POLITECNICO DI TORINO, SORBONNE UNIVERSITÉ

Master's Degree in Physics of Complex Systems

Master's Degree Thesis

Higher-order structures in face-to-face interaction networks

Supervisors

Prof. Alain BARRAT

Prof. Mathieu GÉNOIS

Prof. Luca DALL'ASTA

Candidate

Thomas ROBIGLIO

July 2023

Summary

Face-to-face interactions in human gatherings have a significant impact in various contexts, including disease spreading and opinion dynamics. In this thesis, we investigate the temporal properties of group interactions in different settings using data recorded using the SocioPatterns platform. Our analysis focuses on higher-order structures, revealing that the distributions of group durations exhibit large tails, indicating the absence of a typical time scale for higher-order interactions in human gatherings. By examining the accompanying metadata associated with the contact data, we explore the role of homophily, which refers to the tendency of individuals to interact with others with whom share similar attributes, in face-to-face interactions. Interestingly, our findings demonstrate that the presence of higher-order homophily is possible even in social settings where the corresponding low-order homophily is absent. To better understand these dynamics, we present a simple model for human face-to-face interactions. This initial model fails to accurately reproduce the higher-order temporal statistics observed in the data. As a solution, we present a modified version of the model that successfully captures both levels of the empirical temporal statistics. The insights gained from this research provide a valuable foundation for future studies aiming to uncover the fundamental properties of human interactions. The exploration of higher-order structures and homophily holds great potential in deepening our understanding of the complex dynamics inherent in face-to-face interactions.

Acknowledgements

This thesis is the result of many months of research, coding, conversations, writing and proof-reading. I could not have reached this point without the unwavering support and contributions of many wonderful people, and I want to take a moment to express my heartfelt gratitude to each of them.

First and foremost, I want to thank my supervisors in Marseille, Alain Barrat, and Mathieu Génois, for guiding me throughout this research journey. I also want to extend my thanks to my supervisor in Torino, Luca Dall'Asta, who followed my project with interest and was always there to help me navigate the administrative challenges of this MSc. thesis.

To all the teachers of the PCS program, thank you for your dedication to our learning and for putting up with my occasional stubbornness and spirited debates. I have learned a great deal from all of you, and I appreciate your professional and personal commitment to this unique program.

I am incredibly grateful for the opportunity to collaborate with so many talented researchers during this thesis. My colleagues at CPT in Marseille, Giovanni and the NPL group in Torino, and the amazing XGI team from around the world - your insights, camaraderie and teachings shaped the way I want to do reaserch in my future.

To all my friends who were with me during these months of research and these two years of university, thank you for being there. From my family at the Collège Néerlandais in Paris to all the PCS students who shared the ups and downs of these two years, as well as my close friends in Torino who always made me feel at home - your friendship means the world to me.

There are no words to express how grateful I am for the care and encouragement of those closest to me. I deeply thank my parents Isabelle and Matteo and my brother Pierre for their support, presence and love in all these years of studies. I thank my grandparents and the rest of my family whose life examples I treasure. I thank my girlfriend Elena for bearing with me and bringing joy and love in all my days.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The complex nature of various biological, social, and technological systems stems from the many interactions between their components [1]. In recent years, various complex systems have been effectively characterized using network representations, in which interconnected nodes represent interacting pairs of elements of the system. The study of networks has a long history with roots in sociology [2, 3] and graph theory [4, 5], over the last two decades, the interdisciplinary field of network science has seen a significant expansion. Network science borrows from graph theory the formalism to deal with the graphs drawn by the relations among the elements of a system and from statistical physics the conceptual framework to deal with randomness and seek universal organizing principles [6]. This approach allows us to study the underlying network structure of different complex and disordered systems in terms of fundamental laws that limit and determine their behaviour [7, 1].

The study of complex systems through network science has far-reaching implications in various fields, including medicine, engineering, and economics. In medicine, it has been used to identify key genes involved in disease modules and pathways [8] and to predict the spread of infectious diseases [9, 10]. In engineering, it has been used to study localized failures in communication networks and power grids and design more robust and resilient technical systems that can withstand disruptions [11, 12]. In economics, it has been used to model financial interactions, such as co-ownership, and debtor-creditor relations, and leverage network scientific tools to predict and mitigate financial risk [13, 14]. As such, Network Science has become an important interdisciplinary field that is essential in understanding complex systems and finding solutions to real-world problems.

The relation between network science and social sciences is twofold [15, 16]. Firstly, many of the fundamental concepts and tools used by scientists for studying complex networks stem from social sciences. Straightforward examples of this are different network measures such as node centrality or clustering coefficient that were initially proposed in sociometry to quantify the social importance of

each individual [17]. One of the first representations of a complex system using the language of graph theory - predating its formal definition - is the so-called sociogram, developed in the pioneering studies by Moreno and Jennings [2, 18], to depict relations among a group of people graphically. Secondly, the study of social networks is one of the most relevant applications of network science. The topological structure of the graph drawn by the social interactions within a system - that might be defined in different ways *e.g.* social proximity, exchange of emails, engagements on social media... - has been proved to play a crucial role in the dynamic of socially relevant dynamical processes occurring on the network. For example in dynamical systems describing contagion processes the heavy-tailed distribution in the number of contacts within a population causes the epidemic threshold to vanish [9, 19] or the localized breakdown of a group of nodes can cause systemic damages in large communication systems - modeling for instance the World Wide Web - displaying processes known as cascading failures [11].

1.1 Temporal networks

As the field of network science has advanced, a growing necessity to move beyond the conventional graph representation for networks has emerged. A significant development in this regard is the introduction of temporal networks, which enable the evolution of network's topology over time [20].

Consider a system comprising of N individual nodes that engage in intermittent dyadic interactions observed within a specific time frame ranging from $t = t_{\min}$ to $t = t_{\max}$. For instance, in a social network, these nodes could represent individuals, in an ecological network they could describe species, and in a transport network, they could denote different locations. A temporal network serves as our representation of such observed interactions.

Definition 1 *A temporal network* $\mathcal{G}_T = (V_T, E_T, T)$ *is defined by a set of nodes* $\{V_T = 1, ..., N\}$*, a set of contact events* $E_T = e_1, e_2, ...$ *and a time-window* T*. A contact event* $e = (i, j, t, d)$ *represents an interaction between nodes* $i, j \in V_T$ *at time* t *which lasted for a duration* d*. The time window is defined so that for all events in* E_T *we have* $0 \leq t \leq T$ *and* $0 \leq d \leq T$*.*

Temporal networks offer a valuable framework for capturing the dynamic nature of complex systems, allowing us to analyze the changes in connectivity patterns and study the influence of time on network properties. By incorporating temporal aspects, we gain a deeper understanding of the underlying mechanisms that shape the network's structure and behavior [21].

Temporal networks have found wide-ranging applications in various fields. For instance, they have been employed to study human face-to-face interactions and

physical proximity, allowing researchers to analyze the dynamics of social networks and, for instance, gain insights into contagion phenomena [22, 23]. In neuroscience, temporal networks have been instrumental in investigating functional connections in the brain. By capturing the temporal dynamics of neuronal activity, researchers gain insights into how different brain regions interact and form complex networks underlying cognitive processes [24, 25]. Another notable application of temporal networks is in the analysis of mobile phone calls and text messages. By considering the temporal ordering and patterns of communication, temporal networks reveal valuable information about social relationships and communication dynamics or insights in dynamical processes such as the spread of information within a population [26, 27].

1.2 Higher-order structures in networks

An important limitation in the use of networks to model various complex systems is that only pairwise interactions can be represented [28]. This means that the evolution of the system under study can only come from dyads of elements influencing each other. In this picture, interactions between groups of agents are either neglected or projected down as combinations of pairwise interactions. Considering a simple Ising model, a network can successfully represent a Hamiltonian of the form:

$$H = -\sum_{i,j} J_{ij} s_i s_j \tag{1.1}$$

In this case, the nodes of the networks would represent the spins, being in the two possible states $\uparrow = +1$ or $\downarrow = -1$, and the weighted edges of the network would represent the existence and strength of the interactions. If instead we consider an Ising with plaquette (group) interactions between three elements, described by the Hamiltonian:

$$H = -\sum_{i,j} J_{ij} s_i s_j - \sum_{i,j,k} K_{ijk} s_i s_j s_k \tag{1.2}$$

we see that this system is impossible to represent as a network without assumptions or loss of information on the plaquette interactions.

In several real-world systems, ranging from biology, neuroscience, ecology and social sciences, group interactions are present and cannot be reduced to pairwise relations [28, 29]. For instance, if we consider the interactions between different scientists being co-authored papers, it is evident that a single paper with three authors is a completely different entity with respect to three papers written by the different pairs contained in the triplet [30]. In neuroscience, neuronal dynamics display synergistic behaviors that require interactions among multiple neurons to be predicted [31]. In ecosystems, three or more species routinely compete for food

and territory in complex ecosystems [32]. The presence of a third species influences the interaction between the other two, by affecting directly the interaction (the link) rather than the species involved (the nodes).

In the context of social networks, when considering opinion formation, rumor spreading and similar dynamics the role played by group interactions can be significant [33].

Peer pressure (sometimes referred to in the literature as peer influence) is a classic example that highlights the importance of studying group interactions in the context of social networks. Peer pressure refers to the influence exerted by a group of people on an individual's thoughts, beliefs, and behaviors. It can significantly impact decision-making processes, ranging from simple choices like fashion preferences to more consequential decisions like substance abuse or academic performance [34, 35].

In recent years there has been a growing interest of the network scientific community towards finding and analyzing explicit representations of group interactions in interconnected systems [36]. The natural candidates for providing such descriptions are hypergraphs and simplicial complexes (see Figure 1.1).

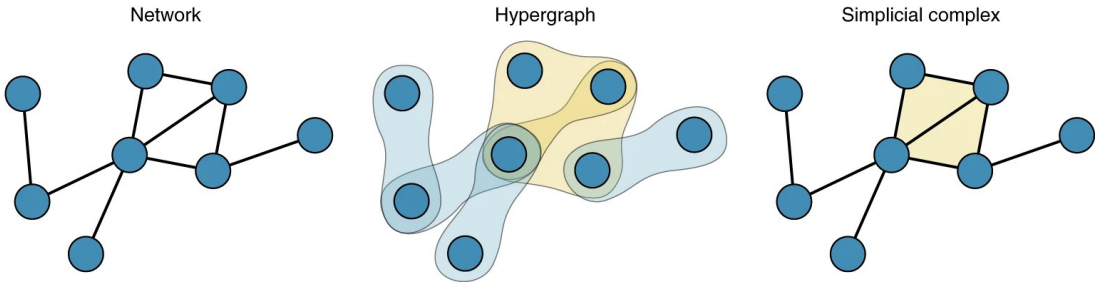

Figure 1.1: Group interactions in complex systems are not taken into account when representing systems as networks (left). Explicit representations of higher-order interactions are provided by hypergraphs (center) and simplicial complexes (right). From Battiston *et al.*, 2021 [36].

Hypergraphs are the straightforward generalization of graphs, allowing to encode interactions among arbitrary numbers of nodes.

Definition 2 *A* ***hypergraph*** *is defined by a set V, whose elements are known as vertices or nodes, and by a family E of subsets of V, known as hyperedges. A hypergraph is denoted by $\mathcal{H} = (V, E)$.*

Simplicial complexes offer another approach. Although more constrained than hypergraphs - all subfaces of a simplex (for example, the edges of a triangle) need to be included -, they provide access to powerful mathematical formalisms coming from algebraic topology.

Definition 3 *A* k***-simplex*** σ *is a set of* $k+1$ *nodes* $\sigma = [p_0, p_1, ..., p_k]$.

Definition 4 *A* ***simplicial complex*** $\mathcal{K}$ *on a given set of nodes* V*, with* $|V| = N$ *is a collection of simplices with the additional condition:*

$$\sigma \in \mathcal{K} \rightarrow \forall \nu \subset \sigma,\ \nu \in \mathcal{K}$$

where ν *is a sub-simplex of* σ*.*

In this thesis, we will use the less constrained representations of systems with group interactions given by hypergraphs.

Previous studies have demonstrated that higher-order interactions can greatly influence the behavior of networked systems, from diffusion and synchronization to social contagion and consensus emergence processes, potentially resulting in explosive phase transitions between states [37, 38, 39, 40]. These transitions, driven by interactions and connectivity, can lead to sudden and significant changes in the behavior and functioning of these systems. Particularly interesting to our scope is the cases of social contagion.

Phenomena that can be described as social contagions are widely present in nature and in human societies: examples of such phenomena are information diffusion and the propagation of social behaviors. These social contagion phenomena have been modeled on simplicial complexes [39] and the same model has been later extended to hypergraphs [41]. In these models, the standard susceptible-infected-susceptible (SIS) model for contagion [10] is extended to account for group interactions. In the SIS model on a network, the agents of a population are modeled as the nodes of a network, accordingly, the edges of the network represent the interactions among the agents. The nodes of the network can be in two states: susceptible or infected. Susceptible agents become infected with a certain probability per unit of time if in contact with at least one infected agent, while infected agents recover independently with a certain probability per unit of time. The extension of this model to higher-order structures is straightforward: if an agent participates in a group interaction (simplex or hyperedge depending on the model) of a certain order where all the other nodes are infected it gets infected with a certain probability per unit of time which depends on the order of the interaction. The striking result in these model is that when the effect of the group interactions is strong enough (*i.e.* above a certain value of the infection probability of the groups) the transition between the disease-free state (*i.e.* the stationary state of the model where the fraction of infected agents is zero) and the endemic state (*i.e.* the stationary state with a finite fraction of infected agents) becomes discontinuous. The distinction between the second-order phase transition observed in the simple dyadic model and the first-order phase transition found

in the higher-order model is of critical importance. The presence of an abrupt phase transition means that infinitesimal changes in the parameter controlling the infectivity of the spreading process can lead to finite changes in the stationary state reached by the system under study. Moreover the study of social contagion phenomena is important not only in the context of social sciences but also when considering biological pathogens - usually described with simple contagion models - as a simple spreading process if driven by a higher-order one can display the macroscopic features of a complex contagion [42].

1.3 Homophily in human gatherings

Homophily is a commonly observed phenomenon in social networks where interactions occur frequently among individuals that share common features [43, 44]. This phenomenon can be determined by the individuals' personal preferences to make contact with similar others (choice homophily), structural opportunities to interact with similar others (induced homophily), or a combination of these two types of causes. A simple example of choice homophily is the fact - statistically observed by Stehle *et al.*, 2013 [45] - that for temporally strong ties students in primary schools display gender preference that increases with grade, *i.e.* boys tend to interact more with boys than they do with girls and *viceversa*, and this tendency increases with the grade they are in.

Homophily might play an important role also when considering dynamical processes on networks such as social contagion. The presence of homophily implies that cultural, behavioral, or material information that flows through networks will tend to be localized. Another consequence of homophily is that distance in terms of social characteristics translates into network distance, the number of relationships or interactions through which a piece of information must travel to connect two individuals. Homophily also implies that any social entity that depends to a substantial degree on networks for its transmission will tend to be localized in social space [43].

An important example of how homophily might play an important role in social systems modeled as networks is given by Ref. [46]. The authors showed using a social network model with tunable homophily that homophily can put minority groups at a disadvantage by restricting their ability to establish links with a majority group or to access novel information.

1.4 Outline of the thesis

In this thesis, we propose a novel analysis of the temporal properties of face-to-face interactions in human gatherings. This kind of analysis is possible thanks

to the high temporal resolution of the measurements of face-to-face interactions performed using the SocioPatterns platform [47, 48]. Previous works have studied the probability distributions of contact duration in several contexts, ranging from scientific conferences [49] to workplaces [50], hospitals [51] and schools [52, 53], and have shown that the shape of this distribution is the same across the great variety of contexts under study. The robustness of this probability distribution over several empirical datasets uncovers some universal features of face-to-face interactions. Our analysis is focused on group or higher-order interactions that are overlooked when considering only the pairwise interactions. We investigate, thanks to the rich metadata accompanying some of the contacts datasets, the role played by homophily in face-to-face contacts. We show that in some of the datasets it is possible to observe higher-order homophily (*i.e.* homophily at the group level) not encoded at the level of simple, pairwise contacts between the participants. Afterward, we present an existing simple stochastic model [54] able to reproduce some of the temporal properties obtained from the empirical temporal network. We show the limitations of this model in reproducing the temporal statistics of group interactions and homophily effects. Finally, we explore some modifications of the model aiming to reproduce the higher-order structures observed in the empirical datasets.

This thesis is articulated as follows:

- In Chapter 2 we present the datasets that have been used in our work and discuss the results obtained in the empirical analysis.
- In Chapter 3 we present and discuss the concept of higher-order homophily and present the results on this subject obtained in our datasets.
- In Chapter 4 we present an existing stochastic model for face-to-face interactions in human gatherings, present the results obtained with this model and discuss its limitations in describing the temporal properties of group interactions and the homophily effects observed in the empirical measurements. We propose a modified version of the model that successfully reproduces the higher-order temporal properties of the data and present a modified version of the model accounting for homophily effects.
- Finally, in Chapter 5 we summarise the work and the major contributions of the thesis.

Chapter 2

Measuring social interactions

A social network is defined as a set of actors and their relationships of various kinds. The actors can represent individuals or social groups, while the relationships that define the edges of the network can take different forms, such as friendships, business or political relations, sexual encounters, local proximity, or social media interactions.

In this thesis, we focus on the social network that describes face-to-face interactions among people in different contexts of human gatherings. Following the approach used in measurements that follow the SocioPatterns platform (see further in §2.1) we define a face-to-face interaction between two individuals as occurring when they are physically closer than a certain distance (in our case, approximately 1.5 meters) and facing each other. This SocioPatterns infrastructure uses radio-frequency devices with short-range signals, which are screened by the human body. Thus, an interaction event is recorded only when the above conditions for defining a face-to-face interaction are met.

The study of social interactions has a long history in sociology and anthropology [15], starting from the already cited pioneering contributions predating the formal mathematical definition of graph theory [2]. Nevertheless, historically the findings were drawn from data collected via subjective reports (*i.e.* the subjects of the study reporting their interactions or their relations) or via external observations. For example in the previously cited seminal work by J. L. Moreno, the author studied the networks of friendships between children from kindergarten through eighth grade by interviewing the children and observing and annotating their interactions over a certain period of time.

In recent years, the study of social networks has seen significant progress due to the availability of large amounts of data on human interactions recorded in a variety of contexts: phone calls, social media interactions, co-location recorded using Bluetooth and Wi-fi technologies *etc.* [55]. For face-to-face interactions, the state of the art for collecting data with high spatial and temporal resolution is

represented by measurements performed using the SocioPatterns platform [47, 48]. This recording framework uses unobtrusive wearable devices that are able to track close ($\lesssim$ 1.5 m) face-to-face contacts between the participants with a temporal resolution of 20 s, and have been deployed in a wide variety of contexts. The availability of such datasets has allowed researchers to uncover and model various properties of face-to-face interactions [56, 45, 57, 58].

Studying the network of face-to-face interactions in human gatherings is important, for instance, in the context of epidemiology, specifically for understanding the diffusion of respiratory diseases transmitted through close contact [59, 60]. By examining the patterns of interactions and connections between individuals, researchers can gain insights into how diseases like COVID-19 or influenza spread within communities. For example, analyzing the network of contacts at a workplace can help to identify employees who are more likely to transmit a disease due to their high number of close contacts [50]. This knowledge can inform targeted intervention strategies such as implementing quarantines, promoting hygiene practices, or prioritizing vaccination efforts to mitigate the spread of the disease under study [61, 62].

In the following section, first we present in detail the SocioPatterns measuring platform and the datasets employed in the thesis. Then we outline how we move from the pairwise description of human interactions to the underlying higher-order structure and present our results regarding the empirical time distributions of groups in human gatherings.

2.1 The SocioPatterns platform

The structure of the network representing the interactions between individuals participating in a human gathering has a direct impact on the diverse phenomena that can take place on top of it. This interplay between structure and dynamics of human contacts is not only interesting from the physical and network scientific point of view but has crucial importance in various research areas. Straightforward examples of this come from epidemiology, where the mixing patterns among individuals determine the transmission of infectious diseases by the respiratory or close-contact route, and social sciences, where in-person interactions between individuals can shape the emergence of collective opinions or the spreading of rumors, trends and fake news.

An important contribution to the availability of representative data of such face-to-face interactions has been conducted over the last 15 years by the SocioPatterns collaboration [47, 48]. The datasets used in this thesis consist of longitudinal data on the physical proximity and face-to-face contacts of individuals in diverse real-world environments collected following the SocioPatterns infrastructure.

A social interaction can include several different human behaviors, such as conversation, and physical or eye contact. The SocioPatterns platform is based on the broader and more straightforward definition of a contact as physical proximity event between two individuals.

The equipment for the SocioPatterns measurements consists of sensors attached to the participants' chests (*e.g.* on their name tags in the case of scientific conferences) and antennas to collect contact data from the sensors covering the area of the venue where the experiment takes place. Each sensor carries an RFID (Radio Frequency IDentification) chip that can detect other sensors in the vicinity within a distance of $\sim$ 1.5 m. The human body blocks the emitted signal, hence the detection occurs only when two individuals are facing each other (see Figure 2.1). Detected events are defined as contacts. Contacts are recorded with a 20 s temporal resolution. The antennas are required to collect data from the sensors continuously as their memory is limited.

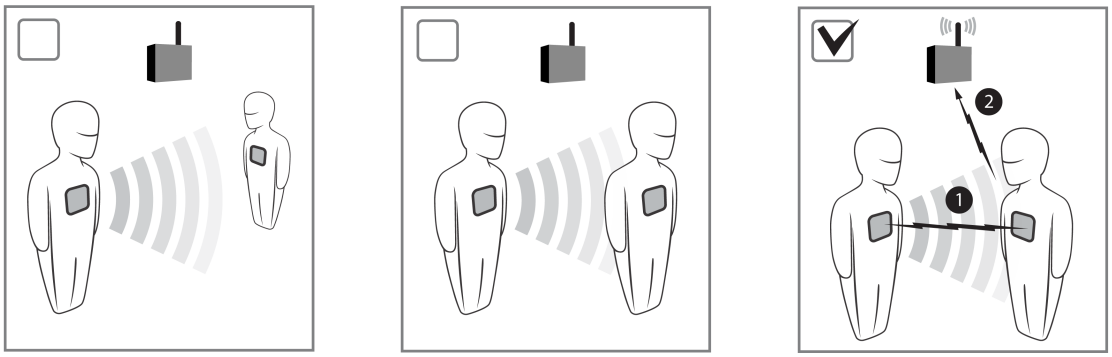

Figure 2.1: Schematic illustration of the RFID sensor system. RFID chips are worn as badges by the individuals participating in the deployments. A face-to-face contact is detected when two persons are close and facing each other. The interaction signal is then sent to the antenna. Figure from Cattuto *et al.*, 2015 [48].

The data resulting from measurements performed using the SocioPatterns platform is in the form of a temporal network in which the nodes are the participants in the gathering and the links represent contacts, appearing and disappearing as time passes. In the traditional network formalism the interaction between two agents a and b at time t, which lasted for a duration d is represented by a temporal link $e = (a, b, t, d)$. The sequence of contact events builds a temporal network $\mathcal{G}_T = (V_T, E_T, T)$, where V_T is the set of agents (*i.e.* the nodes), E_T is the set of contact events (*i.e.* the temporal edges) and the graph evolves in a time-window T so that $0 \leq t \leq T$ and $0 \leq d \leq T$. As the contacts are recorded with a temporal resolution of 20 s the temporal network evolves in discrete time.

2.2 Overview of the datasets

Six different datasets have been used in this thesis. The first one (from now on `primaryschool`) regards the contacts in a primary school (6–12 years old children) in France [52]. The second one (`thiers2011`) contains the temporal network of contacts between students in a high school in Marseilles, France [53]. The other four datasets (`WS16`, `ICCSS17`, `ECSS18` and `ECIR19`) are more recent [49] and have been collected in the context of international scientific conferences. The four datasets about the scientific conferences collected by Génois *et al.*, 2022 [49] are accompanied by rich metadata about the socio-demographical characteristics of the participants (*e.g.* gender, age, academic seniority, country of origin and several others), as well as survey answers about their perception of the conference and about their motivations to participate and several other potentially relevant aspects. The richness of this metadata is described in depth in the original paper [49]. Part of the socio-demographical metadata has been employed in our analysis of higher-order homophily in human gatherings (see Chapter 3).

In Figure 2.2 we display the evolution in time of the number of interactions between the participants. This evolution is similar for all the datasets: we observe a circadian rhythm, with active days and inactive nights. In addition to the circadian alternation of days and nights, we see the alternation of high and low-activity periods. It is intuitive that for the conferences' datasets high-activity periods are "social times" such as registration, coffee/lunch breaks, or poster sessions and low-activity periods are talk sessions. Note that the plenary/parallel talk sessions, where in general the majority of the participants sit in rows facing a distant speaker, are not expected to produce a lot of face-to-face close contact activity as the human body screens the signal of the RFID sensor. Analogously for the two school datasets the high-activity periods correspond to break times while low-activity periods correspond to class hours.

In the following subsections, we present the analysis of the aggregated networks and the study of the contact statistics obtained from the data.

2.2.1 Aggregated network analysis

The first analysis of each dataset that can be carried out is a simple analysis of the aggregated network resulting from the data. The aggregated network is obtained by flattening the temporal network described in §2.1 along the temporal dimension. We obtain a network in which nodes are the participants, and a link exists between two nodes if the participants have interacted at least once during the event.

For each dataset, we explore the following basic properties of the aggregated network:

- Number of participants, N.

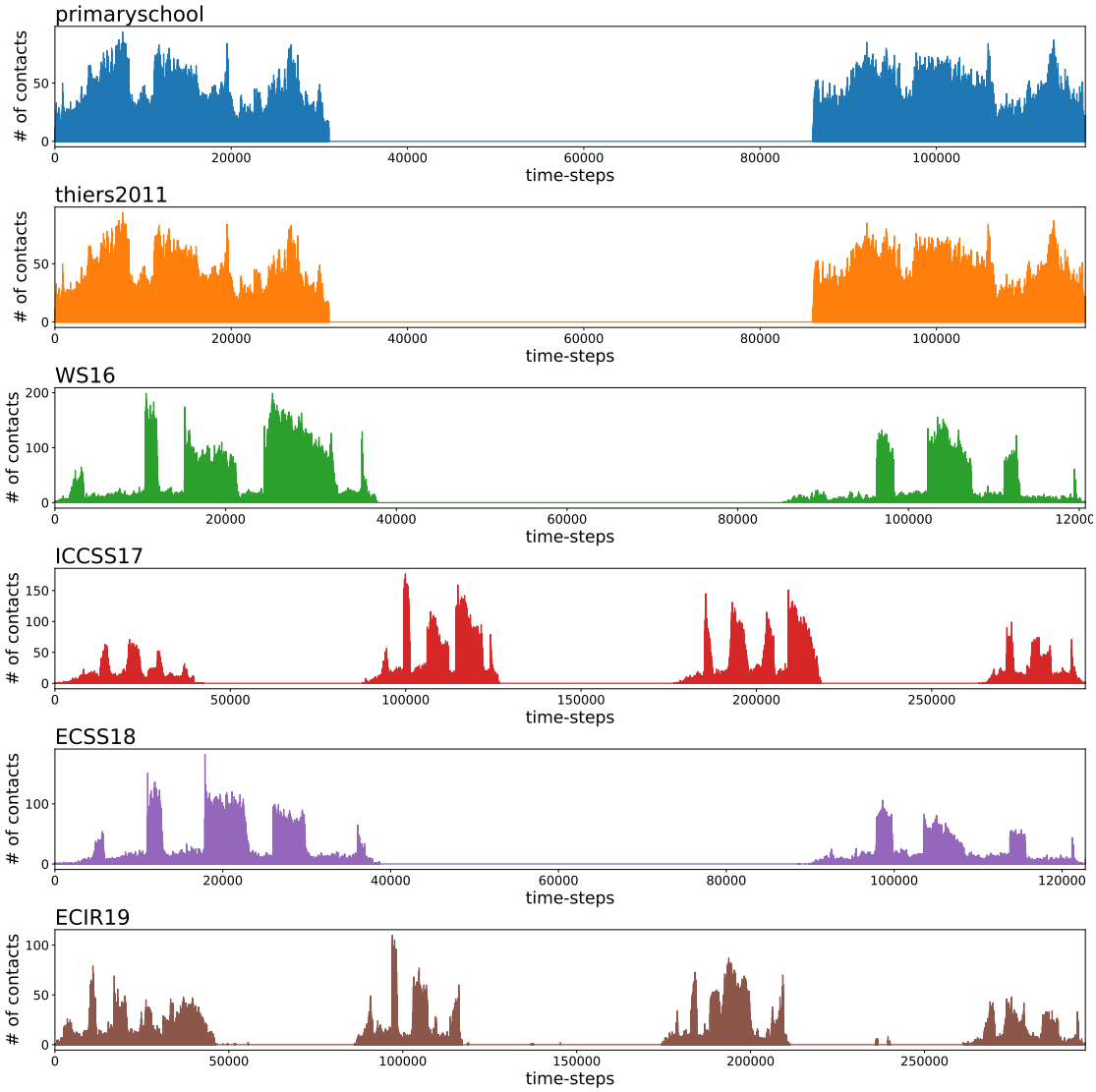

Figure 2.2: Total number of contacts occurring in each 20-second time-step.

- Total number of instantaneous contacts recorded, C.
- Density of the aggregated network, ρ, (*i.e.* the fraction of possible connections that occurred during the recording). This quantity is given by the formula:

$$\rho = \frac{2m}{N(N-1)} \tag{2.1}$$

where m is the number of edges in the aggregated network.

- Average degree of the aggregated network, $\langle k \rangle$, (*i.e.* the average number of persons one participant met during the event). This quantity is given by the formula:
$$\langle k \rangle = \frac{m}{N} \tag{2.2}$$

- Average clustering of the aggregated network, $\langle c \rangle$. This quantity describes on average the ratio between the number of triangles a node participates in and the number of all possible triangles. The clustering coefficient of a node u is given by the formula:
$$c_u = \frac{2T_u}{k_u(k_u - 1)}, \tag{2.3}$$
where T_u is the number of triangles node u participates in and k_u is the degree of node u.

The basic properties of each aggregated network are displayed in Table 2.1

	primaryschool	thiers2011	WS16	ICCSS17	ECSS19	ECIR19
N	242	126	138	274	164	172
C	125773	28540	153371	229536	96362	132949
ρ	0.285	0.217	0.794	0.495	0.567	0.550
$\langle k \rangle$	68.7	27.1	108.7	135.2	92.4	94.1
$\langle c \rangle$	0.526	0.576	0.868	0.694	0.717	0.746

Table 2.1: General properties of the aggregated contact networks

These basic properties of the datasets and of the aggregated networks described by them are in general heterogeneous, reflecting the different situations in which the face-to-face interaction data have been recorded.

2.2.2 Contact statistics

To explore the fundamental dynamical features of the datasets we evaluate two basic statistics regarding the contacts (see Figure 2.3).

We define any instantaneous contact occurring sequentially without in-between gaps as a continuous contact with a duration of τ (*i.e.* an interaction). With this definition, we can then explore the overall temporal properties of the interactions. Additionally, we can examine the inter-contact durations, denoted $\Delta\tau$, between two consecutive interactions between the same participants. These two quantities are relevant for the study of causal processes that can occur on the dynamical contact network, such as for example information diffusion or epidemic spreading.

Analyzing the empirical distributions of these variables, we find well-known, large-tail-shaped distributions. These results suggest that the majority of contacts and

inter-contact intervals last for 20 seconds, meaning that most pairs of participants interact only once and for a duration of 20 seconds. However, there are instances where each of these properties lasts for an extremely long time with a small but noteworthy probability, as evidenced by the power-law-like aspect of the distributions. This behavior describing the bursty nature of human interactions is observed in all 6 datasets. Moreover, the shape of the two distributions is very robust as observed in previous works [63, 48]. This robustness over several empirical datasets uncovers a universal nature of face-to-face interactions. The behavioral data on face-to-face proximity lack any intrinsically characteristic time scale, *i.e.*, no typical duration can be defined for any type of contact.

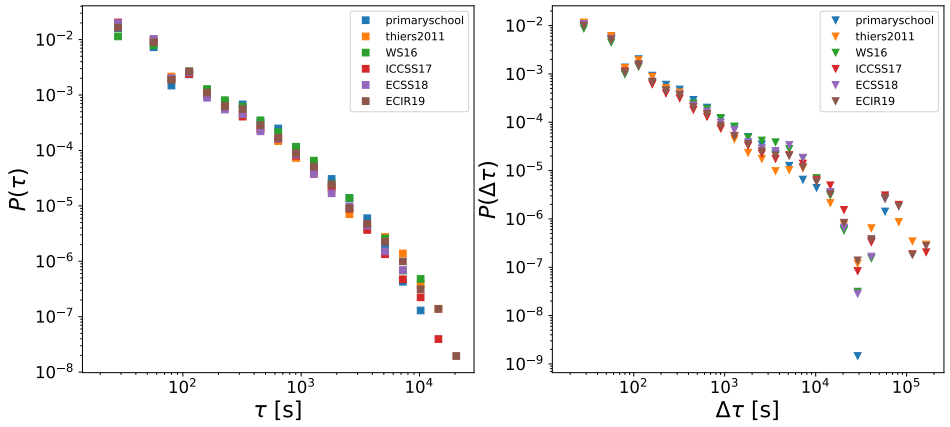

Figure 2.3: Distributions of the temporal features of contacts. τ (left) are the contiguous contact durations; $\Delta\tau$ (right) are the inter-contact durations.

2.3 Temporal higher-order social interactions

As outlined in §1.2 networks are limited to represent dyadic interactions. However, individuals might interact in larger groups of 3 or more individuals and the presence of such higher-order interaction calls for a different representation.

The straightforward generalization to account for higher-order social interaction is to consider a temporal hypergraph whose building blocks are known as temporal hyperedges. A n-hyperedge, or hyperedge of size n, describes the group interaction between n agents. A temporal n-hyperedge at time t and of duration d is then defined as $e_n = (i_1, ..., i_n, t, d)$.

In the considered datasets each interaction is stored only through simple links. Nevertheless, those dyadic interactions intuitively represent only the low-order projection of group interactions. In order to study the temporal properties of group

interactions it is necessary to attempt at reconstructing the original higher-order features of the social interactions. Following Cencetti *et al.*, 2021 [57] we construct the temporal hypergraph from the dyadic structure in the following way: if at time t there are $n \times (n+1)/2$ dyads between the members of a set of n nodes such that they form a fully connected clique, we promote the group of $n \times (n+1)/2$ links to a n-hyperedge (see Figure 2.4).

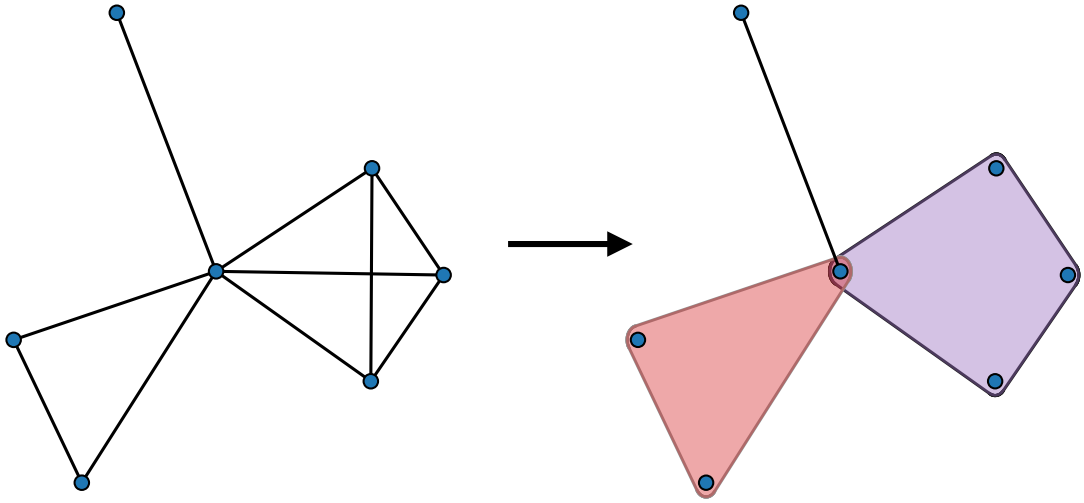

Figure 2.4: Promotion of cliques to hyperedges. This and all the following drawings of higher-order structures are drawn using the `XGI` Python library [64].

A simple example of the complete pipeline that translates the SocioPatterns data into a temporal hypergraph is displayed in Figure 2.5. We can see that the contact between Node 1 and Node 2 lasts for the first six time-steps, first as a group of size 3 and then as part of a group of three agents. We can also see that the group of four agents present time-step 8 is destroyed as soon as one of the contacts composing it (namely the contact between Node 1 and Node 2) disappears.

It is important to notice that with this formalization of the notion of group interactions a group of two people (*i.e.* a 2-clique in the temporal network) is different from the contact between them (*i.e.* a link in the temporal network). Hence we expect the dynamical features of these two types of events to be different.

In Figure 2.6 we show the timely occurrences of group interactions of sizes 2, 3 and 4 in the `ECIR19` dataset (similar patterns are observed across different days and different datasets). It is possible to observe that the emergence of higher-order structures is strongly heterogeneous in time, with the alternation between high and low-activity periods. This visualization highlights the existence of bursty patterns in higher-order interactions that are not independent among various

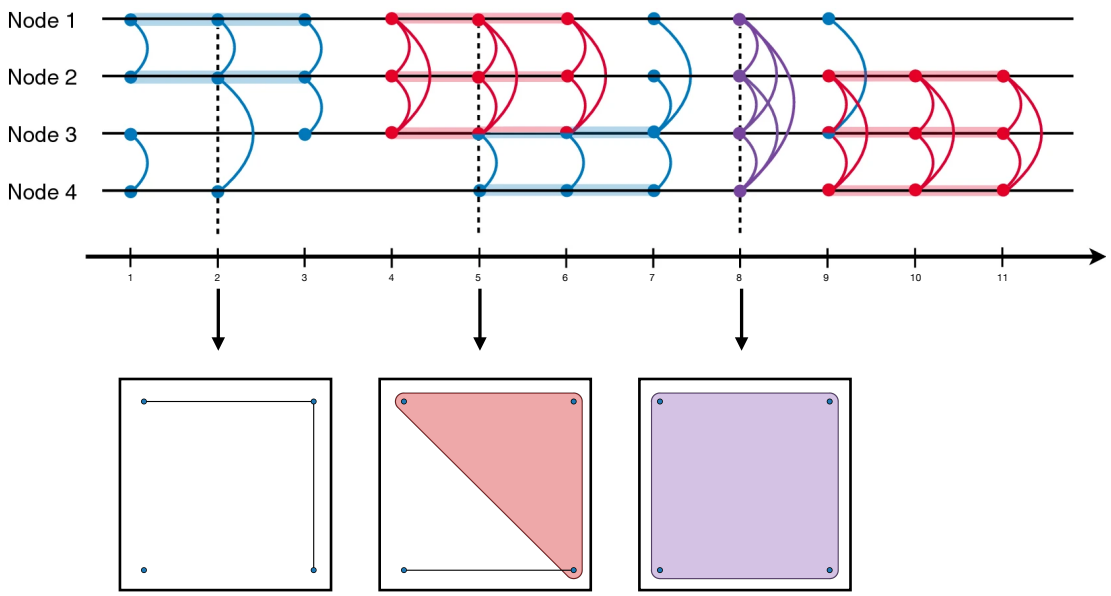

Figure 2.5: From SocioPatterns data to higher-order temporal structures. Example of higher-order interactions among four people. The horizontal lines represent the temporal behavior of each individual.

orders. This result is not unexpected since higher-order events are in our case face-to-face interactions constructed from lower-order structures. Nevertheless, their heterogeneous dynamics and short-term recurrence are far from being obvious. In the following sub-section, we will provide a more formal inspection of these features by analyzing the distributions of durations of such higher-order interactions.

2.3.1 Statistics of higher-order interactions

Similarly to what has been done in §2.2.2, we study the distributions of durations of groups of different sizes. We define any instantaneous interaction between a group of size n occurring sequentially without in-between gaps as a continuous group interaction of size n with duration τ_n.

The distributions of the durations of groups of different sizes (see Figure 2.7) display fat-tail shapes, similar to power-law distributions.

For all different sizes the majority of groups last 20 s, but the fat-tail distributions show that groups can last much longer - up to around 15 minutes for groups of 2 and 3 people and up to around 7-8 minutes for groups of 4 people. Also in this case it is possible to recognise some regularity in the shape of distributions of durations of groups of different sizes. It is worth noticing that the results obtained for the distributions of groups durations in Figure 2.7 seem to be less robust

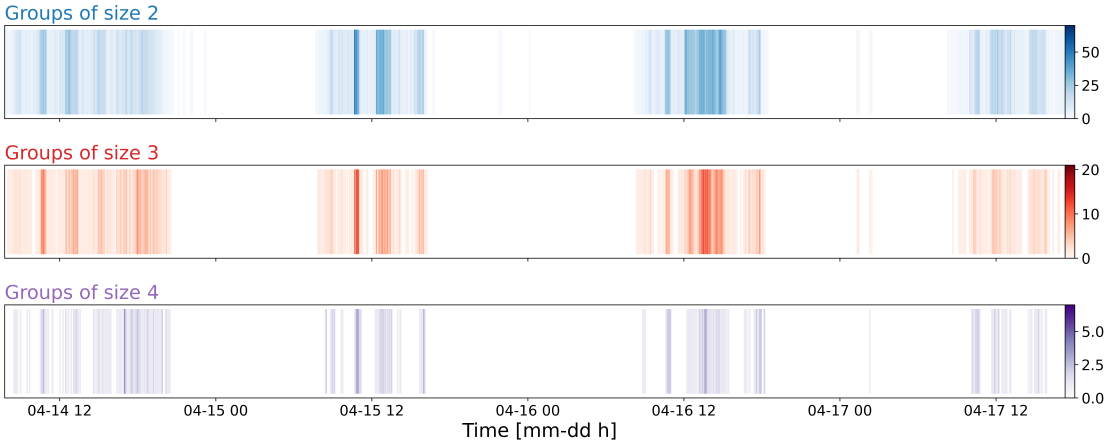

Figure 2.6: Timely occurrences of interactions of sizes 2, 3 and 4 in the `ECIR19` dataset. It is clearly possible to identify the four different days of the conference. The other datasets display similar behavior.

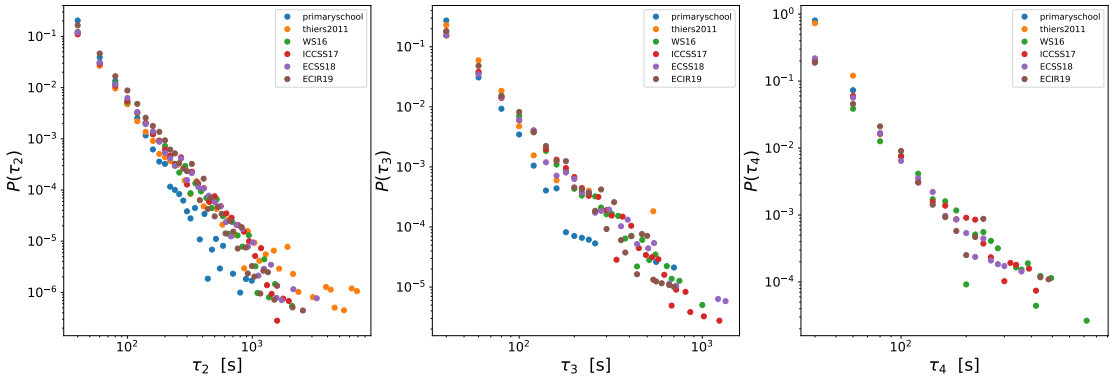

Figure 2.7: Distributions of the durations of groups of size 2 (left), 3 (center) and 4 (right).

across the datasets than the shapes of distributions $P(\tau)$ and $P(\Delta\tau)$ displayed in Figure 2.3. While we are unable to say whether this is due to the fact that contact interactions are caused by underlying mechanisms that are common in human gathering across a great variety of contexts and group interactions are not we put forward the hypothesis that this effect might also be related to finite size effects. The way in which we define the higher-order structures from the pairwise dyadic contact events poses some combinatorial limits to the number of these structures that can be observed in limited datasets. In Table 2.2 we present for all six of the datasets under study the total number of instantaneous contacts recorded, C, and three other quantities C_2, C_3 and C_4 where C_k, with $k = 2,3,4$, is

the total number of instantaneous groups of size k recorded. We speculate that some of the heterogeneity observed in the distributions of group durations for the different datasets might be ascribed to the limited number of such groups that are observed empirically. This is particularly evident for the groups of size 4 in the `primaryschool` and `thiers2011` datasets that also display different shapes of the distribution in Figure 2.7.

	`primaryschool`	`thiers2011`	`WS16`	`ICCSS17`	`ECSS19`	`ECIR19`
C	125773	28540	153371	229536	96362	132949
C_2	96954	25141	63414	142270	54908	84733
C_3	9257	983	21575	25194	11092	13625
C_4	471	75	5723	3299	1934	1782

Table 2.2: Counts of the total number of instantaneous groups event recorded for the different datasets. C_k, with $k = 2,3,4$, is the total number of instantaneous groups of size k recorded.

The shape of these distributions, compatible with power-law-like functions, makes it evident that there is no characteristic time scale for the duration of groups. This fact might play a significant role in the dynamic of processes in which the presence and duration of group contact events play a significant role. Examples of such phenomena are the aforementioned social contagion and consensus emergence processes.

Comparing the $P(\tau)$ distributions for the groups of different sizes of the same dataset (see Figure 2.8 for the `ECIR19` dataset, the data recorded in the other venues display the same behavior) we observe that there is no clear difference between the slopes of the power-law describing the distribution. The only apparent feature is that the cut-off time (*i.e.* the duration over which we do not empirically observe groups with that duration) is larger for groups of size 2 ($\sim 2 \times 10^3$ s) and smaller for groups of size 3 ($\sim 9 \times 10^2$ s) and 4 ($\sim 6 \times 10^2$ s).

Zhao *et. al.*, 2011 [56] observed that like pairwise contacts also higher-order social interactions have a power-law duration with an exponent that is steeper the larger the dimension of the interacting group. In the original paper, they showed that this phenomenology is captured by the simple model of higher-order temporal social network proposed by the same authors in an earlier contribution [65]. They argued that the higher volatility of social interactions involving larger groups is explained by the fact that in larger groups of people, the size of the group changes if any one of its members leaves the group. Therefore assembles that involve more individuals have a steeper distribution of contact duration than gatherings involving fewer people. Although this explanation is sound we were not able to reproduce the empirical result of groups of increasing size displaying increasing power-law

exponents. The difference between our results and those of the above-mentioned paper might be related to slight differences in the way group interactions are defined from the recorded pairwise contacts or to the preprocessing of the data (*e.g.* some kind of triadic closure at the contact level before extracting the cliques). Nevertheless, we speculate that the interpretation of large groups being more fragile than small groups is present also in our results, explaining the decreasing cut-off durations in the distributions.

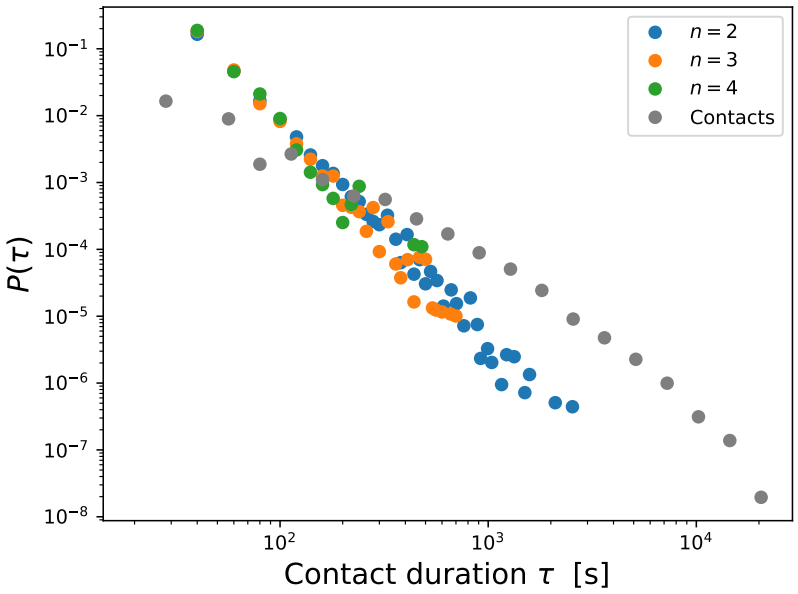

Figure 2.8: Distributions of the durations of groups of size 2, 3 and 4 for the `ECIR19` dataset. We also display (in gray) the distribution of contact durations. The different datasets display the same behavior.

Chapter 3

Higher-order homophily in human gatherings

Homophily is the sociological principle that describes the tendency of individuals to form social connections and interact with other individuals who share similar characteristics or attributes [43, 44]. The concept of homophily has been extensively studied in sociology and network analysis as it is considered to be an important factor that shapes human relationships and social connections.

As discussed in §1.3, gender homophily has been observed in primary school contacts, where students tend to form friendships with others of the same gender [45]. However, homophily extends beyond just gender and has been shown to be correlated with other demographic factors such as age, ethnicity, and socioeconomic status. Furthermore, acquired characteristics such as education level, political affiliation, and religion have also been found to be important determinants of homophily.

Overall, homophily is a key guiding principle in sociological research and network analysis, as it helps to explain the patterns of social connections and interactions among individuals. By understanding the mechanisms underlying homophily, researchers can gain insight into how social networks form and evolve, and how social inequality and segregation can arise as a result [43].

Despite the evidence that group interactions are ubiquitous in social settings and play a significant role in dynamical processes on networks such as contagion or opinion formation (see §1.2) the more commonly used measures of homophily rely only on dyadic representations. In order to measure group homophily these approaches reduce group participation to pairwise relationships based on co-participation in groups.

Following what has been proposed by Veldt *et al.*, 2023 [66] we consider a way to measure explicitly higher-order homophily, *i.e.* homophily defined at the hypergraph level. Using this framework we show that in our datasets we find that

it is possible to observe higher-order homophily in the absence of the corresponding low-order homophily.

3.1 Higher-order affinities

To introduce in all generality our measure of higher-order homophily, we consider a hypergraph $\mathcal{H} = (V, E)$, where the set of nodes V represents the agents of a population and the hyperedges in E represents the group interactions among members of the population. We want to quantify to which extent the nodes belonging to a class $X \subseteq V$ tend to interact among themselves in group settings. The class $X \subseteq V$ represents all the nodes sharing a certain feature of interest, for example age or gender.

The standard pairwise affinity measure is based on the relative importance of in-class connections over all the connections displayed by the members of the class. For example, the graph homophily index is defined by Altenburger & Ugander, 2018 [44] as:

$$\hat{h}_X = \frac{\sum_{i \in X} d_{i,\text{in}}}{\sum_{i \in X} d_{i,\text{in}} + \sum_{i \in X} d_{i,\text{out}}} \tag{3.1}$$

where $d_{i,\text{in}}$ denotes the observed in-class degree and $d_{i,\text{out}}$ denotes the out-class degree. Veldt *et al.*, 2023 [66] have generalized this concept to group interactions of size k (*i.e.* k-hyperedges) by defining for each positive integer $t \in 1, ..., k$ a so-called type-t affinity score. This affinity score expresses the extent to which individuals in class X participate in groups of size k where exactly t members of the group are in the class. The type-t affinity score is defined by the formula:

$$\mathbf{h}_t^k(X) = \frac{D_t^k(X)}{D^k(X)} = \frac{\sum_{v \in X} d_t^k(v)}{\sum_{v \in X} d^k(v)} \tag{3.2}$$

where $d^k(v)$ is the total number of k-hyperedges to which node v participates and $d_t^k(v)$ is the number of k-hyperedges with exactly t nodes belonging to class X to which node v participates. It is straightforward to see that for $k = t = 2$ the type-t affinity index in Eq. 3.2 reduces to the graph homophily index in Eq. 3.1.

3.1.1 Null model for group affinities

In order to quantify the relevance of a measurement obtained using the type-t affinity score (Eq. 3.2) we need to compare it with a baseline score representing a null probability for type-t interactions in groups of size k. If the observed $\mathbf{h}_t^k(X)$ is larger than the corresponding baseline we will have that type-t group interaction are overexpressed for class X.

At the graph level, several null models have been proposed, thus several generalizations to higher-order structures are possible. Following again Veldt *et al.*, 2023 [66], we define the simplest possible null model that takes into account only the fraction of nodes belonging to the class $\alpha = |X|/|V|$. In this uniform null model, the baseline type-t affinity score is given by the probability that a node of class X participates in a group of size k where exactly t members are from X if the other $k-1$ nodes in the group are selected uniformly at random:

$$\mathbf{b}_t^k = \frac{\binom{|X|-1}{t-1}\binom{n-|X|}{k-t}}{\binom{n-1}{k-1}} \tag{3.3}$$

where n is the total number of agents in the population under study. The baseline score in the above equation is the type-t affinity score for a class $X \subseteq V$ in a complete k-uniform hypergraph $\mathcal{H}^*_{k,n}$ of n nodes - *i.e.* a hypergraph with n nodes where all the possible k-hyperedges are present.

3.2 Higher-order homophily in scientific conferences

The `WS16`, `ICCSS17`, `ECSS18` and `ECIR19` are accompanied by rich metadata about participants[1]. Upon registration at the conference venue the participants were informed about the planned SocioPatterns data recording session and asked whether they were willing to participate, participants were also asked to fill out a survey. In such survey participants were asked for socio-demographical pieces of information, such as their age, gender, country of origin, academic status *etc.*, questions about their role in the conference, their perception of the crowd and questions useful to reconstruct some of their psychological traits.

Thanks to the socio-demographic metadata accompanying the contact datasets it is possible to explore higher-order homophily using the formalism presented in §3.1. The procedure we follow is very simple. Considering a single contact dataset we filter out short interactions by constructing from the temporal hypergraph described in §2.3 an aggregated hypergraph containing all the groups with a total duration larger than a certain cut-off. On the aggregated hypergraph it is then possible to perform the analysis described in §3.1 considering classes of participants looking at different entries of the metadata. The choice of the cut-off threshold is arbitrary but the results we present below have shown to be relatively robust with respect to variations of this choice.

[1]The metadata and the questionnaires are described in detail in the original paper Ref. [49] and are available upon request at `https://search.gesis.org/research_data/SDN-10.7802-2352`.

We are also interested in exploring the relation between low-order homophily, *i.e.* homophily at the contact level, and higher-order homophily, *i.e.* homophily at the level of groups. To this goal, we also apply Eq. 3.1 to quantify the graph homophily in the aggregated network of contacts. Again we filter out short contacts by setting a cut-off duration and keeping in the aggregated network only contacts with a duration larger than the cut-off.

Our results show that higher-order homophily can be observed also in the absence of the corresponding homophily. Nevertheless, comparing the four different datasets we recognize that the results regarding higher-order homophily are not at all robust across different venues and across different sizes. This shows that higher-order homphily is very context-dependent and cannot be considered a general feature of face-to-face interactions in human gatherings.

In the following two subsections, we present our results for higher-order gender and age homophily. The results for other socio-demographical features are presented in Appendix A.

3.2.1 Higher-order gender homophily

In Table 3.1 we display the results for the graph gender homophily in the `ECIR19` for different values of the cut-off for the duration. The other datasets display similar behavior.

	20 s	**60 s**	**120 s**	**180 s**	**300 s**
Male	1.01	1.02	1.02	1.00	0.98
Female	0.97	0.91	0.91	0.91	0.91

Table 3.1: Graph gender homophily indices (see Eq. 3.1) for the `ECIR19` dataset, at different cut-offs of the duration of the edges. The other datasets display similar behavior.

We see that there is no homophily effect at the level of contacts as the graph homophily indices are very close to the value 1 corresponding to the absence of homophily. The only noticeable feature is a little heterophily effect for women at larger values of the cut-off. This result confirms a known fact in sociology for which while there is an evident gender homophily effect in kids [67, 45], this effect disappears in adults [68].

In Figure 3.1 we show the results for the higher-order gender homophily in the four scientific conferences. Each plot refers to a group of a specific size for a given dataset. For each size k we can compute the type-t affinity scores for $t = 1, ..., k$ that quantify the extent to which individuals in a given class X participate in groups of size k where exactly t members of the group are in the class. The shaded

areas are the 95% confidence intervals obtained with a bootstrap procedure [69].

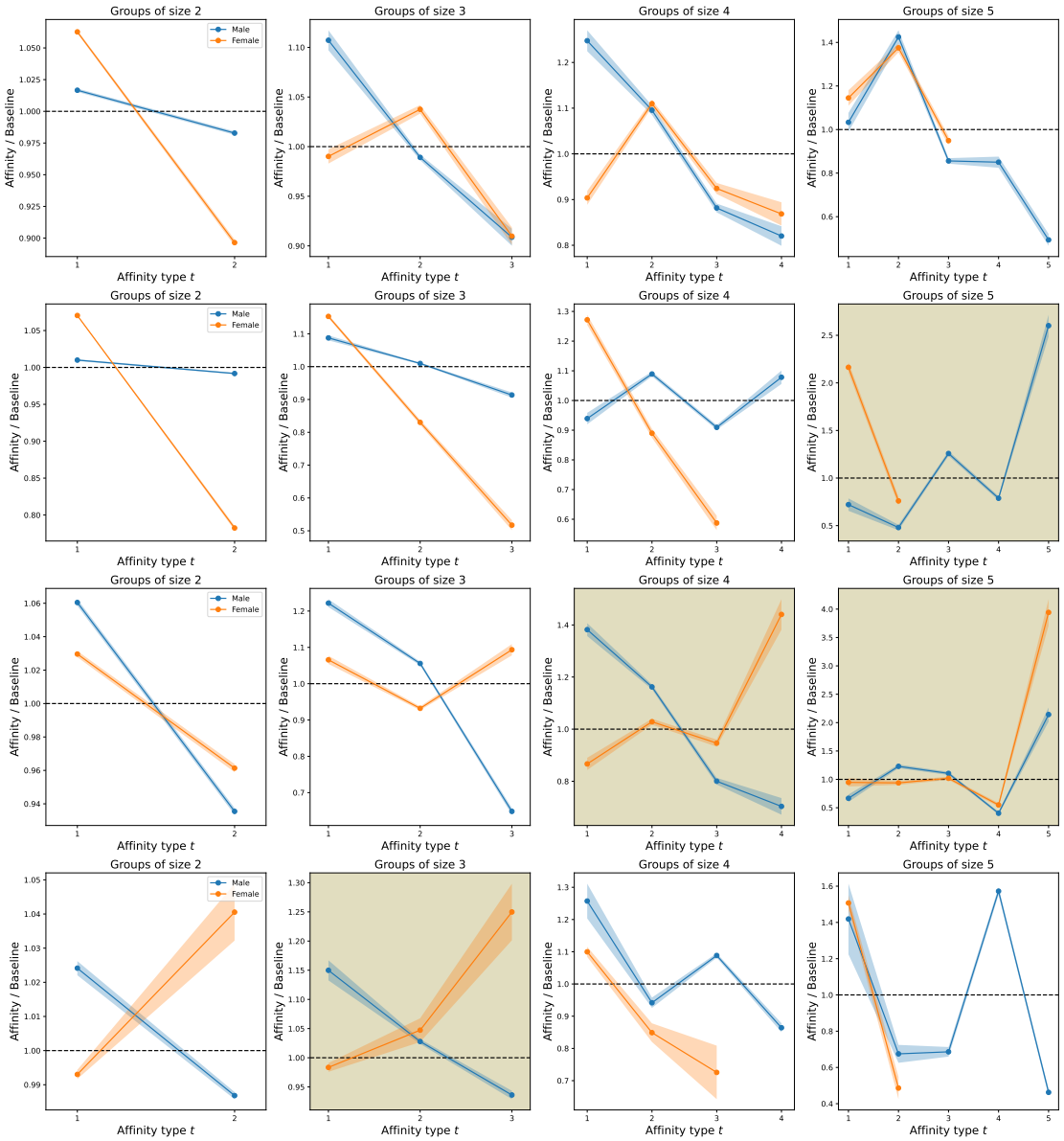

Figure 3.1: Ratio between the type-t affinity score and the baseline to quantify higher-order gender homophily in the `WS16`, `ICCSS17`, `ECSS18` and `ECIR19` (from top to bottom) datasets. The aggregated hypergraph over which we compute the affinity scores is obtained aggregating the temporal hypergraphs keeping only the groups with an aggregated duration ≥ 60 s, similar results are obtained varying this cut-off threshold. The highlighted plots are those displaying higher-order homophily.

We see that the different datasets and different group sizes show different behaviors. We have highlighted the plots displaying a significant higher-order homophily effect, where for a given class in groups of size k the ratio between the type-t affinity and the baseline at $t = k$ is above 1 and larger than the same quantity for values of $t < k$. For example in the second plot from the left in the bottom row of Figure 3.1 we display the ratio between the type-t affinity score and the baseline for the `ECIR19` dataset, there we see that women display a higher-order homophily effect for groups of size 3. This means that, in that specific dataset, women despite showing a little heterophily effect at the level of contact (see Table 3.1), in groups of size 3 they tend to participate more than expected in group interactions of size 3 where the majority of the participant are women as well and this propensity increases as the group in completely formed by women.

3.2.2 Higher-order age homophily

In table 3.2 we display the results for the graph age homophily in the `ECIR19` for different values of the cut-off for the duration. The other datasets display similar behavior. In Figure 3.2 we show the results for the higher-order gender homophily in the four scientific conferences.

	20 s	**60 s**	**120 s**	**180 s**	**300 s**
< 30 y.o.	0.99	0.99	1.01	1.01	1.05
30-39 y.o.	0.96	0.96	0.96	0.97	0.96
> 40 y.o.	0.97	0.83	0.74	0.67	0.65

Table 3.2: Graph age homophily indices (see Eq. 3.1) for the `WS16` dataset, at different cut-offs of the duration of the edges. The other datasets display similar behavior.

The considerations to be made on these results are analogous to the ones concerning gender homophily. Apart from the effect of heterophily in the $30 - 39$ y.o. and > 40 y.o. groups there are no relevant effects at the contact level. Despite this absence of low-order homophily effects for some sizes and some datasets (highlighted in the figure), we can observe some effects of higher-order homophily.

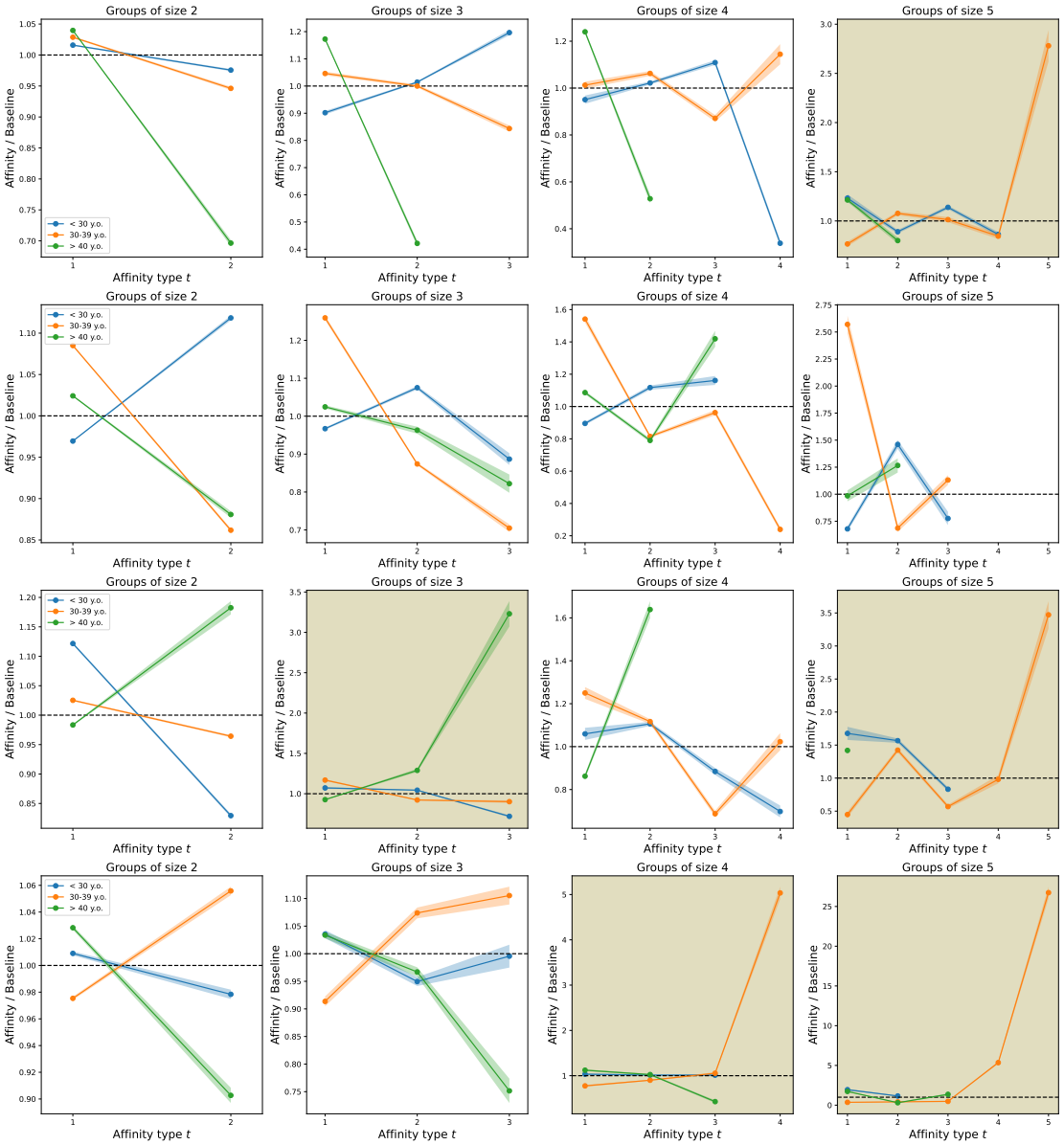

Figure 3.2: Ratio between the type-t affinity score and the baseline to quantify higher-order age homophily in the `WS16`, `ICCSS17`, `ECSS18` and `ECIR19` (from top to bottom) datasets. The aggregated hypergraph over which we compute the affinity scores is obtained aggregating the temporal hypergraphs keeping only the groups with an aggregated duration ≥ 60 s, similar results are obtained varying this cut-off threshold. The highlighted plots are those displaying higher-order homophily.

Chapter 4

Attractiveness model for face-to-face interactions

The complex dynamic of face-to-face interactions has been shown to have a profound impact on the properties of the temporal evolving network defined by the pairwise contacts and has been proven to affect also the behavior of dynamical processes taking place on top of those networks. Understanding the origin and development of such characteristics and effects requires the development of theoretical models able to reproduce them.

4.1 Random walkers biased by attractiveness

In this work, we present and follow a simple model of mobile agents proposed by Starnini *et al.* [54]. This model is based on two aspects influencing the dynamic of human interactions. The first aspect takes into consideration the fact that - as observed in several cases - different individuals have different degrees of social appeal or *attractiveness*. This attractiveness might be related to the personality traits of the individuals, their social status or the role they play in the social gathering under study. The effect of this social appeal will be to bias in some way the interactions in which the particular individual takes part. The second aspect of human interactions taken into account in the model is the fact - evident from the recordings performed with the SocioPatterns platform (see §2.1) - that not all participants in a human gathering are always present at the same time in the venue in which the gathering takes place. The propensity of each individual to be present or absent in the social event might be related to complex socio-demographic or psychological characteristics distinguishing each person and influencing the overall interaction dynamic in the gathering.

The model (see Figure 4.1) is defined as follows. N different agents are placed

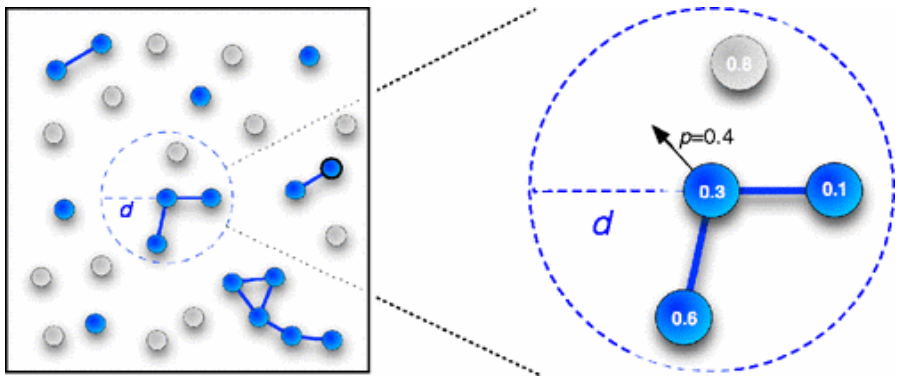

Figure 4.1: Drawing of the dynamics of the attractiveness model. Blue-colored agents are active, and gray-colored agents are inactive meaning that they do not move nor interact. Interacting agents, within a distance d are connected by a link. Each agent is characterized by its attractiveness. The probability for the central agent to move is $p = 1 - 0.6 = 0.4$ as the inactive agent with an attractiveness of 0.8 is not taken into account. Figure from Starnini *et al.*, 2013 [54].

in a square box of side L with periodic boundary conditions. Each agent i is characterized by its attractiveness $a_i \in [0,1]$ which is a random variable extracted from a distribution $\eta(a)$. Whenever two agents are at a distance smaller than d they interact and the interaction lasts as long as the mutual distance is smaller than d, the distance d is a parameter of the model. The free motion of each agent (*i.e.* its motion when they are not interacting with any other agent) is modeled as a random walk where at each time-step each agent picks a direction $\xi \in [0,2\pi)$ at random and moves in that direction with a step of length v. The effect of the interactions is that the random walk of each agent is biased by the presence of the most attractive individual present in the neighborhood of radius d surrounding their position - *i.e.* the most attractive individual with whom they are interacting; this is taken into account introducing a moving probability given by:

$$p_i(t) = 1 - \max_{j \in \mathcal{N}_i(t)} \left\{ a_j \right\} \tag{4.1}$$

With the complementary probability $1-p_i$ the agent does not move. To consider the individuals' inclinations to be active in the social gathering we further characterize each agent with a second random variable - representing their *activeness* - $r_i \in [0,1]$ sampled from a distribution $\zeta(r)$. At each time step an inactive agent i becomes active with a probability r_i, and an active agent j, if isolated (*i.e.* not in contact with any agent), becomes inactive with a probability $1 - r_j$. Only active agents perform the random walk and take part in the interaction dynamic.

In the model described above each one of the N agents performs a random walk in a two-dimensional space, this random walk is interrupted by the possible

interactions with the other agents and eventually by the de-activation of the agent. In order to mimic as accurately as possible the recording procedure of the SocioPatterns measurements (see §2.1) in the simulation of the attractiveness model the movement of the agents is performed in parallel, meaning that at each time-step each one of the active agents performs the routine depicted in Figure 4.2.

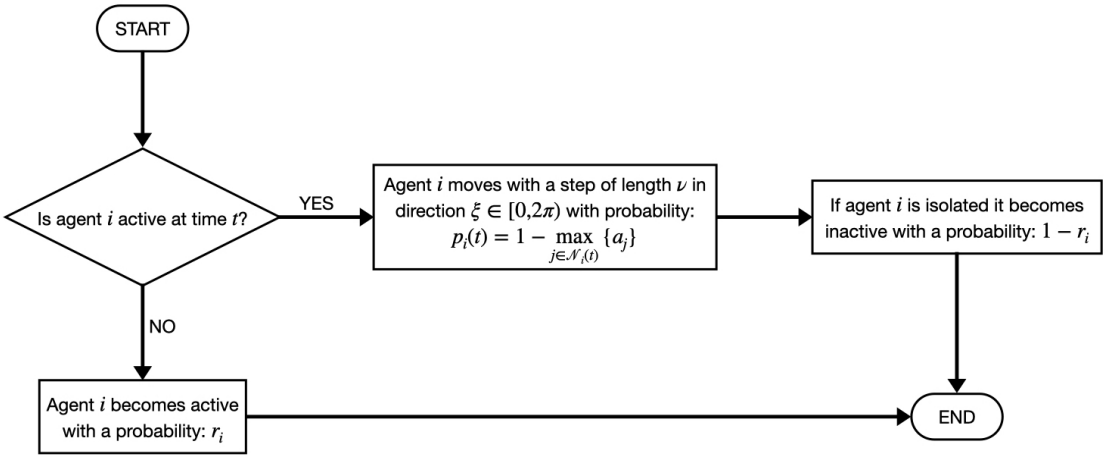

Figure 4.2: Diagram depicting the routine performed by each agent in the attractiveness model at each time step.

It is straightforward to recognize that the model is Markovian: at each time step the agents do not have any memory of their previous movements and interactions. The dynamic of the model is completely encoded by the collision probability, given by:

$$p_c = \rho \pi d^2 \tag{4.2}$$

where $\rho = N/L^2$, the attractiveness distribution $\eta(a)$ and the activeness distribution $\zeta(r)$.

The appeal that a person might have to other people when participating in a human gathering - represented in our model by the attractiveness parameter a_i - is a relational variable that might depend on the unknown combination of different psychological, socio-demographic and environmental factors that vary among different situations and different persons. The same can be said about the propensity of each person to be active at different times in the gathering - modeled with the activeness parameter r_i. Moreover, those quantities and the factors that shape them are hardly accessible empirically and cannot be measured. For these reasons, in order to avoid any speculation on these factors and mechanisms, we assume the attractiveness distribution $\eta(a)$ and the activeness distribution $\zeta(r)$ to both be uniform distributions over the interval [0,1].

In the following sections we will present the results of the numerical simulations of the model contrasted with the results from the empirical datasets (see §2.2).

4.2 Contact statistics of the attractiveness model

The temporal properties of contacts between the agents are the most distinctive feature of empirical face-to-face interactions. In §2.2.2 we have shown that the distribution of the duration τ of contacts and that of the duration of intervals between consecutive contacts $\Delta\tau$ uncover the bursty dynamics of human interaction. These distributions have fat-tailed forms that can be described in terms of power-law functions. Moreover, it is striking how the shape of these distributions is very robust across different datasets, recorded in different conditions and in different social contexts. This robustness across different datasets hints in the direction of some kind of universal nature of the mechanisms governing the dynamic of face-to-face interactions.

Another interesting feature of a temporal network describing face-to-face interactions in human gathering resides in the distribution of links' weights in the aggregated network. The data measured with the SocioPatterns platform (see §2.1) is naturally described by a temporal network where links describe the interactions between the people participating in the gathering. Those interactions evolve in time and thus at different time steps, the corresponding links are present in the temporal network accordingly. Integrating the information contained in the instantaneous networks over a given time window (in our case this is always gonna be the whole duration of the recording session), we obtain an aggregated weighted network. In this aggregated network the weight w_{ij} of the link between nodes i and j is given by the total duration of the contacts between agents i and j.

If Figure 4.3 we show the results obtained with the attractiveness model for the distributions of contact durations, intercontact durations and weights in the aggregated network. The model has been simulated by choosing the parameters $v = d = 1$, $L = 100$ and $N = 200$. At the beginning of the simulation, the agents are placed at randomly chosen positions and are active with probability $1/2$.

The numerical and experimental results match. This is a striking result, a model based on two simple assumptions on the mechanisms regulating human face-to-face interaction is able to reproduce with great accuracy the statistical properties - at least at the level of contacts - of complex social phenomena.

The distribution of links' weights in the aggregated network is - as observed before for the other two distributions considered - broad. This result demonstrates that the heterogeneity in the duration of individual contacts, which was initially observed through analyzing the distribution of contact durations, persists when the duration of a single contact is aggregated over a more extended period of time.

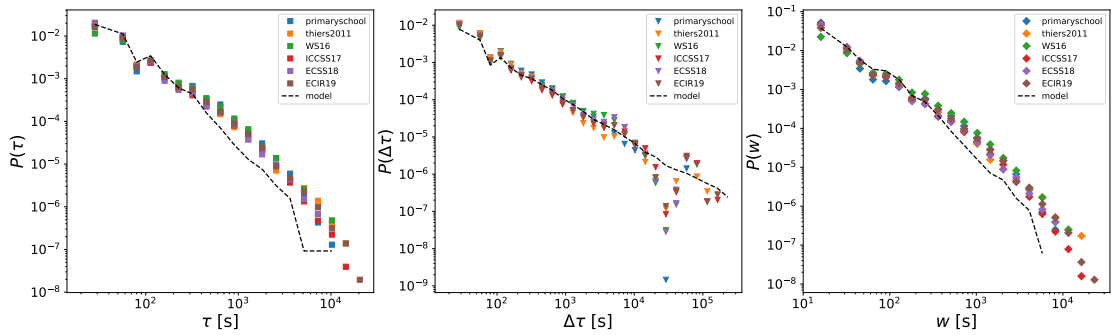

Figure 4.3: Distribution of the contact durations (left), intercontact durations (right) and weights (right) for the datasets and the attractiveness model. The numerical results are obtained with a single simulation with $v = d = 1$ and $L = 100$ and of duration $T = 2 \times 10^4$ time-steps..

4.2.1 Role of the density of agents

The numerical results obtained for the attractiveness model are rather robust with respect to variations in the collision probability p_c defined by Eq. 4.2. In Figure 4.4 we display the numerical results of the attractiveness model for different values of the density of agents $\rho = N/L^2$ obtained by varying the number of agents.

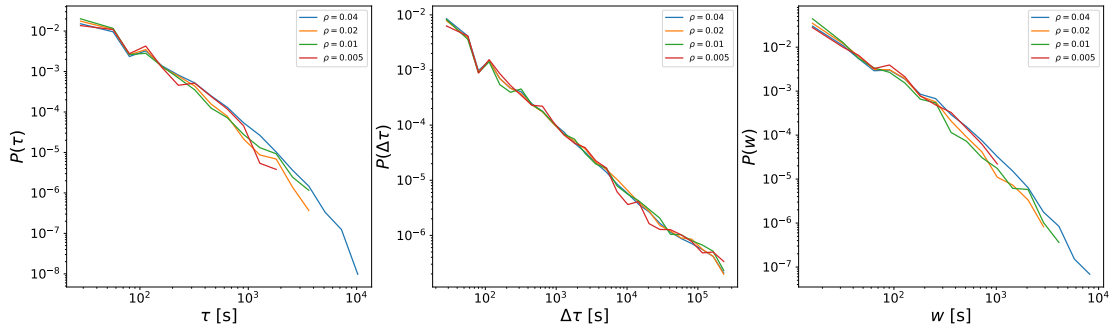

Figure 4.4: Distribution of the contact durations (left), intercontact durations (right) and weights (right) for the attractiveness model for various densities of agents. The numerical results are obtained with a single simulation with $v = d = 1$ and $L = 100$ and of duration $T = 2 \times 10^4$ time-steps.

The fact that the numerical results are roughly independent of the collision probability p_c suggests that the relevant mechanisms in the model are the other two ingredients governing the dynamic of the model, namely the attractiveness distribution $\eta(a)$ and the activeness distribution $\zeta(r)$.

4.3 Failure of the attractiveness model in reproducing group duration statistics

Performing the same analysis done in §2.3 on the results of the simulation of the attractiveness model we obtain the results displayed in Figure 4.5.

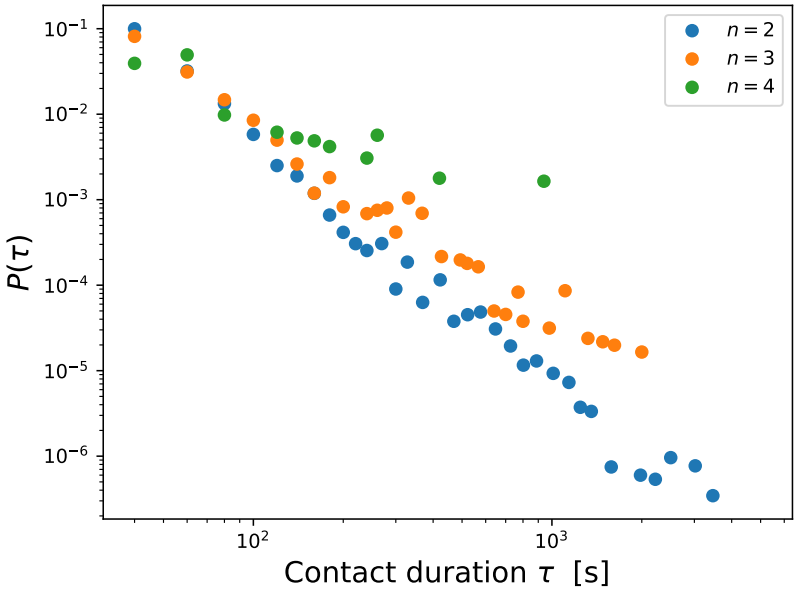

Figure 4.5: Distributions of the durations of groups of size 2, 3 and 4 for the attractiveness model. The numerical results are obtained with a single simulation with $v = d = 1$ and $L = 100$ and of duration $T = 2 \times 10^4$ time-steps.

It is evident that these results are not reproducing what has been observed in the empirical data. We see that the shapes of the three probability distributions are broad and well-described by power-law-like functions. As seen in Figure 4.5 it appears that if we were to write the probability distributions in the form:

$$P(\tau) \propto \tau^{-\alpha} \tag{4.3}$$

the exponent α would be decreasing with the size of the groups. This is the opposite of what is observed in the data and in the model proposed by Zhao *et al.*, 2011 [56], where large groups are more fragile than small groups as in larger groups of people, the size of the group changes if any one of its members leaves the group.

The inadequacy of the attractiveness model in reproducing the statistics of the temporal durations of groups observed in the data hints in the direction that even

if empirically we build the groups from pairwise interactions the group structures in the face-to-face interactions are related to higher-order mechanisms not reducible to the composition of pairwise effects. This is because the rules in the attractiveness model are based only on pairwise interactions - as the moving probability is controlled only by the attractiveness of the most attractive neighbor. We have seen that this pairwise-based model reproduces correctly the low-order temporal features of the datasets - namely the statistics of contact and intercontact durations and the weight distribution - but not the statistics of the durations of higher-order structures. This suggests that the higher-order structures of face-to-face interaction detected with the procedure proposed by Cencetti *et al.*, 2021 [57] that we have followed in §2.3 are not reducible to the composition of low-order features.

4.4 Modified attractiveness model

The fact that the distribution of durations shows smaller exponents for larger groups (see Figure 4.5) due to the phenomenon - observed in the numerical simulation - that in the attractiveness model there is a tendency to form clusters of agents in the vicinity of highly attractive agents. Intuitively this phenomenon is not natural for group interactions in the context of face-to-face interactions in human gatherings. Especially in the context of scientific conferences, we could argue that in informal settings - such as coffee breaks or poster sessions - there is an interplay between the benefit of interacting with highly attractive agents and the drawbacks of interacting in a group that is too large.

In order to account for this interplay between group size and attractiveness and reproduce the statistic of group duration we modify the moving probability given by Eq. 4.1 and propose a modified version of the attractiveness model. In our modified version the moving probability is given by:

$$p_i(t) = 1 - \max_{j \in \mathcal{N}_i(t)} \frac{a_j}{|\mathcal{N}_j(t)|^\gamma} \tag{4.4}$$

We have introduced a new parameter γ that controls how much large groups are penalized in the modified version of the attractiveness model. For $\gamma = 0$ we recover the probability of moving of the original attractiveness model.

If Figure 4.6 we show the results for the distributions of the contact durations, intercontact durations and weights for the modified attractiveness model and the datasets.

We can see that the numerical results of the model are in good agreement with the empirical data. The fact that minor modifications in the rule controlling the probability of moving do not compromise the results for the statistics of contacts is in agreement with what is presented by Starnini *et al.*, 2013 [54] in the original paper presenting the attractiveness model. The authors argued that the key feature

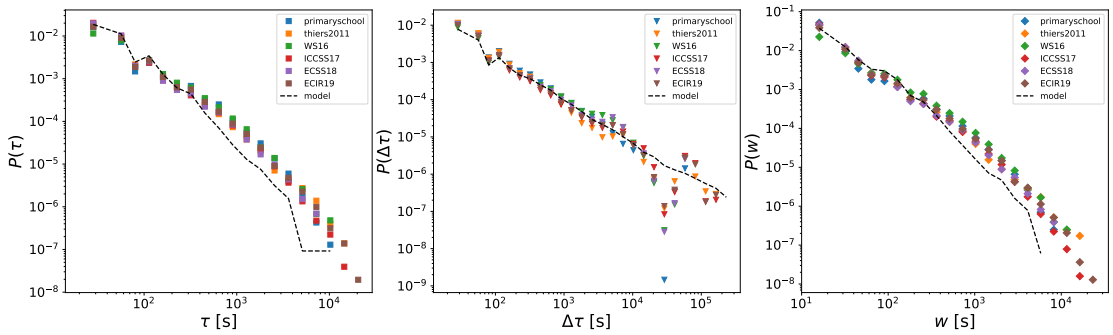

Figure 4.6: Distribution of the contact durations (left), intercontact durations (right) and weights (right) for the datasets and the modified attractiveness model. The numerical results are obtained with a single simulation with $v = d = 1$, $\gamma = 0.1$, $L = 100$ and $N = 400$ of duration $T = 4 \times 10^4$ time-steps.

of the model is the heterogeneity in the distribution of attractiveness, and minor modification in the moving probability such as considering the average of the attractiveness of the neighbors lead substantially to the same behavior produced by Eq. 4.1.

In Figure 4.7 we show the results for the analysis of the statistics of group durations obtained with the modified attractiveness model. Comparing the $P(\tau)$ distributions for the groups of different sizes we see that there is no clear difference between the slopes of the power-law describing the distribution. The only feature is that the cut-off time (*i.e.* the duration over which we do not empirically observe groups with that duration) is larger for groups of size 2 and smaller for groups of size 3 and 4. This is the same behavior that we observed in the empirical dataset in §2.3.

In Figure 4.8 we show the empirical distributions of group durations and the numerical results of the modified attractiveness model. There is a good agreement between the model and the datasets. This result suggests that the modified rule for the probability of moving reproduces more realistically the formation and disgregation of groups in human interactions such as scientific conferences or schools.

4.5 Homophily in the attractiveness model

The original attractiveness model [54] shows limitations in reproducing some complex features of face-to-face interaction networks that are present in the empirical data. We have pointed out how - in its original formulation - it fails to reproduce the distribution of durations of group interactions.

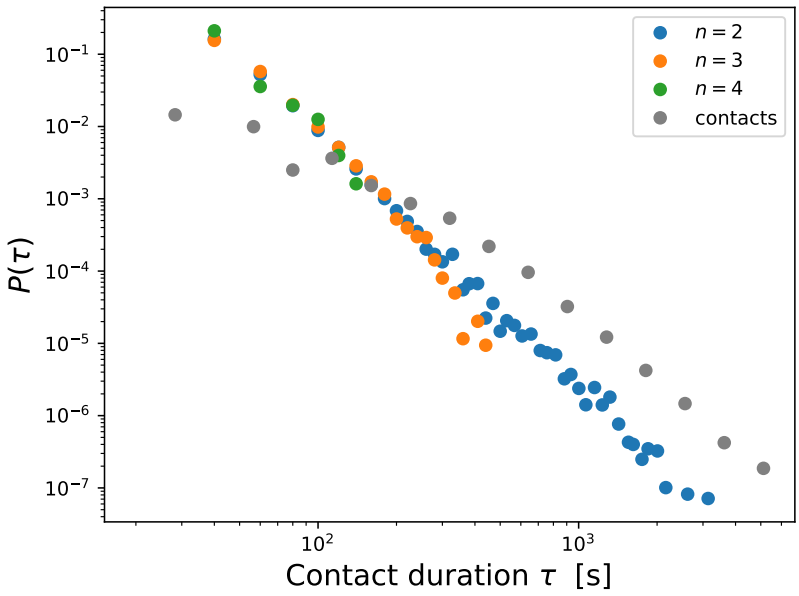

Figure 4.7: Distributions of the durations of groups of size 2, 3 and 4 for the modified attractiveness model. The numerical results are obtained with a single simulation with $v = d = 1$, $\gamma = 0.1$, $L = 100$ and $N = 400$ of duration $T = 4 \times 10^4$ time-steps

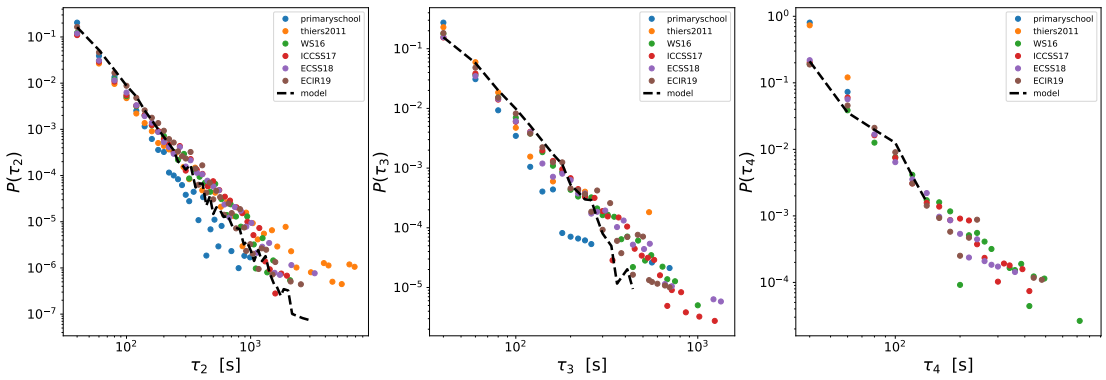

Figure 4.8: Distributions of the durations of groups of size 2 (left), 3 (center) and 4 (right) for the modified attractiveness model for various densities of agents. The numerical results are obtained with a single simulation with $v = d = 1$, $\gamma = 0.1$, $L = 100$ and $N = 400$ of duration $T = 4 \times 10^4$ time-steps.

Oliveira *et al.*, 2022 [58] have explored how the attractiveness model can be modified in order to reproduce the degree inequalities that are sometimes present in the aggregated network of SocioPatterns measurements. What is observed in

the empirical data is that different classes of agents (*e.g.* men and women) if represented in different numbers in the gathering - hence allowing for the definition of a minority in the social gathering - show different values of average degree in the aggregated network. Though this degree inequality is present in the empirical data, the attractiveness model fails to explain the group differences because it ignores group mixing in social gatherings. In the original paper, the authors define group mixing as the systematic preference of group members to interact with individuals from specific social groups. This definition includes both homophily, as we have defined it §1.3, and heterophily *i.e.* the tendency of a specific class of interacting more with members of other classes.

The paper's authors argued that the attractiveness paradigm fails to generate group mixing due to its absence of relational attributes. This means that differences in individual attractiveness alone cannot account for the observed degree inequality and group mixing in the data.

To replicate the degree inequalities found in the data, the authors suggested an altered version of the attractiveness model. This revised model, referred to as the attractiveness-mixing model, incorporates both the intrinsic attractiveness of individuals and the relational attributes between groups. Additionally to the other features already present in the attractiveness model, in the modified model an individual i is also characterized by a group label $b_i \in [0, B-1]$, where B is the number of groups. The mixing patterns in the system are encoded in a $B \times B$ mixing matrix $\mathbf{H}$. Each row of $\mathbf{H}$ can be seen as a probability mass function that weighs the likelihood of group interaction. The dynamic of the model is the same as in the original version described in §4.1 except for the fact that when agent i does not move it interacts with its neighbors of highest mixing likelihood with probability:

$$\beta_i(t) = \max_{j \in \mathcal{N}_i(t)} \{h_{b_i,b_j}\} \tag{4.5}$$

where h_{b_i,b_j} is an element of the matrix $\mathbf{H}$ that encodes the mixing probability between the group to which agents i and j belong.

In the original paper, the authors show the rich dynamic of the attractiveness-mixing model and argue that the two ingredients of individual attractiveness and relational attributes between groups are sufficient to explain the degree inequality observed in social dynamics with minority groups.

An alternative - and complementary - approach to what has been proposed by Oliveira *et al.*, 2022 [58] to have the original attractiveness model displaying some homophily effect is to explore the relationship between individual attractiveness and group membership. This would imply that individuals belonging to different social classes (that may be defined in different ways *e.g.* gender, age, country of origin, *etc.*) have different dynamics of interactions. A possible way of doing this is to divide the agents in the attractiveness model into two groups depending on

their attractiveness.

In Figure 4.9 we show the results for the higher-order homophily in the attractiveness model constructing a high-attractiveness class and a low-attractiveness one. The high-attractiveness (low-attractiveness, respectively) class is constituted by all the agents with attractiveness > 0.5 (< 0.5, respectively).

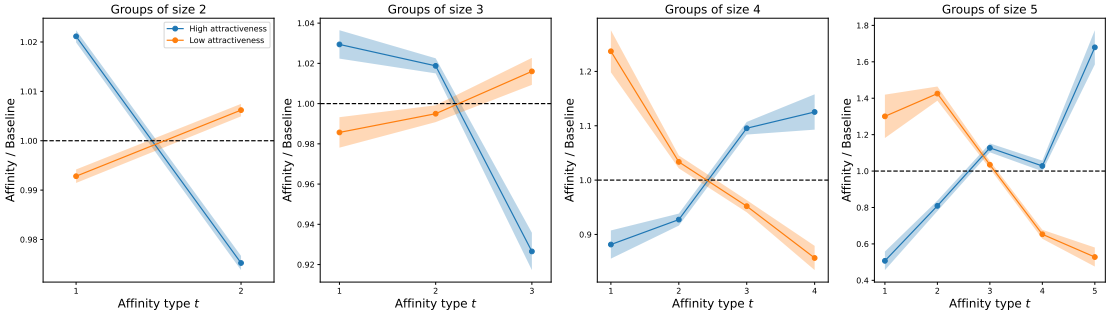

Figure 4.9: Ratios between the type-t affinity scores and the baselines to quantify higher-order attractiveness homophily. The numerical results are obtained with a single simulation with $v = d = 1$, $\gamma = 0.1$, $L = 100$ and $N = 400$ of duration $T = 4 \times 10^4$ time-steps. The high-attractiveness (low-attractiveness, respectively) class is constituted by all the agents with attractiveness > 0.5 (< 0.5, respectively).

Interestingly, our findings regarding higher-order attractiveness homophily align with the expectations set by the model. It is observed that agents with high attractiveness tend to attract a significant number of other agents, irrespective of their attractiveness levels, resulting in the formation of large groups. This phenomenon elucidates why, in small-sized groups (consisting of 2 or 3 individuals), the effect of homophily is predominantly exhibited by low-attractiveness agents, while high-attractiveness individuals do not demonstrate a pronounced homophily effect.

This observation can be explained by considering the inherent nature of high-attractiveness agents. It is highly improbable for a highly attractive individual to remain isolated in pairs or triplets, as they are more likely to gravitate towards forming larger groups. Conversely, it is more common for low-attractiveness agents to remain in small groups due to the relative absence of attractiveness-driven social interactions.

Although these preliminary results concerning higher-order homophily in terms of attractiveness are promising, further exploration with a refined and rigorous approach is warranted. For instance, it would be valuable to investigate how different attractiveness distributions or alternative criteria for constructing the classes of agents might influence the results. By considering such variations, we can delve deeper into the mechanisms underlying social interactions in human gatherings,

potentially uncovering additional insights into the dynamics of attractiveness-based homophily in human gatherings.

Chapter 5

Conclusions

This thesis has delved into various aspects of face-to-face interactions in human gatherings, shedding light on important findings and uncovering novel insights.

By examining the temporal dynamics of social interactions in different empirical datasets, we have reaffirmed previously known results regarding the shape of empirical distributions of contacts and intercontact durations.

Our exploration was focused on higher-order structures the statistics of group interactions, which are often overlooked when relying on the dyadic-only representation of complex systems given by networks. The results regarding the distribution of group durations have unveiled the absence of a typical time scale for higher-order interactions of different sizes. The shape of these distributions is robust across different datasets recording face-to-face interactions in different social contexts and the cut-off in the durations decreases as the group size increases. The higher volatility of social interactions involving larger groups is explained by the fact that in larger groups of people, the size of the group changes if any one of its members leaves the group.

We have presented a method to quantify higher-order homophily by comparing the generalization of graph homophily to hypergraphs with a baseline model Our results have demonstrated that higher-order homophily can be achieved even in the absence of the corresponding low-order homophily. However, we have also discovered that higher-order homophily is not a universal feature of face-to-face interactions, as its presence depends heavily on contextual factors and on the size of the groups that are considered.

To better understand and replicate the observed empirical data, we have presented an existing simple stochastic model for face-to-face interactions in human gatherings. This model is based on the pivotal role played by each individual's attractiveness in the interaction dynamic. While this model successfully reproduces some of the low-order statistics of face-to-face interaction temporal networks, namely the contact durations, intercontact time duration and aggregated weight

distribution, it falls short in capturing the temporal dynamics of group interactions. In response, we have proposed a modified version of the model that not only preserves its efficacy at the contact level but also replicates the higher-order features observed in the empirical data. Additionally, we have also presented an existing modified version of the model that accounts for group mixing, further enhancing its ability to capture the complexity of face-to-face interactions and reproducing the degree inequalities observed in face-to-face gatherings with minorities. As a preliminary result, we have put forward a way to encode homophily in the original attractiveness model by linking attractiveness and class membership.

In conclusion, this comprehensive study contributes to our understanding of face-to-face interactions in human gatherings by investigating their temporal dynamics, higher-order structures, and homophily effects. The insights gained from this research provide a valuable foundation for future studies in this field, enabling us to unravel the intricacies of face-to-face social interactions and their implications in various contexts.

Bibliography

[1] Stefano Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. «Complex networks: Structure and dynamics». In: *Physics Reports* 424 (4-5 Feb. 2006), pp. 175–308. ISSN: 03701573. DOI: 10.1016/j.physrep.2005.10.009 (cit. on p. 1).

[2] J.L. Moreno. *Who shall survive? A new approach to the problem of human interrelations.* Washington: Nervous and Mental Disease Publishing Company, 1934 (cit. on pp. 1, 2, 9).

[3] Mark Granovetter. «The Strength of Weak Ties: A Network Theory Revisited». In: *Sociological Theory* 1 (1983), pp. 201–233. ISSN: 07352751. URL: http://www.jstor.org/stable/202051 (cit. on p. 1).

[4] Paul L. Erdos and Alfréd Rényi. «On random graphs. I.» In: *Publicationes Mathematicae Debrecen* (1959) (cit. on p. 1).

[5] Douglas B. West. *Introduction to Graph Theory*. 2nd ed. Prentice Hall, Sept. 2000. ISBN: 0130144002 (cit. on p. 1).

[6] Albert László Barabási. «The network takeover». In: *Nature Physics* 8 (1 Jan. 2012), pp. 14–16. ISSN: 17452473. DOI: 10.1038/nphys2188 (cit. on p. 1).

[7] Réka Albert and Albert-László Barabási. «Statistical mechanics of complex networks». In: *Reviews of Modern Physics* 74 (1 2002), pp. 47–93 (cit. on p. 1).

[8] Albert László Barabási, Natali Gulbahce, and Joseph Loscalzo. «Network medicine: A network-based approach to human disease». In: *Nature Reviews Genetics* 12 (1 Jan. 2011), pp. 56–68. ISSN: 14710056. DOI: 10.1038/nrg2918 (cit. on p. 1).

[9] Romualdo Pastor-Satorras and Alessandro Vespignani. «Epidemic spreading in scale-free networks». In: *Physical Review Letters* 86 (14 Apr. 2001), pp. 3200–3203. ISSN: 00319007. DOI: 10.1103/PhysRevLett.86.3200 (cit. on pp. 1, 2).

[10] Alessandro Vespignani. «Modelling dynamical processes in complex socio-technical systems». In: *Nature Physics* 8 (1 Jan. 2012), pp. 32–39. ISSN: 17452473. DOI: 10.1038/nphys2160 (cit. on pp. 1, 5).

[11] Paolo Crucitti, Vito Latora, and Massimo Marchiori. «Model for cascading failures in complex networks». In: *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 69 (4 2004), p. 4. ISSN: 1063651X. DOI: 10.1103/PhysRevE.69.045104 (cit. on pp. 1, 2).

[12] Franz Kaiser, Vito Latora, and Dirk Witthaut. «Network isolators inhibit failure spreading in complex networks». In: *Nature Communications* 12.1 (May 25, 2021), p. 3143. ISSN: 2041-1723. DOI: 10.1038/s41467-021-23292-9. URL: https://www.nature.com/articles/s41467-021-23292-9 (cit. on p. 1).

[13] Stefano Battiston, Michelangelo Puliga, Rahul Kaushik, Paolo Tasca, and Guido Caldarelli. «DebtRank: Too central to fail? Financial networks, the FED and systemic risk». In: *Scientific Reports* 2 (2012). ISSN: 20452322. DOI: 10.1038/srep00541 (cit. on p. 1).

[14] Marco Bardoscia, Paolo Barucca, Stefano Battiston, Fabio Caccioli, Giulio Cimini, Diego Garlaschelli, Fabio Saracco, Tiziano Squartini, and Guido Caldarelli. «The physics of financial networks». In: *Nature Reviews Physics* 3 (7 July 2021), pp. 490–507. ISSN: 25225820. DOI: 10.1038/s42254-021-00322-5 (cit. on p. 1).

[15] Linton Freeman. «The development of social network analysis». In: *A Study in the Sociology of Science* 1.687 (2004), pp. 159–167 (cit. on pp. 1, 9).

[16] Stephen P. Borgatti, Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. «Network Analysis in the Social Sciences». In: *Science* 323.5916 (2009), pp. 892–895. DOI: 10.1126/science.1165821. eprint: https://www.science.org/doi/pdf/10.1126/science.1165821. URL: https://www.science.org/doi/abs/10.1126/science.1165821 (cit. on p. 1).

[17] Ulrik Brandes. *Network analysis: methodological foundations*. Vol. 3418. Springer Science & Business Media, 2005 (cit. on p. 2).

[18] J. L. Moreno and H. H. Jennings. «Statistics of Social Configurations». In: *Sociometry* 1.3/4 (1938), pp. 342–374. ISSN: 00380431. URL: http://www.jstor.org/stable/2785588 (cit. on p. 2).

[19] Marián Boguñá, Romualdo Pastor-Satorras, and Alessandro Vespignani. «Absence of Epidemic Threshold in Scale-Free Networks with Degree Correlations». In: *Physical Review Letters* 90 (2 2003), p. 4. ISSN: 10797114. DOI: 10.1103/PhysRevLett.90.028701 (cit. on p. 2).

[20] Petter Holme and Jari Saramäki. «Temporal networks». In: *Physics reports* 519.3 (2012), pp. 97–125 (cit. on p. 2).

[21] Laetitia Gauvin, Mathieu Génois, Márton Karsai, Mikko Kivelä, Taro Takaguchi, Eugenio Valdano, and Christian L Vestergaard. «Randomized reference models for temporal networks». In: *SIAM Review* 64.4 (2022), pp. 763–830 (cit. on p. 2).

[22] Laetitia Gauvin, André Panisson, Ciro Cattuto, and Alain Barrat. «Activity clocks: spreading dynamics on temporal networks of human contact». In: *Scientific reports* 3.1 (2013), p. 3099 (cit. on p. 3).

[23] Taro Takaguchi, Naoki Masuda, and Petter Holme. «Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics». In: *PloS one* 8.7 (2013), e68629 (cit. on p. 3).

[24] Miguel Valencia, J Martinerie, Samuel Dupont, and M Chavez. «Dynamic small-world behavior in functional brain networks unveiled by an event-related networks approach». In: *Physical Review E* 77.5 (2008), p. 050905 (cit. on p. 3).

[25] William Hedley Thompson, Per Brantefors, and Peter Fransson. «From static to temporal network theory: Applications to functional brain connectivity». In: *Network Neuroscience* 1.2 (2017), pp. 69–99 (cit. on p. 3).

[26] Giovanna Miritello, Esteban Moro, and Rubén Lara. «Dynamical strength of social ties in information spreading». In: *Physical Review E* 83.4 (2011), p. 045102 (cit. on p. 3).

[27] Lauri Kovanen, Kimmo Kaski, János Kertész, and Jari Saramäki. «Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences». In: *Proceedings of the National Academy of Sciences* 110.45 (2013), pp. 18070–18075 (cit. on p. 3).

[28] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. «Networks beyond pairwise interactions: Structure and dynamics». In: *Physics Reports* 874 (Aug. 2020), pp. 1–92. ISSN: 03701573. DOI: 10.1016/j.physrep.2020.05.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0370157320302489 (cit. on p. 3).

[29] Anna Ritz, Allison N Tegge, Hyunju Kim, Christopher L Poirel, and TM Murali. «Signaling hypergraphs». In: *Trends in biotechnology* 32.7 (2014), pp. 356–362 (cit. on p. 3).

[30] Alice Patania, Giovanni Petri, and Francesco Vaccarino. «The shape of collaborations». In: *EPJ Data Science* 6 (2017), pp. 1–16 (cit. on p. 3).

[31] Shan Yu, Hongdian Yang, Hiroyuki Nakahara, Gustavo S Santos, Danko Nikolić, and Dietmar Plenz. «Higher-order interactions characterized in cortical activity». In: *Journal of neuroscience* 31.48 (2011), pp. 17514–17526 (cit. on p. 3).

[32] Eyal Bairey, Eric D Kelsic, and Roy Kishony. «High-order species interactions shape ecosystem diversity». In: *Nature communications* 7.1 (2016), p. 12285 (cit. on p. 4).

[33] Damon Centola, Joshua Becker, Devon Brackbill, and Andrea Baronchelli. «Experimental evidence for tipping points in social convention». In: *Science* 360.6393 (2018), pp. 1116–1119 (cit. on p. 4).

[34] Brian Borsari and Kate B Carey. «Peer influences on college drinking: A review of the research». In: *Journal of substance abuse* 13.4 (2001), pp. 391–424 (cit. on p. 4).

[35] Whitney A Brechwald and Mitchell J Prinstein. «Beyond homophily: A decade of advances in understanding peer influence processes». In: *Journal of research on adolescence* 21.1 (2011), pp. 166–179 (cit. on p. 4).

[36] Federico Battiston et al. «The physics of higher-order interactions in complex systems». In: *Nature Physics* 17.10 (Oct. 2021), pp. 1093–1098. ISSN: 1745-2473, 1745-2481. DOI: 10.1038/s41567-021-01371-4. URL: https://www.nature.com/articles/s41567-021-01371-4 (cit. on p. 4).

[37] Timoteo Carletti, Duccio Fanelli, and Renaud Lambiotte. «Random walks and community detection in hypergraphs». In: *Journal of Physics: Complexity* 2 (1 Mar. 2021). ISSN: 2632072X. DOI: 10.1088/2632-072X/abe27e (cit. on p. 5).

[38] Ana P. Millán, Joaquín J. Torres, and Ginestra Bianconi. «Explosive Higher-Order Kuramoto Dynamics on Simplicial Complexes». In: *Physical Review Letters* 124 (21 May 2020). ISSN: 10797114. DOI: 10.1103/PhysRevLett.124.218301 (cit. on p. 5).

[39] Iacopo Iacopini, Giovanni Petri, Alain Barrat, and Vito Latora. «Simplicial models of social contagion». In: *Nature Communications* 10 (1 Dec. 2019). ISSN: 20411723. DOI: 10.1038/s41467-019-10431-6 (cit. on p. 5).

[40] Iacopo Iacopini, Giovanni Petri, Andrea Baronchelli, and Alain Barrat. «Group interactions modulate critical mass dynamics in social convention». In: *Communications Physics* 5 (1 Dec. 2022). ISSN: 23993650. DOI: 10.1038/s42005-022-00845-y (cit. on p. 5).

[41] Guilherme Ferraz de Arruda, Giovanni Petri, and Yamir Moreno. «Social contagion models on hypergraphs». In: *Physical Review Research* 2.2 (Apr. 10, 2020), p. 023032. ISSN: 2643-1564. DOI: 10.1103/PhysRevResearch.2.023032. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.2.023032 (cit. on p. 5).

[42] Maxime Lucas, Iacopo Iacopini, Thomas Robiglio, Alain Barrat, and Giovanni Petri. «Simplicially driven simple contagion». In: *Phys. Rev. Res.* 5 (1 Mar. 2023), p. 013201. DOI: 10.1103/PhysRevResearch.5.013201. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.5.013201 (cit. on p. 6).

[43] Miller McPherson, Lynn Smith-Lovin, and James M Cook. «Birds of a feather: Homophily in social networks». In: *Annual review of sociology* 27.1 (2001), pp. 415–444 (cit. on pp. 6, 21).

[44] Kristen M. Altenburger and Johan Ugander. «Monophily in social networks introduces similarity among friends-of-friends». In: *Nature Human Behaviour* 2.4 (Mar. 19, 2018), pp. 284–290. ISSN: 2397-3374. DOI: 10.1038/s41562-018-0321-8. URL: https://www.nature.com/articles/s41562-018-0321-8 (cit. on pp. 6, 21, 22).

[45] Juliette Stehlé, François Charbonnier, Tristan Picard, Ciro Cattuto, and Alain Barrat. «Gender homophily from spatial behavior in a primary school: A sociometric study». In: *Social Networks* 35.4 (2013), pp. 604–613. ISSN: 0378-8733. DOI: https://doi.org/10.1016/j.socnet.2013.08.003. URL: https://www.sciencedirect.com/science/article/pii/S0378873313000737 (cit. on pp. 6, 10, 21, 24).

[46] Fariba Karimi, Mathieu Génois, Claudia Wagner, Philipp Singer, and Markus Strohmaier. «Homophily influences ranking of minorities in social networks». In: *Scientific reports* 8.1 (2018), p. 11077 (cit. on p. 6).

[47] Alain Barrat Ciro Cattuto. *SocioPatterns*. 2008. URL: http://www.sociopatterns.org/ (cit. on pp. 7, 10).

[48] Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. «Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks». In: *PLOS ONE* 5.7 (July 2010), e11596. DOI: 10.1371/journal.pone.0011596 (cit. on pp. 7, 10, 11, 15).

[49] Mathieu Génois, Maria Zens, Marcos Oliveira, Clemens Lechner, Johann Schaible, and Markus Strohmaier. «Combining sensors and surveys to study social contexts: Case of scientific conferences». In: arXiv:2206.05201 (June 10, 2022). arXiv: 2206.05201[physics]. URL: http://arxiv.org/abs/2206.05201 (cit. on pp. 7, 12, 23, 51).

[50] Mathieu Génois, Christian L. Vestergaard, Julie Fournet, André Panisson, Isabelle Bonmarin, and Alain Barrat. «Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers». In: *Network Science* 3 (03 Sept. 2015), pp. 326–347. ISSN: 2050-1250. DOI: 10.1017/nws.2015.10. URL: http://journals.cambridge.org/article_S2050124215000107 (cit. on pp. 7, 10).

[51] Philippe Vanhems, Alain Barrat, Ciro Cattuto, Jean-François Pinton, Nagham Khanafer, Corinne Régis, Byeul-a Kim, Brigitte Comte, and Nicolas Voirin. «Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors». In: *PLoS ONE* 8.9 (Sept. 2013), e73970. DOI: 10.1371/journal.pone.0073970. URL: http://dx.doi.org/10.1371%5C%2Fjournal.pone.0073970 (cit. on p. 7).

[52] Juliette Stehlé et al. «High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School». In: *PLOS ONE* 6.8 (Aug. 2011), e23176. DOI: 10.1371/journal.pone.0023176. URL: http://dx.doi.org/10.1371/journal.pone.0023176 (cit. on pp. 7, 12).

[53] Julie Fournet and Alain Barrat. «Contact Patterns among High School Students». In: *PLOS ONE* 9.9 (Sept. 2014), pp. 1–17. DOI: 10.1371/journal.pone.0107878. URL: https://doi.org/10.1371/journal.pone.0107878 (cit. on pp. 7, 12).

[54] Michele Starnini, Andrea Baronchelli, and Romualdo Pastor-Satorras. «Modeling Human Dynamics of Face-to-Face Interaction Networks». In: *Physical Review Letters* 110.16 (Apr. 15, 2013), p. 168701. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.110.168701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.110.168701 (cit. on pp. 7, 29, 30, 35, 36).

[55] Nathan Eagle and Alex Pentland. «Reality mining: sensing complex social systems». In: *Personal and ubiquitous computing* 10 (2006), pp. 255–268 (cit. on p. 9).

[56] Kun Zhao, Juliette Stehlé, Ginestra Bianconi, and Alain Barrat. «Social network dynamics of face-to-face interactions». In: *Physical Review E* 83.5 (May 12, 2011), p. 056109. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.83.056109. URL: https://link.aps.org/doi/10.1103/PhysRevE.83.056109 (cit. on pp. 10, 19, 34).

[57] Giulia Cencetti, Federico Battiston, Bruno Lepri, and Márton Karsai. «Temporal properties of higher-order interactions in social networks». In: *Scientific Reports* 11.1 (Mar. 29, 2021), p. 7028. ISSN: 2045-2322. DOI: 10.1038/s41598-021-86469-8. URL: https://www.nature.com/articles/s41598-021-86469-8 (cit. on pp. 10, 16, 35).

[58] Marcos Oliveira, Fariba Karimi, Maria Zens, Johann Schaible, Mathieu Génois, and Markus Strohmaier. «Group mixing drives inequality in face-to-face gatherings». In: *Communications Physics* 5.1 (May 27, 2022), p. 127. ISSN: 2399-3650. DOI: 10.1038/s42005-022-00896-1. URL: https://www.nature.com/articles/s42005-022-00896-1 (cit. on pp. 10, 37, 38).

[59] Simon Cauchemez et al. «Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza». In: *Proceedings of the National Academy of Sciences* 108.7 (2011), pp. 2825–2830 (cit. on p. 10).

[60] Gerardo Chowell and Cécile Viboud. «A practical method to target individuals for outbreak detection and control». In: *BMC medicine* 11.1 (2013), pp. 1–3 (cit. on p. 10).

[61] Timo Smieszek and Marcel Salathé. «A low-cost method to assess the epidemiological importance of individuals in controlling infectious disease outbreaks». In: *BMC medicine* 11.1 (2013), pp. 1–8 (cit. on p. 10).

[62] Marco Ajelli, Piero Poletti, Alessia Melegaro, and Stefano Merler. «The role of different social contexts in shaping influenza transmission during the 2009 pandemic». In: *Scientific reports* 4.1 (2014), pp. 1–7 (cit. on p. 10).

[63] Albert-László Barabási. «The origin of bursts and heavy tails in human dynamics». In: *Nature* 435.7039 (May 2005), pp. 207–211. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature03459. URL: http://www.nature.com/articles/nature03459 (cit. on p. 15).

[64] Nicholas W. Landry, Maxime Lucas, Iacopo Iacopini, Giovanni Petri, Alice Schwarze, Alice Patania, and Leo Torres. «XGI: A Python package for higher-order interaction networks». In: *Journal of Open Source Software* 8.85 (May 2023), p. 5162. DOI: 10.21105/joss.05162. URL: https://joss.theoj.org/papers/10.21105/joss.05162 (cit. on p. 16).

[65] Juliette Stehlé, Alain Barrat, and Ginestra Bianconi. «Dynamical and bursty interactions in social networks». In: *Phys. Rev. E* 81 (3 Mar. 2010), p. 035101. DOI: 10.1103/PhysRevE.81.035101. URL: https://link.aps.org/doi/10.1103/PhysRevE.81.035101 (cit. on p. 19).

[66] Nate Veldt, Austin R. Benson, and Jon Kleinberg. *Combinatorial Characterizations and Impossibilities for Higher-order Homophily*. Jan. 11, 2023. arXiv: 2103.11818[cs]. URL: http://arxiv.org/abs/2103.11818 (cit. on pp. 21–23).

[67] Eleanor E Maccoby. «Gender and group process: A developmental perspective». In: *Current directions in psychological science* 11.2 (2002), pp. 54–58 (cit. on p. 24).

[68] David Laniado, Yana Volkovich, Karolin Kappler, and Andreas Kaltenbrunner. «Gender homophily in online dyadic and triadic relationships». In: *EPJ Data Science* 5.1 (2016), p. 19 (cit. on p. 24).

[69] Bradley Efron and Robert Tibshirani. «The bootstrap method for assessing statistical accuracy». In: *Behaviormetrika* 12 (1985), pp. 1–35 (cit. on p. 25).

Appendix A

Higher-order homophily in scientific conferences - additional material

We present in the following tables and figures the results for the graph homophily and the higher-order homophily based on different categories present in the metadata accompanying the contact data of the four scientific conferences `WS16`, `ICCSS17`, `ECSS18` and `ICCSS17` [49].

In the absence of corresponding graph homophily, our findings demonstrate the presence of higher-order homophily is possible (highlighted plots). However, upon comparing the distinct datasets, we observe that the results regarding higher-order homophily lack consistency across various venues and sizes. This indicates that the concept of higher-order homophily is highly dependent on the context and cannot be regarded as a universal characteristic of face-to-face interactions in human gatherings.

Where only three datasets (`WS16`, `ECSS18` and `ECIR19`) are shown it is because in the fourth dataset (`ICCSS17`) the possible answers of the survey were organized differently hence not allowing the comparison with the other datasets.

		20 s	**60 s**	**120 s**	**180 s**	**300 s**
WS16	**Language 1**	0.96	0.95	0.96	0.99	0.96
	Language 2	0.95	0.94	0.93	0.89	0.95
	Others	1.01	1.03	1.04	1.06	1.07
ECSS18	**Language 1**	0.96	0.95	0.95	0.97	0.97
	Language 2	1.11	1.17	1.20	1.02	1.04
	Others	0.98	0.93	0.91	0.91	0.89
ECIR19	**Language 1**	1.10	1.17	1.19	1.17	1.24
	Language 2	0.84	1.06	0.80	0.98	1.31
	Others	0.98	0.97	0.97	0.96	0.98

Table A.1: Graph language homophily indices for different datasets at different cut-offs of the duration of the edges.

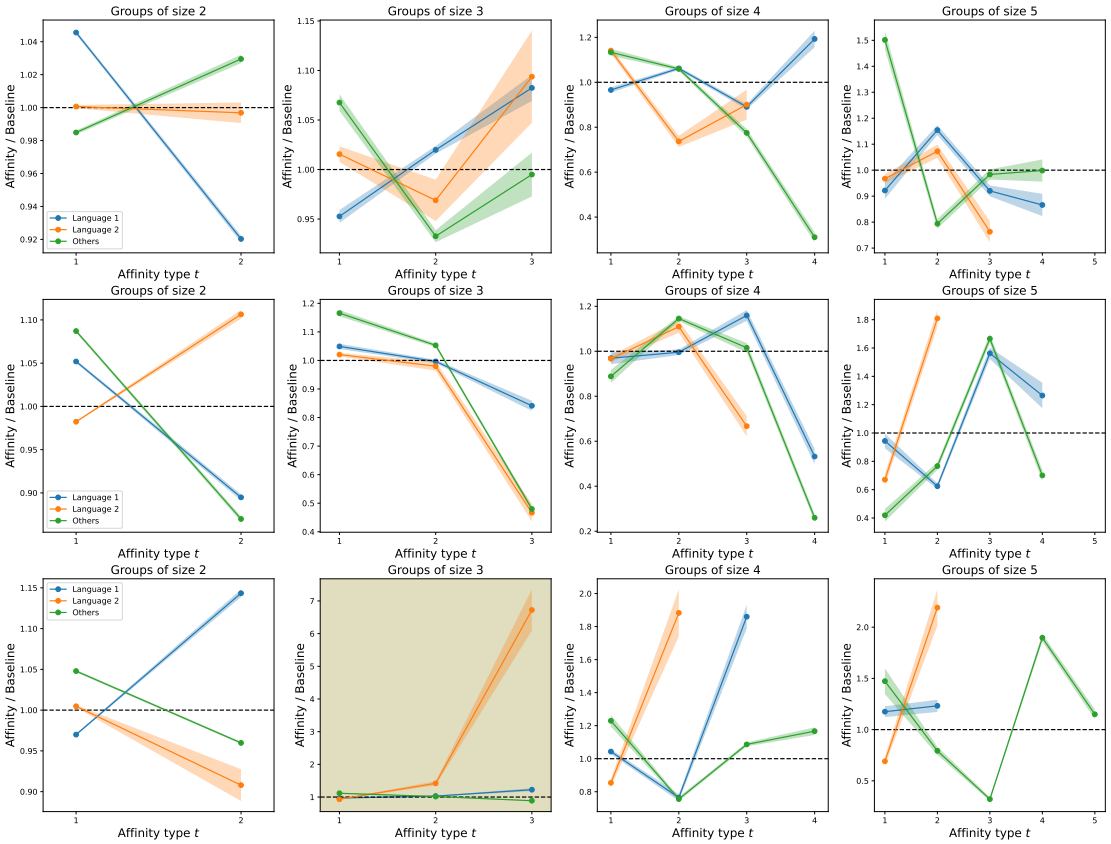

Figure A.1: Ratio between the type-t affinity score and the baseline to quantify higher-order language homophily in the `WS16`, `ECSS18` and `ECIR19` (from top to bottom) datasets. The aggregated hypergraph over which we compute the affinity scores is obtained aggregating the temporal hypergraphs keeping only the groups with an aggregated duration ≥ 60 s, similar results are obtained varying this cut-off threshold. The highlighted plots are those displaying higher-order homophily.

		20 s	**60 s**	**120 s**	**180 s**	**300 s**
WS16	**Social/Political sciences**	1.00	0.99	0.96	0.94	0.95
	Comp./Inform. science	0.96	0.94	0.96	0.97	0.98
	Math/Physics	1.02	1.09	1.03	0.87	0.96
	Media/Comm. science	1.00	1.13	1.26	1.44	1.39
ICCSS17	**Social/Political sciences**	1.02	1.02	1.07	1.04	1.05
	Comp./Inform. science	0.99	0.93	0.91	0.87	0.83
	Math/Physics	0.94	0.81	0.74	0.77	0.81
	Media/Comm. science	0.92	0.99	1.02	1.13	1.14
ECSS18	**Social/Political sciences**	0.99	1.01	1.02	1.03	0.98
	Comp./Inform. science	1.01	0.99	0.96	1.00	0.97
	Math/Physics	1.08	1.41	1.67	1.93	2.38
	Media/Comm. science	0.94	0.89	0.87	0.99	1.10
ECIR19	**Social/Political sciences**	/	/	/	/	/
	Comp./Inform. science	1.02	1.02	1.02	1.03	1.02
	Math/Physics	/	/	/	/	/
	Media/Comm. science	1.06	1.07	1.01	0.91	0.99

Table A.2: Graph discipline homophily indices for different datasets at different cut-offs of the duration of the edges.

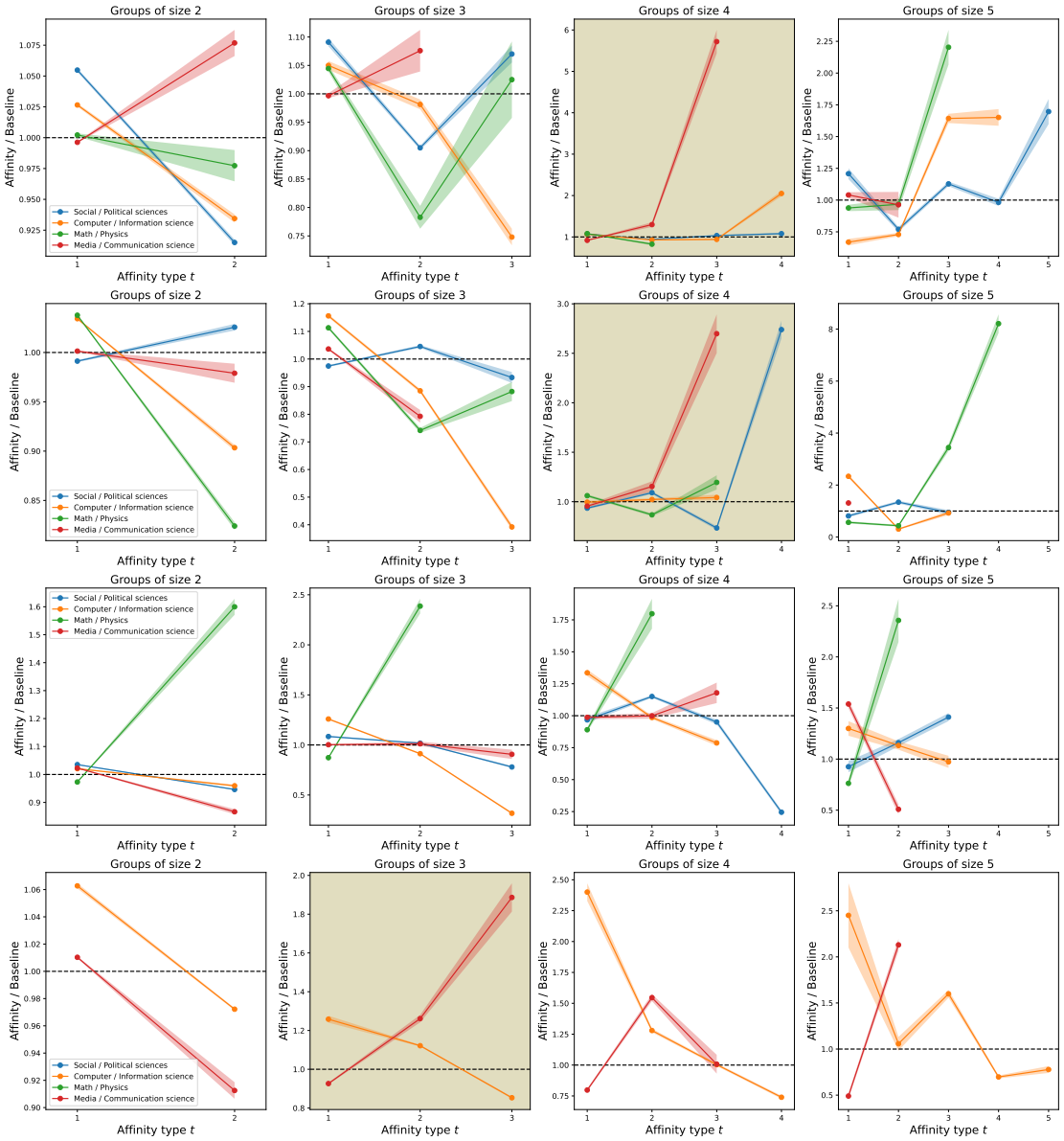

Figure A.2: Ratio between the type-t affinity score and the baseline to quantify higher-order scientific discipline homophily in the `WS16`, `ICCSS18`, `ECSS18` and `ECIR19` (from top to bottom) datasets. The aggregated hypergraph over which we compute the affinity scores is obtained aggregating the temporal hypergraphs keeping only the groups with an aggregated duration ≥ 60 s, similar results are obtained varying this cut-off threshold. The highlighted plots are those displaying higher-order homophily.

		20 s	**60 s**	**120 s**	**180 s**	**300 s**
WS16	**BA./MA./PhD student**	0.94	0.89	0.86	0.87	0.88
	Postdoc./Senior researcher	1.06	1.11	1.15	1.17	1.23
	Ass./Assoc. prof./Lecturer	0.99	1.10	1.29	1.27	1.28
	Full prof./Chair	1.07	0.97	0.91	0.83	0.96
ECSS18	**BA./MA./PhD student**	0.94	0.92	0.91	0.90	0.92
	Postdoc./Senior researcher	1.07	1.07	1.03	0.96	0.96
	Ass./Assoc. prof./Lecturer	1.14	1.21	1.22	1.13	1.19
	Full prof./Chair	/	/	/	/	/
ECIR19	**BA./MA./PhD student**	1.00	1.02	1.01	1.00	1.01
	Postdoc./Senior researcher	0.95	0.98	1.02	1.17	0.83
	Ass./Assoc. prof./Lecturer	1.11	1.23	1.23	1.08	0.98
	Full prof./Chair	1.08	1.10	1.08	1.03	1.06

Table A.3: Graph academic status homophily indices for different datasets at different cut-offs of the duration of the edges.

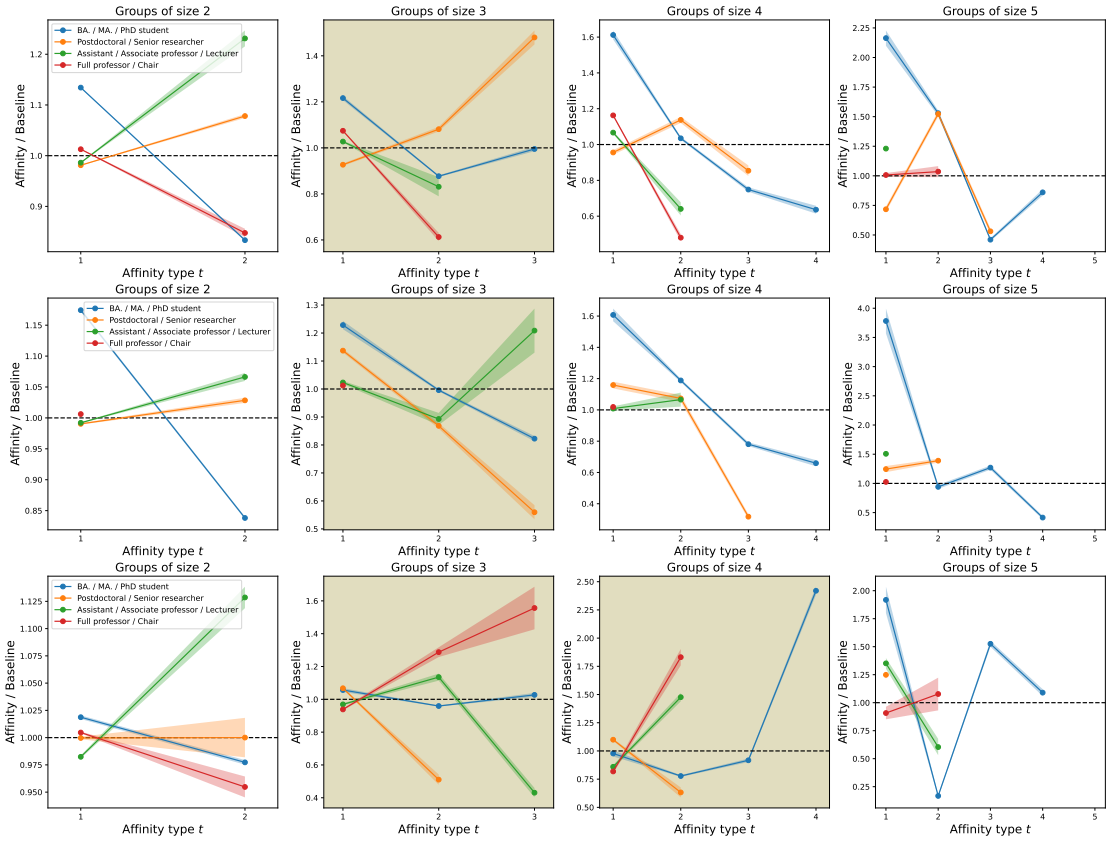

Figure A.3: Ratio between the type-t affinity score and the baseline to quantify higher-order academic status homophily in the `WS16`, `ECSS18` and `ECIR19` (from top to bottom) datasets. The aggregated hypergraph over which we compute the affinity scores is obtained aggregating the temporal hypergraphs keeping only the groups with an aggregated duration ≥ 60 s, similar results are obtained varying this cut-off threshold. The highlighted plots are those displaying higher-order homophily.