atoms". It is our opinion that this new perspective may help develop more effective models in the future.

9.1.3 Beyond humans: application to primates

Finally, in the third part of this thesis, we have extended the analysis to non-human primates. We have used data from chimpanzees living at the Chimfunshi Orphanage in Zambia, and by applying the models we have developed for humans, we have demonstrated that the organisation of their social relationships is very similar.

In chapter 8, we have demonstrated that chimpanzees organise their grooming time in a way that aligns with the continuum resource allocation theory applied to humans. This suggests that, like humans, chimpanzees allocate their social resources among group members based on relationship intensity. In consequence, these findings support the idea that grooming is a resource allocation problem and serves as a means of expressing friendship in non-human primates. Furthermore, we have revealed that all the relationships that form chimpanzees' ego-networks show a circle-based organisation. Therefore, chimpanzees exhibit similar social strategies to humans, with their social capital usage differing depending on group size. This demonstrates the adaptability of chimpanzees to their social environment, investing social resources widely among group members or focusing only on a few individuals.

9.2 Future work

9.2.1 Data collection and software development

One of the challenges we continue to work on beyond this thesis is the collection of data through surveys on friendship and enmity relationships among high school students. Although the results presented here only use the responses collected up to the 2021/2022 academic year, we have carried out three additional data collection waves at IES Blas de Otero during this last year, corresponding to the 2022/2023 course. In this way, we have been able to continue monitoring the responses of the students at this centre. For some of them, we already have information from up to eight waves spread over three years, which will allow us to characterise their evolution and determine if there are differences in the structure of their relationships over time. Additionally, we have started to collaborate with new educational institutions, both in Spain and internationally, which will allow us to expand the scope of our study, gather a more diverse and comprehensive dataset and potentially uncover new universal patterns and trends.

Some of the innovations we have added in the latest surveys we have conducted are new questions that, besides the ones related to friendship and enmity relationships, allow us to understand each student's academic performance better and create a comprehensive profile of their personality. This is very interesting, as it will help us explore the social differences that exist between people with complementary personalities and identify potential correlations between personality traits, academic performance and the nature of their social relationships. Therefore, we can gain a deeper understanding of how individual characteristics might influence friendship and enmity dynamics among high school students by examining all these aspects together. Furthermore, all this innovative information can provide valuable

insights for educators and school staff to develop more effective strategies for fostering a positive learning environment and promoting social cohesion among their students.

Moreover, we are collaborating with the company Kampal Data Solutions, a spin-off from the University of Zaragoza, which also includes members from other universities, especially Carlos III University of Madrid. This company has developed a new software for collecting and visualising survey data called Kampal Schools. In the latest data collections, we are using this software, as it offers a more comprehensive set of functionalities compared to the SAND tool that we have employed in this thesis. However, for the groups we were already working with previously, we continue to use SAND since we have no way of connecting student IDs between the two applications, and we do not want to break down the available time series data.

The most interesting aspect of the comparison between these two software programs is that, although they ask exactly the same questions about students' social relationships, their order, the structure and the layout of the applications are different. Therefore, we can compare the results obtained and verify if there is any bias in the application design that affects the results or, on the contrary, if they are consistent. Our preliminary results show that the patterns are identical in both cases, suggesting that they are universal, although we continue analysing them. These findings highlight the robustness of the survey methods and the reliability of the data collected, which strengthens the validity of our research. As we continue to use both Kampal Schools and the SAND tool, we can further explore the intricacies of social relationships in high school students and gain valuable insights from the data. Furthermore, the collaboration with the company Kampal Data Solutions and the involvement of experts from different universities incentives a multidisciplinary approach, which can lead to innovative perspectives and solutions in addressing the challenges associated with understanding and improving social dynamics among high school students. As our research progresses, we hope to build upon these preliminary results and contribute to the development of educational interventions that effectively promote positive social interactions in high school students.

Finally, we are planning to expand our research beyond high school students. Our research is progressing well, and we are looking to broaden our scope beyond high school students. The development, testing, and implementation of our software have been the most challenging aspects, but once completed, we hope to gain insight into the structure of social relationships across different age groups. Our initial plan is to involve first-year engineering students in the project and compare their results with the former ones in high schools. This will enable us to examine the impact of age on social relationships and identify differences between various stages, such as high school and university. Furthermore, we can try to understand the factors that contribute to the formation and evolution of social networks and explore how the transition between different stages may influence the development and dynamics of social relationships. This broader perspective can help us identify potential patterns, trends and key differences that may be specific to certain age groups or educational settings.

9.2.2   More insights into social structure and dynamics

One of the certainties we conclude from this thesis is that the results we present are a relevant contribution to the social physics discipline, particularly in what concerns the structure and dynamics of social relationships. However, the number of open and unanswered questions that need to be addressed in the future is significant. The data we have collected (and continue collecting) can be very useful for managing all these questions using a data-driven approach.

Firstly, we aim at expanding and elaborating on the "social atom" concept. We want to verify that the patterns we have detected in human social behaviour are still being observed by using data from the new waves, as well as continue collecting a significant amount of data over the years that renders statistical support to this new paradigm. The new data we have gathered correspond to the responses of students who have previously participated in the study (those who have been in high school for the longest time) and some who have not (those who are new to the school). Therefore, detecting the same patterns in all responses lends significant consistency to the results. We plan to analyse the data thoroughly, considering factors such as age, gender or itinerary of the participants. This will allow us to understand any potential biases in the data better and to make necessary adjustments.

In particular, we want to focus on the comparison by gender. Some studies suggest that the way men and women organise their social relationships is different. However, these conclusions are more related to group-level dynamics than individual-level differences. Our preliminary results indicate that no significant difference exists between genders, but more statistics are required to figure out whether there is any gender difference in group dynamics. Furthermore, by incorporating the new personality data we have collected in the latest waves, we can achieve a more accurate characterisation of the participants. This will enable us to examine the potential influence of personality traits on the patterns observed in social relationships and determine to what extent any differences that may exist between individuals are linked to them or even predict future relationships based on them.

Another open question arising from the results presented in this thesis is related to the reciprocity of relationships. We have observed that this property is an essential factor in determining social structure, and it maintains a constant value of around 60% across all waves. Moreover, for homophilic pairs of relationships, such as those sharing a group or of the same sex, the reciprocity rate is slightly higher. However, we are unable to explain the reason behind the constancy of this value through all the waves. Our goal is to continue analysing the data, utilising new techniques and developing new models that enable us to understand this phenomenon. Therefore, we plan to explore possible explanations for the observed reciprocity rate to determine its implications for social dynamics.

Additionally, we want to delve deeper into the importance of enmities in social dynamics. One of the most significant findings we have obtained in this thesis, using a data-driven approach, is that the structure of friendship and enmity relationships is completely different. While friendships are more abundant and stable over time, enmity relationships are fewer and less lasting. For this reason, when explaining the structure and dynamics of social relationships, much more importance is often given to analysing friendships than enmities.

We have already shown that the community structure within the system considerably varies depending on whether negative relationships are taken into account or not. However, we believe that enmities also play a crucial role in explaining the dynamics of relationships, especially those that disappear from one wave to the next.

Our hypothesis is closely related to the social balance theory. If we consider transitivity in a signed network, those triangles with a negative number of links are unstable and are likely to disappear from one wave to the following, as it is not common that an individual will maintain a friendship with someone who has an enmity relationship with one of their friends. Therefore, using different data analysis techniques, we aim to study the connection between the negative relationships present in one data collection and the friendships that disappear in subsequent data collections to determine if they are related. By doing so, we hope to gain a more comprehensive understanding of how both positive and negative relationships shape social dynamics and determine the role of balance in maintaining social harmony.

Finally, another area of research we are working on is determining the structure of personal networks. The personal network associated with an individual is the network that includes all the relationships between people who have been identified as a friend by the ego, but without considering the relationships they maintain with the ego. This type of network allows us to understand the ego's environment and the relationships that appear in it. Using network theory metrics, such as different centralities, number of communities or number of connected components, along with unsupervised machine learning techniques, we aim to classify personal networks into different categories and compare their properties with existing literature. Ultimately, we try to find a quantitative classification related to specific values of known metrics that allows us to characterise personal networks. This research can provide valuable insights into the factors that influence their structure and dynamics and help to comprehend how individuals organise their social environments.

9.2.3 Refining statistical mechanics models

Another area of research that emerges from the results presented in this thesis is the study of the structure and dynamics of social relationships using models derived from statistical mechanics. The "social fluid" concept introduces a new perspective for understanding these systems, which is completely different from previous approaches. Traditionally, networks have been the tool used to model social relationships. Within this discipline, numerous techniques have been developed to simulate systems increasingly similar to those observed in the real world. Some examples of these techniques include weighted networks (for relationships of varying intensity), signed networks (friendships or enmities) and multiplex networks (different groups or communities). These complex techniques, developed over decades for the study of gases and liquids, have proven to be very helpful in understanding their behaviour. The use of statistical mechanics in the study of social systems may be very useful in gaining insights into them.

The models we have presented in the second part of this thesis use exponential random graphs and density functional theory to reproduce macroscopic features of social systems, such as reciprocity and transitivity. Despite our effort, they may be used more as examples of applying these techniques than as realistic representations of the complete structure of social

relationships among high school students. Our models focus solely on understanding each group at the individual level, without considering inter-group relationships. Additionally, different phenomena derived from the data analysis, such as homophilic relationships (e.g., between people of the same gender) being more probable, are not considered. To include all these properties in our models, it is necessary to add a parameter associated with each of them in the Hamiltonian of the system. This would result in a larger parameter spectrum, more complex relationships and complicated calculations. Our objective is to reflect in our models all these properties observed in real systems, even if it is not possible to obtain analytical results, and we have to resort to numerical methods to determine the parameters.

Moreover, we want to continue studying the correspondence between models derived using statistical mechanics and Bayesian methods, such as parameter estimation using Monte Carlo simulations. The results presented in this thesis suggest that very similar maximum likelihood values for the parameters are obtained using both methods. If we were able to confirm this hypothesis, it would lend significant power to the introduction of statistical mechanics techniques for network analysis. Although the associated calculations are often complicated in these models, the computational power required compared to Bayesian methods is much lower. As computational power is one of the main limitations for analysing large systems, models based on statistical mechanics would enable us to study more complex social networks with reduced computational resources. Furthermore, this interdisciplinary approach could lead to the development of novel techniques, further enhancing the understanding of social systems and their dynamics.

9.2.4 Comparison with other species of non-human primates

Finally, one of the most interesting and innovative research areas that emerge from our work is to continue exploring whether the models we have developed to understand the structure of human social relationships can be extrapolated to other species, especially to other primate groups. In this thesis, we have worked with grooming data from four chimpanzee groups at the Chimfunshi Orphanage in Zambia, and we have demonstrated using a continuous version of Dunbar's circles that their social behaviour is similar to that observed in humans.

At present, thanks to the ongoing collaboration with our anthropologist co-authors, we have access to more grooming data among different primate species. Specifically, we have analysed grooming data from various bonobo and chimpanzee groups. Some of these groups live in captivity in different European zoos, while others reside at the Lola Bonobo sanctuary in Africa. Again, we have applied the continuous model to this data, and our results suggest that the social behaviour of all these primate groups shares many similarities. This research offers a unique opportunity to investigate further the commonalities in social structure across different primate species.

Additionally, these datasets are very rich and provide specific demographic data for each of the individuals we have analysed. This has allowed us to study how some of these factors, such as species, habitat, or age, affect the way in which these animals structure their social (grooming) relationships. With this in mind, we have used machine learning techniques, such as boosting, to determine the influence of each of these factors on the model. Our preliminary results suggest that this social structure changes slightly between bonobos and chimpanzees.

9. CONCLUSIONS AND FUTURE WORK

Bonobos are a more sociable species than chimpanzees, so they apply their social capital to maintain a larger number of relationships, although most of them are less intense. In contrast, chimpanzees form fewer relationships, but they are more intense. The same is observed for age as a factor. The younger the primates are, the more relationships they maintain simultaneously, and this number decreases over time. Elderly individuals maintain only a very limited number of relationships, but all of them are very intense. Finally, other factors, such as habitat, do not affect this structure, suggesting that it is more closely tied to the nature of the species and their cognitive abilities than to the enclosure they live in. Therefore, our findings suggest that the social structure of these primates is primarily influenced by inherent species characteristics and their cognitive capabilities. These insights contribute to a better understanding of the evolutionary underpinnings of social behaviour among primates and may even help us understand better human social dynamics.

Appendices

A

Appendix for Chapter 2

A.1 Supplementary figures

A.1.1 Number of respondents across groups
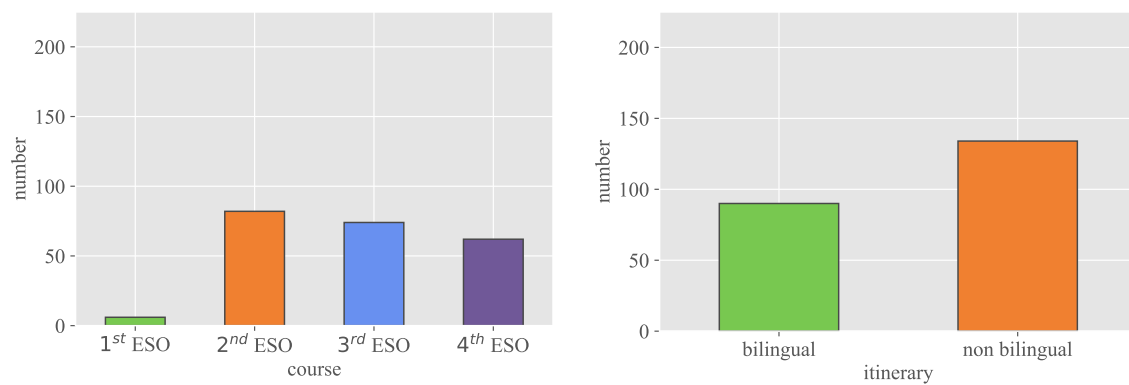
Figure A.1: Distribution of the students participating in all five waves - Left: by course. Right: by itinerary.
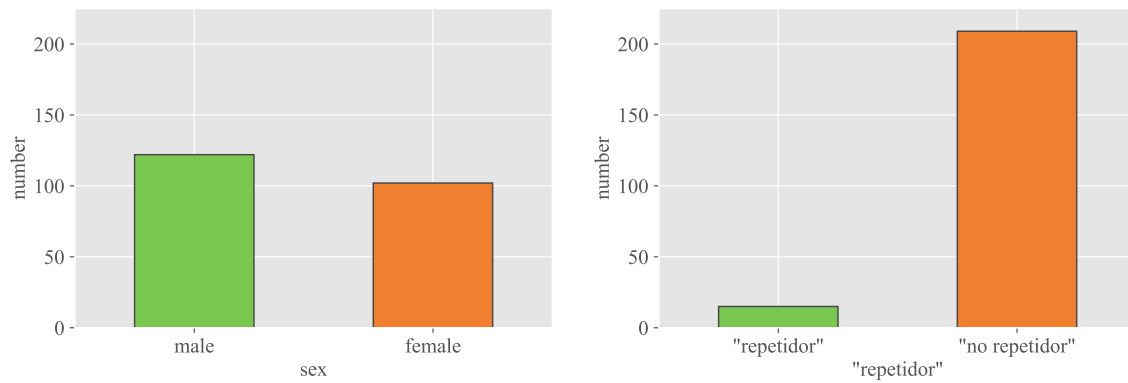
Figure A.2: Distribution of the students participating in all five waves - Left: by sex. Right: by "repetidores" or not.

A.1.2   Number of friendships
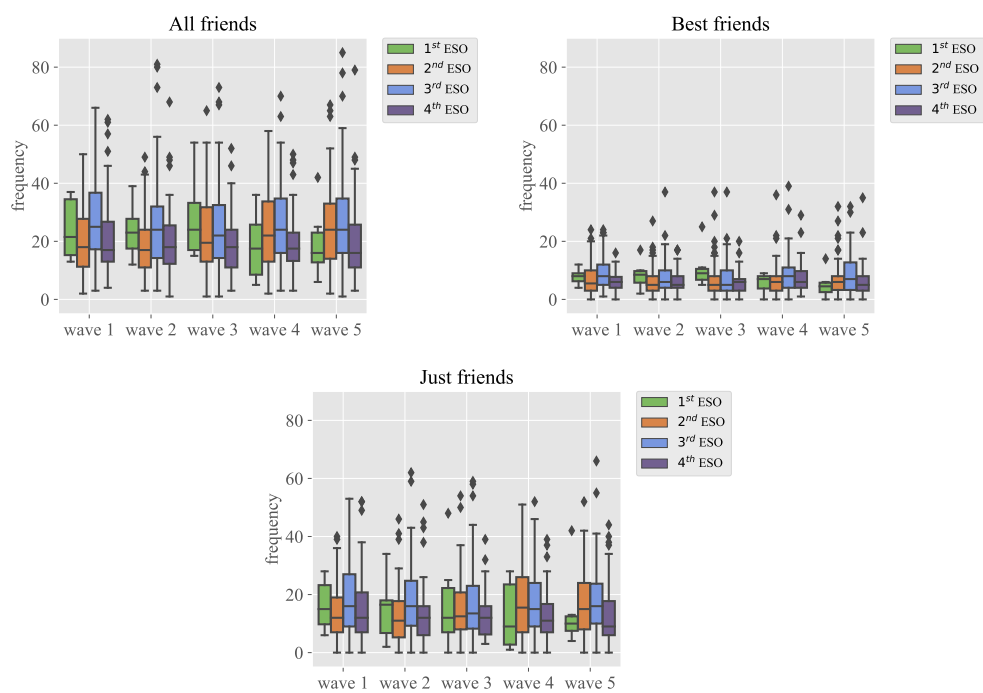
Figure A.3: Number of friendships declared by course - Left: all friends. Right: best friends. Bottom: just friends.
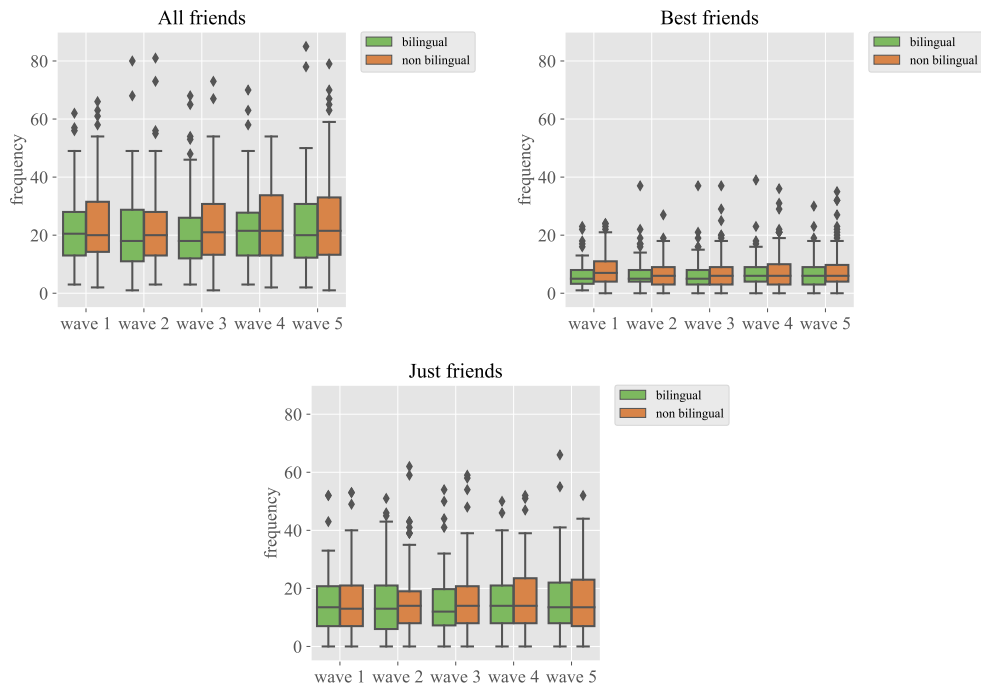
Figure A.4: Number of friendships declared by itinerary - Left: all friends. Right: best friends. Bottom: just friends.
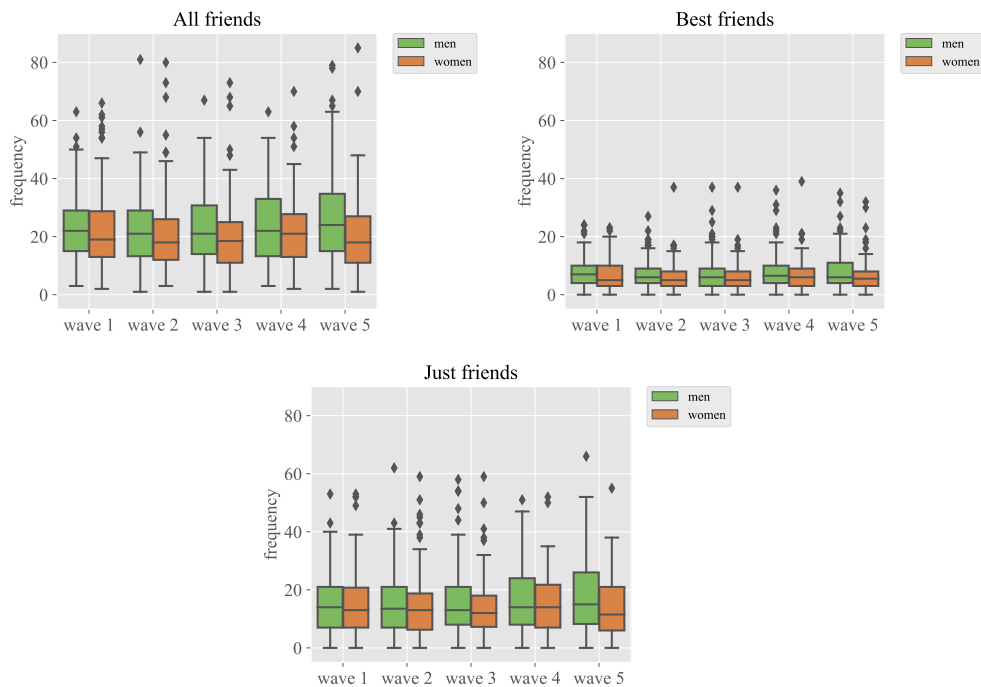
Figure A.5: Number of friendships declared by gender - Left: all friends. Right: best friends. Bottom: just friends.
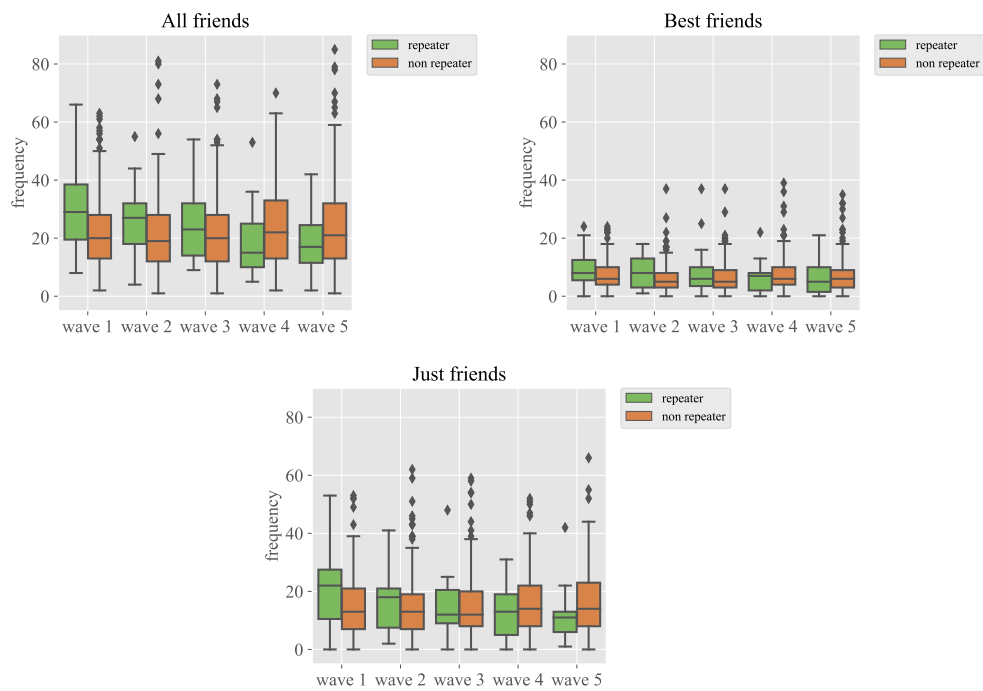
Figure A.6: Number of friendship declared by "repetidores" or not - Left: all friends. Right: best friends. Bottom: just friends.

A.1.3 μ parameter
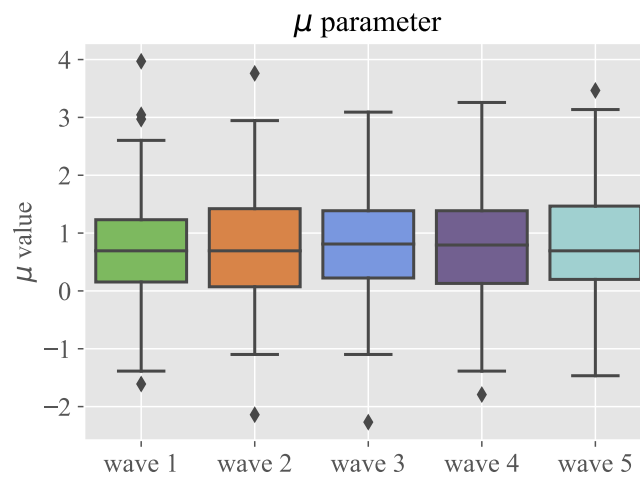
Figure A.7: μ **parameter distribution across waves**
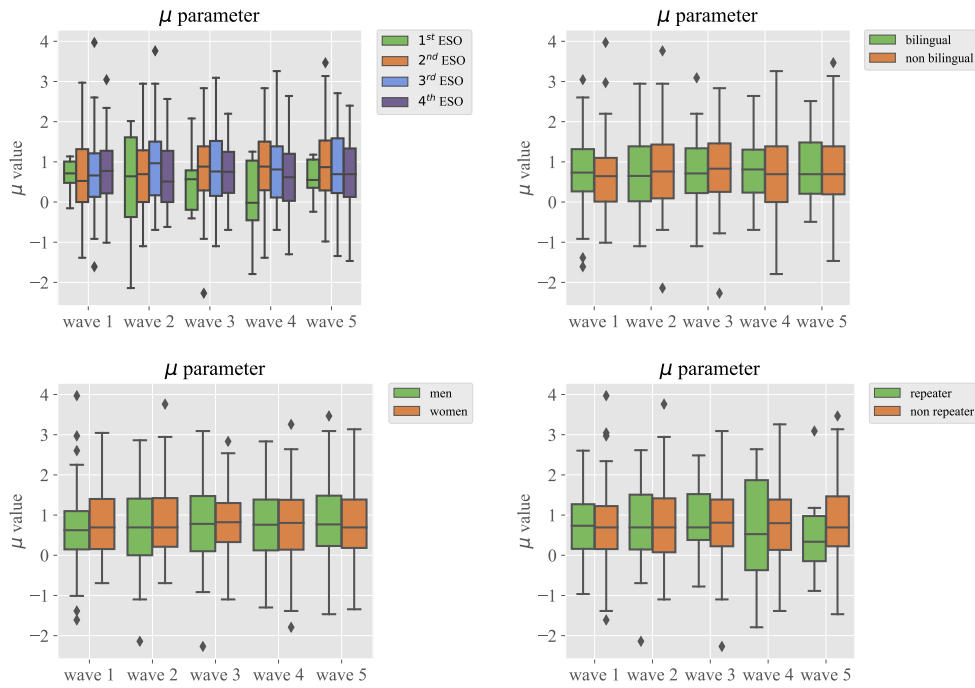
Figure A.8: μ parameter distribution with division by groups - Top left: by course. Top right: by itinerary. Bottom left: by gender. Bottom right: by "repetidores" or not.

A.1.4 Slopes of the linear fit to aggregate evolution
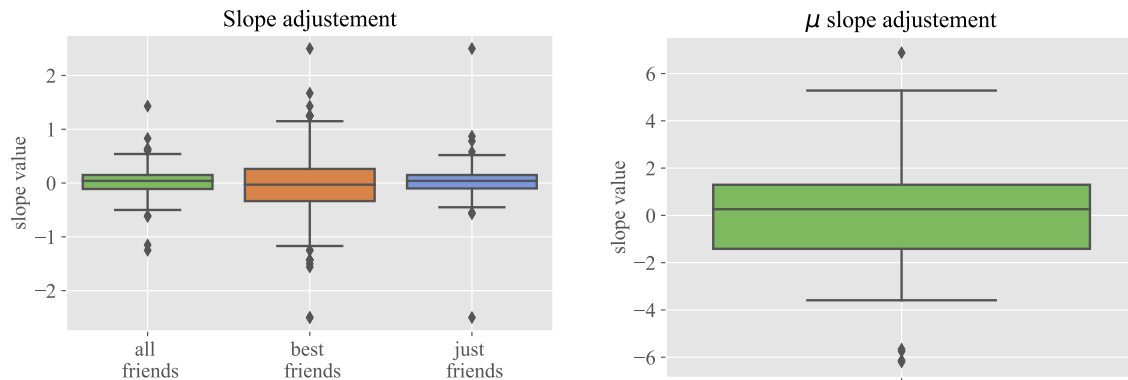
Figure A.9: Slope of linear fits to the evolution - Left: number of friendships. Right: μ parameter.

A.1.5 Transitions of relationships between waves
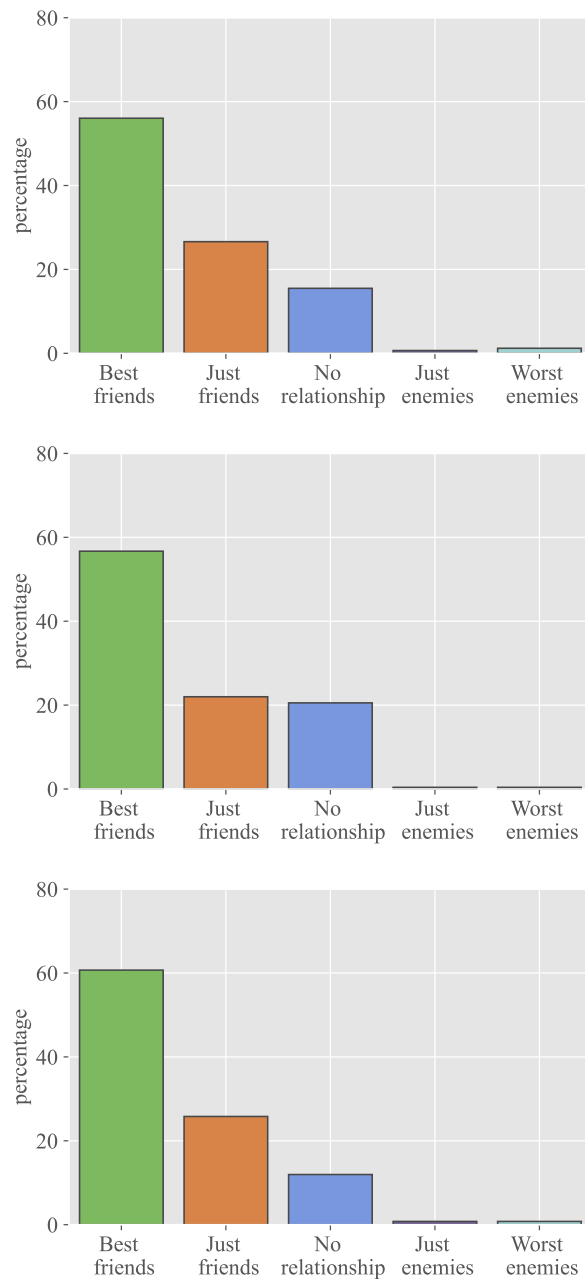
Figure A.10: Best friends transitions - Percentage of individuals that ended up in a given category in wave n, when they were marked as "best friend" in the previous wave (conditional probability $P(x, w_n| + 2, w_{n-1})$). Top: from wave 1 to wave 2. Middle: from wave 2 to wave 3. Bottom: from wave 3 to wave 4.
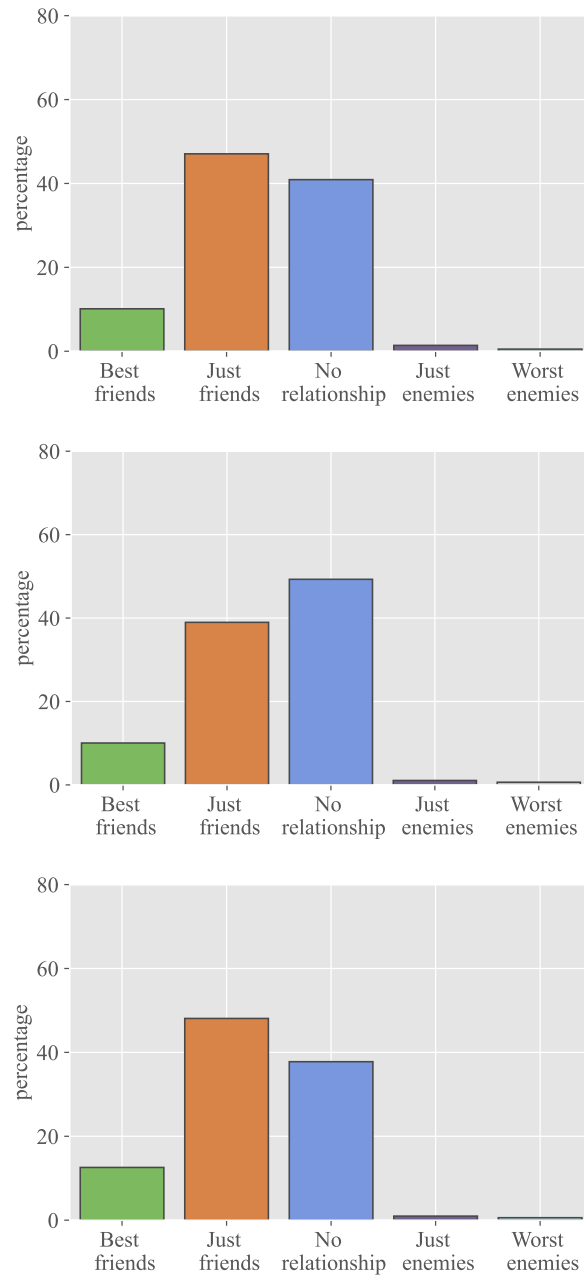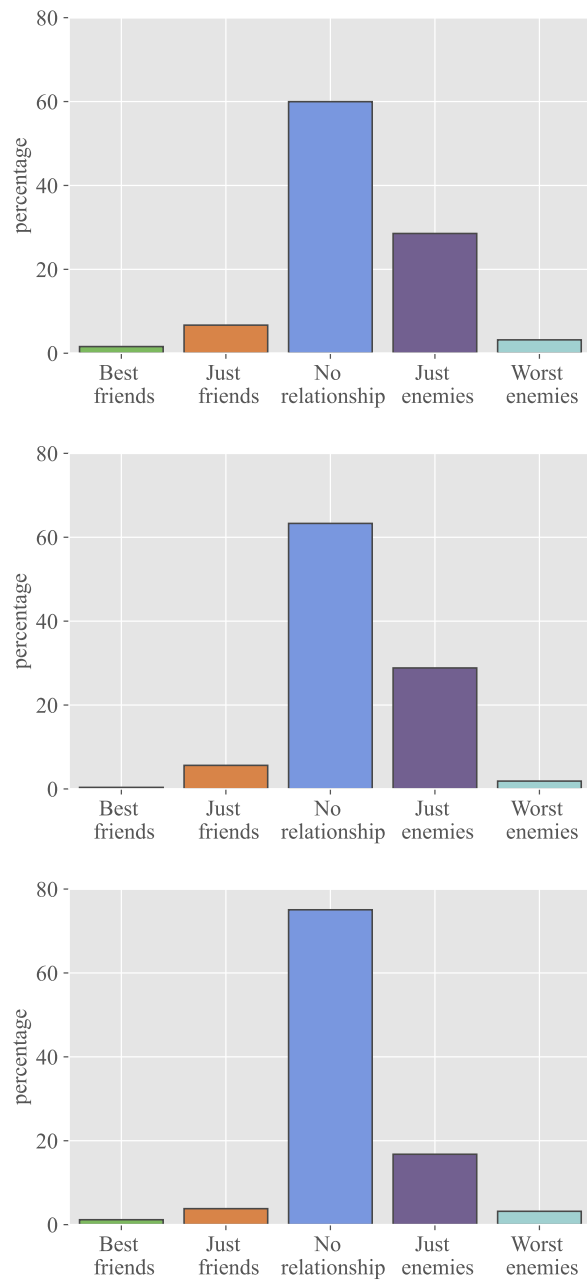
Figure A.11: Friends transitions - Percentage of individuals that ended up in a given category in wave n, when they were marked just as "friend" in the previous wave (conditional probability $P(x, w_n| + 1, w_{n-1})$). Top: from wave 1 to wave 2. Middle: from wave 2 to wave 3. Bottom: from wave 3 to wave 4.

Figure A.12: Enemies transitions - Percentage of individuals that ended up in a given category in wave n, when they were marked just as "enemy" in the previous wave (conditional probability $P(x, w_n| - 1, w_{n-1})$). Top: from wave 1 to wave 2. Middle: from wave 2 to wave 3. Bottom: from wave 3 to wave 4.
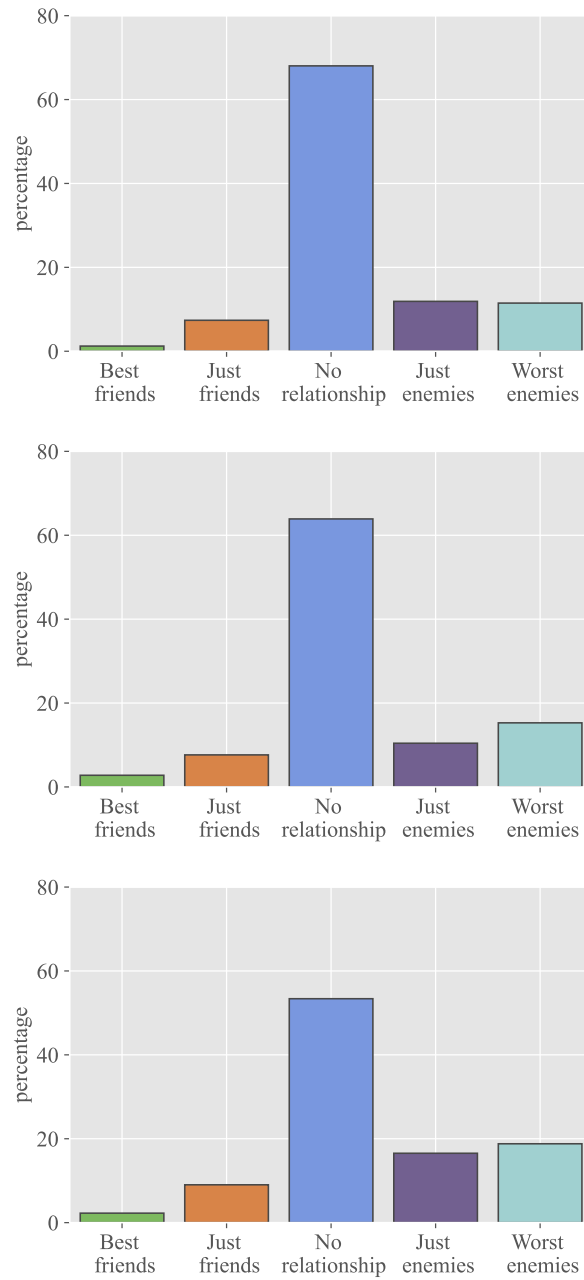
Figure A.13: Worst enemies transitions - Percentage of individuals that ended up in a given category in wave n, when they were marked as "worst enemy" in the previous wave (conditional probability $P(x, w_n | -2, w_{n-1})$). Top: from wave 1 to wave 2. Middle: from wave 2 to wave 3. Bottom: from wave 3 to wave 4.
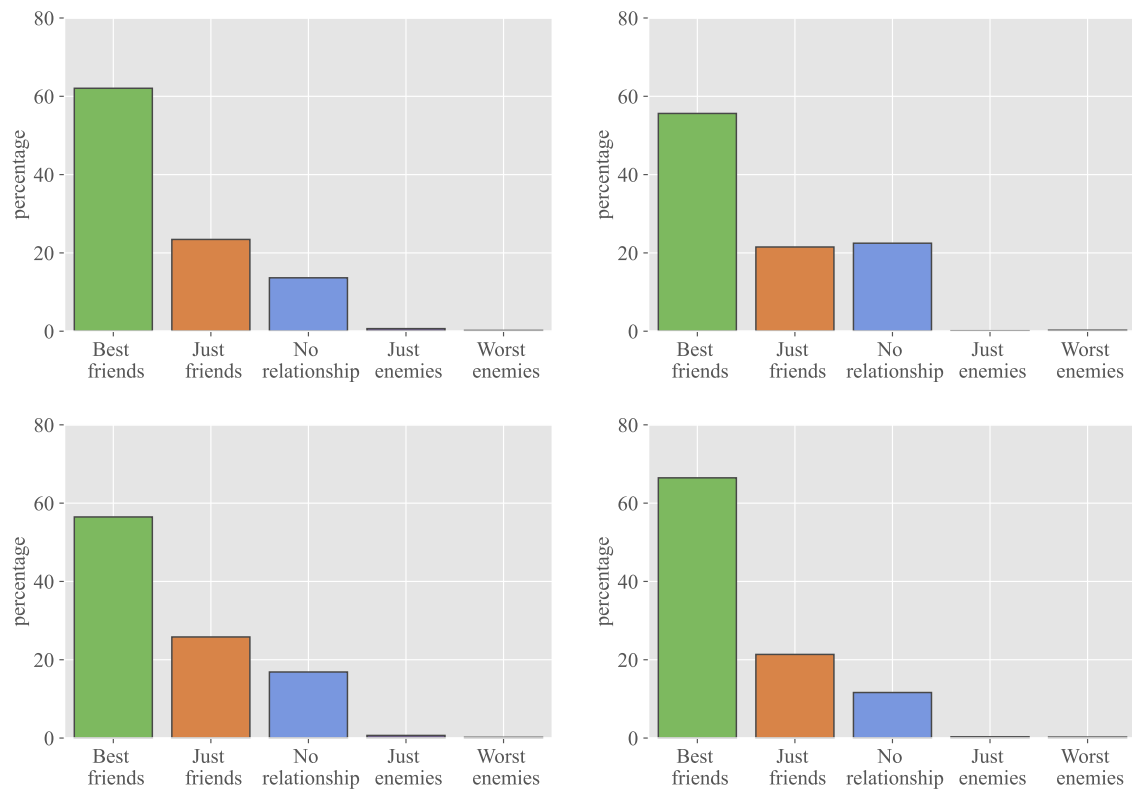
Figure A.14: Best friends origin - Percentage of individuals that end up as "best friends" in wave n and were marked in any category in the previous wave (conditional probability $P(x, w_{n-1}|2, w_n)$). Top left: from wave 1 to wave 2. Top right: from wave 2 to wave 3. Bottom left: from wave 3 to wave 4. Bottom right: from wave 4 to wave 5.
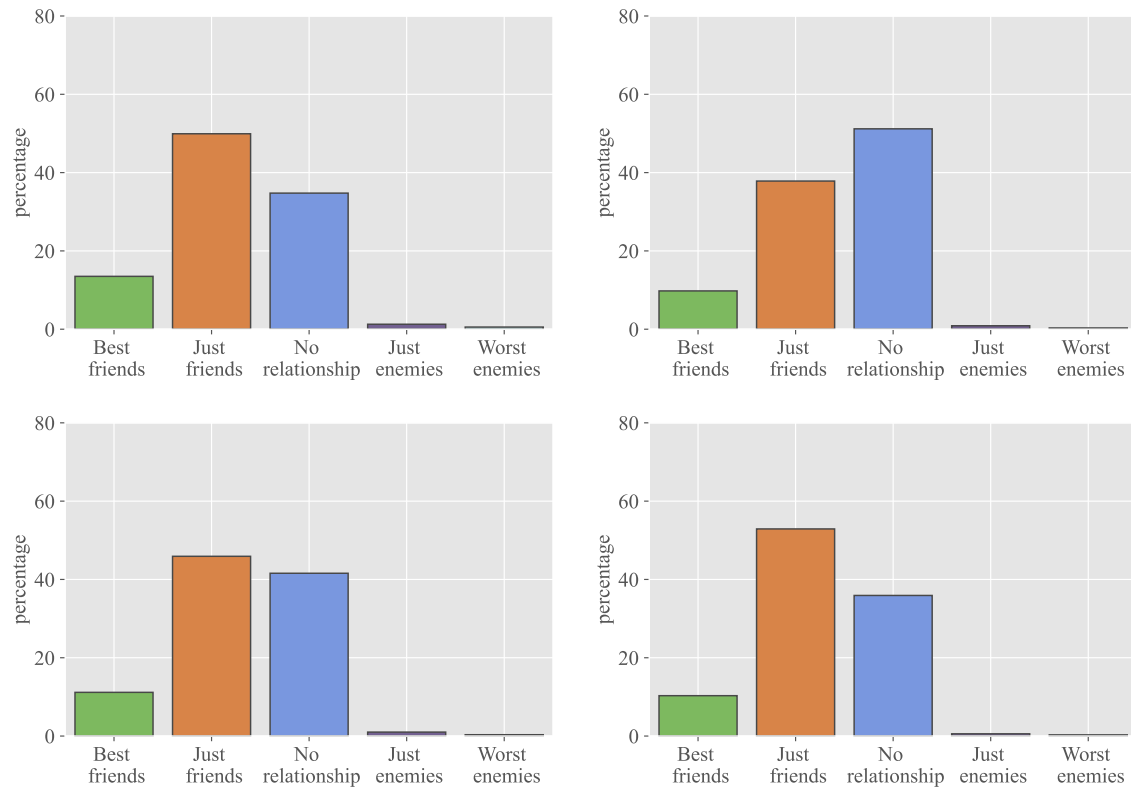
Figure A.15: Friends origin - Percentage of individuals that end up as "friends" in wave n and were marked in any category in the previous wave (conditional probability $P(x, w_{n-1}|1, w_n)$). Top left: from wave 1 to wave 2. Top right: from wave 2 to wave 3. Bottom left: from wave 3 to wave 4. Bottom right: from wave 4 to wave 5.
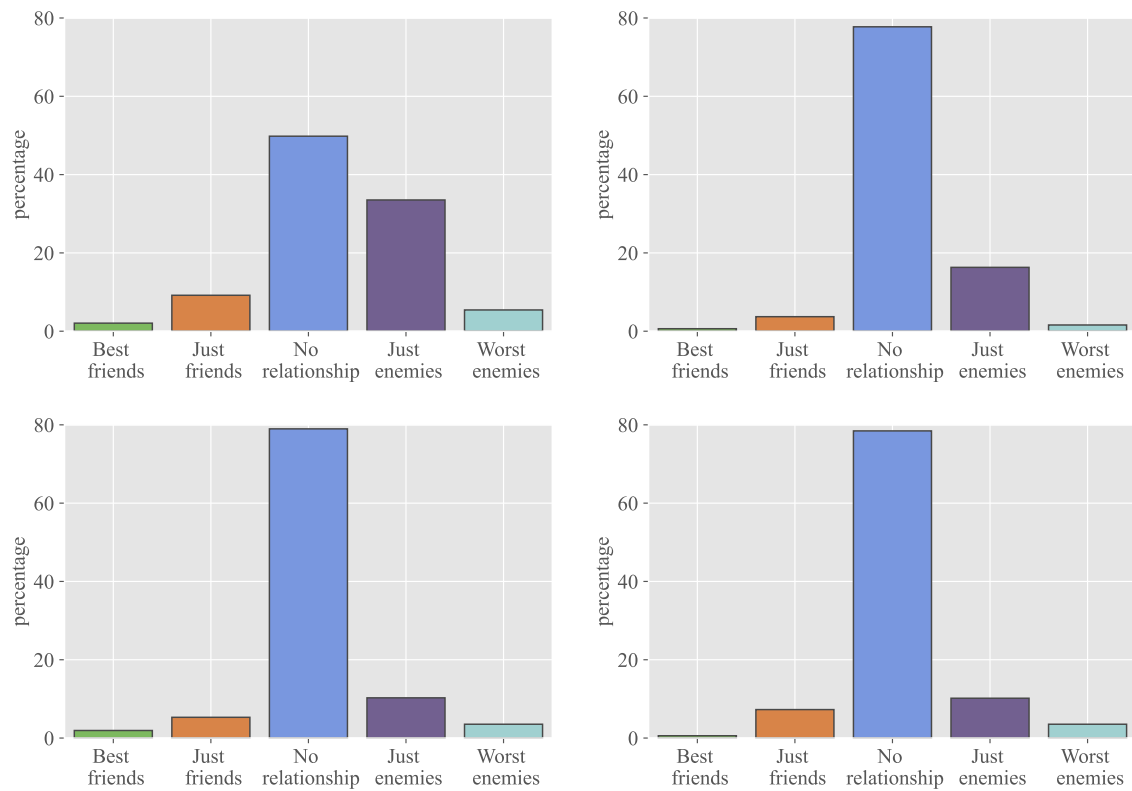
Figure A.16: Enemies origin - Percentage of individuals that end up as "enemies" in wave n and were marked in any category in the previous wave (conditional probability $P(x, w_{n-1}| - 1, w_n)$). Top left: from wave 1 to wave 2. Top right: from wave 2 to wave 3. Bottom left: from wave 3 to wave 4. Bottom right: from wave 4 to wave 5.
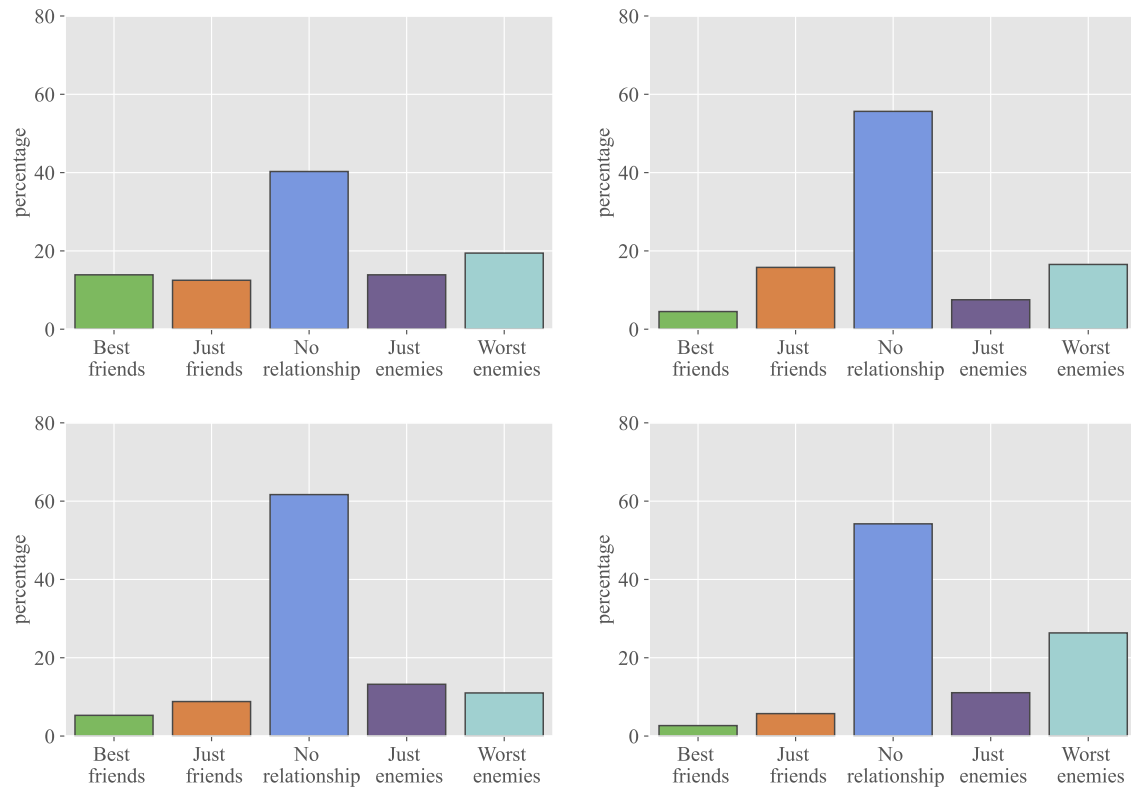
Figure A.17: Worst enemies origin - Percentage of individuals that end up as "worst enemies" in wave n and were marked in any category in the previous wave (conditional probability $P(x, w_{n-1}|-2, w_n)$). Top left: from wave 1 to wave 2. Top right: from wave 2 to wave 3. Bottom left: from wave 3 to wave 4. Bottom right: from wave 4 to wave 5.

A.1.6 Number of S-S, S-D, D-S, and D-D relationships

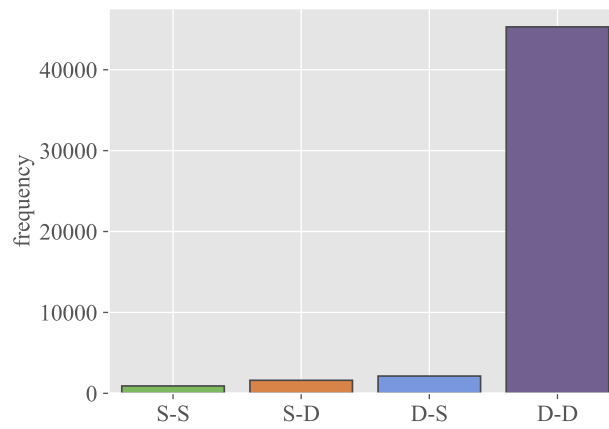**Figure A.18: Number of pairs of relationships** - S-S: same class in both academic years. S-D: same class the first year and different the second. D-S: different class the first year and same the second. D-D: different classes in both academic years.

A.1.7 Reciprocity
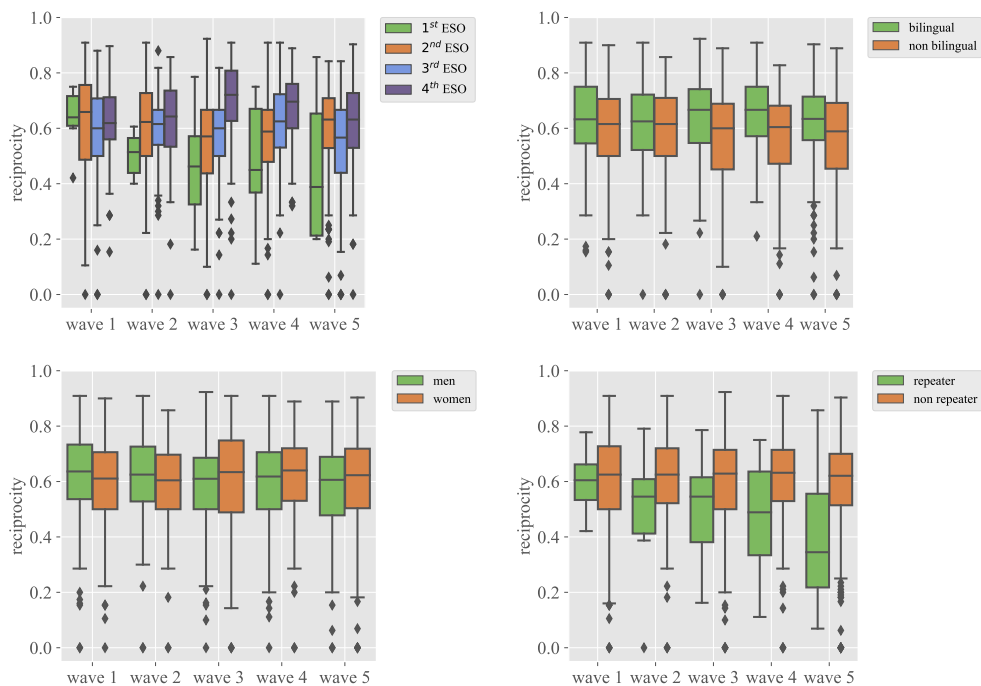
Figure A.19: Reciprocal relationships - Top left: by course. Top right: by itinerary. Bottom left: by gender. Bottom right: by "repetidores" or not.
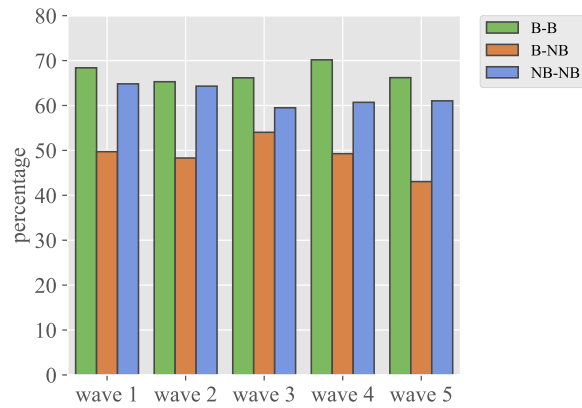
Figure A.20: Reciprocity per itinerary - B-B (bilingüe-bilingüe), B-NB (bilingüe-no bilingüe), NB-NB (no bilingüe-no bilingüe).
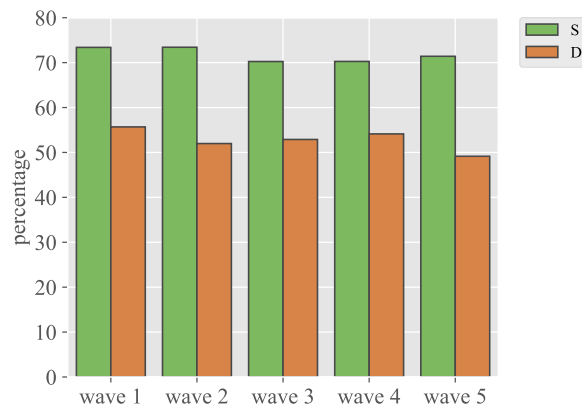
Figure A.21: Reciprocity per group - S (same group), D (different group).
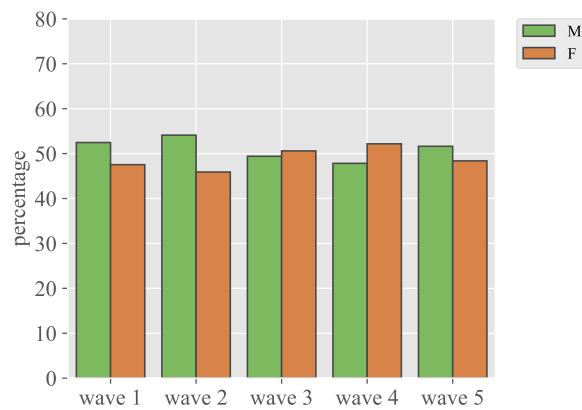
Figure A.22: Reciprocity per sex - Percentage of non-reciprocal relationships that are directed from men to women (green) and from women to men (orange) in each wave.

141

B

APPENDIX FOR CHAPTER 3

B.1   Supplementary tables

B.1.1   Distribution of the students

|	Male	Female	Total
A	9	20	29
B	11	20	31
C	15	15	30
D	19	11	30
E	19	12	31
Total	73	78	151

Table B.1: Distribution of the students by group and gender - The table contains information on the distribution of the students by group and gender. It is valid for the two waves, December 2018 and May 2019.

B.1.2   Community structure

	C1	C2	C3	C4
A	27	2	0	0
B	28	3	0	0
C	0	3	27	0
D	0	10	20	0
E	0	0	2	29
Total	55	18	49	29

	C1	C2	C3
A	28	1	0
B	31	0	0
C	0	8	22
D	2	10	18
E	0	2	29
Total	61	21	69

Table B.2: Distribution of the communities by group in the complete school - The table contains information on wave 1 (left) and wave 2 (right).

	C1	C2	C3	C4
Male	15	16	23	19
Female	40	2	26	10
Total	55	18	49	29

	C1	C2	C3
Male	20	19	34
Female	41	2	35
Total	61	21	69

			May		
		C1	**C2**	**C3**	**Total**
	C1	55	0	0	55
	C2	5	12	1	18
December	**C3**	1	7	41	49
	C4	0	2	27	29
	Total	61	21	69	

Table B.3: Distribution of the communities by gender in the complete school - The table contains information on wave 1 (left), wave 2 (right) and transitions between communities (bottom).

	C1	C2	C3
Male	2	4	3
Female	11	6	3
Total	13	10	6

	C1	C2	C3
Male	0	8	1
Female	12	3	5
Total	12	11	6

		May			
		C1	C2	C3	Total
	C1	5	4	4	13
December	C2	6	4	0	10
	C3	1	3	2	6
	Total	12	11	6	

Table B.4: Distribution of the communities by gender in group A - The table contains information on wave 1 (left), wave 2 (right) and transitions between communities (bottom).

	C1	C2	C3
Male	2	5	4
Female	13	7	0
Total	15	12	4

	C1	C2	C3
Male	1	8	2
Female	10	2	8
Total	11	10	10

		May			
		C1	C2	C3	Total
	C1	10	4	1	15
December	C2	1	2	9	12
	C3	0	4	0	4
	Total	11	10	10	

Table B.5: Distribution of the communities by gender in group B - The table contains information on wave 1 (left), wave 2 (right) and transitions between communities (bottom).

	C1	C2
Male	12	3
Female	5	10
Total	17	13

	C1	C2
Male	12	3
Female	5	10
Total	17	13

		May		
		C1	C2	Total
	C1	16	1	17
December	C2	1	12	13
	Total	17	13	

Table B.6: Distribution of the communities by gender in group C - The table contains information on wave 1 (left), wave 2 (right) and transitions between communities (bottom).

	C1	C2	C3
Male	10	9	0
Female	1	1	9
Total	11	10	9

	C1	C2	C3
Male	10	9	0
Female	1	1	9
Total	11	10	9

		May			
		C1	C2	C3	Total
	C1	2	9	0	11
December	C2	8	1	1	10
	C3	1	0	8	9
	Total	11	10	9	

Table B.7: Distribution of the communities by gender in group D - The table contains information on wave 1 (left), wave 2 (right) and transitions between communities (bottom).

	C1	C2	C3
Male	9	8	2
Female	3	2	7
Total	12	10	9

	C1	C2	C3
Male	11	2	6
Female	0	8	4
Total	11	10	10

		May			
		C1	C2	C3	Total
	C1	5	2	5	12
December	C2	4	4	2	10
	C3	2	4	3	9
	Total	11	10	10	

Table B.8: Distribution of the communities by gender in group E - The table contains information on wave 1 (left), wave 2 (right) and transitions between communities (bottom).

B.1.3 Social balance

	wave 1				wave 2			
# positive links	0	1	2	3	0	1	2	3
option a	0	15	106	2974	1	31	195	4774
option b	341	1403	1696	2974	1737	4974	2501	4774

Table B.9: Distribution of the triangles in the network, considering positive and negative links - The table contains information on the number of observed triangles formed by a mix of positive and negative links in the first wave (left) and in the second wave (right) for the complete course depending on the criteria.

B.2 Supplementary figures
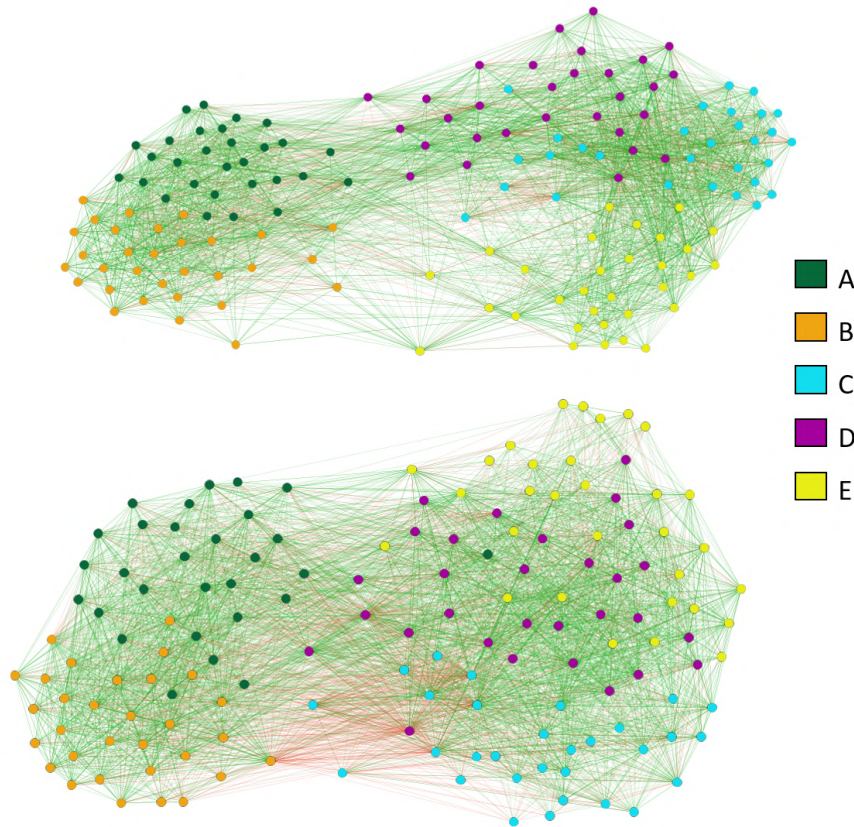
Figure B.1: Network representation of all the social relationships - The figure represents the network corresponding to all the social relationships in wave 1 (top) and wave 2 (bottom) with division by group.
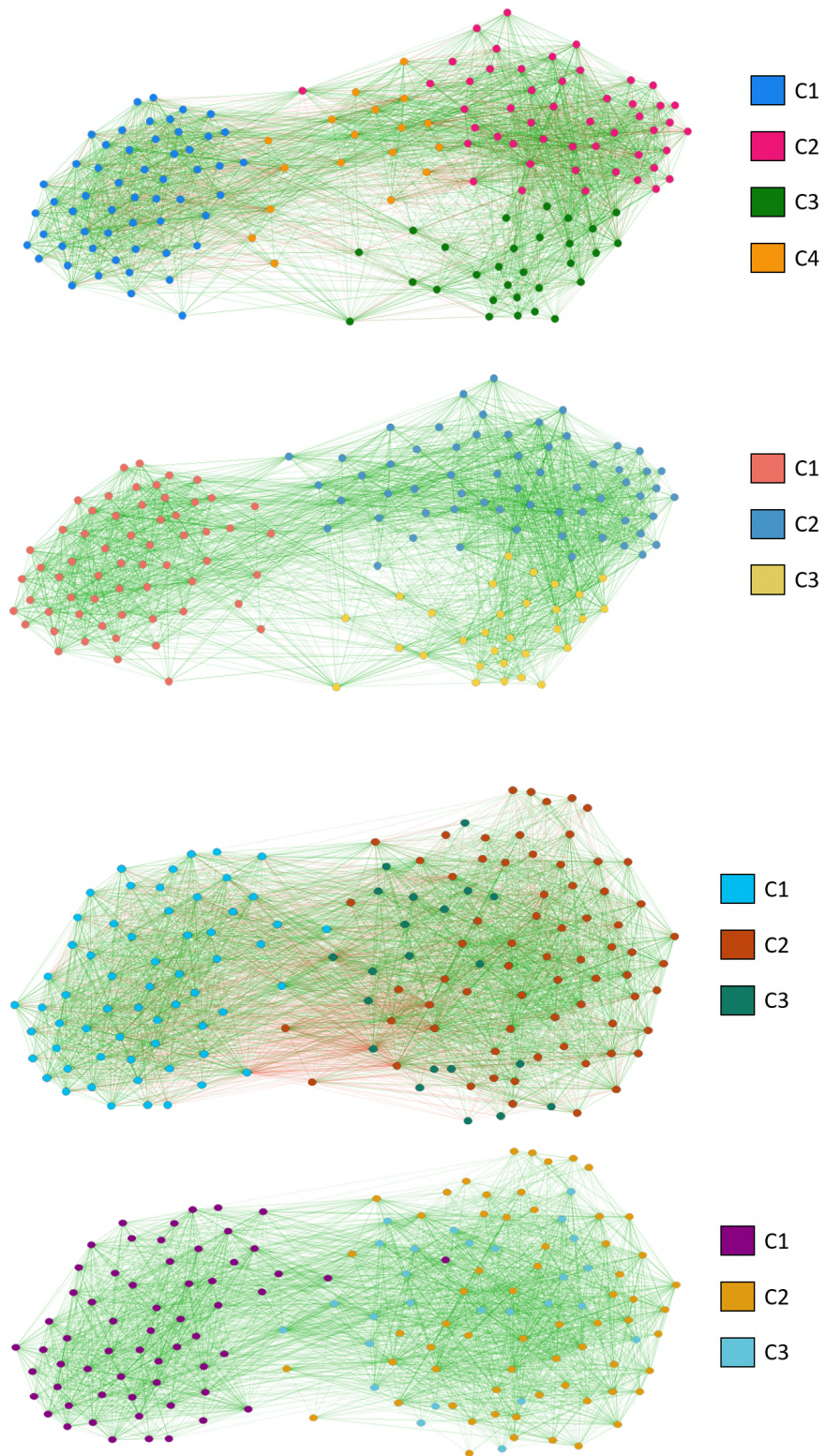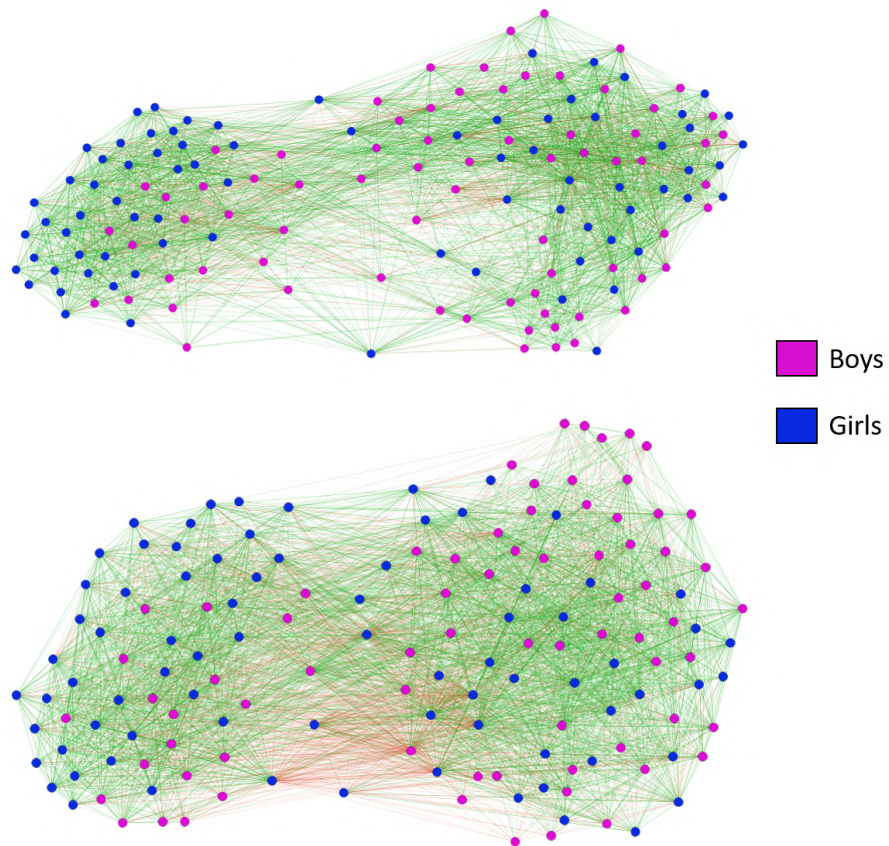
Figure B.2: Network representation of the communities - The figure represents the community analysis results when considering only positive links and all links. From top to bottom, all links in wave 1, only positive links in wave 1, all links in wave 2 and only positive links in wave 2.

Figure B.3: Network representation of all the social relationships with division by gender - The figure represents the network corresponding to all the social relationships in wave 1 (top) and wave 2 (bottom) with division by gender.

C

Appendix for Chapter 4

C.1 Supplementary information

C.1.1 Original emails with the feedback received

In this section, we present the main emails received with feedback on our application from the guidance team at the high school. The first of these was received on October 12th, 2021, two weeks after we first granted them access to the application. Its original content, written in Spanish, is the following:

Hola a todos,

En primer lugar, daros la enhorabuena por vuestro trabajo.

La aplicación me parece de mucha utilidad porque nos da mucha información acerca de las relaciones de los alumnos. Para mí, en orientación, que no doy clase, me da muchísima información de los alumnos y de cómo se relacionan en su grupo y en otros grupos.

He usarla observando a alumnos que tenemos en seguimiento por varios motivos.

Me ha resultado de utilidad saber que un alumno absentista es bien valorado por un compañero, y yo no lo sabía, ya que podré buscar ayuda en él de cara a la motivación para que asista con más frecuencia al IES.

También he confirmado, que tenía dudas, del rechazo que provoca un alumno respecto a sus

compañeros y me ha sorprendido ver que hay otro alumno más aceptado de lo que yo pensaba.

Me resulta de utilidad ver que hay alumnas con situaciones emocionales graves que tienen más apoyos de los que pensaba, y saber con quién hay reciprocidad me resulta de mucha utilidad para poder buscar alumnos acompañantes para ellas.

También he podido ver cómo han ido mejorando algunas relaciones del curso pasado a este en algún alumno concreto que me apetecía saber si había mejorado sus relaciones.

Respecto a los alumnos de 1º ESO, que son nuevos en el centro, me parece muy gráfico poder verlo con los tutores y prevenir situaciones porque ya se detectan personas vulnerables según vuestro cuestionario.

Estoy segura de que con más tiempo (solo he dedicado esta tarde) podré sacarle mucha más utilidad. Como por ejemplo, tener en cuenta los datos para hacer agrupaciones. Eso no sé cómo se haría.

Reitero mi enhorabuena y mi agradecimiento por compartir conmigo los datos, porque cada vez que entreviste a un alumno, voy a tener la posibilidad de contrastar los datos ofrecidos en la entrevista con la dinámica de sus relaciones y me va a ofrecer mucha información.

Muchas gracias a todos.

The second email is more recent, dated January 29th, 2023. It reflects the usage of the application by the institute and its usefulness in improving coexistence within it. Like the previous one, it is also written in Spanish, as communication has been conducted in this language:

GRACIAS.

Y quiero comenzar así, dándoos las gracias.

Mi mayor interés en esta segunda toma era por dos casos, principalmente. Luego veré más alumnos, pero mi preocupación eran dos alumnos porque estamos preocupados y, gracias a vuestras visualizaciones, me da mucha información, tanta información, que he estado más de dos horas solo con estos dos alumnos. Me he hecho una tabla de resultados para poder verlo todo en detalle.

Me alegra confirmar que debemos seguir ocupados con uno de ellos, pero bajar el nivel de preocupación. Y tenemos que seguir preocupados con el otro caso.

Me da pistas sobre a quién debo poner mis ojos en el patio y en clase (le diré al equipo educativo que vigile) y sobre a quién debo reforzar relaciones, porque ambos alumnos en cuestión no se dan cuenta de que son bien valorados por más gente de lo que creen. Por supuesto, intentaré reforzar esas relaciones sin dar detalles de estas encuestas, le diré al equipo educativo que en los trabajos en equipo pongan juntos a quienes yo estoy viendo que se reforzarían, en las parejas de educación física, incluso en las mesas de clase... Hay muchas formas de intentar reforzar relaciones sin que ellos sepan que es por lo que han puesto en las encuestas... De ahí la frase típica de "que parezca un accidente...".

Así que MUCHAS GRACIAS.

La pena es que no puedo dedicar todo el tiempo que quiero a la aplicación.

Os seguiré diciendo.

Un saludo.

Appendix for Chapter 5

D.1 Supplementary calculus

D.1.1 Confidence intervals for the parameters

We define the variable $\ell(G)$ as

$$\ell(G) = -\log P(G) = -H(G) + \Omega(G). \tag{D.1}$$

The first derivative of this function with respect to each parameter, evaluated at the maximum likelihood values, satisfies

$$\left(\frac{\partial \ell(G)}{\partial E_{ij}} \bigg|_{E=E_m} \right) = 0. \tag{D.2}$$

The Hessian matrix $\mathcal{H}(G)$ is constructed by computing the second derivatives of $\ell(G)$ with respect to the parameters analytically and evaluating them at the empirical efficacy values E_m, that is

$$\mathcal{H}(G) = \left(\frac{\partial^2 \ell(G)}{\partial E_{ij} \, \partial E_{kl}} \bigg|_{E=E_m} \right). \tag{D.3}$$

The second derivatives satisfy

$$\frac{\partial^2 \ell(G)}{\partial E_{ij} \, \partial E_{kl}} = \left[-\frac{\partial^2 H(G)}{\partial E_{ij} \, \partial E_{kl}} + \frac{\partial^2 \Omega(G)}{\partial E_{ij} \, \partial E_{kl}} \right] = \frac{\partial^2 \Omega(G)}{\partial E_{ij} \, \partial E_{kl}}, \tag{D.4}$$

since the Hamiltonian depends linearly on the efficacies.

Finally, the variance-covariance matrix $V(G)$ is

$$V(G) = [-\mathcal{H}(G)]^{-1}. \tag{D.5}$$

The diagonal entries of this matrix correspond to the variances associated with the parameters E_{ij}.

For the linear model with reciprocity, where the efficacies are the sum of other parameters, the following properties are considered when finding the variances of each parameter:

1. $\mathrm{Var}(X + Y + Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z) + 2\,\mathrm{Cov}(X, Y) + 2\,\mathrm{Cov}(X, Z) + 2\,\mathrm{Cov}(Y, Z)$,

2. $\mathrm{Var}(aX) = a^2\,\mathrm{Var}(X)$,

3. $\mathrm{Cov}(a\,X, b\,Y) = a\,b\,\mathrm{Cov}(X, Y)$.

We assume that the parameters follow a normal distribution (due to it being the most unbiased hypothesis). Therefore, this expression for the confidence intervals holds:

$$\hat{E}_{ij} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} < \langle E_{ij} \rangle < \hat{E}_{ij} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}, \tag{D.6}$$

where $\hat{E}_{ij}$ represents the estimated value of the parameter, $\langle E_{ij} \rangle$ is the mean value, and $z_{\alpha/2}$ is the value of a normal distribution with 0 means and variance 1 that leaves a percentage of $\alpha/2$ on its right side.

Therefore, calculating the confidence intervals of the efficacy parameters in different models is reduced to computing the Hessian matrix in each and applying the procedure outlined in this section.

D.1.1.1   Hessian matrix for the pairwise model

In the pairwise model, the entries of the Hessian matrix, evaluated at the maximum likelihood parameters, satisfy

$$\left(\left. \frac{\partial^2 \Omega}{\partial E_{ij}\, \partial E_{kl}} \right|_{E = E_m} \right) = -\langle R_{ij} \rangle \left(\delta_{i,k} \delta_{j,l} - \frac{\langle R_{kl} \rangle}{\binom{N}{2}} \right). \tag{D.7}$$

D.1.1.2   Hessian matrix for the linear model with reciprocity

In the linear model with reciprocity, the entries of the Hessian matrix are defined by

$$\frac{\partial^2 \Omega}{\partial \lambda^2} = -2\binom{N}{2} \frac{(e^{-\mu} + 1)\,e^{-(\lambda + 2\mu + \beta)}\left(e^{-(2\lambda + 4\mu + \beta)} + (e^{-\mu} + 1)(2e^{-(\lambda + 2\mu)} + e^{-\mu} + 1)\right)}{\left((1 + e^{-\mu})^2 + e^{-(\lambda + 2\mu + \beta)}\left(2 + 2e^{-\mu} + e^{-(\lambda + 2\mu)}\right)\right)^2},$$

$$\frac{\partial^2 \Omega}{\partial \lambda \partial \mu} = \frac{\partial^2 \Omega}{\partial \mu \partial \lambda} = -2\binom{N}{2} \frac{(e^{-\mu} + 2)\,e^{-(\lambda + 2\mu + \beta)}\left(e^{-(2\lambda + 4\mu + \beta)} + (e^{-\mu} + 1)(2e^{-(\lambda + 2\mu)} + e^{-\mu} + 1)\right)}{\left((1 + e^{-\mu})^2 + e^{-(\lambda + 2\mu + \beta)}\left(2 + 2e^{-\mu} + e^{-(\lambda + 2\mu)}\right)\right)^2},$$

$$\frac{\partial^2\Omega}{\partial\lambda\partial\beta} = \frac{\partial^2\Omega}{\partial\beta\partial\lambda} = -2\binom{N}{2}\frac{\left(e^{-\mu}+1\right)^2\left(e^{-(\lambda+2\mu)}+e^{-\mu}+1\right)e^{-(\lambda+2\mu+\beta)}}{\left(\left(1+e^{-\mu}\right)^2+e^{-(\lambda+2\mu+\beta)}\left(2+2e^{-\mu}+e^{-(\lambda+2\mu)}\right)\right)^2},$$

$$\frac{\partial^2\Omega}{\partial\mu^2} = -2\binom{N}{2}\frac{e^{-\mu}(4e^{-(\lambda+\mu+\beta)}+11e^{-(\lambda+2\mu+\beta)}+2e^{-(2\lambda+4\mu+2\beta)}+8e^{-(\lambda+3\mu+\beta)}}{\left(\left(1+e^{-\mu}\right)^2+e^{-(\lambda+2\mu+\beta)}\left(2+2e^{-\mu}+e^{-(\lambda+2\mu)}\right)\right)^2}$$

$$+\frac{8e^{-(2\lambda+3\mu+\beta)}e^{-(\lambda+4\mu+\beta)}+9e^{-(2\lambda+4\mu+\beta)}+2e^{-(2\lambda+5\mu+\beta)}+4e^{-(3\lambda+5\mu+2\beta)}}{\left(\left(1+e^{-\mu}\right)^2+e^{-(\lambda+2\mu+\beta)}\left(2+2e^{-\mu}+e^{-(\lambda+2\mu)}\right)\right)^2}$$

$$+\frac{e^{-(3\lambda+6\mu+2\beta)}+\left(1+e^{-\mu}\right)^2)}{\left(\left(1+e^{-\mu}\right)^2+e^{-(\lambda+2\mu+\beta)}\left(2+2e^{-\mu}+e^{-(\lambda+2\mu)}\right)\right)^2},$$

$$\frac{\partial^2\Omega}{\partial\mu\partial\beta} = \frac{\partial^2\Omega}{\partial\beta\partial\mu} = -2\binom{N}{2}\frac{\left(e^{-\mu}+1\right)\left(e^{-\mu}+2\right)\left(e^{-(\lambda+2\mu)}+e^{-\mu}+1\right)e^{-(\lambda+2\mu+\beta)}}{\left(\left(1+e^{-\mu}\right)^2+e^{-(\lambda+2\mu+\beta)}\left(2+2e^{-\mu}+e^{-(\lambda+2\mu)}\right)\right)^2},$$

$$\frac{\partial^2\Omega}{\partial\beta^2} = -\binom{N}{2}\frac{\left(e^{-\mu}+1\right)^2\left(2e^{-(\lambda+2\mu+\beta)}+2e^{-(\lambda+3\mu+\beta)}+e^{-(2\lambda+4\mu+\beta)}\right)}{\left(\left(1+e^{-\mu}\right)^2+e^{-(\lambda+2\mu+\beta)}\left(2+2e^{-\mu}+e^{-(\lambda+2\mu)}\right)\right)^2}.$$

D.1.2 Estimation and confidence intervals for the number of triangles

D.1.2.1 Mean and variance of triangles with homogeneous links

To compute the mean and variance of the number of triangles in a network, we first define $\mathcal{N} = \{1,\ldots,N\}$ as the set of nodes of the network and the variable

$$X_{ijk} = \begin{cases} 1, & \text{if the nodes } i,j,k \text{ form a triangle,} \\ 0, & \text{otherwise.} \end{cases} \tag{D.8}$$

The total number of triangles in the network is represented by

$$T = \sum_{\{i,j,k\}\subset\mathcal{V}^3} X_{ijk}, \tag{D.9}$$

where the sum is taken over all subsets of three indices from $\mathcal{N}$.

If a link between nodes i and j in set $\mathcal{N}$ has a probability p, which is independent of other links, then X_{ijk} is a Bernoulli random variable with probability p^3. This results in an expected value of T defined by

$$\langle T \rangle = \binom{N}{3}p^3, \tag{D.10}$$

as p^3 is the probability of a triangle forming between three nodes and $\binom{N}{3}$ is the number of subsets of three indices in $\mathcal{N}$.

To calculate the variance, we need to obtain the average of T^2, which is

$$\langle T^2 \rangle = \sum_{\{i,j,k\} \subset \mathcal{V}^3} \sum_{\{i',j',k'\} \subset \mathcal{V}^3} \langle X_{ijk} X_{i'j'k'} \rangle, \tag{D.11}$$

a sum we have to calculate taking into account that when the indices match in both subsets, the values of X_{ijk} y $X_{i'j'k'}$ are not independent.
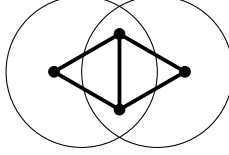
Therefore, we group them as follows:

(a) $\{i,j,k\} = \{i',j',k'\}$. In this case $X_{ijk} = X_{i'j'k'}$, so

$$\langle X_{ijk} X_{i'j'k'} \rangle = p^3. \tag{D.12}$$

The total number of such terms in the sum is $\binom{N}{3}$.

(b) The intersection of the sets $\{i,j,k\}$ and $\{i',j',k'\}$ contains exactly two indices when two triangles share one of their links, as illustrated in the figure:

In this case

$$\langle X_{ijk} X_{i'j'k'} \rangle = p^5, \tag{D.13}$$

and the number of terms in the sum is equal to $\binom{N}{4} \times 12$, which represents the number of ways to choose four indices to form the two triangles multiplied by the number of ways to place two of the four indices of the intersection ($\binom{4}{2} = 6$) and the number of ways of placing the two remaining indices in the other two nodes.

(c) The intersection of the sets $\{i,j,k\}$ and $\{i',j',k'\}$ contains exactly one index when the two triangles share a vertex, as illustrated in the figure:
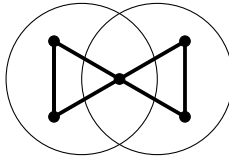
In this case

$$\langle X_{ijk} X_{i'j'k'} \rangle = p^6, \tag{D.14}$$

and the number of terms in the sum is equal to $\binom{N}{5} \times 30$, which represents the number of ways to choose five indices to form the two triangles multiplied by the number of ways to pick one index for the intersection (5) and two more indices for one of the sets ($\binom{4}{2} = 6$). The remaining two indices go to the other set in a unique way.

(d) The intersection between $\{i,j,k\}$ and $\{i',j',k'\}$ is empty, meaning the triangles are not connected. In this case

$$\langle X_{ijk} X_{i'j'k'} \rangle = p^6, \tag{D.15}$$

and the number of terms in the sum is equal to $\binom{N}{6} \times 20$, which represents the number of ways to choose six indices to form the two triangles multiplied by the number of ways to distribute 6 elements into two sets such that the sets do not share any element ($\binom{6}{3} = 20$).

To sum up,

$$\langle T^2 \rangle = \binom{N}{3}p^3 + 12\binom{N}{4}p^5 + 30\binom{N}{5}p^6 + 20\binom{N}{6}p^6. \tag{D.16}$$

It can be verified (using induction, for example) that by setting $p = 1$ in the formula above

$$\binom{N}{3} + 12\binom{N}{4} + 30\binom{N}{5} + 20\binom{N}{6} = \binom{N}{3}^2, \tag{D.17}$$

which is the total number of possible ways to choose two pairs of subsets, each consisting of three arbitrary indices from $\mathcal{N}$.

As previously discussed, $\langle T \rangle$ is determined by Equation D.10, so

$$\langle T \rangle^2 = \binom{N}{3}^2 p^6 = \left[\binom{N}{3} + 12\binom{N}{4} + 30\binom{N}{5}, +20\binom{N}{6}\right]p^6, \tag{D.18}$$

using Equation D.17.

Thus, the variance of T turns out to be

$$\langle T^2 \rangle - \langle T \rangle^2 = \binom{N}{3}p^3(1 - p^3) + 12\binom{N}{4}p^5(1 - p). \tag{D.19}$$

It is noteworthy that the initial term of this sum represents the variance of the combination of $\binom{N}{3}$ Bernoulli variables, each having a value of 1 with a probability p^3 and 0 otherwise. The additional term provides an adjustment that does not seem negligible.

D.1.2.2   Mean and variance of triangles with heterogeneous links

Now, the links can be classified into two categories, 1 and 2, and each category can occur with distinct probabilities, p_1 and p_2. First, we will determine the number of triangles that have two links of type 1 and one link of type 2, that is

$$X_{ijk} = \begin{cases} 1, & \text{if } i, j, k \text{ form a triangle with two links of type 1 and one of type 2,} \\ 0, & \text{otherwise.} \end{cases} \tag{D.20}$$

The expected value of T is

$$\langle T \rangle = 3\binom{N}{3}p_1^2 p_2, \tag{D.21}$$

where the factor 3 accounts for the three possible positions where the link of type 2 could be placed within the triangle.

Concerning the variance, we repeat the same group of terms that were used in the sum given by Equation D.11:
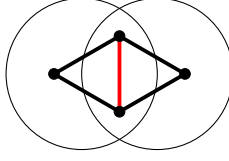
(a) $\{i, j, k\} = \{i', j', k'\}$. In this case $X_{ijk} = X_{i'j'k'}$, so that

$$\langle X_{ijk} X_{i'j'k'} \rangle = p_1^2 p_2. \tag{D.22}$$

The number of terms within the sum is $\binom{N}{3} \times 3$.

(b) The intersection of the sets $\{i, j, k\}$ and $\{i', j', k'\}$ includes only two indices. Depending on where this type 2 link appears in the triangles, there are two possible scenarios. The first is illustrated in the figure:
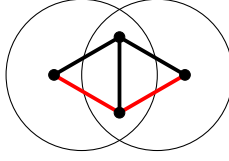

In this case

$$\langle X_{ijk} X_{i'j'k'} \rangle = p_1^4 p_2, \tag{D.23}$$

and the number of terms in the sum is $\binom{N}{4} \times 12$, just as it is in the case of homogeneous triangles.

The second scenario is illustrated in the following figure:

In this case

$$\langle X_{ijk} X_{i'j'k'} \rangle = p_1^3 p_2^2, \tag{D.24}$$

and the number of terms in the sum is $\binom{N}{4} \times 12$, as in the case of homogeneous triangles, multiplied by the 4 different ways of positioning the type 2 links in the two triangles.

(c) The intersection of the sets $\{i, j, k\}$ and $\{i', j', k'\}$ includes only one element. This is the situation where a single vertex connects two triangles, as shown in the figure:
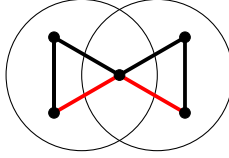
In this case

$$\langle X_{ijk} X_{i'j'k'} \rangle = p_1^4 p_2^2, \tag{D.25}$$

and the number of terms in the sum is equal to $\binom{N}{5} \times 30$, as in the case of homogeneous triangles, multiplied by the 9 different ways of positioning the type 2 links in the two triangles.

(d) The sets $\{i, j, k\}$ and $\{i', j', k'\}$ have no common elements. The triangles are not connected in any way. In this scenario

$$\langle X_{ijk} X_{i'j'k'} \rangle = p_1^4 p_2^2, \tag{D.26}$$

and the number of terms in the sum is $\binom{N}{6} \times 20$, as in the case of homogeneous triangles, multiplied by the 9 different ways of positioning the type 2 links in the two unconnected triangles.

To sum up,

$$\langle T^2 \rangle = 3\binom{N}{3} p_1^2 p_2 + 12\binom{N}{4} p_1^4 p_2 + 48\binom{N}{4} p_1^3 p_2^2 + 270\binom{N}{5} p_1^4 p_2^2 + 180\binom{N}{6} p_1^4 p_2^2. \tag{D.27}$$

By subtracting the square of the formula in Equation D.21 and using Equation D.17,

$$\langle T^2 \rangle - \langle T \rangle^2 = \binom{N}{3} 3 p_1^2 p_2 (1 - 3 p_1^2 p_2) + 12\binom{N}{4} p_1^3 p_2 (p_1 + 4 p_2 - 9 p_1 p_2). \tag{D.28}$$

Once more, the initial term of this sum corresponds to the variance of the total of $\binom{N}{3}$ Bernoulli variables, which each take on a value of 1 with probability $3 p_1^2 p_2$ and a value of 0 otherwise, and the second term adds a correction.

D.1.2.3   Confidence intervals for the number of triangles

We assume that the links in the triangles follow a normal distribution. The confidence intervals are determined by

$$\hat{T} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \langle T \rangle < \hat{T} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \tag{D.29}$$

where $\hat{T}$ represents the estimated value of the parameter, $\langle T \rangle$ is the mean value, and $z_{\alpha/2}$ is the value of a normal distribution that leaves a percentage of $\alpha/2$ on its right side.

D.1.3   Transformation for the linear model with reciprocity

We are working with the Hamiltonian on the graph G

$$-H(G) = \sum_{\langle ab \rangle} \left[\lambda(2 - \delta_{\sigma_{ab},0} - \delta_{\sigma_{ba},0}) + \mu(\sigma_{ab} + \sigma_{ba}) + \beta(1 - \delta_{\sigma_{ab},0})(1 - \delta_{\sigma_{ba},0}) \right] + \frac{\gamma}{N} \sum_{\langle abc \rangle} T_{abc},$$

where $\sigma_{ab} \in \{0, 1, ..., r\}$, $\langle ab \rangle$ denotes links in G, $\langle abc \rangle$ denotes triangles in G and $T_{abc} = 1$ if there is a reciprocal triangle spanning the three nodes a, b and c of the triangle and $T_{abc} = 0$ otherwise. In other words,

$$T_{abc} \equiv (1 - \delta_{\sigma_{ab},0})(1 - \delta_{\sigma_{ba},0})(1 - \delta_{\sigma_{ac},0})(1 - \delta_{\sigma_{ca},0})(1 - \delta_{\sigma_{bc},0})(1 - \delta_{\sigma_{cb},0}). \tag{D.30}$$

The grand partition function Ξ is calculated by summing over all possible configurations of the system, which means

$$\Xi = \sum_{\{\sigma_{ab},\sigma_{ba}\}} e^{-H(G)}. \tag{D.31}$$

Let us now define on every pair $\langle ab \rangle$ the new variables

$$\tau_{ab} \equiv (1 - \delta_{\sigma_{ab},0})(1 - \delta_{\sigma_{ba},0}). \tag{D.32}$$

The variables in question can take only two values, namely $\tau_{ab} = 0$ and $\tau_{ab} = 1$. The value of τ_{ab} depends on whether $\sigma_{ab}\sigma_{ba} = 0$ or $\sigma_{ab}\sigma_{ba} \neq 0$. If we consider the sum over the pair in the grand partition function defined by Equation D.31, then:

$$S_{ab}(\boldsymbol{\tau}) \equiv \sum_{\sigma_{ab}=0}^{r} \sum_{\sigma_{ba}=0}^{r} \exp\left\{\lambda(2 - \delta_{\sigma_{ab},0} - \delta_{\sigma_{ba},0}) + \mu(\sigma_{ab} + \sigma_{ba}) + \tau_{ab}\left[\beta + J_{ab}(\boldsymbol{\tau})\right]\right\}, \tag{D.33}$$

where, for simplicity, we have introduced the short-hand

$$J_{ab}(\boldsymbol{\tau}) \equiv \frac{\gamma}{N} \sum_{c \neq a,b} \tau_{ac}\tau_{bc}. \tag{D.34}$$

We can divide this sum into two parts as follows:

$$S_{ab}(\boldsymbol{\tau}) = \begin{cases} \displaystyle\sum_{\sigma_{ab},\sigma_{ba}:\tau_{ab}=0} \exp\left\{\lambda(2 - \delta_{\sigma_{ab},0} - \delta_{\sigma_{ba},0}) + \mu(\sigma_{ab} + \sigma_{ba})\right\}, & \text{if } \tau_{ab} = 0, \\ \displaystyle\sum_{\sigma_{ab},\sigma_{ba}:\tau_{ab}=1} \exp\left\{\lambda(2 - \delta_{\sigma_{ab},0} - \delta_{\sigma_{ba},0}) + \mu(\sigma_{ab} + \sigma_{ba}) + \left[\beta + J_{ab}(\boldsymbol{\tau})\right]\right\}, & \text{if } \tau_{ab} = 1, \end{cases}$$

and after performing the sums

$$S_{ab}(\boldsymbol{\tau}) = \begin{cases} 1 + 2e^{\lambda+\mu}\displaystyle\sum_{k=0}^{r-1} e^{k\mu}, & \text{if } \tau_{ab} = 0, \\ e^{2\lambda+2\mu+\beta+J_{ab}(\boldsymbol{\tau})}\left(\displaystyle\sum_{k=0}^{r-1} e^{k\mu}\right)^2, & \text{if } \tau_{ab} = 1, \end{cases} \tag{D.35}$$

Therefore, if we define the variable

$$\phi = \beta + 2\lambda + 2\mu - \log\left(1 + 2e^{\lambda+\mu}\frac{1 - e^{r\mu}}{1 - e^{\mu}}\right) + 2\log\left(\frac{1 - e^{r\mu}}{1 - e^{\mu}}\right), \tag{D.36}$$

we can rewrite

$$S_{ab}(\boldsymbol{\tau}) = \left(1 + 2e^{\lambda+\mu}\frac{1 - e^{r\mu}}{1 - e^{\mu}}\right)\sum_{\tau_{ab}=0,1} \exp\left\{\tau_{ab}\left[\phi + J_{ab}(\boldsymbol{\tau})\right]\right\}. \tag{D.37}$$

As a result, we obtain an alternative expression for Equation D.31, which is given by

$$\Xi = \left(1 + 2e^{\lambda+\mu}\frac{1 - e^{r\mu}}{1 - e^{\mu}}\right)^{\binom{N}{2}}\sum_{\boldsymbol{\tau}} e^{-\tilde{H}(G)}, \tag{D.38}$$

where

$$-\tilde{H}(G) \equiv \phi \sum_{\langle ab \rangle} \tau_{ab} + \frac{\gamma}{N} \sum_{\langle abc \rangle} \tau_{ab} \tau_{ac} \tau_{bc}. \tag{D.39}$$

This Hamiltonian corresponds to Strauss's model (Strauss, 1986) for graph G and has been analysed using mean-field techniques by Park and Newman (2005).

D.2 Supplementary tables

	No reciprocal links		1 reciprocal link		2 reciprocal links		3 reciprocal links	
	Real	Model	Real	Model	Real	Model	Real	Model
1 ESO A	17	17 (1;33)	82	71 (37;105)	78	99 (64;134)	**85**	**46 (16;76)**
1 ESO B	5	5 (0;10)	35	30 (8;52)	79	58 (29;87)	56	38 (11;65)
1 ESO C	29	21 (4;38)	81	56 (26;86)	47	50 (22;78)	28	15 (1;29)
1 ESO D	68	74 (33;115)	264	251 (177;325)	277	284 (210;358)	**208**	**107 (54;160)**
1 ESO E	23	11 (0;22)	34	19 (4;34)	6	10 (0;20)	**9**	**2$(-1;5)$**
1 ESO F	21	26 (5;47)	113	103 (60;146)	140	134 (90;178)	73	58 (23;94)
1 ESO G	25	27 (6;48)	55	36 (14;58)	20	16 (2;30)	5	2$(-3;7)$

	No reciprocal links		1 reciprocal link		2 reciprocal links		3 reciprocal links	
	Real	Model	Real	Model	Real	Model	Real	Model
2 ESO A	13	17 (2;32)	103	81 (41;121)	100	129 (84;173)	**137**	**69 (29;109)**
2 ESO B	16	16 (1;31)	67	48 (21;74)	52	48 (21;75)	29	16 (1;31)
2 ESO C	11	12 (0;24)	43	38 (15;61)	54	42 (18;66)	**29**	**15 (1;29)**
2 ESO D	15	14 (0;28)	75	60 (30;90)	50	83 (53;113)	78	38(10;66)
2 ESO E	21	25 (5;45)	83	68 (37;99)	55	61 (31;91)	29	18 (2;34)
2 ESO F	23	17 (2;32)	69	53 (24;82)	63	56 (26;86)	**48**	**19 (3;35)**
2 ESO G	27	28 (7;49)	117	94 (54;134)	101	103 (62;144)	**74**	**38 (11;65)**

	No reciprocal links		1 reciprocal link		2 reciprocal links		3 reciprocal links	
	Real	Model	Real	Model	Real	Model	Real	Model
3 ESO A	89	70 (32;108)	255	209 (140;278)	251	211 (140;278)	**136**	**70 (32;108)**
3 ESO B	5	6 (0;14)	39	25 (7;43)	43	33 (13;53)	**39**	**15 (1;28)**
3 ESO C	2	4 (0;8)	10	12 (1;23)	12	13 (1;25)	**10**	**5 (1;9)**
3 ESO D	55	45 (17;73)	111	88 (50;127)	84	58 (27;89)	**40**	**13 (1;24)**
3 ESO E	4	5 (0;10)	27	27 (8;47)	42	50 (23;77)	**105**	**31 (9;53)**

Table D.1: Distribution of triangles - Comparative between the real number of triangles and those expected when performing a random distribution of the links. The numbers in parentheses correspond to the confidence intervals. Each of the tables above is associated with a different course in wave 1.

D.3   Supplementary figures
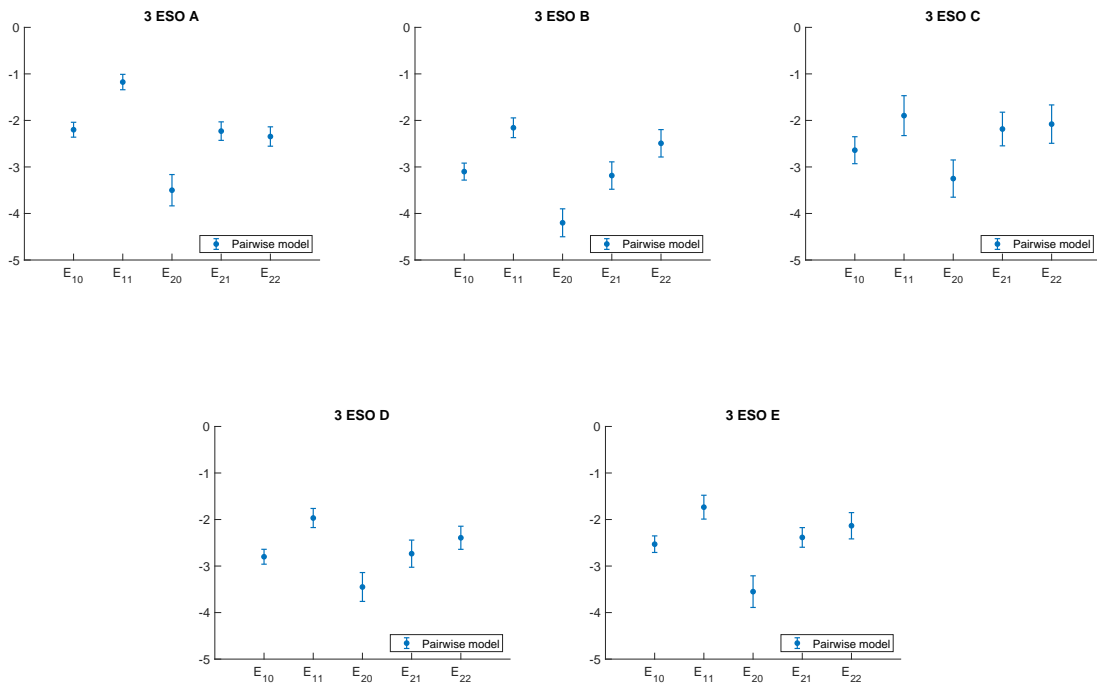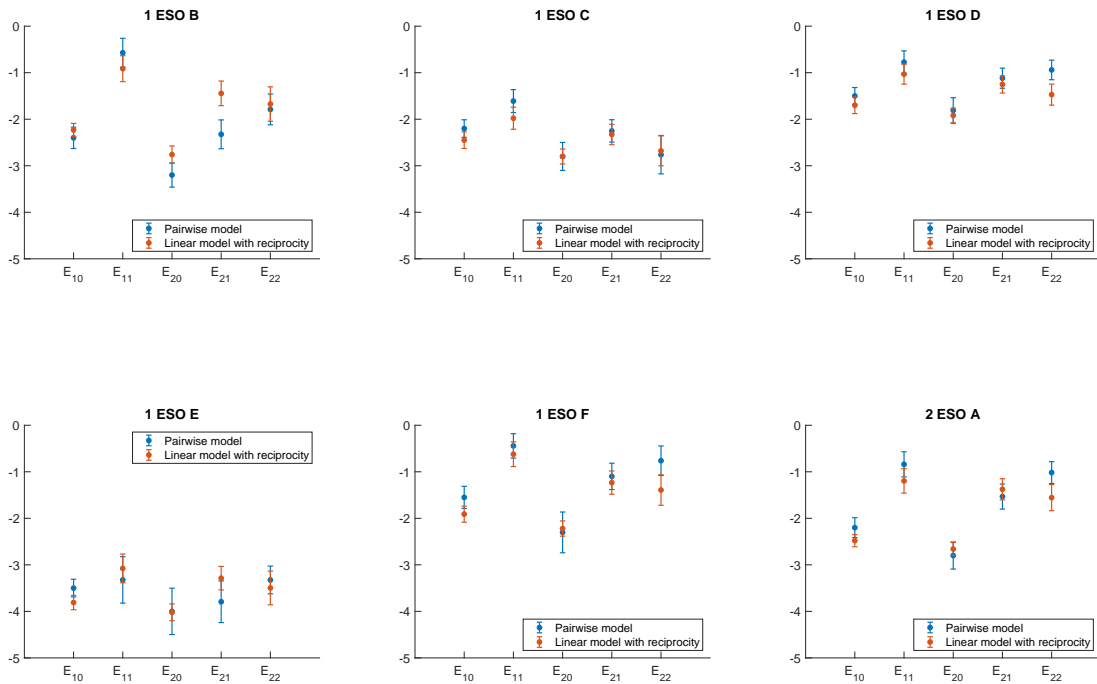
D.3.1   Pairwise model

Figure D.2: Parameters of the pairwise model - Values obtained for the parameters of the pairwise model for each of the remaining groups in wave 1.
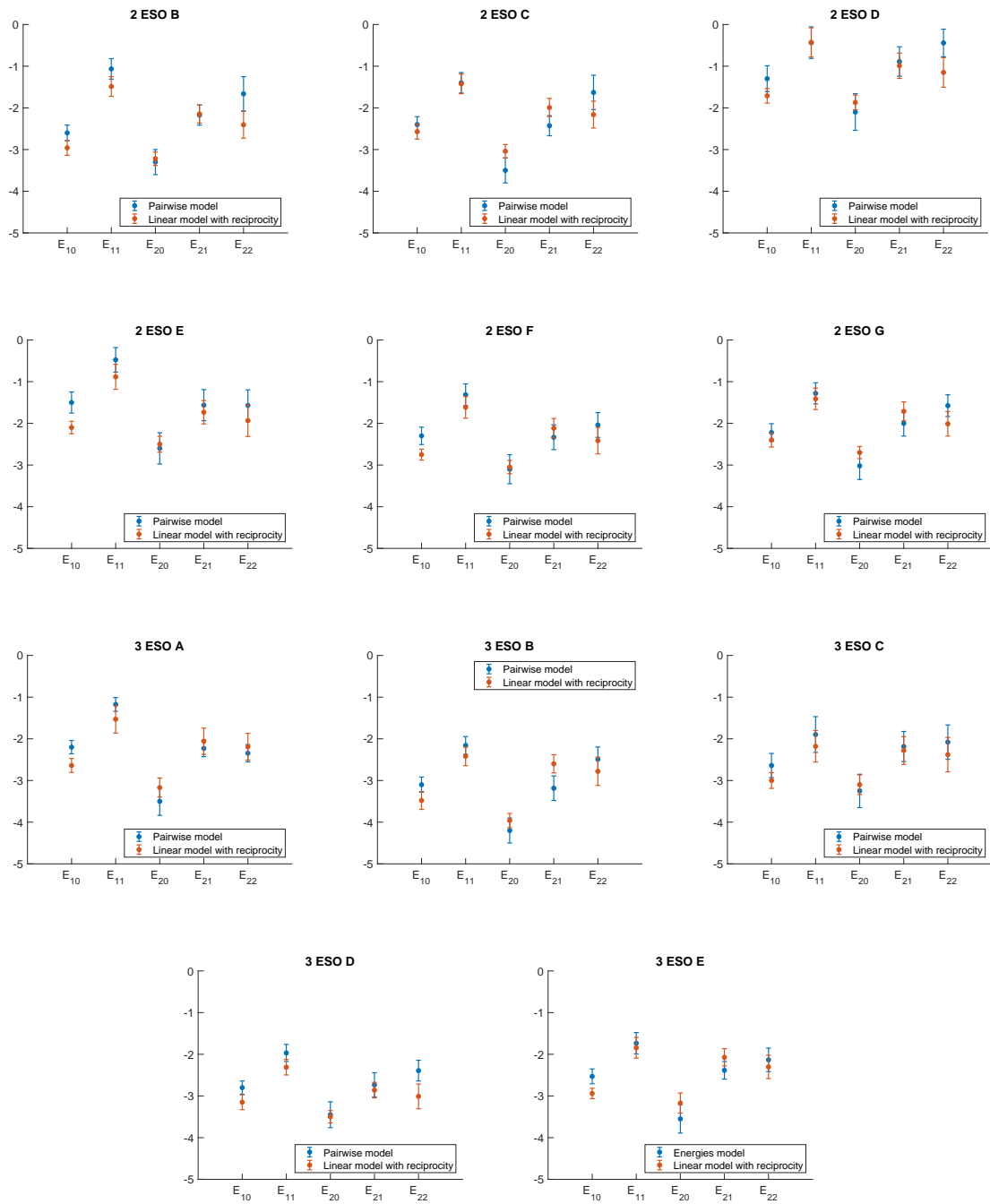
D.3.2 Linear model with reciprocity

Figure D.3: Parameters of the linear model with reciprocity - Values obtained for the parameters of the linear model with reciprocity for each of the remaining groups in wave 1.

APPENDIX FOR CHAPTER 6

E.1 Supplementary calculus

E.1.1 Free energy of a single triangle

For a triangle,

$$\Xi_3 = (1 + z_{ij})(1 + z_{jk})(1 + z_{ki}) + \zeta_{ijk} z_{ij} z_{jk} z_{ki}, \tag{E.1}$$

hence

$$
\begin{aligned}
\rho_{ij} &= \frac{z_{ij}(1 + z_{jk})(1 + z_{ki}) + \zeta_{ijk} z_{ij} z_{jk} z_{ki}}{\Xi_3}, \\
\rho_{jk} &= \frac{z_{jk}(1 + z_{ij})(1 + z_{ki}) + \zeta_{ijk} z_{ij} z_{jk} z_{ki}}{\Xi_3}, \\
\rho_{ki} &= \frac{z_{ki}(1 + z_{ij})(1 + z_{jk}) + \zeta_{ijk} z_{ij} z_{jk} z_{ki}}{\Xi_3}.
\end{aligned}
\tag{E.2}
$$

It will prove convenient to introduce

$$\rho_{ijk} \equiv \frac{\zeta_{ijk} z_{ij} z_{jk} z_{ki}}{\Xi_3}. \tag{E.3}$$

Now, dividing Equation E.1 by Ξ_3 we get

$$\frac{(1 + z_{ij})(1 + z_{jk})(1 + z_{ki})}{\Xi_3} = 1 - \rho_{ijk}. \tag{E.4}$$

On the other hand, Equation E.2 can be rewritten as

$$
\begin{aligned}
\rho_{ij} - \rho_{ijk} &= \frac{z_{ij}(1 + z_{jk})(1 + z_{ki})}{\Xi_3}, \\
\rho_{jk} - \rho_{ijk} &= \frac{z_{jk}(1 + z_{ij})(1 + z_{ki})}{\Xi_3}, \\
\rho_{ki} - \rho_{ijk} &= \frac{z_{ki}(1 + z_{ij})(1 + z_{jk})}{\Xi_3}.
\end{aligned}
\tag{E.5}
$$

Multiplying them out and using Equation E.3 and Equation E.4 leads to Equation 6.12. Also, using Equation E.4 in Equation E.5 we obtain

$$
\begin{aligned}
\rho_{ij} - \rho_{ijk} &= \frac{z_{ij}}{1 + z_{ij}}(1 - \rho_{ijk}), \\
\rho_{jk} - \rho_{ijk} &= \frac{z_{jk}}{1 + z_{jk}}(1 - \rho_{ijk}), \\
\rho_{ki} - \rho_{ijk} &= \frac{z_{ki}}{1 + z_{ki}}(1 - \rho_{ijk}),
\end{aligned}
\tag{E.6}
$$

whose solutions are

$$
z_{ij} = \frac{\rho_{ij} - \rho_{ijk}}{1 - \rho_{ij}}, \quad z_{jk} = \frac{\rho_{jk} - \rho_{ijk}}{1 - \rho_{jk}}, \quad z_{ki} = \frac{\rho_{ki} - \rho_{ijk}}{1 - \rho_{ki}}.
$$

Substituting these expressions in Equation E.3 and using Equation 6.12 we obtain

$$
\Xi_3 = \frac{(1 - \rho_{ijk})^2}{(1 - \rho_{ij})(1 - \rho_{jk})(1 - \rho_{ki})}.
\tag{E.7}
$$

Thus $\Phi_3 = \rho_{ij} \log z_{ij} + \rho_{jk} \log z_{jk} + \rho_{ki} \log z_{ki} - \log \Xi_3$ becomes Equation 6.11.

E.1.2 Probability of forming a triangle

The probability that nodes i, j, k form a triangle can be obtained from the grand potential as

$$
T_{ijk} \equiv \langle \tau_{ij} \tau_{jk} \tau_{ki} \rangle = -N \frac{\partial \Omega}{\partial \gamma_{ijk}}.
\tag{E.8}
$$

But inverting the Legendre transform defined by Equation 6.5,

$$
\Omega = F - \sum_{\{ij\}} \phi_{ij} \rho_{ij},
\tag{E.9}
$$

where ρ_{ij} depends $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$ through Equation 6.7. Thus,

$$
T_{ijk} = \sum_{\{lm\}} \phi_{lm} N \frac{\partial \rho_{lm}}{\partial \gamma_{ijk}} - N \frac{\partial F}{\partial \gamma_{ijk}}.
\tag{E.10}
$$

Now,

$$
\frac{\partial F}{\partial \gamma_{ijk}} = \sum_{\{lm\}} \frac{\partial F}{\partial \rho_{lm}} \frac{\partial \rho_{lm}}{\partial \gamma_{ijk}} + \left(\frac{\partial F}{\partial \gamma_{ijk}} \right)_{\boldsymbol{\rho}},
$$

where the last partial derivative is taken at constant $\boldsymbol{\rho}$. Thus, substituting into Equation E.10 and using Equation 6.7 we obtain

$$T_{ijk} = -N \left(\frac{\partial F}{\partial \gamma_{ijk}} \right)_{\boldsymbol{\rho}} = -(1 + \zeta_{ijk}) \left(\frac{\partial F}{\partial \zeta_{ijk}} \right)_{\boldsymbol{\rho}}. \tag{E.11}$$

Notice that the free energy depends on ζ_{ijk} only through ρ_{ijk} via Equation 6.12, therefore

$$
\begin{aligned}
T_{ijk} = & -(1 + \zeta_{ijk}) \left(\frac{\partial F}{\partial \rho_{ijk}} \right)_{\boldsymbol{\rho}} \left(\frac{\partial \rho_{ijk}}{\partial \zeta_{ijk}} \right)_{\boldsymbol{\rho}} \\
= & (1 + \zeta_{ijk}) \rho_{ijk} \left[\frac{1 - 3\rho_{ijk}}{\rho_{ijk}(1 - \rho_{ijk})} + \frac{1}{\rho_{ij} - \rho_{ijk}} \right. \\
& \left. + \frac{1}{\rho_{jk} - \rho_{ijk}} + \frac{1}{\rho_{ki} - \rho_{ijk}} \right] \left(\frac{\partial \rho_{ijk}}{\partial \zeta_{ijk}} \right)_{\boldsymbol{\rho}}.
\end{aligned}
$$

On the other hand, taking logarithms in Equation 6.12 and differentiating with respect to ζ_{ijk} at constant $\boldsymbol{\rho}$ we get

$$
\begin{aligned}
\frac{1}{\zeta_{ijk}} = & \left[\frac{1 - 3\rho_{ijk}}{\rho_{ijk}(1 - \rho_{ijk})} + \frac{1}{\rho_{ij} - \rho_{ijk}} \right. \\
& \left. + \frac{1}{\rho_{jk} - \rho_{ijk}} + \frac{1}{\rho_{ki} - \rho_{ijk}} \right] \left(\frac{\partial \rho_{ijk}}{\partial \zeta_{ijk}} \right)_{\boldsymbol{\rho}},
\end{aligned}
$$

which finally leads to Equation 6.13.

E.1.3   Strauss's model for small networks

Let $\boldsymbol{\tau}$ denote a vector of components τ_{ν}, where ν is a subset of 2 elements of $\{1, 2, \ldots, N\}$, and $W \equiv \{0,1\}^N$. The grand partition function is defined as

$$
\begin{aligned}
\Xi = & \sum_{\boldsymbol{\sigma} \in W} \exp \left(\phi \sum_{i<j}^{N} \tau_{ij} + \frac{\gamma}{N} \sum_{i<j<k}^{N} \tau_{ij} \tau_{jk} \tau_{ik} \right) \\
= & \sum_{L=0}^{\binom{N}{2}} \sum_{T=0}^{\binom{N}{3}} Q(L,T) e^{\phi L + \gamma T/N},
\end{aligned}
\tag{E.12}
$$

where $Q(L,T)$ is the number of configurations $\boldsymbol{\tau}$ with L links and T triangles.

Let us set $N = 4$ and compute the values of $Q(L,T)$. Clearly, $Q(L,0) = \binom{6}{L}$ for $L = 0, 1, 2$, but $Q(L,T) = 0$ otherwise. Furthermore, $Q(5,2) = 6$, $Q(6,4) = 1$, and $Q(5,T) = Q(6,T) = 0$ otherwise. As for $L = 3, 4$, there are configurations with either $T = 0$ or $T = 1$. Thus, $Q(3,1) = 4$ and $Q(3,0) = \binom{6}{3} - 4 = 16$. On the other hand, $Q(4,0) = 3$ and $Q(4,1) = \binom{6}{4} - 3 = 12$. Accordingly, if we denote $x \equiv e^{\phi}$, $y \equiv e^{\gamma/4}$,

$$
\begin{aligned}
\Xi = & 1 + 6x + 15x^2 + 16x^3 + 3x^4 + 4x^3(1 + 3x)y \\
& + 6x^5 y^2 + x^6 y^4.
\end{aligned}
\tag{E.13}
$$

The density can be obtained as

$$\rho = \frac{1}{6}\frac{\partial}{\partial \phi}\log \Xi = \frac{x}{6}\frac{\partial}{\partial x}\log \Xi, \qquad (E.14)$$

and from that,

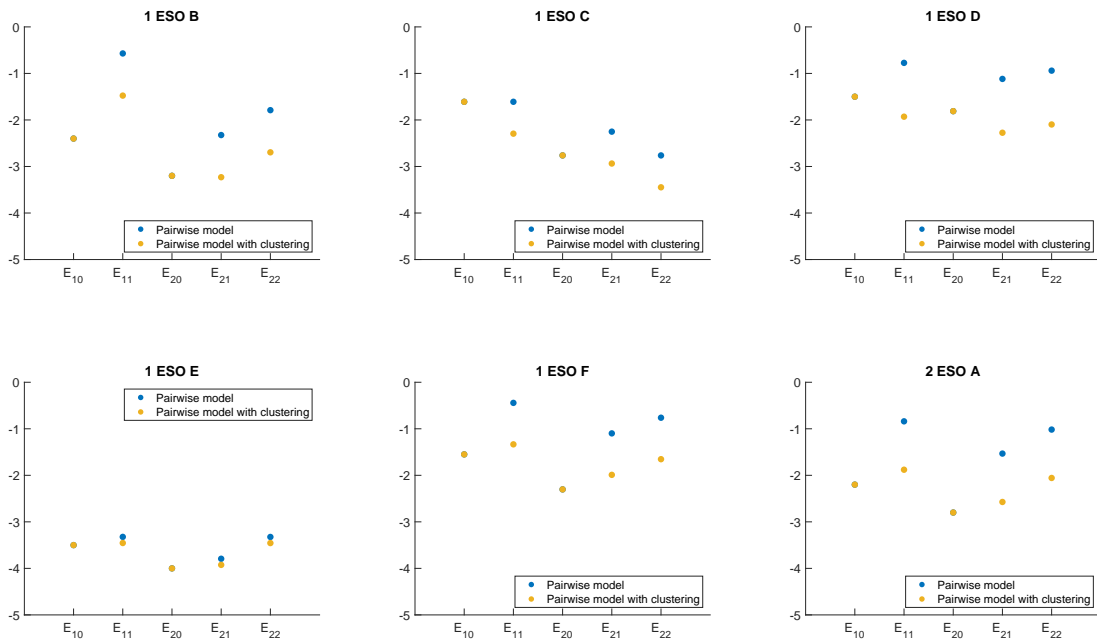$$f = \frac{F}{6} = \rho \log x - \frac{1}{6}\log \Xi. \qquad (E.15)$$

Thus, fixing the interaction γ (i.e., fixing y) we can obtain parametrically, as $0 < x < \infty$, the curve $f(\rho, \gamma)$.

F

APPENDIX FOR CHAPTER 7

F.1 Supplementary figures

F.1.1 Pairwise model with clustering
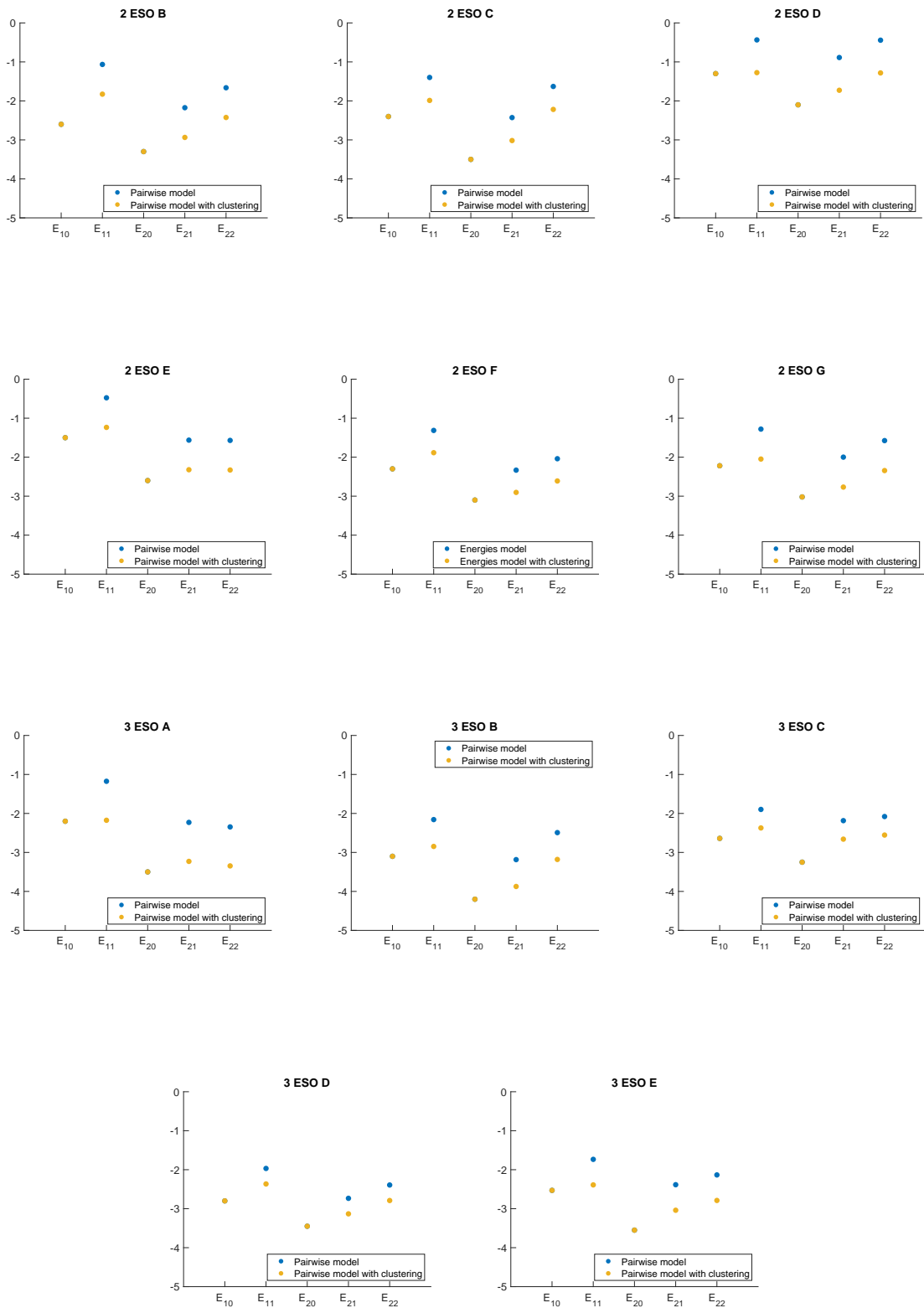
Figure F.1: Parameters of the pairwise model with clustering - Values obtained for the parameters of the pairwise model with clustering for each of the remaining groups in wave 1.

F.1.2   Linear model with reciprocity and clustering

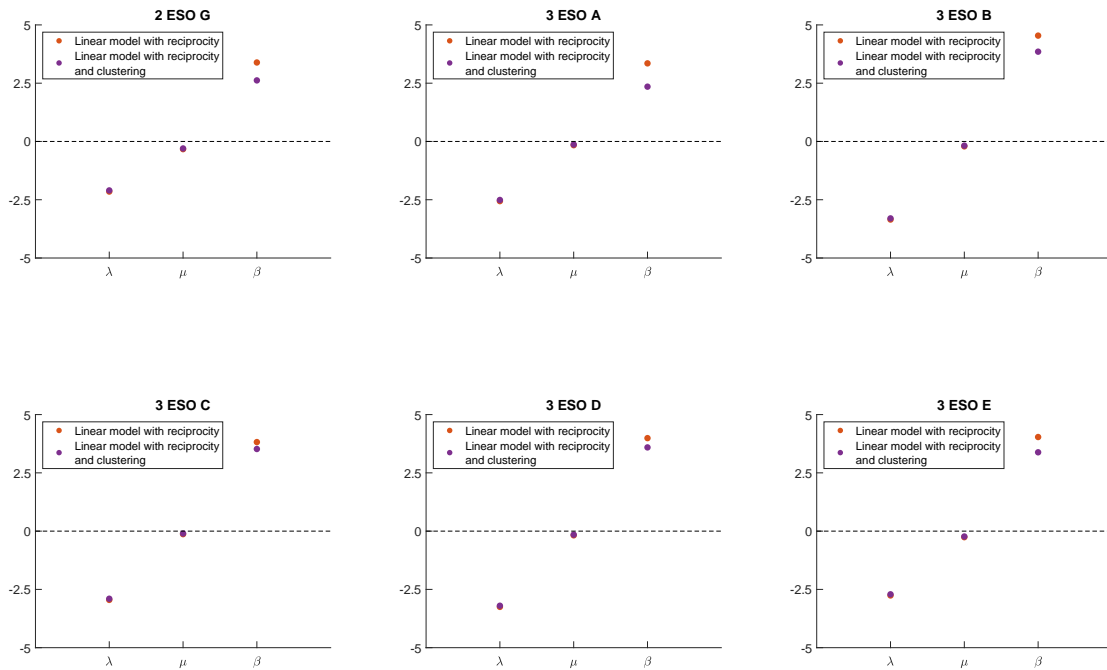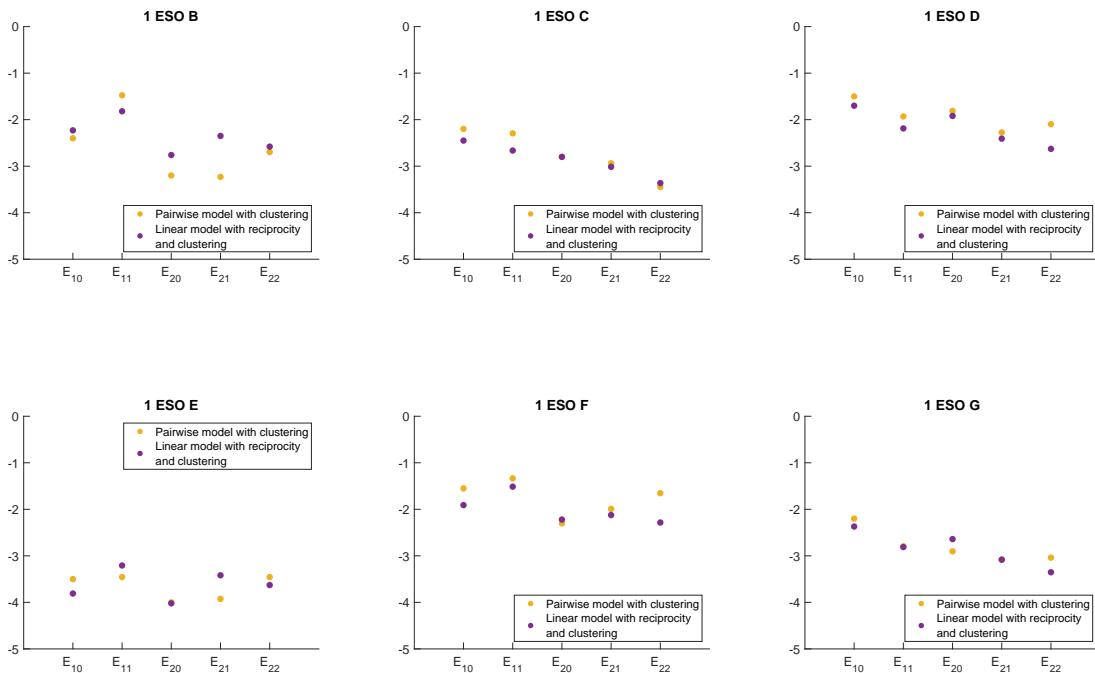### F.1.2.1   Comparative with the linear model with reciprocity
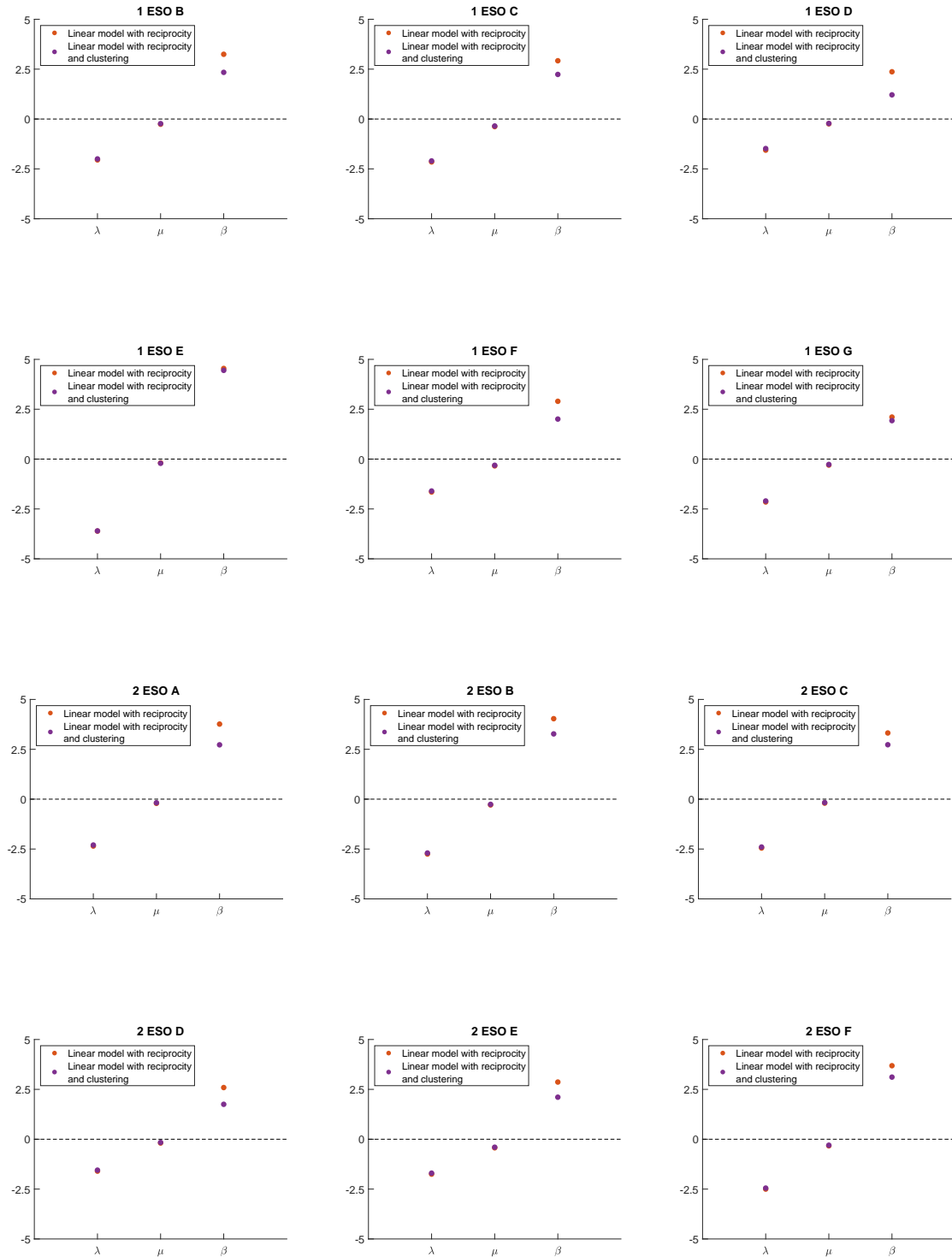
Figure F.2: Parameters of the linear model with reciprocity and its comparative with the linear model with reciprocity - Values obtained for the parameters of the linear model with reciprocity for each of the remaining groups in wave 1.

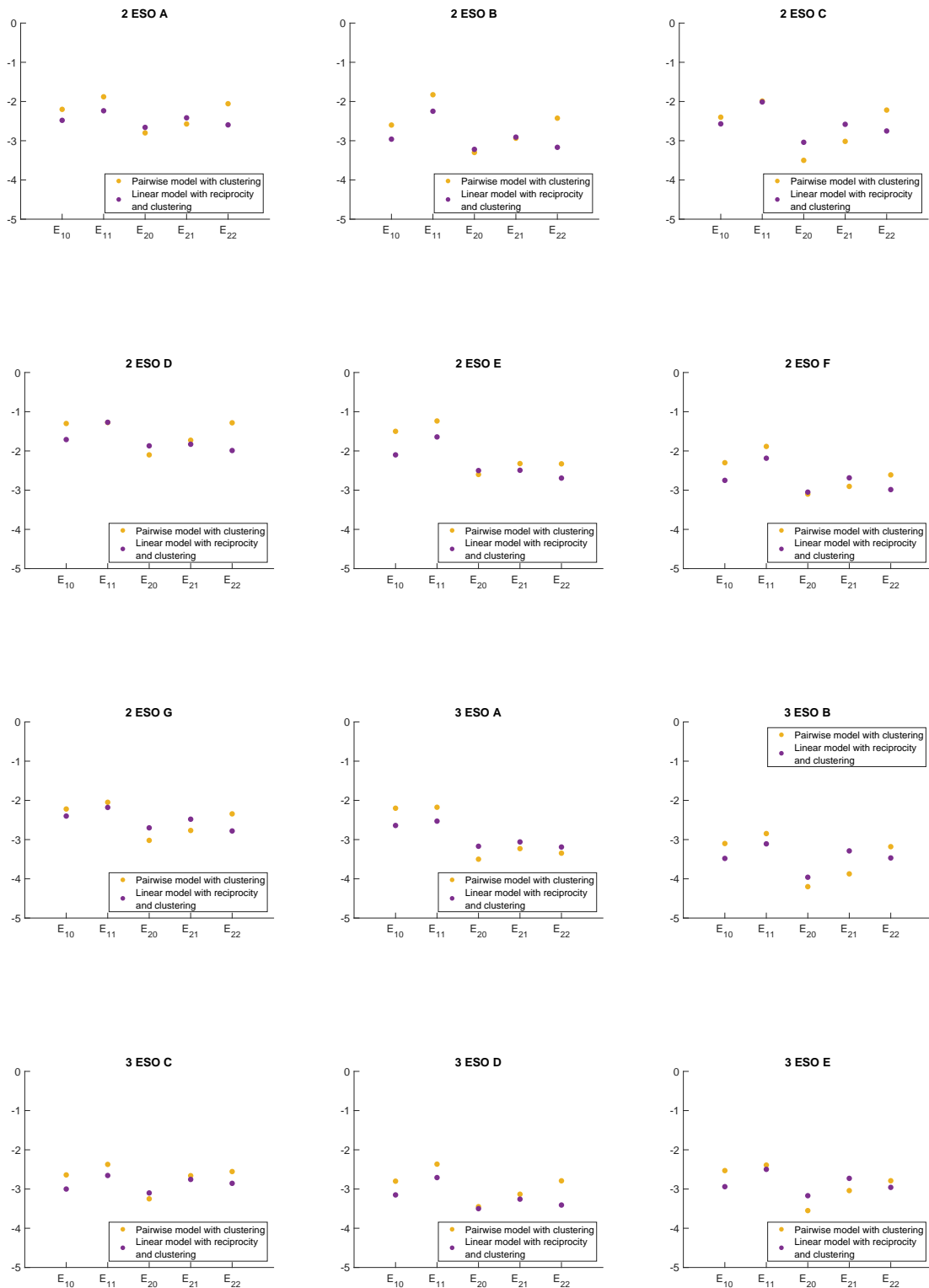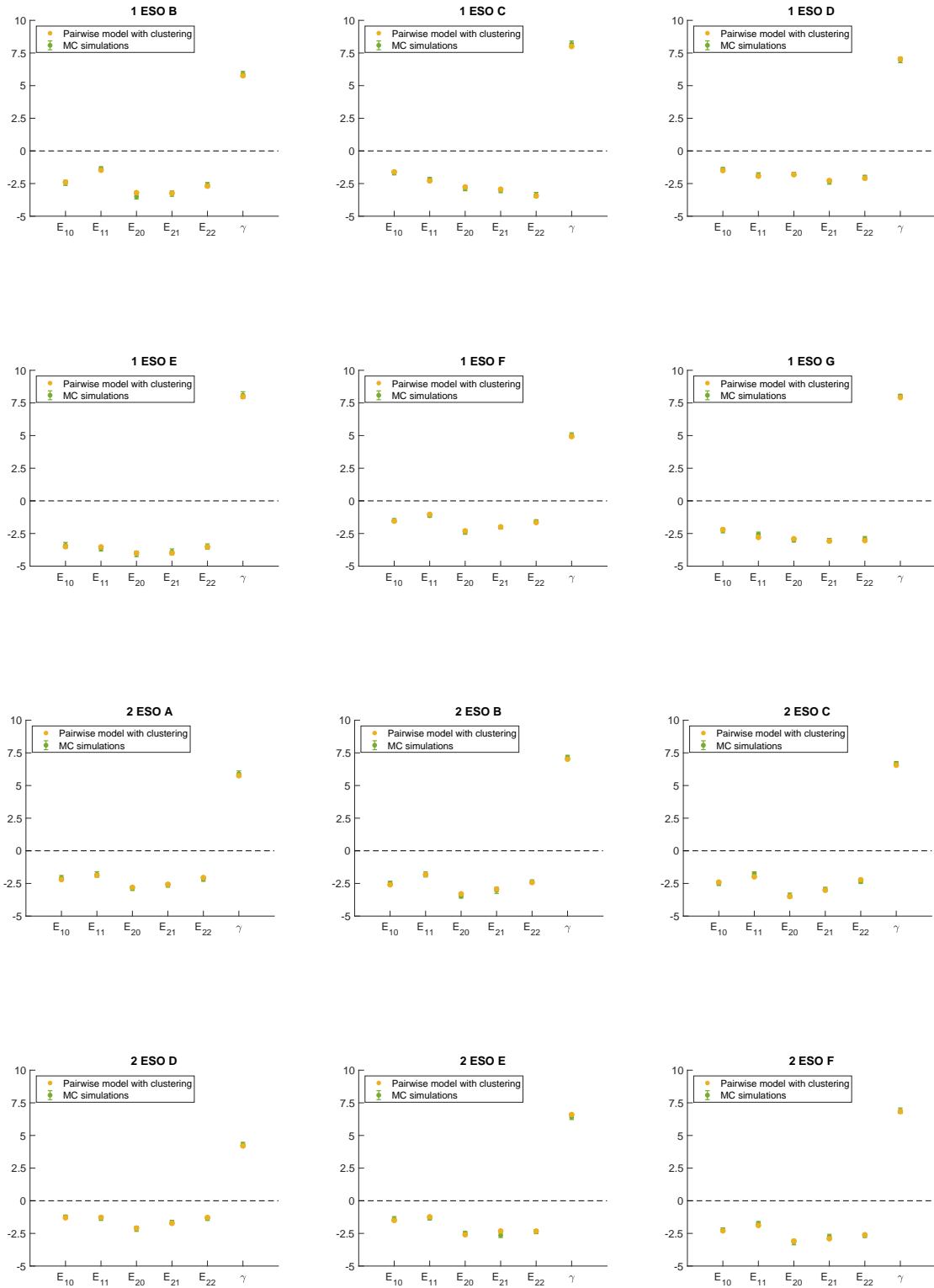### F.1.2.2   Comparative with the pairwise model with clustering

Figure F.3: Parameters of the linear model with reciprocity and its comparative with the pairwise model with clustering - Values obtained for the parameters of the linear model with reciprocity for each of the remaining groups in wave 1.

F.1.3   Monte Carlo simulations for the pairwise model with clustering
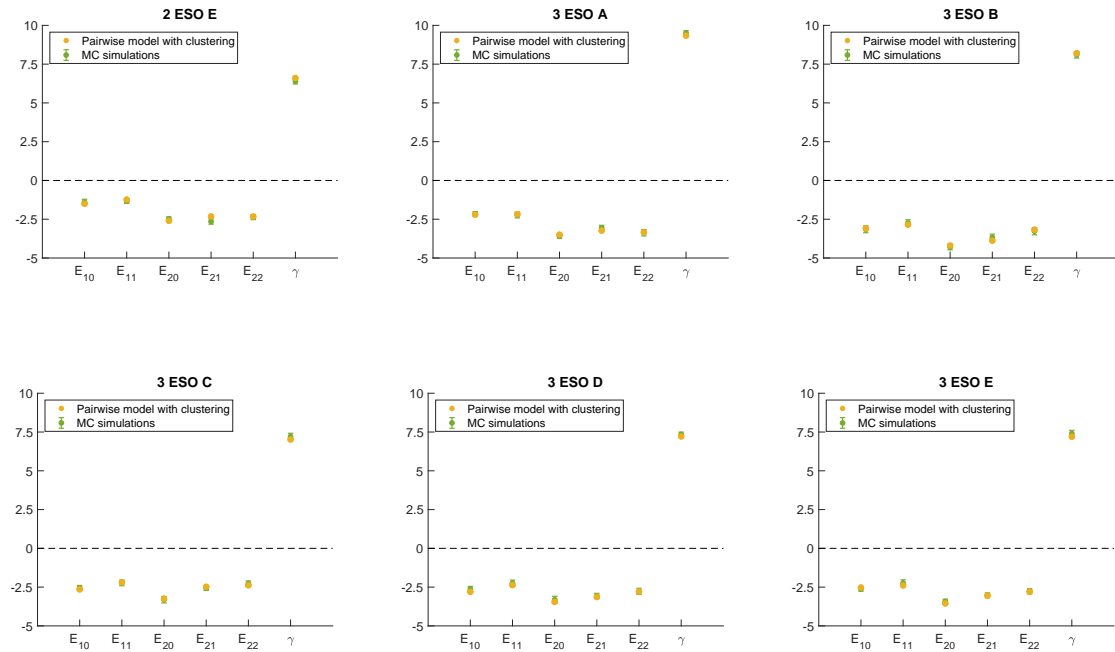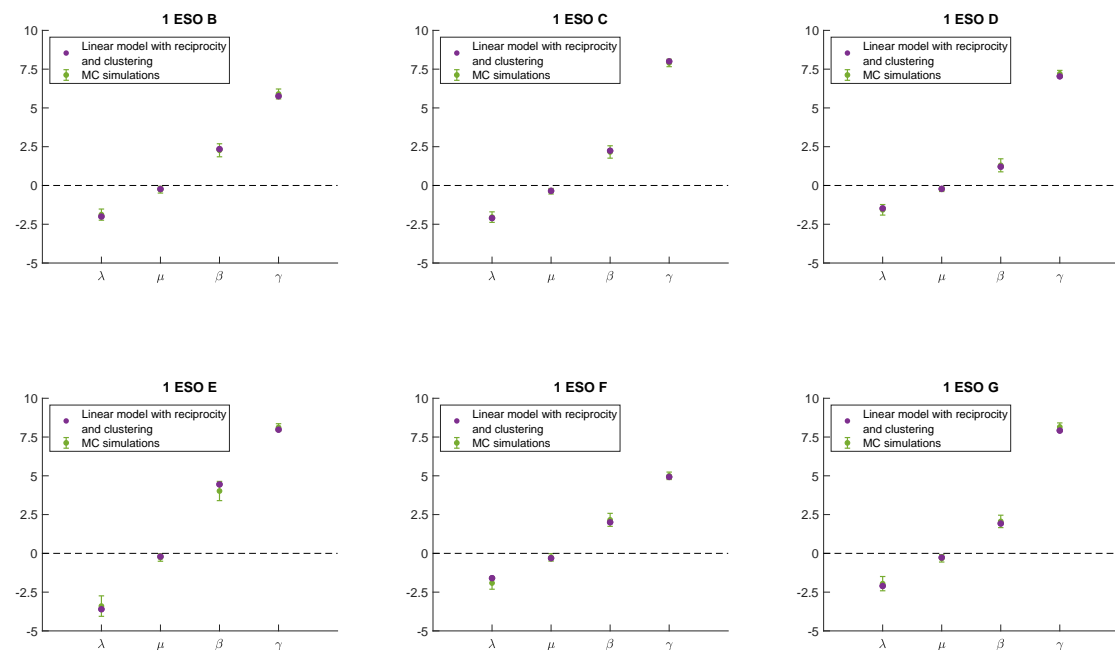
Figure F.4: Comparative between the parameters associated with the pairwise model with clustering and Monte Carlo simulations - Values obtained for the parameters of the pairwise model with clustering and its comparative with Monte Carlo simulations for each of the remaining groups in wave 1.

F.1.4 Monte Carlo simulations for the linear model with reciprocity and clustering
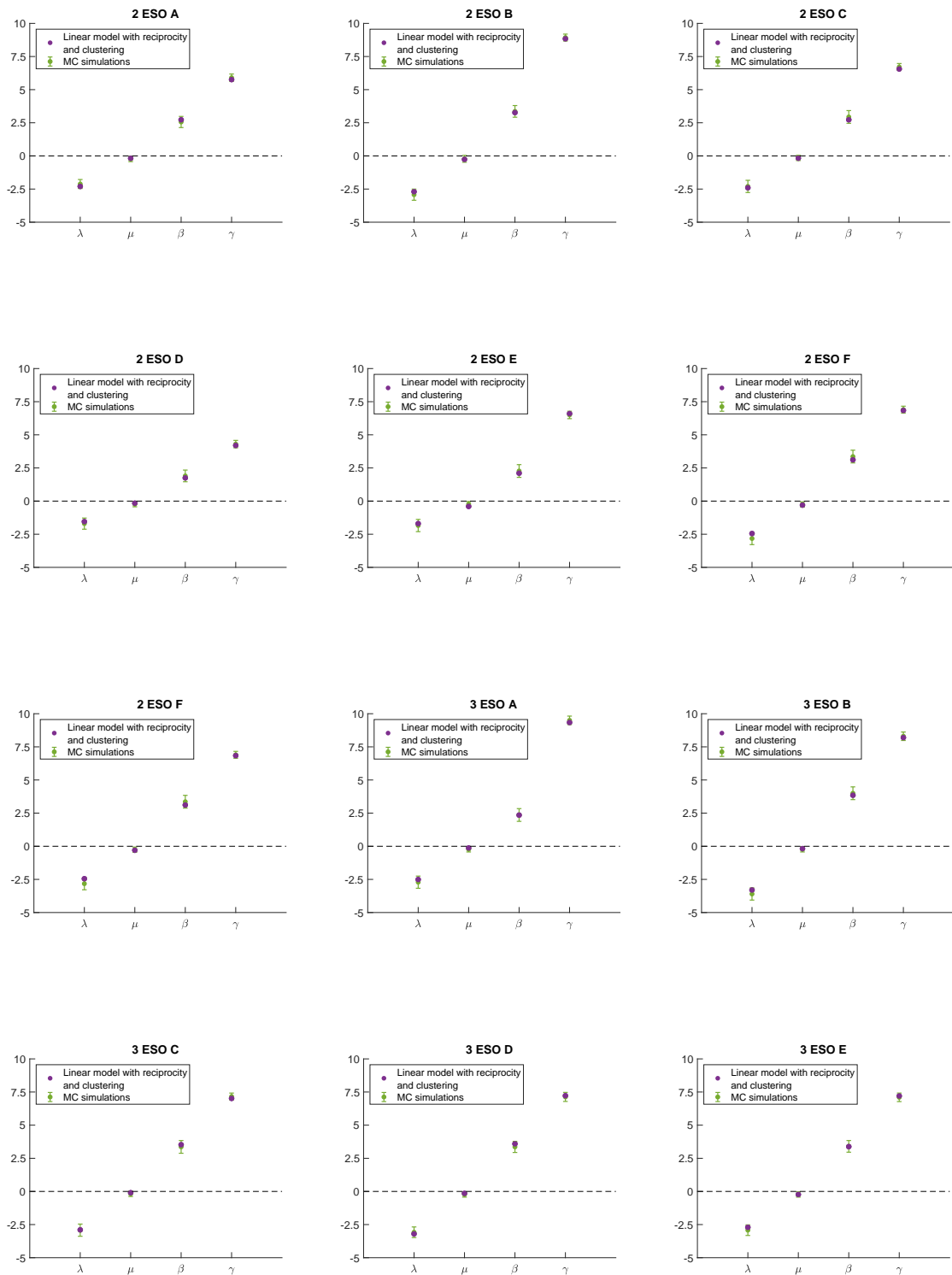
Figure F.5: Comparative between the parameters associated with the linear model with reciprocity and clustering and Monte Carlo simulations - Values obtained for the parameters of the linear model with reciprocity and clustering and its comparative with Monte Carlo simulations for each of the remaining groups in wave 1.

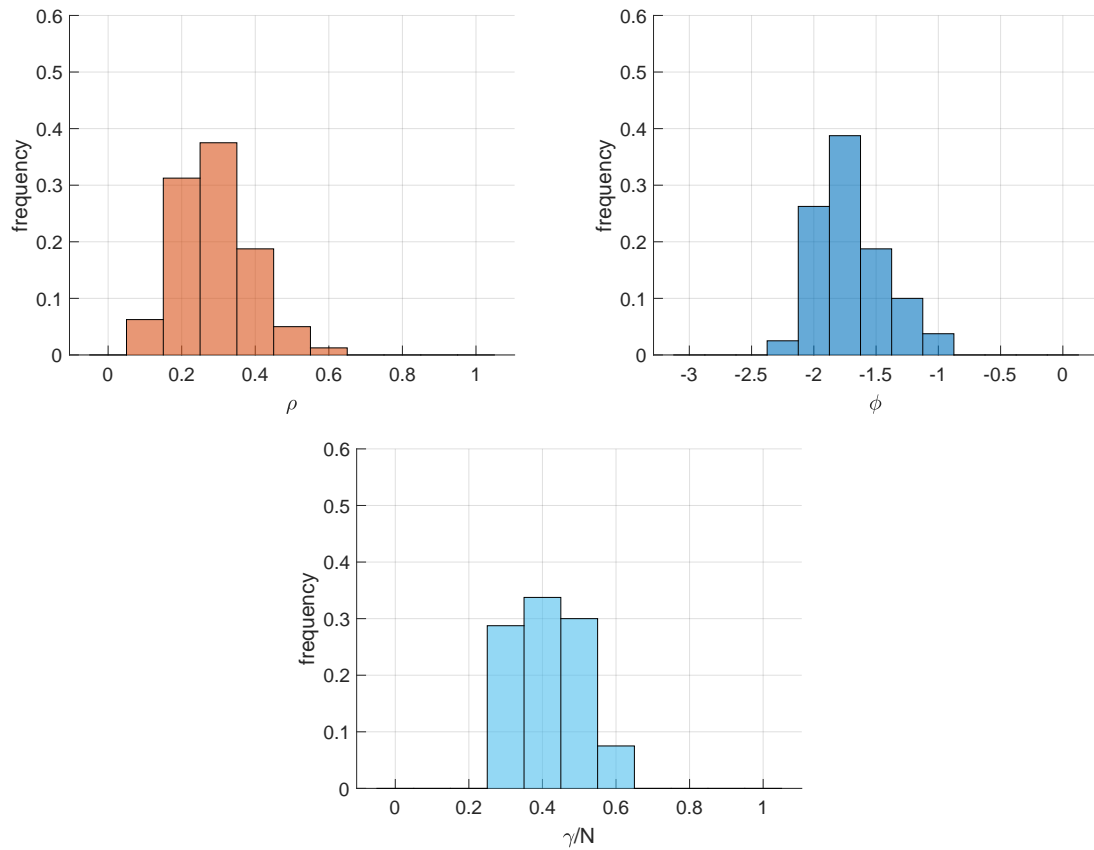## F.1.5 Distribution of the parameters

Figure F.7: Histograms representing the distribution of the parameters - Histograms representing the distribution of the density of relationships ρ, the parameter ϕ and the parameter γ/N for all the groups in the five waves.

G

Appendix for Chapter 8

G.1 Supplementary tables

G.1.1 Information on the chimpanzees

Group	Individual	Sex	Age*	Origin	Parameter η
1	Pal	Male	39	Wild born	Yes
1	Booboo	Male	38	Wild born	Yes
1	Tobar	Male	38	Wild born	Yes
1	Girly	Female	38	Wild born	Yes
1	Rita	Female	37	Wild born	Yes
1	Tara	Male	37	Wild born	Yes
1	Ingrid	Female	29	Captive born	Yes
1	Brenda	Female	24	Captive born	Yes
1	Genny	Female	23	Captive born	Yes
1	Renate	Female	23	Captive born	Yes
1	Bob	Male	19	Captive born	Yes
1	Gerard	Male	18	Captive born	Yes
1	Ilse	Female	17	Captive born	Yes
1	Chrissie	Female	13	Captive born	Yes
1	Rusty	Male	13	Captive born	Yes
1	Regina	Female	13	Captive born	Yes
1	Innocentia	Female	13	Captive born	Yes

1	BJ	Female	12	Captive born	Yes
1	Gonzaga	Female	12	Captive born	Yes
1	Irene	Female	8	Captive born	Yes
1	Rachel	Female	7	Captive born	Yes
1	Richard	Male	5	Captive born	No
1	Ian	Male	5	Captive born	No
1	Genny's baby	Female	4	Captive born	No
1	Ida	Female	4	Captive born	No
1	Ricky	Male	1	Captive born	No

Table G.1: Group 1 demographics - *Age at the end of the observation window (late 2019).

Group	Individual	Sex	Age*	Origin	Parameter η
2	Noel	Female	43	Wild born	Yes
2	Donna	Female	36	Wild born	Yes
2	Little Jane	Female	35	Wild born	Yes
2	Coco	Female	35	Wild born	Yes
2	Maggie	Female	34	Wild born	Yes
2	Misha	Female	32	Wild born	Yes
2	Mikey	Male	32	Wild born	No
2	Pan	Male	31	Wild born	Yes
2	Pippa	Female	31	Wild born	Yes
2	Dora	Female	31	Wild born	Yes
2	Trixie	Female	30	Wild born	Yes
2	Zsabu	Male	30	Wild born	Yes
2	Masya	Female	29	Wild born	Yes
2	Violet	Female	29	Wild born	No
2	Diana	Female	29	Wild born	Yes
2	Little Judy	Female	24	Captive born	Yes
2	Dolly	Female	23	Captive born	Yes
2	Carol	Female	23	Captive born	Yes
2	Nikkie	Female	22	Captive born	Yes
2	Tess	Female	21	Captive born	Yes
2	Tilly	Female	19	Captive born	Yes
2	Debbie	Female	18	Captive born	Yes
2	David	Male	18	Captive born	Yes
2	Maxine	Female	18	Captive born	No
2	Doug	Male	17	Captive born	Yes
2	Nina	Female	17	Captive born	Yes
2	Claire	Female	17	Captive born	Yes
2	Toni	Female	17	Captive born	No
2	Vis	Male	16	Captive born	Yes
2	Taylor	Female	16	Captive born	No
2	Daisey	Female	15	Captive born	Yes
2	Mary	Female	14	Captive born	No

2	Long John	Male	13	Captive born	Yes
2	Little Jenkins	Female	13	Captive born	Yes
2	Max	Male	13	Captive born	Yes
2	Darwin	Male	13	Captive born	No
2	Dizzy	Female	12	Captive born	Yes
2	Moyo	Male	12	Captive born	Yes
2	Charity	Female	12	Captive born	Yes
2	Little Jones	Female	9	Captive born	No
2	Little Jacky	Male	8	Captive born	No
2	Danny	Male	8	Captive born	No
2	Martin	Male	8	Captive born	No
2	May	Female	7	Captive born	No
2	Masya's baby	Female	7	Captive born	No
2	Mavis	Female	7	Captive born	No
2	Chitalu	Female	6	Captive born	No
2	Debbie's baby	Male	5	Captive born	No
2	Tom	Male	5	Captive born	No
2	Tina	Female	4	Captive born	No
2	Don	Male	4	Captive born	No
2	Toni's baby	Female	4	Captive born	No
2	Nina's baby2	Female	4	Captive born	No
2	Mumba	Male	3	Captive born	No
2	Nancy	Female	3	Captive born	No
2	Merial	Female	2	Captive born	No
2	Little Joey	Male	2	Captive born	No
2	Camilla	Female	2	Captive born	No
2	Muriel	Female	2	Captive born	No
2	Victoria	Female	2	Captive born	No

Table G.2: Group 2 demographics - *Age at the end of the observation window (late 2019).

Group	Individual	Sex	Age*	Origin	Parameter η
3	Buffy	Female	35	Wild born	Yes
3	Clement	Male	27	Wild born	Yes
3	Brian	Male	26	Wild born	Yes
3	E.T.	Female	25	Wild born	Yes
3	Roxy	Female	25	Wild born	Yes
3	Barbie	Female	24	Wild born	Yes
3	Bussie	Male	15	Captive born	Yes
3	Bruce	Male	10	Captive born	Yes
3	Lods	Female	9	Captive born	Yes
3	Brent	Female	6	Captive born	Yes
3	Bill	Male	1	Captive born	No

Table G.3: Group 3 demographics - *Age at the end of the observation window (late 2019).

Group	Individual	Sex	Age*	Origin	Parameter η
4	Nicky	Male	29	Wild born	Yes
4	Sinkie	Male	26	Wild born	Yes
4	Bobby	Male	26	Wild born	Yes
4	Kambo	Female	24	Wild born	Yes
4	Kathy	Female	21	Wild born	Yes
4	Miracle	Female	20	Captive born	Yes
4	Val	Male	20	Wild born	Yes
4	Commander	Male	20	Wild born	Yes
4	Kit	Male	15	Captive born	Yes
4	Jack	Male	12	Captive born	No
4	Leila	Female	9	Wild born	No
4	Ken	Male	8	Captive born	Yes
4	Jewel	Male	6	Captive born	No
4	Grace	Female	5	Wild born	No

Table G.4: Group 4 demographics - *Age at the end of the observation window (late 2019).

G.2 Supplementary figures

G.2.1 Grooming networks
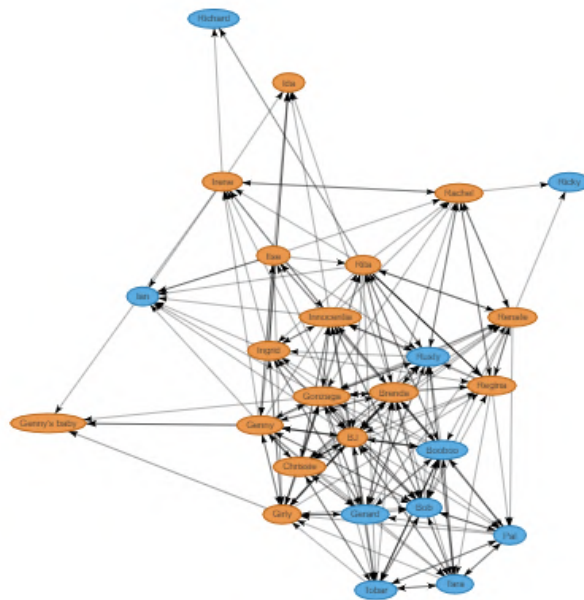
Figure G.1: Grooming network - Group 1 - Orange ovals represent females and blue ones males.
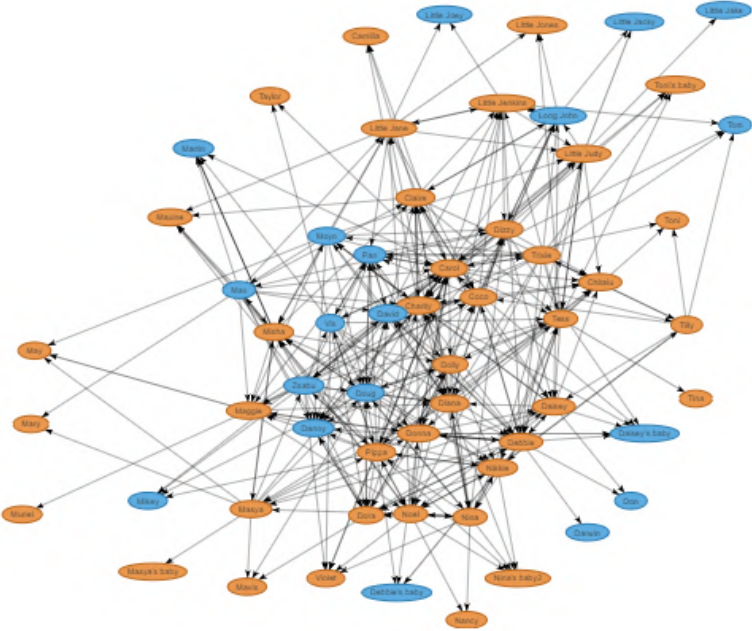
Figure G.2: Grooming network - Group 2 - Orange ovals represent females and blue ones males.
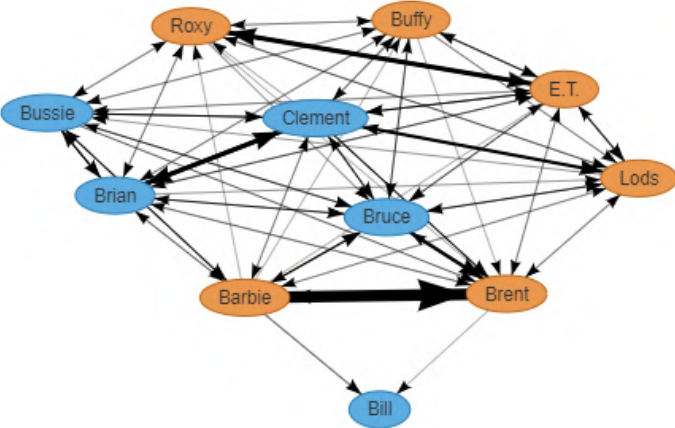

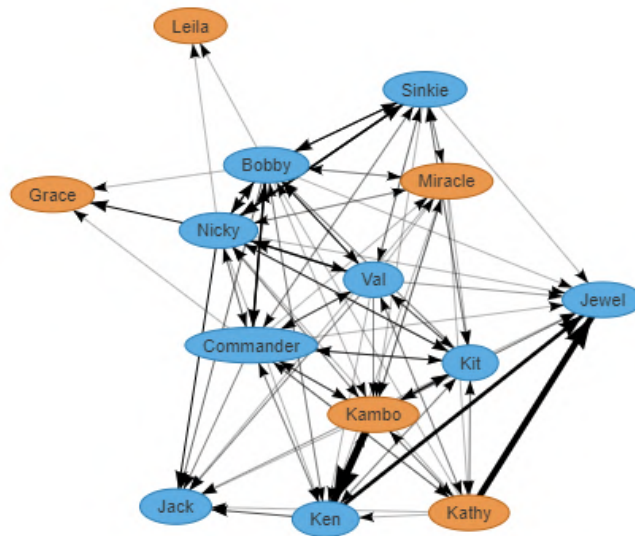**Figure G.3: Grooming network - Group 3** - Orange ovals represent females and blue ones males.

Figure G.4: Grooming network - Group 4 - Orange ovals represent females and blue ones males.
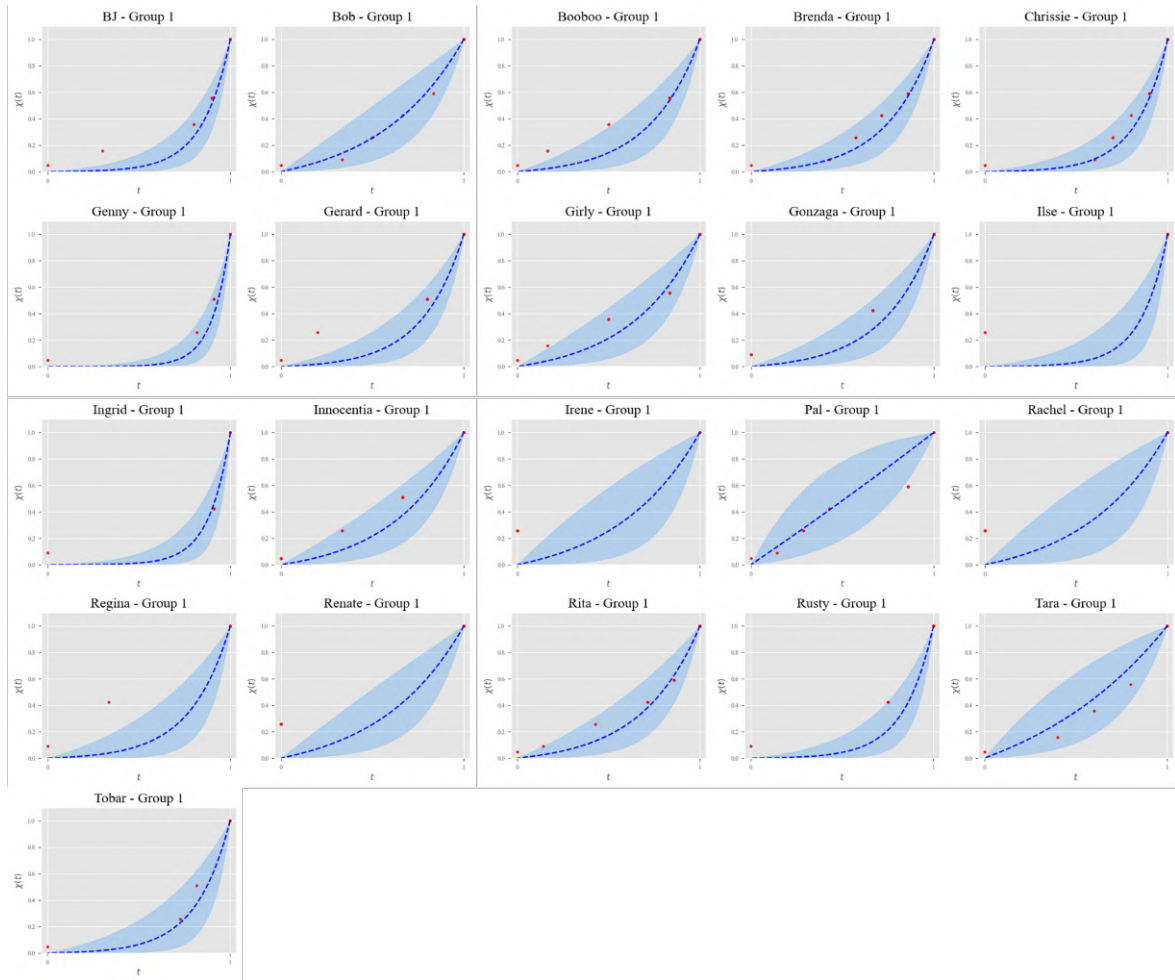
G.2.2   Individual fits of the distribution of ties

Figure G.5: Individual fits of the distribution of ties - Group 1
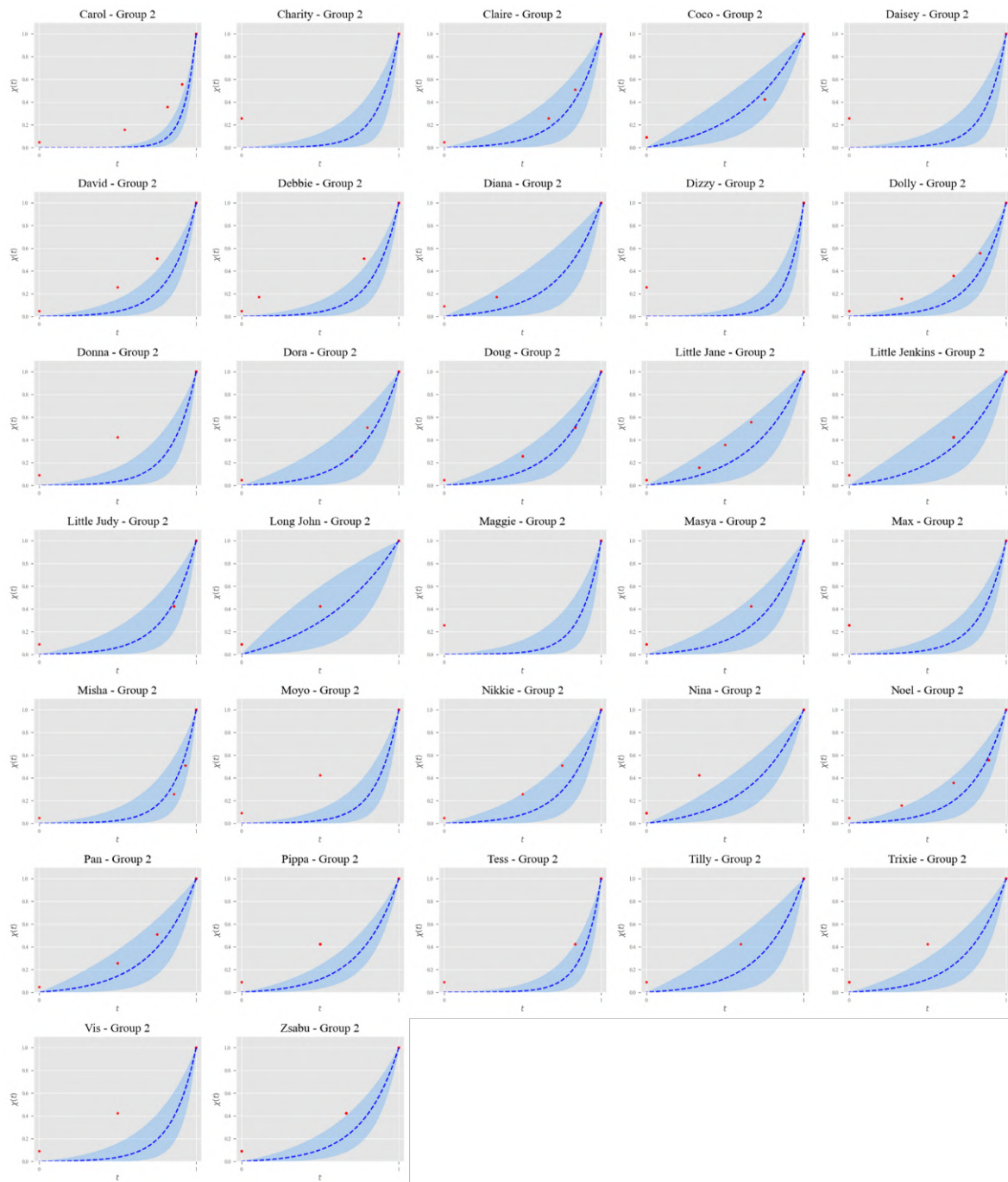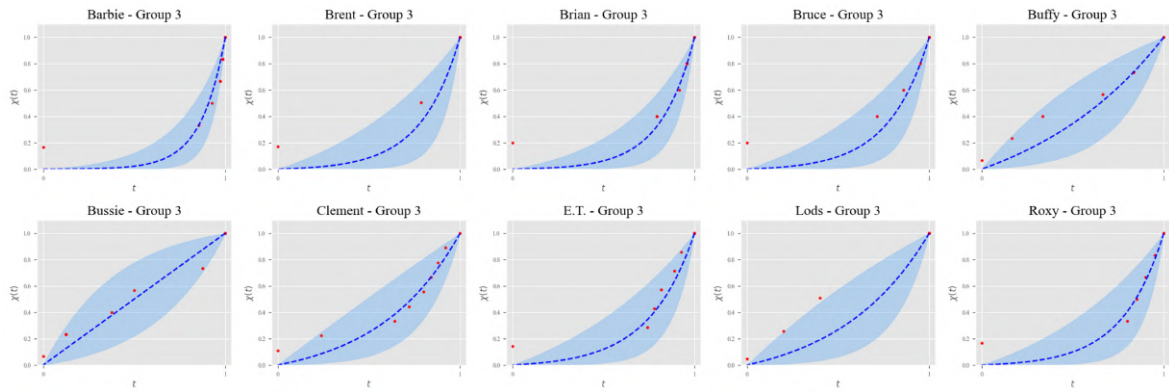
Figure G.6: Individual fits of the distribution of ties - Group 2

Figure G.7: Individual fits of the distribution of ties - Group 3

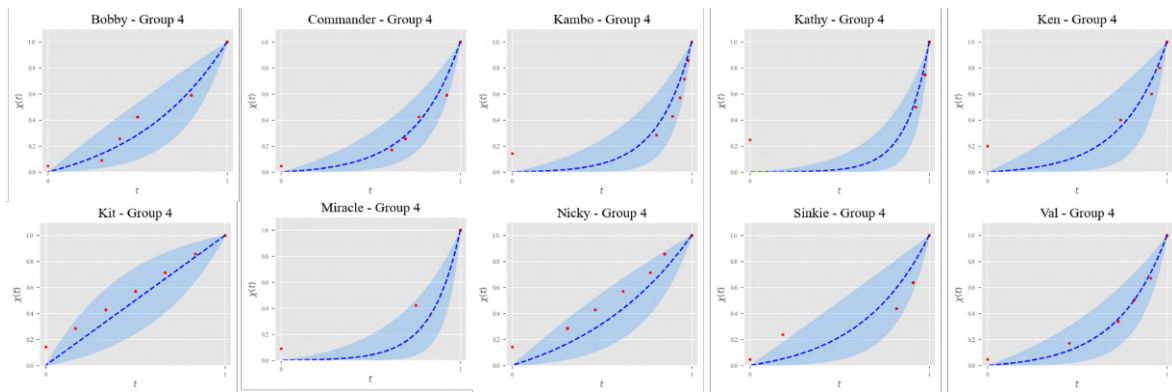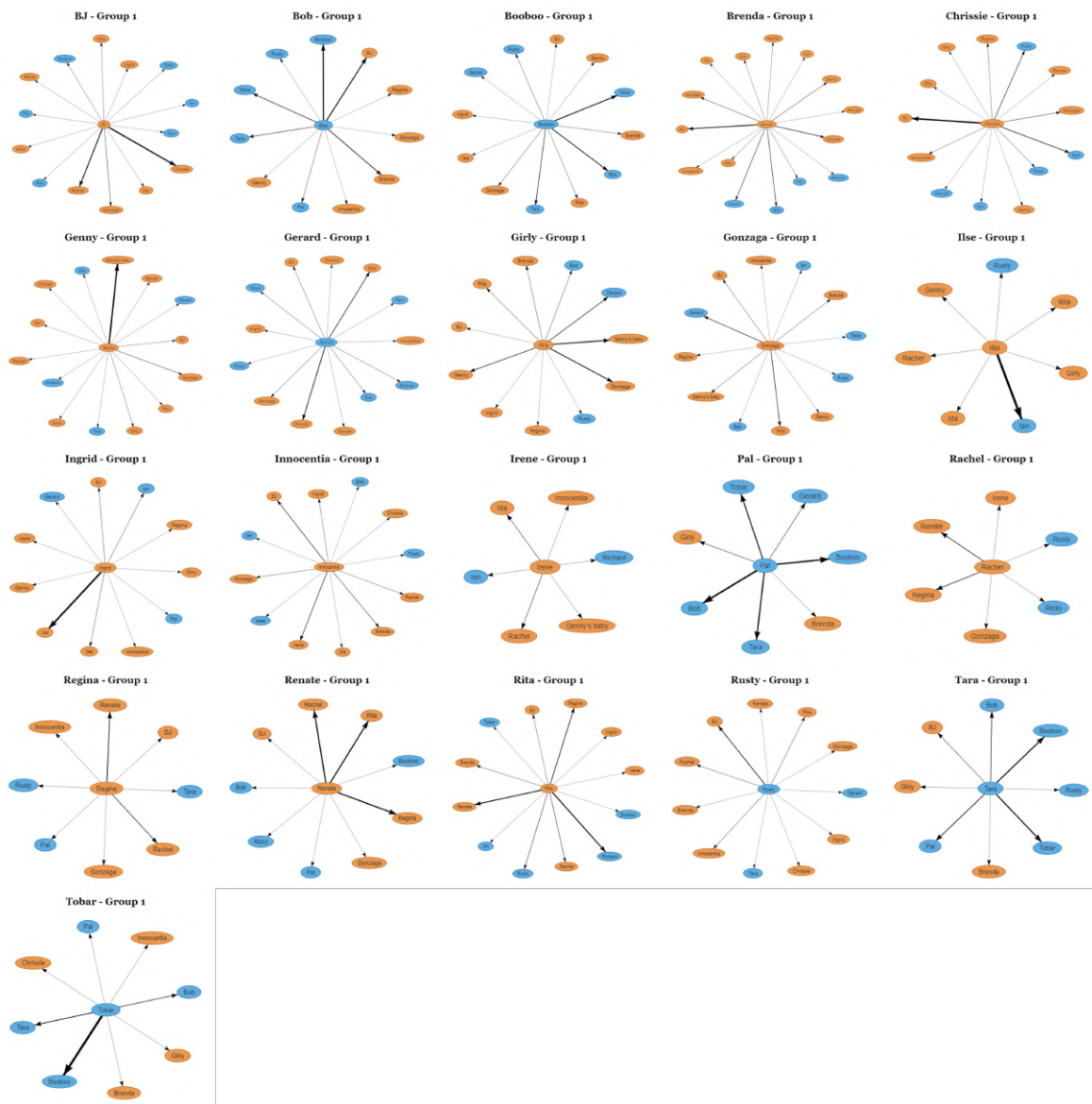Figure G.8: Individual fits of the distribution of ties - Group 4

G.2.3 Ego-networks

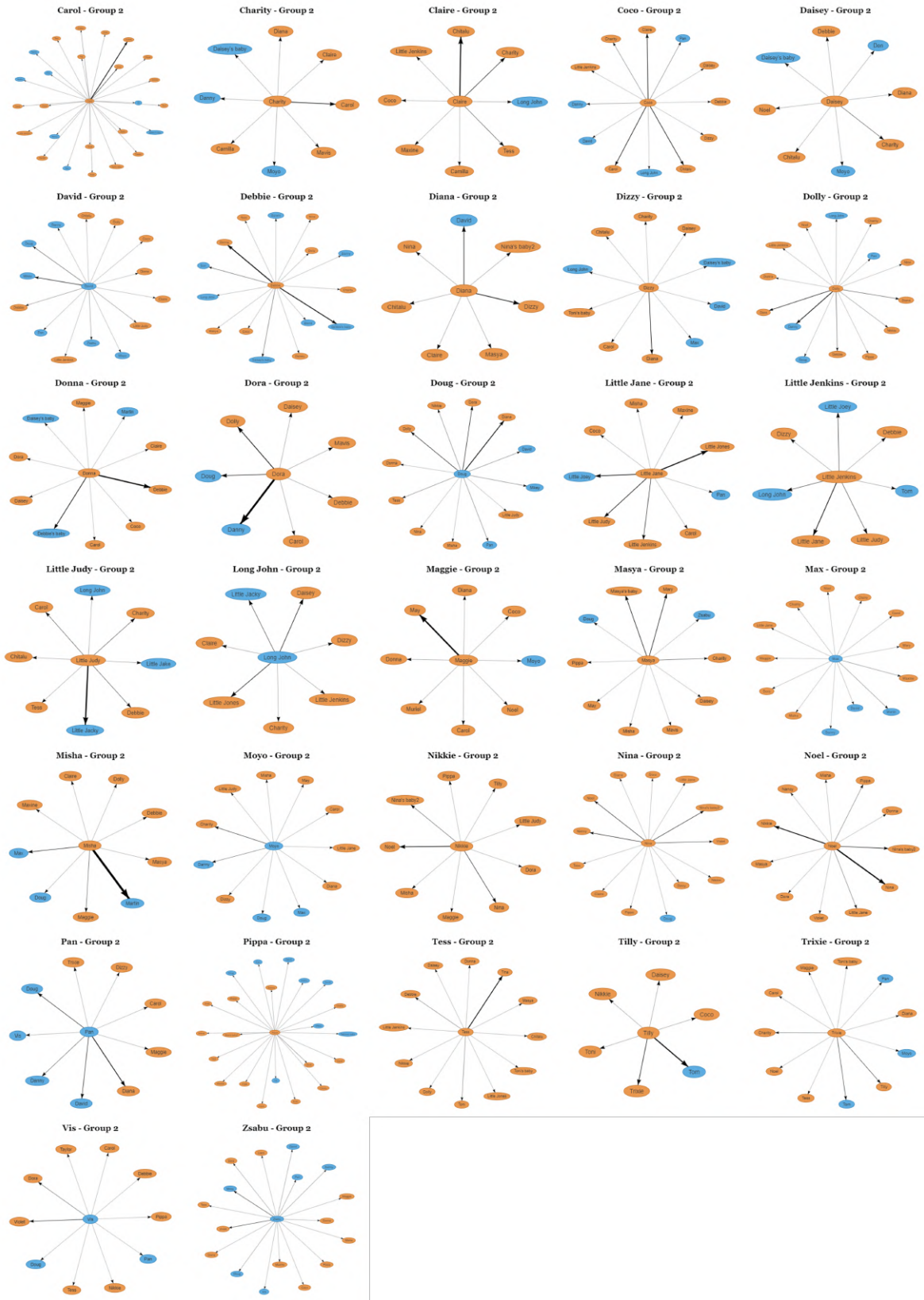**Figure G.9: Ego-networks - Group 1** - Orange ovals represent females and blue ones males.

Figure G.10: Ego-networks - Group 2 - Orange ovals represent females and blue ones males.
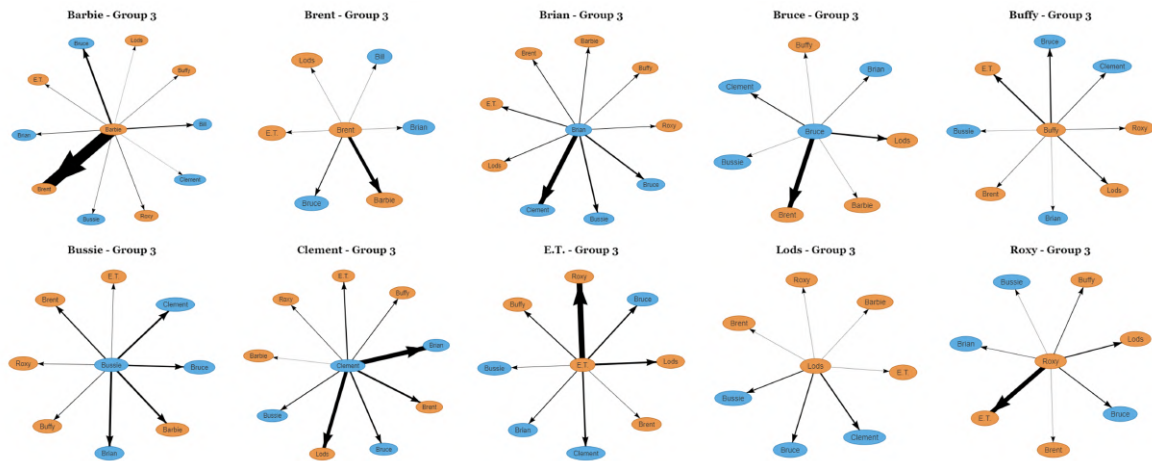
Figure G.11: Ego-networks - Group 3 - Orange ovals represent females and blue ones males.
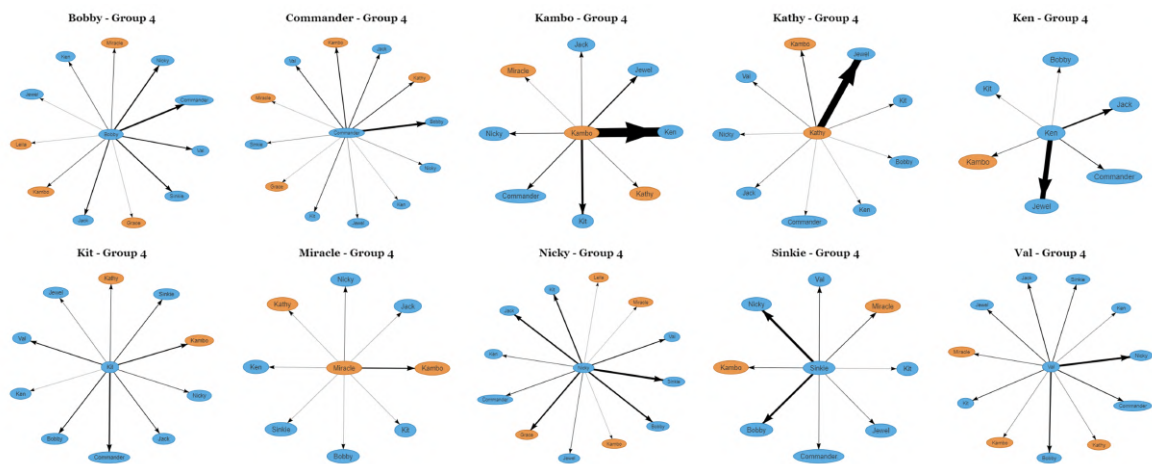

Figure G.12: Ego-networks - Group 4 - Orange ovals represent females and blue ones males.

References

Adams, A. M., Madhavan, S., & Simon, D. (2002). Women's social networks and child survival in Mali. *Social Science & Medicine*, *54*, 165–178.

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, *74*, 47.

Alm, E., & Arkin, A. P. (2003). Biological networks. *Current Opinion in Structural Biology*, *13*(2), 193–202.

Almaatouq, A., Radaelli, L., Pentland, A., & Shmueli, E. (2016). Are you your friends' friend? Poor perception of friendship ties limits the ability to promote behavioral change. *PLoS One*, *11*, e0151588.

Anderson, R. M., & May, R. M. (1992). *Infectious diseases of humans: Dynamics and control.* Oxford University Press.

Arnaboldi, V., Conti, M., Passarella, A., & Dunbar, R. I. M. (2017). Online social networks and information diffusion: The role of ego networks. *Online Social Networks and Media*, *1*, 44–55.

Arnaboldi, V., Conti, M., Passarella, A., & Pezzoni, F. (2012). Analysis of ego network structure in online social networks. *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, 31–40.

Arnaboldi, V., Passarella, A., Conti, M., & Dunbar, R. I. M. (2015). *Online social networks: human cognitive constraints in Facebook and Twitter personal graphs.* Elsevier.

Aubin, D. (2014). Principles of Mechanics that are susceptible of application to society: An unpublished notebook of Adolphe Quetelet at the root of his social physics. *Historia Mathematica*, *41*, 204–223.

Avetisov, V., Gorsky, A., Maslov, S., Nechaev, S., & Valba, O. (2018). Phase transitions in social networks inspired by the Schelling model. *Physical Review E*, *98*, 032308.

Avetisov, V., Hovhannisyan, M., Gorsky, A., Nechaev, S., Tamm, M., & Valba, O. (2016). Eigenvalue tunneling and decay of quenched random network. *Physical Review E*, *94*, 062313.

Ball, P. (2003). The physical modelling of human social systems. *Complexus*, *1*, 190–206.

Ball, P. (2012). *Why society is a complex matter: Meeting twenty-first century challenges with a new kind of science.* Springer Science & Business Media.

REFERENCES

Barrett, L. F., Lane, R. D., Sechrest, L., & Schwartz, G. E. (2000). Sex differences in emotional awareness. *Personality and Social Psychology Bulletin, 26*, 1027–1035.

Barrett, L., Henzi, S. P., Weingrill, T., Lycett, J. E., & Hill, R. A. (1999). Market forces predict grooming reciprocity in female baboons. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 266*, 665–670.

Bernard, H. R., & Killworth, P. D. (1973). On the social structure of an ocean-going research vessel and other important things. *Social Science Research, 2*, 145–184.

Bernard, H. R., & Killworth, P. D. (1979). Why are there no social physics? *Journal of the Steward Anthropological Society, 11*, 33–58.

Bernard, H. R., & Killworth, P. D. (1997). The search for social physics. *Connections, 20*, 16–34.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B, 36*(2), 192–225.

Bhattacharya, K., Ghosh, A., Monsivais, D., Dunbar, R. I. M., & Kaski, K. (2016). Sex differences in social focus across the life cycle in humans. *Royal Society Open Science, 3*, 160097.

Bianconi, G., Darst, R. K., Iacovacci, J., & Fortunato, S. (2014). Statistical physics of exchangeable sparse network ensembles. *Physical Review E, 90*, 042806.

Binder, J. F., Roberts, S. G. B., & Sutcliffe, A. G. (2012). Closeness, loneliness, support: Core ties and significant ties in personal communities. *Social Networks, 34*, 206–214.

Birkhoff, G., & Mac Lane, S. (1997). *A survey of modern algebra.* CRC Press.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science, 323*, 892–895.

Brask, J. B., Ellis, S., & Croft, D. P. (2021). Animal social networks: An introduction for complex systems scientists. *Journal of Complex Networks, 9*, cnab001.

Brechwald, W. A., & Prinstein, M. J. (2011). Beyond homophily: A decade of advances in understanding peer influence processes. *Journal of Research on Adolescence, 21*, 166–179.

Burda, Z., Jurkiewicz, J., & Krzywicki, A. (2004). Network transitivity and matrix models. *Physical Review E, 69*, 026106.

Buys, C. J., & Larson, K. L. (1979). Human sympathy groups. *Psychological Reports, 45*, 547–553.

Candia, C., Oyarzún, M., Landaeta, V., Yaikin, T., Monge, C., Hidalgo, C., & Rodriguez-Sickert, C. (2022). Reciprocity heightens academic performance in elementary school students. *Heliyon, 8*, e11916.

Carstensen, L. L. (1991). Selectivity theory: Social activity in life-span context. *Annual Review of Gerontology and Geriatrics, 11*, 195–217.

Cartwright, D., & Harary, F. (1956). Structural balance: a generalization of Heider's theory. *Psychological Review*, *63*, 277.

Carvalho, R., Buzna, L., Bono, F., Gutiérrez, E., Just, W., & Arrowsmith, D. (2009). Robustness of trans-European gas networks. *Physical Review E*, *80*, 016106.

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, *81*, 591.

Cattuto, C., den Broeck, W. V., Barrat, A., Colizza, V., Pinton, J.-F., & Vespignani, A. (2010). Dynamics of person-to-person interactions from distributed rfid sensor networks. *PLoS One*, *5*, e11596.

Chatterjee, S., & Diaconis, P. (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics*, *41*(5), 2428–2461.

Clutton-Brock, T. H., & Harvey, P. H. (1980). Primates, brains and ecology. *Journal of Zoology*, *190*, 309–323.

Cranmer, S. J., & Desmarais, B. A. (2011). Inferential network analysis with exponential random graph models. *Political Analysis*, *19*(1), 66–86.

Cronin, K. A., Leeuwen, E. J. C. V., Vreeman, V., & Haun, D. B. M. (2014). Population-level variability in the social climates of four chimpanzee societies. *Evolution and Human Behavior*, *35*, 389–396.

Curry, O. S., & Dunbar, R. I. M. (2013a). Do birds of a feather flock together? The relationship between similarity and altruism in social networks. *Human Nature*, *24*, 336–347.

Curry, O. S., & Dunbar, R. I. M. (2013b). Sharing a joke: The effects of a similar sense of humour on affiliation and altruism. *Evolution and Human Behavior*, *34*, 125–129.

Demır, M., & Weitekamp, L. A. (2007). I am so happy'cause today I found my friend: Friendship and personality as predictors of happiness. *Journal of Happiness Studies*, *8*, 181–211.

Desmarais, B. A., & Cranmer, S. J. (2012). Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and its Applications*, *391*, 1865–1876.

Dominguez, S., & Arford, T. (2010). It is all about who you know: Social capital and health in low-income communities. *Health Sociology Review*, *19*, 114–129.

Donnelly, K. (2015). *Adolphe Quetelet, social physics and the average men of science, 1796–1874*. Routledge.

Dorogovtsev, S. N., Goltsev, A. V., & Mendes, J. F. F. (2008). Critical phenomena in complex networks. *Reviews of Modern Physics*, *80*, 1275.

Dunbar, R. I. M. (1991). Functional significance of social grooming in primates. *Folia Primatologica*, *57*, 121–131.

Dunbar, R. I. M. (1992a). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, *22*, 469–493.

REFERENCES

Dunbar, R. I. M. (1992b). Time: A hidden constraint on the behavioural ecology of baboons. *Behavioral Ecology and Sociobiology*, *31*, 35–49.

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, *16*, 681–694.

Dunbar, R. I. M. (1998a). *Grooming, gossip, and the evolution of language*. Harvard University Press.

Dunbar, R. I. M. (1998b). The Social Brain Hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *6*, 178–190.

Dunbar, R. I. M. (2016). Sexual segregation in human conversations. *Behaviour*, *153*, 1–14.

Dunbar, R. I. M. (2018). The anatomy of friendship. *Trends in Cognitive Sciences*, *22*, 32–51.

Dunbar, R. I. M. (2020). Structure and function in human and primate social networks: Implications for diffusion, network stability and health. *Proceedings of the Royal Society A*, *476*, 20200446.

Dunbar, R. I. M. (2021). *Friends: Understanding the power of our most important relationships*. Hachette UK.

Dunbar, R. I. M., Arnaboldi, V., Conti, M., & Passarella, A. (2015). The structure of online social networks mirrors those in the offline world. *Social Networks*, *43*, 39–47.

Dunbar, R. I. M., MacCarron, P., & Shultz, S. (2018). Primate social group sizes exhibit a regular scaling pattern with natural attractors. *Biology Letters*, *14*, 20170490.

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. *Science*, *317*, 1344–1347.

Dunbar, R. I. M., & Shultz, S. (2021). Social complexity and the fractal structure of group size in primate social evolution. *Biological Reviews*, *96*, 1889–1906.

Dunbar, R. I. M., & Sosis, R. (2018). Optimising human community sizes. *Evolution and Human Behavior*, *39*, 106–111.

Dunbar, R. I. M., & Spoors, M. (1995). Social networks, support cliques, and kinship. *Human Nature*, *6*, 273–290.

Erdös, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae Debrecen*, *6*, 290–297.

Escribano, D. (2019). *Statistical mechanics of social relationships* (Master's thesis). Universidad Carlos III de Madrid (UC3M).

Escribano, D., & Cuesta, J. A. (2022). Free-energy density functional for Strauss's model of transitive networks. *Physical Review E*, *106*, 054305.

Escribano, D., Doldán-Martelli, V., Cronin, K. A., Haun, D. B., van Leeuwen, E. J., Cuesta, J. A., & Sánchez, A. (2022). Chimpanzees organize their social relationships like humans. *Scientific Reports*, *12*, 16641.

Escribano, D., Doldán-Martelli, V., Lapuente, F. J., Cuesta, J. A., & Sánchez, A. (2021). Evolution of social relationships between first-year students at middle school: From cliques to circles. *Scientific Reports*, *11*, 11694.

Everett, M. G., & Borgatti, S. P. (2014). Networks containing negative ties. *Social Networks*, *38*, 111–120.

Feng, D., Altmeyer, R., Stafford, D., Christakis, N. A., & Zhou, H. H. (2022). Testing for balance in social networks. *Journal of the American Statistical Association*, *117*, 156–174.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*, 75–174.

Foster, D., Foster, J., Paczuski, M., & Grassberger, P. (2010). Communities, clustering phase transitions, and hysteresis: Pitfalls in constructing network ensembles. *Physical Review E*, *81*, 046115.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, *81*, 832–842.

Freeman, L. (2004). The development of social network analysis. *A Study in the Sociology of Science*, *1*, 159–167.

Fronczak, A. (2014). Exponential random graph models. *Encyclopedia of Social Network Analysis and Mining*, 500–517.

Fronczak, A. (2018). Exponential Random Graph Models. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of social network analysis and mining* (pp. 810–826). Springer.

Fruteau, C., Lemoine, S., Hellard, E., van Damme, E., & Noë, R. (2011). When females trade grooming for grooming: Testing partner control and partner choice models of cooperation in two primate species. *Animal Behaviour*, *81*, 1223–1230.

Fung, H. H., Carstensen, L. L., & Lang, F. R. (2001). Age-related patterns in social networks among European Americans and African Americans: Implications for socioemotional selectivity across the life span. *The International Journal of Aging and Human Development*, *52*, 185–206.

Galam, S. (2014). *When humans interact like atoms*. Psychology Press.

Gibson, K. R. (1986). Cognition, brain size, and the extraction of embedded food resources. *Primate Ontogeny, Cognition, and Social Behavior*.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences (USA)*, *99*, 7821–7826.

Godfrey, S. S., Moore, J. A., Nelson, N. J., & Bull, C. M. (2010). Social network structure and parasite infection patterns in a territorial reptile, the tuatara (Sphenodon punctatus). *International Journal for Parasitology*, *40*, 1575–1585.

Gonçalves, B., Perra, N., & Vespignani, A. (2011). Modeling users' activity on Twitter networks: Validation of Dunbar's number. *PLoS One*, *6*, e22656.

REFERENCES

Goyal, S. (2007). *Connections: An introduction to the economics of networks*. Princeton University Press.

Grandy Jr, W. T. (2012). *Foundations of Statistical Mechanics: Equilibrium Theory* (Vol. 19). Springer Science & Business Media.

Granell, C., Gómez, S., & Arenas, A. (2011). Mesoscopic analysis of networks: Applications to exploratory analysis and data clustering. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *21*, 016102.

Guclu, H., Read, J., Jr, C. J. V., Galloway, D. D., Gao, H., Rainey, J. J., Uzicanin, A., Zimmer, S. M., & Cummings, D. A. T. (2016). Social contact networks and mixing among students in K-12 schools in Pittsburgh, PA. *PLoS One*, *11*, e0151139.

Haerter, J. O., Jamtveit, B., & Mathiesen, J. (2012). Communication dynamics in finite capacity social networks. *Physical Review Letters*, *109*, 168701.

Handcock, M. S., Robins, G., Snijders, T., Moody, J., & Besag, J. (2003). Assessing degeneracy in statistical models of social networks. *Journal of the American Statistical Association*, *76*, 33–50.

Hansen, J.-P., & McDonald, I. R. (2013). *Theory of simple liquids* (4th). Academic Press.

Harary, F. (1959). On the measurement of structural balance. *Behavioral Science*, *4*, 316–323.

Harvey, P. H., & Krebs, J. R. (1990). Comparing brains. *Science*, *249*, 140–146.

Harvey, P. H., & Pagel, M. D. (1991). *The comparative method in evolutionary biology* (Vol. 239). Oxford University Press.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology*, *21*(1), 107–112.

Hill, R. A., Bentley, R. A., & Dunbar, R. I. M. (2008). Network scaling reveals consistent fractal pattern in hierarchical mammalian societies. *Biology Letters*, *4*, 748–751.

Hill, R. A., & Dunbar, R. I. M. (2003). Social network size in humans. *Human Nature*, *14*, 53–72.

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, *76*(373), 33–50.

Holme, P., & Kim, B. J. (2002). Growing scale-free networks with tunable clustering. *Physical Review E*, *65*, 026107.

Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports*, *519*, 97–125.

Honerkamp, J. (1998). *Statistical physics: An advanced approach with applications*. Springer.

Huang, K. (1987). *Statistical mechanics* (2nd). John Wiley & Sons.

Huitsing, G., & Veenstra, R. (2012). Bullying in classrooms: Participant roles from a social network perspective. *Aggressive Behavior*, *38*, 494–509.

Hume, D. (1739). *A treatise of human nature* (Vol. 1).

Humphrey, N. K. (1976). The social function of intellect. *Growing Points in Ethology*, *37*, 303–317.

Ilany, A., Barocas, A., Koren, L., Kam, M., & Geffen, E. (2013). Structural balance in the social networks of a wild mammal. *Animal Behaviour*, *85*, 1397–1405.

Iranzo, J., Krupovic, M., & Koonin, E. V. (2016). The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. *mBio*, *7*, e00978–16.

Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.-F., & den Broeck, W. V. (2011). What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, *271*, 166–180.

Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.

Jackson, M. O. (2019). The friendship paradox and systematic biases in perceptions and social norms. *Journal of Political Economy*, *127*, 777–818.

Jackson, M. O., Rodriguez-Barraquer, T., & Tan, X. (2012). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, *102*, 1857–1897.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One*, *9*, e98679.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press.

Jenkins, R., Dowsett, A. J., & Burton, A. M. (2018). How many faces do people know? *Proceedings of the Royal Society B*, *285*, 20181319.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, *407*, 651–654.

Jerison, J. H. (1973). *Evolution of the brain and intelligence*. New York: Academy Press.

Jonasson, J. (1999). The random triangle model. *Journal of Applied Probability*, *36*, 852–867.

Kaburu, S. S. K., & Newton-Fisher, N. E. (2015). Egalitarian despots: hierarchy steepness, reciprocity and the grooming-trade model in wild chimpanzees, Pan troglodytes. *Animal Behaviour*, *99*, 61–71.

Kanai, R., Bahrami, B., Roylance, R., & Rees, G. (2012). Online social network size is reflected in human brain structure. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 1327–1334.

Kappeler, P. M., & van Schaik, C. P. (2002). Evolution of primate social systems. *International Journal of Primatology*, *23*, 707–740.

Kasper, C., & Voelkl, B. (2009). A social network analysis of primate groups. *Primates*, *50*, 343–356.

Kawasaki, K. (1972). Kinetics of Ising models. *Phase Transitions and Critical Phenomena* (pp. 443–501). Academic Press.

REFERENCES

Kef, S. (1997). The personal networks and social supports of blind and visually impaired adolescents. *Journal of Visual Impairment & Blindness*, *91*, 236–244.

Kiesow, H., Dunbar, R. I. M., Kable, J. W., Kalenscher, T., Vogeley, K., Schilbach, L., Marquand, A. F., Wiecki, T. V., & Bzdok, D. (2020). 10,000 social brains: Sex differentiation in human brain anatomy. *Science Advances*, *6*, eaaz1170.

Killworth, P. D., Bernard, H. R., McCarty, C., Doreian, P., Goldenberg, S., Underwood, C., Harries-Jones, P., Keesing, R. M., Skvoretz, J., & Wemegah, M. V. S. (1984). Measuring patterns of acquaintanceship. *Current Anthropology*, *25*, 381–397.

Koskinen, J. H., Robins, G. L., Wang, P., & Pattison, P. E. (2013). Bayesian analysis for partially observed network data, missing ties, attributes and actors. *Social Networks*, *35*, 514–527.

Krause, J., James, R., Franks, D. W., & Croft, D. P. (2015). *Animal social networks*. Oxford University Press.

Kucharski, A. J., Wenham, C., Brownlee, P., Racon, L., Widmer, N., Eames, K. T. D., & Conlan, A. J. K. (2018). Structure and consistency of self-reported social contact networks in British secondary schools. *PLoS One*, *13*, e0200090.

Kudo, H., & Dunbar, R. I. M. (2001). Neocortex size and social network size in primates. *Animal Behaviour*, *62*, 711–722.

Kummer, H. (1982). Social knowledge in free-ranging primates. *Animal Mind—Human Mind*, 113–130.

Kwak, S., Joo, W., Youm, Y., & Chey, J. (2018). Social brain volume is associated with in-degree social network size among older adults. *Proceedings of the Royal Society B: Biological Sciences*, *285*, 20172708.

Lafuente, L., & Cuesta, J. A. (2004). Density functional theory for general hard-core lattice gases. *Physical Review Letters*, *93*, 130603.

Lafuente, L., & Cuesta, J. A. (2005). Cluster density functional theory for lattice models based on the theory of Möbius functions. *Journal of Physics A: Mathematical and General*, *38*, 7461–7482.

Lamb, S. D., Altobelli, J. T., Easton, L. J., Godfrey, S. S., & Bishop, P. J. (2022). Captive Hamilton's frog (Leiopelma hamiltoni) associates non-randomly under retreat sites: Preliminary insights into their social networks. *New Zealand Journal of Zoology*, *49*, 236–251.

Launay, J., & Dunbar, R. I. M. (2015). Playing with strangers: Which shared traits attract us most to new people? *PLoS One*, *10*, e0129688.

Lavis, D. A., & Bell, G. M. (1999). *Statistical mechanic of lattice systems* (Vol. 1). Springer.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., & Gutmann, M. (2009). Computational social science. *Science*, *323*, 721–723.

Lehmann, E. L., Romano, J. P., & Casella, G. (2005). *Testing statistical hypotheses* (Vol. 3). Springer.

Lewis, P. A., Rezaie, R., Brown, R., Roberts, N., & Dunbar, R. I. M. (2011). Ventromedial prefrontal volume predicts understanding of others and social network size. *Neuroimage*, *57*, 1624–1629.

Lincoln Park Zoo. (2020). Zoomonitor (version 3.2.) [mobile app]. https://zoomonitor.org

Liu, M., & Chen, X. (2003). Friendship networks and social, school and psychological adjustment in Chinese junior high school students. *Psychology in the Schools*, *40*, 5–17.

López, F. A., Barucca, P., Fekom, M., & Coolen, A. C. C. (2018). Exactly solvable random graph ensemble with extensively many short cycles. *Journal of Physics A: Mathematical and Theoretical*, *51*, 085101.

López, F. A., & Coolen, A. C. C. (2020). Imaginary replica analysis of loopy regular random graphs. *Journal of Physics A: Mathematical and Theoretical*, *53*, 065002.

López, F. A., & Coolen, A. C. C. (2021). Transitions in random graphs of fixed degrees with many short cycles. *Journal of Physics: Complexity*, *2*, 035010.

López-Pintado, D. (2016). An overview of diffusion in complex networks. *Complex Networks and Dynamics: Social and Economic Interactions*, 27–48.

Lubbers, M. J., Molina, J. L., & Valenzuela-García, H. (2019). When networks speak volumes: Variation in the size of broader acquaintanceship networks. *Social Networks*, *56*, 55–69.

Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press.

MacCarron, P., Kaski, K., & Dunbar, R. I. M. (2016). Calling Dunbar's numbers. *Social Networks*, *47*, 151–155.

Marsden, P. V. (2003). Interviewer effects in measuring network size using a single name generator. *Social Networks*, *25*, 1–16.

Martin, R. D. (1981). Relative brain size and basal metabolic rate in terrestrial vertebrates. *Nature*, *293*, 57–60.

Mason, O., & Verwoerd, M. (2007). Graph theory and networks in biology. *IET Systems Biology*, *1*(2), 89–119.

Massen, J., Sterck, E., & de Vos, H. (2010). Close social associations in animals and humans: Functions and mechanisms of friendship. *Behaviour*, *147*, 1379–1412.

Mastrandrea, R., Fournet, J., & Barrat, A. (2015). Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS One*, *10*, e0136497.

REFERENCES

McCarty, C., Lubbers, M. J., Vacca, R., & Molina, J. L. (2019). *Conducting personal network research: A practical guide*. Guilford Publications.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415–444.

Migliano, A. B., Battiston, F., Viguier, S., Page, A. E., Dyble, M., Schlaepfer, R., Smith, D., Astete, L., Ngales, M., Gomez-Gardenes, J., Latora, V., & Vinicius, L. (2020). Hunter-gatherer multilevel sociality accelerates cumulative cultural evolution. *Science Advances, 6*, eaax5913.

Miller, J. C. (2009). Percolation and epidemics in random clustered networks. *Physical Review E, 80*, 020901.

Miritello, G., Moro, E., Lara, R., Martínez-López, R., Belchamber, J., Roberts, S. G. B., & Dunbar, R. I. M. (2013). Time as a limited resource: Communication strategy in mobile phone networks. *Social Networks, 35*, 89–95.

Moreno, J. L., & Jennings, H. H. (1938). Statistics of social configurations. *Sociometry*, 342–374.

Newman, M. E. J. (2009). Random graphs with clustering. *Physical Review Letters, 103*, 058701.

Newman, M. E. J. (2018). *Networks: An Introduction*. Oxford University Press.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*, 026113.

Noonan, M. P., Mars, R. B., Sallet, J., Dunbar, R. I. M., & Fellows, L. K. (2018). The structural and functional brain networks that support human social networks. *Behavioural Brain Research, 355*, 12–23.

Page, A. E., Chaudhary, N., Viguier, S., Dyble, M., Thompson, J., Smith, D., Salali, G. D., Mace, R., & Migliano, A. B. (2017). Hunter-gatherer social networks and reproductive success. *Scientific Reports, 7*, 1153.

Pagel, M. D., & Harvey, P. H. (1988). How mammals produce large-brained offspring. *Evolution, 42*, 948–957.

Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2004). Statistical mechanics of topological phase transitions in networks. *Physical Review E, 69*, 046117.

Park, J., & Newman, M. E. J. (2004a). Solution of the two-star model of a network. *Physical Review E, 70*, 066146.

Park, J., & Newman, M. E. J. (2004b). Statistical mechanics of networks. *Physical Review E, 70*, 066117.

Park, J., & Newman, M. E. J. (2005). Solution for the properties of a clustered network. *Physical Review E, 72*, 026136.

Pascual, M. (2005). Computational ecology: From the complex to the simple and back. *PLoS Computational Biology*, *1*, e18.

Pastor-Satorras, R., Vázquez, A., & Vespignani, A. (2001). Dynamical and correlation properties of the Internet. *Physical Review Letters*, *87*, 258701.

Pentland, A. (2014). *Social physics: How good ideas spread the lessons from a new science*. Penguin.

Pollet, T. V., Roberts, S. G. B., & Dunbar, R. I. M. (2011). Extraverts have larger social network layers. *Journal of Individual Differences*.

Pospelov, N., Nechaev, S., Anokhin, K., Valba, O., Avetisov, V., & Gorsky, A. (2019). Spectral peculiarity and criticality of a human connectome. *Physical Life Review*, *31*, 240–256.

Quetelet, A. (1835). *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale* (Vol. 2).

Read, J. M., Eames, K. T. D., & Edmunds, W. J. (2008). Dynamic social networks and the implications for the spread of infectious disease. *Journal of the Royal Society Interface*, *5*, 1001–1007.

Reblin, M., & Uchino, B. N. (2008). Social and emotional support and its implication for health. *Current Opinion in Psychiatry*, *21*, 201.

Roberts, S. B. G., & Dunbar, R. I. M. (2015). Managing relationship decay: Network, gender, and contextual effects. *Human Nature*, *26*, 426–450.

Roberts, S. G. B., & Dunbar, R. I. M. (2011). Communication in social networks: Effects of kinship, network size, and emotional closeness. *Personal Relationships*, *18*, 439–452.

Roberts, S. G. B., Dunbar, R. I. M., Pollet, T. V., & Kuppens, T. (2009). Exploring variation in active network size: Constraints and ego characteristics. *Social Networks*, *31*, 138–146.

Robins, G., Pattison, P., & Elliott, P. (2001). Network models for social influence processes. *Psychometrika*, *66*, 161–189.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, *29*, 173–191.

Ron, T., & McGrew, W. C. (1988). Ecological assessment for a chimpanzee rehabilitation project in Northern Zambia. *Primate Conservation*, *9*, 37–41.

Roy, C., Bhattacharya, K., Dunbar, R. I. M., & Kaski, K. (2022). Turnover in close friendships. *Scientific Reports*, *12*, 11018.

Sah, P., Méndez, J. D., & Bansal, S. (2019). A multi-species repository of social networks. *Scientific Data*, *6*, 44.

Saramäki, J., Leicht, E. A., López, E., Roberts, S. G. B., Reed-Tsochas, F., & Dunbar, R. I. M. (2014). Persistence of social signatures in human communication. *Proceedings of the National Academy of Sciences (USA)*, *111*, 942–947.

REFERENCES

Savic, I., Garcia-Falgueras, A., & Swaab, D. F. (2010). Sexual differentiation of the human brain in relation to gender identity and sexual orientation. *Progress in Brain Research*, *186*, 41–62.

Sawyer, R. K. (2005). *Social emergence: Societies as complex systems*. Cambridge University Press.

Schweitzer, F. (2018). Sociophysics. *Physics Today*, *71*, 40.

Sekara, V., Stopczynski, A., & Lehmann, S. (2016). Fundamental structures of dynamic social networks. *Proceedings of the National Academy of Sciences (USA)*, *113*, 9977–9982.

Silk, J. (2002). Using the 'F'-word in primatology. *Behaviour*, *139*, 421–446.

Smirnov, I., & Thurner, S. (2017). Formation of homophily in academic performance: Students change their friends rather than performance. *PLoS One*, *12*, e0183473.

Smith, K. P., & Christakis, N. A. (2008). Social networks and health. *Annual Review of Sociology*, *34*, 405–429.

Snijders, T. A. B. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, *3*, 1–40.

Snijders, T. A. B. (2011). Statistical models for social networks. *Annual Review of Sociology*, *37*, 131–153.

Snijders, T. A. B., Wang, P., & Steglich, C. E. (2010). Temporal exponential-family random graph models for social networks. *Journal of the American Statistical Association*, *105*(490), 890–903.

South, S. J., & Haynie, D. L. (2004). Friendship networks of mobile adolescents. *Social Forces*, *83*, 315–350.

Spitzer, F. (1971). Markov random fields and Gibbs ensembles. *The American Mathematical Monthly*, *78*, 142–154.

Stadtfeld, C., Takács, K., & Vörös, A. (2020). The emergence and stability of groups in social networks. *Social Networks*, *60*, 129–145.

Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.-F., Quaggiotto, M., den Broeck, W. V., Régis, C., & Lina, B. (2011). High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS One*, *6*, e23176.

Stewart, J. Q. (1948). Concerning "social physics". *Scientific American*, *178*, 20–23.

Stiller, J., & Dunbar, R. I. M. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, *29*, 93–104.

Stivala, A. D., Koskinen, J. H., Rolls, D. A., Wang, P., & Robins, G. L. (2016). Snowball sampling for estimating exponential random graph models for large networks. *Social Networks*, *47*, 167–188.

Strauss, D. (1986). On a general class of models for interaction. *SIAM Review*, *28*, 513–527.

Strogatz, S. H. (2001). Exploring complex networks. *Nature, 410*, 268–276.

Sutcliffe, A. G., Dunbar, R. I. M., Binder, J., & Arrow, H. (2012). Relationships and the social brain: Integrating psychological and evolutionary perspectives. *British Journal of Psychology, 103*, 149–168.

Sutcliffe, A. G., Dunbar, R. I. M., & Wang, D. (2016). Modelling the evolution of social structure. *PLoS One, 11*, e0158605.

Tamarit, I. (2019). *Ego-centred models of social networks: The social atom* (PhD. thesis). Universidad Carlos III de Madrid (UC3M).

Tamarit, I., Cuesta, J. A., Dunbar, R. I. M., & Sánchez, A. (2018). Cognitive resource allocation determines the organization of personal networks. *Proceedings of the National Academy of Sciences (USA), 115*, 8316–8321.

Tamarit, I., Sánchez, A., & Cuesta, J. A. (2022). Beyond Dunbar circles: a continuous description of social relationships and resource allocation. *Scientific Reports, 12*, 1–11.

Tamm, M. V., Shkarin, A. B., Avetisov, V. A., Valba, O. V., & Nechaev, S. K. (2014). Islands of stability in motif distributions of random networks. *Physical Review Letters, 113*, 095701.

Toivonen, R., Kovanen, L., Kivelä, M., Onnela, J.-P., Saramäki, J., & Kaski, K. (2009). A comparative study of social network models: Network evolution models and nodal attribute models. *Social Networks, 31*(4), 240–254.

van Leeuwen, E. J. C., Bruinstroop, B. M., & Haun, D. B. (2022). Early trauma leaves no social signature in sanctuary-housed chimpanzees (Pan troglodytes). *Animals, 13*, 49.

van Leeuwen, E. J. C., Cronin, K. A., & Haun, D. B. M. (2018). Population-specific social dynamics in chimpanzees. *Proceedings of the National Academy of Sciences (USA), 115*, 11393–11400.

van Leeuwen, E. J. C., Cronin, K. A., & Haun, D. B. M. (2019). Reply to Farine and Aplin: Chimpanzees choose their association and interaction partners. *Proceedings of the National Academy of Sciences (USA), 116*, 16676–16677.

van Leeuwen, E. J. C., Cronin, K. A., Haun, D. B. M., Mundry, R., & Bodamer, M. D. (2012). Neighbouring chimpanzee communities show different preferences in social grooming behaviour. *Proceedings of the Royal Society B: Biological Sciences, 279*, 4362–4367.

Waal, F. B. M. D. (1997). The chimpanzee's service economy: Food for grooming. *Evolution and Human Behavior, 18*, 375–386.

Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p. *Psychometrika, 61*, 401–425.

Watts, D. (2002). Reciprocity and interchange in the social relationships of wild male chimpanzees. *Behaviour, 139*, 343–370.

REFERENCES

Weatherall, D., Bell, J., Blakemore, C., Ludlow, R., Lord, & Walport, M. (2006). The use of non-human primates in research. https://royalsociety.org/topics-policy/publications/2006/weatherall-report

Webber, E., & Dunbar, R. (2020). The fractal structure of communities of practice: Implications for business organization. *PLoS One*, *15*, e0232204.

Wellman, B. (2001). Computer networks as social networks. *Science*, *293*(5537), 2031–2034.

Wellman, B., & Berkowitz, S. D. (1988). *Social structures: A network approach* (Vol. 15). CUP Archive.

Wellman, B., & Wortley, S. (1990). Different strokes from different folks: Community ties and social support. *American Journal of Sociology*, *96*, 558–588.

West, B. J., Massari, G. F., Culbreth, G., Failla, R., Bologna, M., Dunbar, R. I. M., & Grigolini, P. (2020). Relating size and functionality in human social networks through complexity. *Proceedings of the National Academy of Sciences (USA)*, *117*, 18355–18358.

Whiten, A., & Byrne, R. W. (1988). *The Machiavellian Intelligence Hypothesis*. Oxford University Press.

Wrzus, C., Hänel, M., Wagner, J., & Neyer, F. J. (2013). Social network changes and life events across the life span: A meta-analysis. *Psychological Bulletin*, *139*, 53.

Yin, M. (2016). A detailed investigation into near degenerate exponential random graphs. *Journal of Statistical Physics*, *164*, 241–253.

Yose, J., Kenna, R., MacCarron, M., & MacCarron, P. (2018). Network analysis of the Viking age in Ireland as portrayed in Cogadh Gaedhel re Gallaibh. *Royal Society Open Science*, *5*, 171024.

Zhou, W.-X., Sornette, D., Hill, R. A., & Dunbar, R. I. M. (2005). Discrete hierarchical organization of social group sizes. *Proceedings of the Royal Society B: Biological Sciences*, *272*, 439–444.