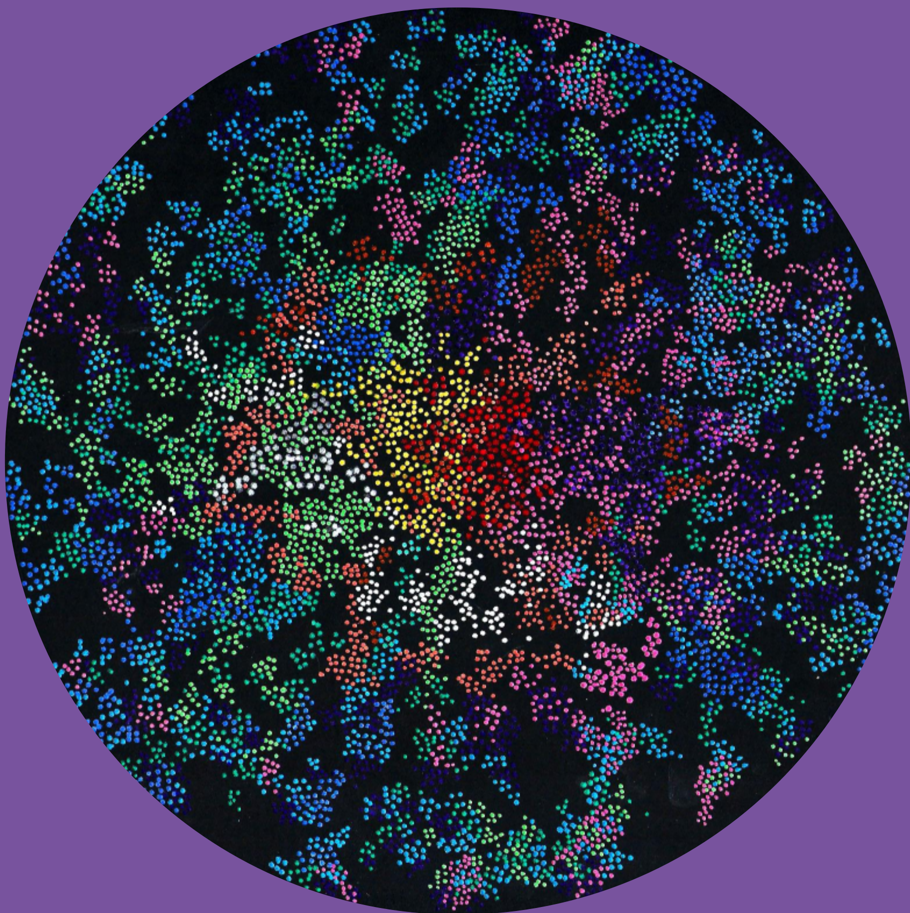
Structured Interactions

Inferring Social Behaviour in Networked Systems

Javier Ureña Carrión



Aalto University

DOCTORAL THESES

Aalto University publication series
DOCTORAL THESES 188/2024

Structured Interactions

Inferring Social Behaviour in Networked Systems

Javier Ureña Carrión

A doctoral thesis completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall T2 of the Department of Computer Science on 15 October 2024 at 12:00.

Aalto University
School of Science
Complex Systems

Supervising professor
Professor Mikko Kivelä, Aalto University, Finland

Thesis advisor
Professor Mikko Kivelä, Aalto University, Finland

Preliminary examiners
Professor Naoki Masuda, University at Buffalo, United States of America
Professor Eduardo Enrique López Atencio, George Mason University, United States of America

Opponent
Professor Esteban Moro Egido, Northeastern University, United States of America

Aalto University publication series
DOCTORAL THESES 188/2024

ISBN 978-952-64-2012-7 (printed)
ISBN 978-952-64-2013-4 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-64-2013-4

Images: "Identity Politics". Acrylic on black cardboard by Javier Ureña Carrion, 2024.

Unigrafia Oy
Helsinki 2024

Finland

NORDIC SWAN ECOLABEL
Printed matter
4041-0619

A!
Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Javier Ureña Carrión

Name of the doctoral thesis
Structured Interactions: Inferring Social Behaviour in Networked Systems

Publisher School of Science

Unit Department of Computer Science

Series Aalto University publication series DOCTORAL THESES 188/2024

Field of research Network Science

Manuscript submitted 12 March 2024 **Date of the defence** 15 October 2024

Permission for public defence granted (date) 20 June 2024 **Language** English

☐ **Monograph** ☒ **Article thesis** ☐ **Essay thesis**

Abstract

Social behaviour permeates our lives. We live locally through our friends, family and acquaintances, and on larger scales by sharing cultural values, identities and political opinions. Sociology has a rich tradition of characterising societal dynamics through concepts such as social roles, group identities and social spaces. At the same time, network science offers a valuable framework for analyzing interactions patterns, integrating tools from physics and data analysis to connect micro-level contacts with large-scale phenomena. Aided by the digital traces left behind by information technologies, both fields have helped us understand how we communicate, how information and epidemics spread in social systems and provided insights into polarization. These advances have revealed that the structures and dynamics of social systems are deeply intertwined, prompting broader questions about how social behaviour affects and is affected by networked phenomena across scales. Our contributions elucidate how networked social behaviour is reflected on empirical communication data and how it can be modeled through network dynamics —two pillars of modern network science.

We evaluate communication patterns through the lenses of longstanding sociological theories, revealing how tie strength and multiplexity —overlapping social roles— manifest in rich temporal data. We analyse features of temporal communication that act as indicators of relationship strength, assessing their prevalence and limitations across various communication mediums. We also show that multiplexity has a temporal expression through the use of different social times during the week, such as worktimes and weekends. We show that such multiplexity has structural effects on networks that align with the expectations of the theory of social foci. We also assess the degree to which these communication patterns mirror intrinsic human behaviour by analyzing a historical dataset of epistolary communication.

Then, we model how mechanistic networked processes such as meeting friends-of-friends or popular people can interact with group identities and lead to salient large-scale phenomena, such as polarization or inequality between groups. We analyse models of choice homophily, triadic closure and preferential attachment, showing that the latter two can impose structural constraints —inducing more in-group ties and leading to core-periphery networks where one group dominates connections, respectively. We use different inferential frameworks to quantify induced homophily in empirical systems, and to infer the presence of these mechanisms in data with core-periphery parings.

Our findings reveal how forms of social behaviour function as both drivers and outcomes within networked systems. The modelling and inferential methodologies we propose uncover a wide spectrum of societal patterns, encompassing tie strength, multiplexity, and structural biases induced by homophily and networked interactions.

Keywords Social networks, Human Communication, Multiplexity, Homophily, Core-Periphery

ISBN (printed) 978-952-64-2012-7 **ISBN (pdf)** 978-952-64-2013-4

ISSN (printed) 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2024

Pages 194 **urn** http://urn.fi/URN:ISBN:978-952-64-2013-4

Preface

As I finish my doctoral journey on how we affect and are affected by our social environment, it feels impossible not to realise that this thesis represents a collective effort. My parents, Laura and Javier, have been such a fundamental part of my life that it'd be unimaginable for me to disentangle the countless ways in which they've shaped and supported me. The same goes to my sister Laura. Even though we all live in different countries, they have been the most persistent source of joy, knowledge and love throughout my life. This is also the work of my chosen family, Raúl and Janne, who fill my every day with love, laughter, and wisdom regarding life, society, and art, and who have put up with me in arduous times. Although our understanding of canine cognition can't provide insights into his mind, Tonttu is also a fundamental part this family.

My supervisor Mikko Kivelä has provided invaluable guidance through this journey, patiently coaching me through difficult projects, and showing me how to turn disparate ideas into testable, scientific insights. Jari Saramäki, Gerardo Íñiguez and Fariba Karimi have also been invaluable in this process. I'd also like to thank Professors Naoki Masuda and Eduardo López, whose comments and insights during pre-examination have made this work much stronger. I could not thank enough my colleagues and friends from the Complex Systems research group, and from Aalto in general: Sara, Talayeh, Silja, Ana, Tarmo, Abbas, Richard, Ali S, Takayuki, Arash, Ted, Ali F, Hasti, Yan, Tolou and many others who have shared their academic journey with me. And of course, I'd like to thank the countless researchers whose work has inspired, shaped and influenced this thesis – I am happy to stand in the shoulders of giants.

I moved to Finland in 2017, happy to pursue a new life, and grateful to have Raúl embark on this adventure with me. Through these years I've come to realise that migrating is exciting, but it's also not an easy feat –it implies leaving part of your life behind, rebuilding parts of it anew, reinventing some others and gaining a new identity as an immigrant. It can be heartbreaking to be away from people you love in a way that no research on the transience of relationships or alter turnover can capture. I'm incredibly grateful to my friends who have shaped me and grown with me, forming a fundamental part of my worldview, and who I am. Sofi, Bernardo, Sam, Ceci, Regina, Gonzalo, César, Alan, you are all loved, and I cherish the moments we've had together, and can't wait to share more.

New social spaces, however, also provide opportunities to make new connections, and I am extremely happy and proud to have wonderful people around me. Paola, thank you for navigating these new spaces with me, you've become one of my closest friends. Arif, you've become a cornerstone of my life in Helsinki, even if we don't always have similar communication strategies. Shubi, Saara, Hanan, and everyone from the Moving Co., you've all deeply shaped and supported me. Alku, Ulla and Frida, serendipity brought us together, but I can't think of better friends and flatmates. To the countless friends from the academic, artist, immigrant, and queer circles in Helsinki, I can't thank you enough for showing me life in community. In the spirit of this work, and in the knowledge that all relationships start, are maintained for some time, and then decay: to *weak ties* past, present and future, you open doors to new worlds.

Helsinki,

Javier Ureña Carrión

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Ureña-Carrion, Javier and Saramäki Jari and Kivelä, Mikko. Estimating tie strength in social networks using temporal communication data. *EPJ Data Science*, 9, 37, December 2020.

II Asikainen, Aili and Iñiguez, Gerardo and Ureña-Carrión, Javier and Kaski, Kimmo and Kivelä, Mikko. Cumulative effects of triadic closure and homophily in social networks. *Science Advances*, 6, 19, May 2020.

III Ureña-Carrion, Javier and Leskinen, Petri and Tuominen, Jouni and van den Heuvel, Charles and Hyvönen, Eero and Kivelä, Mikko. Communication now and then: analyzing the Republic of Letters as a communication network. *Applied Network Science*, 7,1, May 2022.

IV Ureña-Carrión, Javier and Karimi, Fariba and Iñiguez, Gerardo and Kivelä, Mikko. Assortative and preferential attachment lead to core-periphery networks. *Physical Review Research*, 5, 4, p. 043287, December 2023.

V Ureña-Carrion, Javier and Heydari, Sara and Aledavood, Talayeh and Saramäki, Jari and Kivelä, Mikko. Multiplexity is temporal: effects of social times on network structure. Submitted to *Peer Review*, March 2024.

Author's Contribution

Publication I: "Estimating tie strength in social networks using temporal communication data"

JUC and MK designed the study, JUC analysed the data and all authors interpreted the results. All authors contributed to writing and approving the final manuscript.

Publication II: "Cumulative effects of triadic closure and homophily in social networks"

AA, GI, KK and MK conceived the main ideas, AA performed the experiments and AA, GI, KK, and MK wrote the initial manuscript. JUC designed, implemented and evaluated the likelihood-free inference methods. All authors contributed to writing and approving the final manuscript.

Publication III: "Communication now and then: analyzing the Republic of Letters as a communication network"

JUC, EH and MK conceived the main ideas. JUC and MK designed the analysis, and PL, JT and JUC collected the data and contributed with analytic tools. JUC implemented and performed the analysis. CVDH contextualised historical data and designed analysis. JUC took the lead in writing the manuscript to which all authors contributed.

Publication IV: "Assortative and preferential attachment lead to core-periphery networks"

All authors conceived, designed, and developed the study. JUC implemented and analysed the models, including the inferential framework and empirical data studies. JUC wrote the initial manuscript, and all authors contributed to writing and approving the paper.

Publication V: "Multiplexity is temporal: effects of social times on network structure"

All authors conceived the study. JUC designed and developed the methods and analyses. JUC implemented the models and empirical data studies, to which SH also contributed. All authors contributed to the interpretation of the results. JUC wrote the first draft and all authors revised and approved the manuscript.

List of Figures

Abbreviations

ABC Approximate Bayesian Computation

BA Barabási-Albert (network model)

ER Erdős–Rényi (network model)

GCC Giant connected component

JSD Jensen-Shannon Divergence ’

MFE Mean-field equation

ML Maximum Likelihood

NNMF Non-negative matrix factorization

SBM Stochastic Block Model

Symbols

k Node degree

O_{ij} Overlap for tie ij.

ΔT Length of sliding window

$\hat{O}^t$ Mean temporal overlap

$G(\theta)$ Random graph model

θ Model parameter

$\hat{\theta}$ Estimator for θ.

g Observed network

n_k Number of nodes of degree k

w Total contact counts

t_i Time of ith contact

τ Inter-event time

B Burstiness coefficient

E Number of events in a bursty cascade

N^E Number of bursty cascades

TS Temporal stability of contacts within an observation window

$\hat{t}$ Average interaction time

T Temporal biases in observation window

C_i Fraction of events in weekly cluster

J Number of latent components in NNMF

X Population-level matrix of weekly activity

H NNMF matrix of social times

G NNMF matrix of usage of social times

X_i Weekly activity of tie i (random variable)

α_i Coefficient of social times for tie i

$S(\alpha)$ Temporal multiplexity, entropy of α

d_* Distance between social signatures –either for the same ego *self* or for a reference *ref*

s_a Choice homophily of group a

N_a Number of nodes in group a

n_a Relative size of group a

L_{ab} Number of links between groups a and b

P_{ab} Fraction of links between groups a and b

T_{ab} Probability that following a random link from group a will lead to group b

ρ_{ab} Link density between groups a and b

1. Introduction

People interact with each other in countless ways. We participate in conversations in person and through different communication channels, we share common spaces for leisure and work, we inhabit different parts of our towns or cities, and engage in online discussions through social media platforms. Among the people we interact with, we have close friends and family, but we also have colleagues and a wide array of acquaintances. Social contexts vary, yet we might have the impression that the many people around us share the same values to us [135, 25, 138], or be rightly surprised to find that we have common friends with someone we just met – and who lives regions away. Then again, reading newspapers or browsing social media we might get the impression that we live in bubbles [136, 42], and that other people, perhaps geographically close but otherwise distant, have worldviews and lives diametrically opposed to ours [137, 145]. These, and many other social phenomena might seem counter-intuitive, and truly, society is complex. Yet to understand how these and other social phenomena come to be, we need to understand that we live our lives through interactions with others, and that all these different forms interactions are both networked and imbued in layers of social behaviour [190, 25, 121, 76, 137, 120, 191, 59].

Network science provides a useful framework for analyzing systems of interactions by representing them as networks – mathematical objects that we know how to analyse, approximate and manipulate using methods from statistical physics and computer science [3, 152, 130]. In the last twenty or so years, the widespread adoption of new communication technologies and social media platforms has provided unprecedented granularity into human interaction [185, 90, 130]. Such large-scale data sources, analysed under the frameworks of network science and sociology, have provided a wide array of advances on areas such as epidemic spreading [204, 129, 174, 82], social influence [181, 73], mental health [165, 5], communication patterns [142, 184, 80, 93, 81], labor flows [10, 126], social inequality [145, 201, 100], political polarization [188, 42, 39], among many others [185, 21, 98, 121].

The focus on interactions provided by network science allows us to zoom into auto-recorded contacts created by mobile phone calls, emails, social media platforms or proximity data, revealing rich behavioural patterns for individuals and groups [184, 185, 6, 117]. It also allows us to zoom out and gain large-scale perspectives of how these webs of interpersonal contacts scale up to constitute social systems [156, 157, 36, 15] – which are dynamic and in constant dialogue with both the existing patterns of relations and social mechanisms. You might meet someone new because they are your friend's friend [79, 116], but also because you happen to have a common interest [137] or because they work in the same building [59, 191]. Such contexts provide opportunities to interact, but sociality is also an individual trait, with people having different personalities [142, 34], social capacities [184, 6, 180], and strategies for maintaining their relationships [144, 93].

Networked mechanisms of social behaviour have large-scale effects, creating local perceptions of consensus [122], impacting access to information [42], shaping social influence and opinion dynamics [56, 181], affecting the spreading of information and disease [82, 129], or entrenching inequality if they favour particular populations [191, 100]. They affect social media platforms by explicitly or implicitly governing how users explore online social spaces [192, 132, 121]. A user might navigate a platform by befriending a suggested friend-of-a-friend, sharing their friend's post, or simply enjoying suggested content that is either "popular" or suggested to the user based on some underlying socio-demographic profile. Such interaction mechanisms in online platforms can interact with social behaviour in a way that shapes or amplifies structural effects such as echo-chambers [132, 42] or centralization of information flow [18].

The structure and dynamics of social systems are deeply intertwined with social behaviour, affecting and being affected by complex networked patterns [173, 138, 16, 199]. This means that understanding social systems requires understanding different forms of social behaviour [173, 190, 121, 25], such as how people use different strategies to navigate their environments or reflect their social identities. However, neither interpersonal relations nor group identities are physical objects that we can measure. To this end, a central question in computational social science is *how* to characterise such social behaviour and how to detect it and its networked effects.

Sociology provides a rich framework for understanding many societal dynamics, including concepts such as multiplexity – overlapping social roles within relationships – [205, 59], the strength of interpersonal ties that is topologically reflected in networked structures [76, 75], or group identities that drive the creation of new interpersonal ties [137, 191, 190]. These sociological concepts allow us to frame social behaviour in terms of unique combinations of social environments, where individuals create connections by sharing spaces that overlap with group identities, socioeconomic status and broader cultural contexts. Methods for network science, on the other hand, provide statistical frameworks of inference, as well as rich methodologies for modelling patterns of interaction in large systems.

Our contributions build up on the ever-evolving field of social network analysis and provide new characterizations of temporal social behaviour by grounding them in longstanding sociological theories. Using a combination of data analysis, modelling and a wide array of inferential frameworks, our work contributes to two large pillars of contemporary network science – the analysis of massive auto-recorded datasets and mechanistic models of social behaviour.

1.1 Research questions and thesis structure

The publications contained in this thesis examine various methods for inferring social behaviour in networked systems, focusing on two forms of dynamics – communication patterns and structural effects of networked mechanisms of social behaviour. Our overarching research questions are:

Q1 *How do contact patterns reflect sociological concepts of tie strength and multiplexity?*

Building on a large body of research for characterising the dynamics of human communication, in Publications I and III we systematically analyse temporal features of human communication. In Publication I we identify how such features capture topological tie strength, characterised by M. Granovetter through the presence of overlapping groups of friends around ties [76]. We introduce measures that account for social times during the week and biases in observation periods. Our contributions have direct application as they allow for the reconstruction of weighted networks where tie strength is reflected topologically.

In Publication V we extend our work on weekly social times by proposing a methodology for obtaining population-level social times. We reconstruct multilayer networks where ties use one, some or all social times. We propose that multiplexity –overlapping social roles – is reflected temporally through the use of such times. We test our framework under Feld's theory of social foci [59], showing that temporal multiplexity aligns with the expectations of focal multiplexity.

Q2 *How consistent is such social behaviour on different communication channels?*

In Publications I and III we systematically analyse communication patterns in different communication channels. Here, our goal is not to assess individual behaviour in different media, but to analyse the extent to which features of dyadic interactions can serve as proxies for tie strength (in Publication I) or largely reflect human behaviour in historical contexts, such as letters sent between 1500-1900 in Publication III.

We analyse datasets that reflect a myriad social systems, from phone calls and letters to online communication on social media platforms or forums, and work emails. These different datasets help us assess common patterns and system-specific differences, such as work emails reflecting a particular social context [120, 59, 191]. The dataset of epistolary communication is of particular interest, as it is subject to patterns in missing data that we can't fully anticipate. However, knowing that it accounts for figures of historical importance gives hints that some ties and egos have more complete information [91].

Q3 *How can we untangle the behavioural patterns of group identities from their effects on networked interactions?*

In Publications II and IV we model group behaviour through choice homophily – the notion that, when exploring social spaces, people make new connections based on similarity [137]. In Publication II we propose a model for the dynamic interaction of choice homophily and triadic closure, or meeting friend-of-friends. We show that triadic closure can induce more homophily in the system by restricting the available pool of new connections, and quantify the amount of induced homophily in a variety of temporal datasets.

In Publication IV we propose two models of choice homophily and preferential attachment, showing that their interplay can lead to structural imbalances where a group of nodes can become a core that dominates connections, in a structure called core-periphery. We estimate the amounts of choice homophily and preferential attachments present in empirical core-periphery systems using statistical inference.

Our work follows an interdisciplinary approach, borrowing concepts and methodologies from social science, mathematics, statistical physics and data analysis. The structure of this thesis follows a framework where we first explore how sociology and network science can provide insights into societal dynamics, how different methodologies can be used to gain such insights, and follow with detailed descriptions of our contributions.

In Chapter 2 we introduce basic notions of social network analysis, including key sociological frameworks and results from network science. We review some major results related to our two lines of research: how social behaviour is reflected in communication metadata, and how mathematical models of social interactions can help us gain insights about larger structural effects. Chapter 3 covers technical aspects of network analysis, including definitions, approaches to dealing with temporal data, inferential frameworks and tools used to understand large interacting systems. We present our main contributions for the analysis of interaction data in Chapter 4. We cover datasets and methodological choices, tie-level dynamics and methods for reconstructing multiplex networks of social times. We uncover ways in which tie strength and multiplexity are reflected temporally, and show that some of these behavioural patterns are also present in historical contexts. In Chapter 5 we cover network evolution models of social behaviour. We cover different approaches to measuring group-level interactions in social systems, such as homophily and core-peripheries. We introduce and systematically analyse our network evolution models, and show how these mechanisms are present in empirical datasets. We finalize with a short discussion and reflections on the research questions.

2. Background: networks as representations of social systems

2.1 Networks and data

Networks or graphs are mathematical objects that include a set of nodes that represent agents, actors or people, and a set of edges that represent pairwise relations or interactions between the nodes [152, 25]. These open-ended terms such as “interactions” or “pairwise relations” highlight why networks offer such a flexible modelling framework: at the core of network science lies the idea that for similar types of relations, the patterns of interactions themselves can capture essential aspects about the system in which those interactions occur. This flexibility is perhaps one of the reasons why network science has matured into a field of its own: interacting elements exist in neuroscience [198], economies [98, 13], the Internet [16], and —the object of study of this dissertation— society. This, of course, does not mean that *all these networks* representing interactions in *all these systems* look the same. They vary in structures and dynamics, which have distinct meanings and implications for different processes. In fact, this flexibility also smuggles a corollary about social networks: society does not consist of a “single social system”, but of collections of phenomena that involve different forms of human sociality and affect people in overlapping, contextualised, and fluid ways [121, 25].

This opens up a myriad of possibilities about *what* we can query about social systems. And so, while early sociological work focused on describing society through webs of “formal” interpersonal relationships –family, friends, colleagues, acquaintances [11, 128, 212, 50]–, not all social systems are fully defined by those concepts. This does not mean that such interpersonal relationships are not important, they are foundational concepts that have very tangible effects on networks, but they fall short as means of representing human sociality [121, 25]. In social media platforms, for example, friends, family and acquaintances might share the same online space with content creators, politicians, celebrities, corporations or other forms of parasocial relationships that are naturally asymmetric and unidirectional [35, 125]. The joint presence of two people in a restaurant or a café may not necessarily capture interpersonal relationships, but by existing within social and geographical contexts they can reveal, e.g, patterns of social inequality [214, 145], or potentially transmit an airborne disease [92, 123]. The term *tie* usually captures these broad notions of social relations. Now, whether ties attempt to capture formal interpersonal relationships or other forms of interactions is largely contextual, and perhaps more importantly, dependent on data.

Data sources are thus an essential part of social network analysis, as they largely determine what we can ask and understand about the phenomena or social system we want to study [120, 121, 25]. In this light, we'll begin this chapter by briefly describing some major categories of data sets and the social systems they represent. Figure 2.1 depicts an example of how empirical phenomena can be reconstructed as social networks.

Surveys were some of the first sources of social network data, being widely used by sociologists in the late 20th century, and still a valuable data source today [136, 183, 34]. The "name generator" was a common data-collection method, where researchers asked participants to name people they considered important within a social role or context [190, 128, 211]. While these methods are liable to the participant's memory and mood [212, 213], they are a valuable source of information on the nature of relationships as they were perceived by the participants —-including social roles, emotional support or hierarchies.

Communication metadata provided the first glimpse into large-scale social systems [185, 88], particularly from the analysis of mobile phone calls from country-wide mobile operators. These datasets have been used both as descriptors of large-scale communications systems that involve many real-life relationships [156, 118], but also for characterising behavioural patterns at personal and dyadic levels [107, 184]. Other communication datasets include metadata from emails [213, 40], text messages [80, 7]. Publication III analyses a large-scale dataset of historical letters as a communication network [54, 91].

Proximity data is usually collected for research purposes using bluetooth or other proximity sensors. In these cases, the social system corresponds to interactions of physical proximity within some social setting, such as a school or an office [183, 43].

Social media platforms are also rich sources of interaction data. The type of interactions available depend on the platform's infrastructure [121], such as explicit mutual friendships, sharing other user's content.

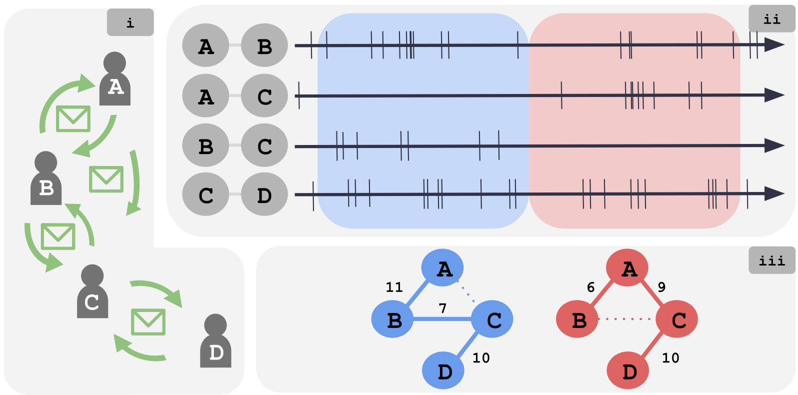


Figure 2.1. Network reconstruction from contact data. Interaction datasets, such as phone calls or proximity data, consist of sequences of contacts that may be reconstructed as networks. Different observation periods (colors) can lead to different structures. From III, licensed under CC 4.0.

The networked structure of social media platforms may differ widely depending on interaction mechanisms [102] or the type of information being shared [192]. Online platforms may be used to assess different phenomena, such as political polarization [39, 42] or exposure to different media sources [73].

Many other data sources do not necessarily fall within a particular category, but capture patterns of affiliation in scientific publications [115], gender dynamics in boards of directors [187], or national records of kinship and formal ties [22], among many others. We leave this list open as social processes can leave a myriad of different traces.

Contemporary network analyses may incorporate several different channels such as phone calls, social media, proximity data and surveys to obtain a more complete description of underlying interpersonal relationships [183]. These combined datasets have revealed, for example, that people use different communication channels at specific times of the day to engage with distinct social circles [7], but also that long-term individual communication patterns are persistent in time and across channels [80, 93].

2.2 The humble tie and the towering social structures on which it stands

In its most simple form, a tie (i.e., an edge in a graph) is a mathematical indicator that is zero if there is no pairwise relationship, and one if there is. If this definition looks deceptively simple, it's because many times it is: it implies mapping the contextualised complexities of human interactions onto a dichotomous indicator, which is an active theoretical choice [32, 190]. Ties can take richer mathematical forms and reflect a wide array of temporal dynamics; yet this does not mean that static, dichotomous ties are not useful. Whether they represent actual lifelong friendships or some minimal online contact (e.g., a couple of "like"s in a social media platform), ties may carry information about the social system they inhabit [24, 61, 138, 191, 121]. To interpret that meaning, however, it makes sense to stand back and examine the social structures that facilitate social interactions. These structures refer to sociological concepts that attempt to formalise major aspects of human sociality. In this section, we'll analyse ties from the perspective of three social structures: (i) as interpersonal relationships, (ii) as interactions that occur in social spaces, (iii) as markers of various forms of identity. In the process, we'll describe how these social structures can be reflected in the structure of a network, which also called topology. Then, we'll discuss how the network itself can be understood as a social structure.

2.2.1 Ties as understood by sociology

Interpersonal relationships are not easy to characterise [128, 212, 50]. They are intuitive when referencing friends, family or colleagues, but they do not have clear-cut borders. Relationships are intangible and expressed through social practice, mutual understandings and interactions – not through measurable physicality [190, 32, 127]. Perhaps for these reasons our understanding of interpersonal relationships is not unified, but there are several approaches. One way to characterise ties is through formal social roles: kinship, friendship, romantic partnerships, schoolmates, or colleagues [59, 24, 38]. Roles tend to refer to particular social orders such as families, or social spaces such as colleagues or classmates [190, 22]. As the study of social networks became more widespread in the second half of the 20th century, sociologists began to characterise ties in terms of how they were perceived [50]. Overall, the framework of perception allowed researchers to enquire about functional roles: whether someone was considered a friend, a source of advice or social support, perceived social status or simply whether they provided company via frequent contacts [50, 127]. Since many of these roles are not mutually exclusive, the term multiplexity usually refers to relationships that fulfill several social roles or needs [205].

Now, ties tend to cluster together [209, 84, 76]. The analysis of the most basic cluster, a triangle with three nodes and three ties, dates back to the beginnings of social network analysis with the work of G. Simmel at the turn of the twentieth century [84]. Psychologist F. Heider proposed that triads emerged as a form of cognitive consistency, which meant that either all three people had a positive attitude about each other, or two people would dislike the third, i.e., friendship needed to be transitive, triads would tend to close [79]. This simple model had a lasting impact as it defined social influence as a simple mathematical model based on ties, and triangles and triads have now become one of the major lenses through which we understand social systems, including as mechanisms through which we meet each other [94, 158], exert influence on one another [181], and as one of the building blocks of community-like structures [210, 94, 164]. The functional roles of ties and the tendency to form clusters became the foundation of the first sociological theory to link micro-level ties to macro-level social structures: the theory of the *strength of weak ties* [76]. Introduced by M. Granovetter in 1973, tie strength unified many of the "functional roles" of ties onto a single qualitative concept. Strong ties were characterised by emotional intensity, reciprocity, time and intimacy, whereas weak ties by the lack thereof [1]. Granovetter argued that strong ties would be located within overlapping circles of friends; but most importantly, that weak ties served as inter-community bridges. The implications of coupling a social concept such as tie strength with a topological concept such as bridges were huge, as they suggested that weak social connections had a prime role for accessing new information, whereas strong ties had a homogenising effect [76]. We now know that the actual role of weak ties is affected by communication patterns [104], social identities [134], or the nature of the diffusion process [36, 197, 13], yet the framework where the strength of interpersonal relationships has a topological role is still influential, and has led to a wide array of efforts to characterise tie strength [128, 127, 166, 213, 149].

S. Feld generalised ideas of transitivity and overlapping circles of friends through the concept of a *social focus*: the physical, institutional or legal spaces that facilitate the appearance of ties [59]. A focus can be a workplace, a gym or a university. Yet the core idea is that the structural roles of ties are linked to the characteristics of their social foci, such as its size and constraining ability, i.e. how likely it is to create ties. A university can be a large and non-constraining focus, as it does not necessarily enforce any two people to meet. A small class in that same university might be constraining if, e.g., it favours groups activities. [59, 61, 191]. This led the way to a unique form of multiplexity, where instead of focusing on fulfilling social roles [205], Feld defined multiplexity in terms of whether ties were present in several foci. He argued that the structural role of a tie within the network, such as its bridging capacity, is linked to the characteristics of the collections of foci in which it is present.

Analyzing how collections of social foci impact network structure can be difficult, as it requires defining imprecise and overlapping social environments [191, 61]. Focusing on particular social spaces can still be very

[1] In the original manuscript Granovetter did not particularly focus on characterising tie strength, but more on how it could be reflected topologically and its potential effect on larger social structures [25].

revealing of their effects. Small analysed how childcare centers impacted the social networks of mothers, and had an impact on well-being and new economic opportunities [191, 46]. From a large-scale perspective, recent advances have shown that labour flows between firms can be revealing of their size and effect on unemployment [126].

Ties can also reflect social identities. *Homophily* refers the notion that similar people tend to be connected [135, 137, 101], with similarity stated in terms of either sociodemographic descriptors such as race, gender and income, or value-based features such as political or ideological affiliation [137, 136, 42]. McPherson *et al.* characterised homophily as resulting in homogeneous personal networks, but most importantly, as “mapping sociodemographic distance to network distance” [138]. Its importance relies not on the fact that any person will be connected to another in the same group, but that it can account for structure and information flow across a demographic space [136, 189, 147]. Together with social foci, homophily is an approach to understanding community structures, and is present in political echo-chambers [42, 72], can lead to structural inequalities between groups [100], or impact epidemic spreading through the propensity of vaccinated and non-vaccinated people to form separate clusters [82].

The contributions of this thesis shed light on how these sociological concepts affect and are reflected in social systems: from how communication patterns can encode notions of tie strength and multiplexity to how homophily can become entrenched through simple mechanisms such as homogeneous social foci or meeting friends-of-friends [191, 116]. Publication I focuses on network reconstruction methods that capture topological tie strength through communication patters, while Publication V approaches network reconstruction by representing social times, such as workweeks and weekends, as layers on which ties interact. In our model, multiplexity is temporally reflected through the use of different social times, and we show that such multiplexity aligns with the expectations of Feld’s focus theory. Publication III compares how different communication systems capture temporal patterns of communication across different channels and historical periods. Under a modelling perspective, in Publication II we analyse how triadic closure can induce more homophily in a social system and result higher patterns of observed homophily. In other words, how the local network structures can constrain the pool of available people you can meet, increasing the number of same-group connections. We further elaborate on how homophily can interact with network structures in Publication IV, where we propose a model of homophily that leads to one group becoming dominant over another in a structure called core-periphery [26, 178]. To understand these processes, however, we’ll first examine how modelling and analyzing networked patterns of interactions can help us decode the composition of society and social systems.

2.2.2 Network topology as a social structure

A central idea within both sociology and contemporary network science is that the structure of a social network itself has an impact on individuals and groups [29, 3, 123, 25]. This means that the patterns of connections can provide sociological insights, such as someone's ability to influence other people [181, 17] or spread a message [36, 197]. The strength of analyzing networks is that shifting the focus from individuals to their interactions provides us with the tools to link the local to the global, such as whether there is a path between any two people in a society [209, 3, 104], or how seemingly trivial features, such as differences in the number of ties that people have, can reflect structural asymmetries [16, 60, 75].

Early attempts to understand society as a network emphasised global connectivity, or how many people are needed to connect any two random persons [35, 140]. S. Milgram's small-world experiment was one of the first studies to show evidence of high-connectivity: he selected random people from the US, and asked half of them to send a letter to someone in the other half, using only their name and state. The idea was that if they didn't know the person, they should send it to someone who they thought would know the target [140]. The popular notion of "six degrees of separation" owes its name to the average number of steps that the surviving letters followed to reach their target [35, 190]. More broadly, small-world networks refer to these coexisting notions of *global sparseness*, as most people don't know each other, *local density* as most people have triangles around them, and *high-connectivity*, as there are short paths between most pairs of nodes [140, 209, 104]. Efforts to understand patterns of global connectivity also came from theoretical network models. Using a probabilistic framework, mathematicians P. Erdős, A. Rényi [55], and E. Gilbert, separately [70], developed mathematical models of random graphs that showed how global connectivity and sparseness could emerge through stochastic processes. If there was a random network with n nodes and each edge between pairs of nodes appeared with probability p, a relatively small connection probability $(p > \frac{\ln n}{n})$ would likely result in most nodes being connected to each other through some path in a giant connected component (GCC) [55, 3, 152]. Now, social networks are not random, at least not in such a simplistic way [209, 3, 108, 123, 98], yet formal network models such as that of Erdős-Rényi (ER) paved the way to understanding how emergent global patterns can be formalised through analytical frameworks. Random network models still serve as a basis for understanding real-world networks through statistical inference [3, 176, 162].

The number of connections of a node, their degree k, is usually associated with the node's relative importance since more connections could imply more opportunities to spread or receive information [35, 152]. This is not always the case, as activity patterns [141, 105, 41], the type of spreading phenomena [36, 197] and the larger-scale network structure [152, 130] also affect spreading processes and overall connectivity. Yet differences in degrees do have an impact on how networks look and behave. A good example of this is the *friendship paradox*, that states that "your friends have more friends than you" [60]. This structural asymmetry stems from differences in node degrees, and its underlying mechanisms can be illustrated under the task of randomly finding a "popular" person when one only has access to a list of nodes and edges. Picking a random edge is more likely to lead to a popular person than picking a random node because popular people have more connections. This way, counting the number of friends of some person A implies picking a random node and then counting their friends. However, counting the number of friends of B, who is a friend of A, implies picking a random edge before observing B, so that on average B will have a higher degree [60, 57]. Such structural biases have direct implications on networks, including how they can entrench patterns of homophily [122, 116] or structural inequality [16, 201].

Global patterns of connectivity and degree heterogeneity are central concepts in network analysis, and efforts to characterise them and their effects have arguably led to network science becoming a field in itself [209, 15, 89, 98]. Degree heterogeneity is assessed by looking at the degree distribution, or the probability $p(k)$ that a node has some degree k. A major breakthrough for network science came from the realization that in many empirical networks' degree distributions are not only diverse – but massively so [16, 3, 45, 194]. An intuitive example of this are social media platforms where some people may have hundreds of connections, many others a few thousands, and some others millions [192, 36]; such broad distributions also exist in networks of interpersonal relationships [156, 22]. For degrees that vary so much, the average value is not very informative of the population, so that a notion of the "scale" of degrees is difficult to assess. Distributions where large numbers are not extremely uncommon are called heavy-tailed, and not without controversy [87, 45, 15], modeled using power-law degree distributions that formalise the notion of scale-free degrees [16, 3].

The importance of scale-free networks, however, also lies in the possible mechanisms that lead to their appearance [52, 15, 98, 89]. Models of preferential attachment, analysed by D. Price [168] and further studied and popularised by A.L. Barabási and R. Albert [16], take degrees at their core by proposing that new connections depend on the existing connections, i.e., new links are created with probabilities proportional to degrees [3, 152]. In an evolving network, this means that higher-degree nodes are cumulatively more likely to increase their degree, a reason why preferential attachment is also called cumulative advantage, or rich-get-richer phenomena [168, 52]. An intuitive example of preferential attachment is how "influencers" or "hubs" in social media have easier possibilities to grow their audience. In other words, having a large audience can translate into more people sharing their content and increasing their visibility, while social media algorithms themselves can increase exposure by explicitly taking popularity into account [192, 195]. Under the Barabási-Albert (BA) model, growing a network infinitely leads to a scale-free network with a power-law degree distribution [3, 15]. Such models of network evolution show that network dynamics and structure are intertwined, so that the mechanisms that drive changes in connections such as preferential attachment can lead to structural asymmetries, such as heavy-tailed degree distributions [87, 15, 216].

2.3 Social behaviour in networks is dynamic

Up to this point, we have largely discussed networks as static objects, a framework where ties can serve as symbolic highways transporting messages or social influence. Our main assertion, in line with this analogy, is that the landscape of highways mirrors social interactions by aligning to the underlying topography of social structures. However, human relationships are not static highways, but dynamic connections that also reflect social behaviour through activity and change, in time. Focusing on the dynamic nature of social systems unveils rich patterns where social circles change, paths appear and disappear, and social structures interact with each other. Network dynamics can be so rich, in fact, that different methodological approaches have arguably become sub-fields of their own [117, 130, 15, 90]. Temporal networks, for example, define interactions as *events* instead of links: brief connections such as phone calls or physical proximity measured through Bluetooth signals [130, 90]. Between events and fully static links, ties can also be thought to be active during a certain period [130, 88], capturing combinations of persistent and transient relationships [144, 149, 81]. Network dynamics may also focus on the mechanisms of change itself, characterising the evolution processes that underlie some static representation – such as how preferential attachment can lead to highly-skewed degree distributions [16, 199, 216].

This thesis focuses on two lines of research that incorporate elements of temporality and staticity from different perspectives. The first one is largely centered on *empirical data analysis* and traces its origins to the widespread use of information technologies. Auto-recorded metadata provided new insights into networked social behaviour, such as different strategies that people use to communicate with each other [142, 184], or detailed characterizations of how such behaviour affects the spread of information [104, 88]. The second framework focuses on *theoretical models* of microscopic social behaviour that can lead to large-scale effects. In other words, such models can shed light on how the local way in which people navigate their social spaces can reflect societal-wide phenomena, such as polarization [147, 195] or structural inequality between groups [52, 122]. Crucially, these two lines of research differ in their handling of dynamics and change: while the computational field handles time as determined by the data, the theoretical framework conceives more as a conduit for evolution and change.

In the following subsections we cover some major insights provided by these two major lines of research, highlighting relevant results to the publications of this thesis. In Sec. 2.3.1 we focus on communication networks, which are reconstructed from contact events produced by calls, messages, emails, among other communication channels. In Sec. 2.3.2 we outline some major results from network evolution models. In each case, we also outline the major theses and hypotheses of our contributions.

2.3.1 Communication patterns and human behaviour

Emails and mobile phones, which became increasingly available at the turn of the millennium, provided both novel channels through which people could interact, and an auto-recorded footprint of the interaction. In it's most basic form, such metadata included a sender, a receiver and a timestamp[2] [156, 185, 21]. These basic ingredients have led to a wide array of results: from describing and modelling how one-on-one communication occurs in time [14, 105] reflecting tie strength and relationship stability [156, 169, 148, 81], to behavioural characterizations of how individuals, called *egos* in network analysis, use their limited time to communicate with other people, called their *alters* [144, 184, 80, 93, 180]. These datasets are also a rich source for modelling information and epidemic spreading, as they capture the timescales and temporal correlations that likely affect such processes [129, 131, 86], with different lines of research focusing on tie-level dynamics [104, 143, 117, 88], individual nodes [141, 41] or correlated group behaviour [117, 197]. For now, we'll focus on tie- and ego-level communication patterns, which we motivate by asking: how long do relationships last?

Ties have a wide array of timescales: some relationships last entire lifetimes, many social interactions are shorter than a minute-long chat, and there's a long in-between where it's not always possible to determine whether an interpersonal relationship has ended or not [25, 23, 142, 81]. Using long observation periods that can range from 18 months to three years, a common approach is to discuss ties in terms of *active* or *dormant* states [142, 144, 88, 146], or as more recently proposed by Vergara Hidd *et al.*, as *stable* and *transient*, the latter of which can last from a few weeks to up to a year [81, 206]. In communication networks transient or temporary

[2] Mobile phone datasets may include more information, such as duration of call or different geolocation resolutions. For examples of geographical analyses see [33, 145, 214].

ties dominate [142, 81], although demographic factors such as the age of the people in the tie [184] and their gender [180] affect their prevalence. In general, persistent ties tend to communicate with some frequency [175, 149], volume [142] and include overlapping circles of friends [30], frequency tends to imply that they are also active early in an observation period [81, 149]. On the other hand, transient ties can behave in a wide array of ways, which generally includes a small call volume within several months [142], but might include high levels of activity for some mid-range period [81]. Relationship stability also involves social times: while new relationships mostly communicate during weekdays, communication during weekends is a good predictor of longer-term stability [206].

The notion that several timescales exist in relationships is mirrored in communication data as *event patterns have their own timescales*. Perhaps the best example is how human communication is *bursty* [14, 71, 97], where sequential calls may be followed by long waiting periods only to be followed by more events; in other words, inter-contact times are also heavy-tailed [33, 204]. Burstiness is perhaps a nesting doll of timescales: besides having long inter-event periods, the *sizes* of bursty cascades –sequences of events where any two contacts don't exceed a small time window – also have their own timescales [105, 106]. Similar results for the length of contacts have been found for proximity data [43, 123]. Naturally, burstiness can strongly affect the assessment of whether a relationship has decayed or not, a reason why long observation windows are used to define latent or dormant ties [142, 185, 81].

Egos have particular ways of dealing with their communication partners. The population-level transience of ties usually means that there is a high degree of alter turnover [142, 184, 80, 180]. Close relationships tend to be persistent [175, 149], but weaker alters leave and arrive at similar rates [142]. Most importantly, the *rankings* of alters with respect to contact volumes is persistent for egos. *Social signatures* capture the way egos distribute their contacts across alters, and they tend to be persistent in both time and across communication channels [80]. More recently, Íñiguez *et al.* proposed that this ego-specific but broad distribution of tie strengths is related to the way egos manage balance old ties through cumulative advantage, and the arrival of new alters [93]. This balance between low and high alter turnover characterises the communication strategies of egos. Some people are *keepers* and have a higher-share of stable connections and lower degrees, while *explorers* tend to have a higher number of connections, albeit weaker and more transient [144, 141].

Contributions

Our contributions in Publication I identify temporal patterns in communication that capture tie strength in the Granovetter sense –social behaviour where strong ties are *temporally* embedded within overlapping circles of friends. Our work is largely practical in the sense that it deals with the question reconstructing communication networks. This way, while the total number of contacts of a tie is commonly used as a proxy for tie strength (see Sec. 4.3), our results highlight its multifaceted nature. While tie strength is indeed related to communication intensity, it is also affected by bursty behaviour and by the temporal stability of communication patterns. We also investigated the use of social times during the week, showing that weekend contacts are good predictors of tie strength.

In Publication III, we analyse whether major elements of dyadic communication are also present in historical letters, and we compare epistolary contact networks to different communication channels, such as phone calls, messages, emails and social media platforms. Although datasets of historical letters lack the high granularity of contemporary auto-recorded data, their large temporal scales offer unprecedented insights into communication patterns, such as decades-long tie decay and persistent social signatures.

In Publication V, we expand on the the effect of social times during the week on network topology. In other words, how natural day-night cycles and social work-leisure cycles can reflect topologically on networks. We propose a method for inferring latent social times during the week from population-wide activity patterns, and use these social times as a basis for characterising ties. Using the latent signals of population-level activity, we classify each tie's weekly activity in terms of how they use combinations of the social times. Our core idea is that ties that are active during, e.g., working hours, might differ from ties that are active during most social times. We reconstruct networks that capture specific social times, and introduce a measure of *temporal multiplexity*, which serves as a temporal analogue of S. Feld's focal multiplexity [59]. Our results show that social time usage is tie-specific and largely resilient to bursty dynamics, with temporal multiplexity affecting local and global connectivity. We also find that social times can behave as foci around egos under particular conditions, and that weekends tend to over-represent overlapping circles of friends [81].

2.3.2 Social behaviour as a mechanism of tie creation

The second line of research of this thesis focuses on modelling and analyzing mechanisms of tie creation. Triangles, tie strength, homophily, among other social structures, do not only describe observed patterns in networks (Sec. 2.2.1) but can also be thought of as mechanisms that drive the creation of ties. Network evolution models allow us to simulate, understand and test minimal characterizations of social behaviour as drivers of connections [199, 158]. They offer plausible explanations for the emergence of large-scale network patterns through the combination of social behaviour and structural constraints in networks [186, 116, 199, 16], helping us understand drivers of inequality in social connections [155, 100], mechanisms that amplify polarization [147, 188], the coevolution of opinions and identities in politicians [56], or reveal strategies of human communication [93]. Understanding how different forms of social behaviour interact with networked structures can help make informed choices when designing policies [155, 100] or social media algorithms [1, 132].

Such models formalise either empirical or sociological observations, proving a framework for testing and analyzing theories of social behaviour. Early models of social behaviour include the seminal work of R. Axelrod, who proposed mechanisms for the dissemination of culture on a grid, where agents had a set of F features that could take q values. Here, dissemination of culture referred to social influence based on homophily: at each step, an active node and one of their neighbors were randomly selected. The active node would adopt the neighbor's value on a differing feature with probability proportional to the number of shared features. Repeating this process led to local convergence and global polarization: nodes tended to become similar to their neighbors locally, yet display polarization on a larger scale [9]. Just as Axelrod's polarised grid, many evolution models lead to a stationary distribution [186, 16, 199], where nodes and links keep changing but macroscopic behaviour becomes stagnant in some sense. Such static representations usually don't contain information on *how* they came to be [164, 139], e.g., knowing that a grid is polarised it is not always easy or feasible to know which mechanisms led to such polarised behaviour. In this sense, the strength of dynamical models is that they provide additional information on how mechanisms interact and lead to cumulative effects and emergent phenomena [15, 216, 199]. In turn, they can also provide different lenses from which to analyse macroscopic behaviour, both through mechanisms – e.g., preferential attachment– and through network topology – e.g., heavy-tailed or power law degree distributions– [89].

Many network evolution models focus on capturing particular networked phenomena, with some major classifications including either growing dynamics, as the BA model were new nodes arrive and make connections [199, 122, 158, 1], or different forms or rewiring dynamics where a link is replaced by another [209, 147, 195]. Other evolution frameworks include, e.g., the co-evolution of opinions and structure [56] or heterogeneous activity patterns for nodes [195]. Rivera *et al.* systematically reviewed social mechanisms of tie creation, which range from assortative mechanisms such as homophily or heterophily, triadic and focal closure, physical proximity in social spaces, dynamics of degree such as preferential attachment, or node fitness [173]. Recently, Overgoor *et al.* proposed a generalised framework for growing networks that could account for a wide array of growth mechanisms, including triadic closure [158]. Other known models of social evolution include the Weighted Social Network model, which focuses on triadic closure and dynamic weights, reproducing the Granovetter effect with bridging weak links and community structures [119, 148, 109]. This model was recently extended by Murase *et al.* using Axelrod's framework of multidimensional social identity [147]. They showed that multidimensional homophily has lower resolution than the number of homophilic features F could suggest. In other words, structural constraints for link creation meant that nodes tended to belong to a single community even if they could theoretically belong to F.

Introduced in 1978, the Schelling segregation model used an agent-based approach to show that mild homophilic preferences could be amplified on grids and lead to overall segregation [186]. More explicitly, for agents that changed spots on a grid if they didn't have at least B_a fraction of same-group neighbors, Schelling showed that $B_a = \frac{1}{3}$ sufficed to lead to segregated grids: the microscopic decisions of each agent could cumulative saturate sections of the grid with same-group agents. Kossinets and Watts, on the other hand, analysed empirical datasets of interactions in a university. They used emails, course registrations and individual attributes, such as gender, age and department, finding that that triadic and focal closure –with foci defined via shared courses– could create structural constraints and lead to induced homophily, or more segregated behaviour: people met others similar to themselves as their social contexts provided homophilous options to connect [116].

Contributions

Publication II is inspired by the Schelling model and the empirical results of Kossinets and Watts. Our core idea is that triadic closure can induce more homophily as the people that you can meet through your friends will tend to contain an outsized fraction of same-group options. We proposed a network evolution model with tunable group homophilies, group sizes and amounts of triadic closure, distinguishing between choice homophily – how likely someone will connect with another based on their group –, and observed homophily, or the actual number of same-group connections. Using a combination of simulations and theoretical analysis (see Sec. 3.4 and 5.1), we show how triadic closure can induce additional homophily in social networks, estimating the effect of triadic closure on empirical social systems. What is more, we found that our model displayed two different regimes: homophily amplification occurred initially, but in the long term one group could dominate connections and become a core, while the other could become a periphery mostly connected to the core.

We explicitly focused on the emergence of such core-periphery structures in Publication IV. Our rationale was that triadic closure served as an implicit form of preferential attachment: finding friends of friends tends to lead to higher-degree nodes, as we saw with the friendship paradox [60]. We show how the combination of choice homophily and explicit preferential attachment can lead to core-periphery structures, and particularly how the different processes of growth and rewiring can lead to different types of core-periphery structures [64]. We also analyse empirical networks of core-periphery, estimating the amounts of choice homophily and preferential attachment in real-world systems, and we analyse the effect of an affirmative action policy in Norway.

Publications II and IV provide a framework for understanding local and global networked mechanisms, with direct applications on social policy and social media design. Suggesting friends based on mutual connections, sharing other's messages or favouring popular content are common features of social media platforms. Our contributions shed light on how such design choices can interact with social behaviour and favour the appearance of both homophilous groups and highly-connected groups of elites.

2.4 Triadic tensions within social network analysis

The threads underlying the works of this thesis concern the inference of social behaviour in networks. All models rely on assumptions [32]; however, combining sociology, mathematics, and empirical phenomena raises particular epistemological questions. We frame them in terms of *triadic tensions* as a means to address how each pair of elements carry different implications.

Sociology and social phenomena

Interpersonal relationships or social identities are intuitive since they constitute significant aspects of our daily lives. However, relationships and identities manifest through contextualised social practices, and are themselves devoid of any measurable physical form [190, 212, 127, 32]. We may observe contacts or survey data, but modelling interpersonal relationships carries assumptions about their ontological status [190, 32]; in other words, to measure tie strength, social roles, or group identities, we first assume that such concepts are valid and they are captured by data [120, 32, 161]. In social science, this means that many sociological concepts are not fixed – part of a reason why ties, identity and culture can capture vast definitions [212, 127]. In practice, we usually measure sociological concepts through proxies, or ground them in sociological theories [215, 128]. This is the approach we use in Publications I and V where we evaluate behavioural proxies for tie strength under the framework of Granovetter's work or Feld's theory of social foci.

These notions also affect the definitions of social systems. Laumann, Marsden and Prensky defined the *boundary specification problem* as a crucial concern when working with social data, as the validity of an analysis relies on the alignment between the data and the social system being studied [120, 90]. They proposed a typology for defining such boundaries, where a *realist* approach to social network analysis centers on perceived membership, encompassing datasets that focus on entities like companies, schools, or nuclear families. Conversely, the *nominalist* approach involves researchers specifying boundaries according to an analytical framework, which may include criteria such as social elites, demographic features, or, for instance, phone calls within a country. The boundary identification problem affects, for example, the persistence of social behaviour across communication mediums in Publication III, or the core and periphery partitions of Publication IV.

Mathematics and sociology

While we may not be able to observe sociological concepts directly or objectively, they arguably capture aspects about human behaviour with tangible effects [171, 193, 191]. A different matter is how to represent them using mathematical language, as many sociological concepts can't be represented in mathematical terms without imposing additional assumptions. [193, 128, 74, 32]. The way we mathematically express a social identity directly impacts the insights that we can gain about such identities [32]. As we'll discuss in 3.3.2, mathematical assumptions are sometimes the only way we can gain insights into sociological concepts, such as when untangling homophily and triadic closure [189, 210, 164]. Although mathematising sociological concepts places constraints, such formal language also allows to test them [208, 162].

A corollary to this problem is not epistemological but interpretative, as the same sociological notions may have contextual mathematical expressions. We'll explore this in Chapter 5 with homophily, core-periphery and triadic closure. Although these notions could seem straightforward, different definitions can lead to seemingly confounding results. For example, Abebe *et al.* found that the interplay of triadic closure and homophily could lead to more inter-group connections [1], which could seem contradictory to our results of Publication II, although their definitions of both homophily and triadic closure differ, with the latter being a form of second-degree search [94].

Applied mathematics and phenomena

The last epistemological question is broadly related to how applied mathematics or the physical sciences relate to natural phenomena. We'll motivate this with the BA model of preferential attachment, which leads to and power-law degree distributions when the process continues infinitely long [16, 3, 152, 89]. This last part is not trivial: infinity is mathematically well-defined in a variety of domains, which include convergence of distributions or when defining infinitely small changes [208, 53, 130]. Infinity has myriads of applications, yet any real-world system is finite in both the data it can produce, and its stability. In the end, the question of whether a theoretical distribution exists in nature or not does not have a straightforward answer [87, 194, 28].

Our work on inferring social behaviour in social systems is inevitably constrained by how we define sociality, our data sources, and the methods we use. In many cases, there are no ground truths [161, 120, 215], and our best bet is to be explicit about the underlying sociological or mathematical frameworks and their assumptions [32, 162]. Despite these challenges, mathematics does provide a robust framework for testing many of these assumptions and quantifying many sources of uncertainty –the entire fields of probability and statistics are devoted to this [208, 68, 162]. We can test scenarios, hypotheses, and quantify uncertainties induced by models or present in the data, which provides us with an enormous toolset to gain insights from social systems. In the following Chapter 3, we introduce some of the frameworks, models and methods we use to represent systems of interacting components, as well as principled ways to gain insights from them through inference.

3. Network-scientific methods in social science

3.1 Static networks

A graph is a pair $G = (V, E)$ where $V = \{v_1, v_2, \ldots, v_N\}$ contains N nodes, and E is a set of unordered pairs (v_i, v_j) where $v_i, v_j \in V$ [152]. The elements of V do not usually have an order, but for simplicity, we can refer to them using natural numbers $V = \{1, 2, \ldots, N\}$. Graphs can be represented in matrix form: A is called the adjacency matrix of graph G if it's a square matrix of size $N \times N$, where element (i, j) takes value one if it's an edge $(i, j) \in E$ and zero otherwise.

Graphs can be generalised in a myriad of ways. We cover only some of these ways here, but introductory materials and comprehensive reviews can be found in [152] and [112]. A graph is called *directed* when the order of the interactions matters, i.e., if $(i, j) \neq (j, i)$. A *weighted* graph assigns numerical values to the edges, which are now of the form (i, j, w). In social networks, the weight can represent the *tie strength* [156, 185], but also the probability that an interaction occurs [88]. Nodes can also have associated values called *attributes*. These can represent, e.g., social groups or features such as gender or age [152, 24, 199]. If the attributes are discrete enough, groups can be visualised in the adjacency matrix as "blocks" [122, 152] –a notion that will become important further when we discuss group-level measures in Sec. 3.1, or probabilistic models such as stochastic block models (SBMs) in Sec. 2.3.

Multilayer networks generalise simple graphs by introducing additional structures called *layers* [111]. Layers can be useful for representing social roles if, e.g., all the edges in a layer represent colleagues; but layers may also capture a snapshot in time, such as the calls placed during a month [185, 87, 130]. Multilayer networks are general enough to account for, e.g., both social roles in *layers* and time in different structures called *aspects*, so that edges may exist within layers, between layers or connect any combination of nodes, layers and aspects [112]. In the context of the present work, where we mainly use multiplex networks without aspects (Publication V), we use the terms multilayer and multiplex interchangeably, and define them as $G = (V, E, L)$, where $L = (l_1, \ldots, l_J)$ are the J layers. In our framework, edges connect nodes only within layers, i.e., they are of the form (i, j, l, w), where i, j are nodes, l is a layer, and w is the link weight in the layer [111]. Other generalizations of networks include higher-order graphs, where edges are not only pairwise, but may connect an arbitrary number of nodes. The articles included in this thesis do not include higher-order frameworks, but more information on this active research area can be found in [19].

Measures of static networks

A major concept within the social sciences is that of local density: the alters around an ego are also connected with each other, and people usually form triads [79, 76, 59]. In simple graphs, this idea is captured by a node's clustering coefficient [209, 152, 24], which measures the ratio of closed triangles around that node out of all the possible pairwise combinations of alters. For e_i closed triangles around a node of degree k_i, the node's clustering coefficient is $cc(i) = \frac{2e_i}{k_i(k_i-1)}$, as there are $\frac{k_i(k_i-1)}{2}$ possible closed triangles [209]. If the clustering coefficient is 1, then all the neighbors of a node are also connected among themselves; if the value is zero, then none of the neighbors have any connections. Topological overlap captures tie-level clustering [156, 149], and is usually associated with local "bridginess" , or whether the two nodes in a tie can reach each other through common friends. Just as with the clustering coefficient, overlap measures the number of closed triangles out of all the possible triangles, or the common friends out of all the set of friends [156, 149]. Given two connected nodes of degree k_i and k_j, the number of possible triangles around a link is simply $(k_i - 1) + (k_j - 1) - n_{ij}$. For n_{ij} neighbors of both i and j (that is, the number of closed triangles), the overlap of a tie is defined as $O_{i,j} = \frac{n_{ij}}{k_i+k_j-2-n_{ij}}$ [156]. Bridginess can be extended to larger ranges [159], and overlap can be extended to account for both weight and topology [133].

Measures of centrality capture the *importance* of nodes in a network [152, 89]. Such importance is relative to what one wishes to measure, such as dissemination of information [36] or tolerance to attacks [3]. A node's degree k —the count of vertices that are incident to the node— is also referred to as degree centrality [152]. From a fully structural perspective, a node's degree might not be reflective of its actual reach [3, 152].

Just as weak links can connect different parts of a network [76, 156], nodes with low clustering can also act as bridges [31]. More generally, the structural importance of both nodes and edges can be assessed through how they facilitate flows on a global scale, typically measured through paths and random walks [3, 152, 131]. Our work does not explicitly focus on information diffusion and global connectivity, yet we briefly mention these concepts as they are both foundational and ubiquitous in network science and stochastic processes. In a nutshell, a path is a sequence of distinct edges that connects two nodes, and shortest paths are those of minimal length [152, 55]. This way, closeness centrality measures the average length of shortest paths from a node to all others, while betweenness centrality measures the ratio of shortest paths that go through a node, over all the possible shortest paths [152]. Random walks are stochastic processes that navigate through a network by moving from one node to another based on probabilistic choices at each step [152, 131]. As an example, PageRank centrality –which was famously used by Google to rank websites–, is based on random walks in directed graphs [152].

Collections of nodes and edges can represent different structures. A *clique* is a subgraph where all nodes are connected [152], whereas a motif is a subset of nodes and edges that appear more frequently than expected under some random model [152, 117]. In social networks, triangles are perhaps the paragon of both cliques and motifs as they represent the principle of transitivity or triadic closure [83, 76]. Cliques and motifs are widely used in random graph models [176, 199], including when performing inference [77, 164]. Clustered groups of triangles can be seen as community structures to the extent that they may confound community detection methods specifically devised to avoid detecting spurious communities [103, 210, 164].

Larger structures such as communities do not have a unique definition in social networks —-or network theory in general [152, 103, 161]. Given a partition of nodes into groups, a community can refer to a group that is strongly connected among itself and weakly towards non-members [103, 152], although in real-life networks such partitions are not trivial to define [161]. In the ideal world of known groups, if N_a/N is the fraction of nodes in group a out of a total of N nodes; the group density ρ_{aa} can be defined in terms of observed connections L_{aa} out of all possible connections [122, 64], i.e.,

$$\rho_{aa} = \frac{2L_{aa}}{N_a(N_a - 1)}. \tag{3.1}$$

Following the terminology of Fortunato and Hric, a weak community exists if a group's density is larger than the inter-group density $\rho_{ab} = \frac{L_{ab}}{N_a(N-N_a)}$, where L_{ab} is the number of links between a and b [62].

Community detection and analysis is a subfield within network science [103, 161, 164]. We'll briefly review how to model communities in random networks such as SBMs in Sec. 3.3.2, but do not further cover methods for community detection or analysis as they are not pertinent to the contents of this work. Groups, however, can interact in other ways than communities. Core-periphery structures capture meso-scale dominance as a core group is highly-connected to itself, while a periphery is mostly connected to the core and sparsely to itself [26, 64, 113, 178]. Just as with communities, core-peripheries have been defined in a myriad of different ways, as "structural dominance" may not only refer to an observed pattern but to the methods used to detect them: core-peripheries can be based on blocks and density when groups are known [26, 64]; however, when detecting core-peripheries other definitions may include cliques [85], paths [49], or other structures [64]. In fact, core-peripheries can't be detected as two groups if the sets of nodes have not been pre-defined, as high-degree nodes can be trivially defined as the core [113]. We use the framework of Gallagher *et al.*, who formalised a typology based on the densities on pre-defined blocks [64]. For groups a and b with intra-group densities ρ_{aa} and ρ_{bb}, and inter-group density ρ_{ab}, a *hub-and-spoke* has ordered densities $\rho_{aa} > \rho_{ab} > \rho_{bb}$ with a as the core; in a *layered* network the periphery is as likely to connect to itself as to the core, so $\rho_{ab} = \rho_{bb}$. We use these definitions based on blocks to assess group interactions on Publications II and IV, in the latter case proposing two evolution models based on homophily that can lead to these two typologies of core-periphery.

3.2 From events to networks

Network reconstruction refers to the process of representing data as networks [185, 118, 163, 37]. When using event data, different approaches to handling temporality may result in different mathematical objects. For example, aggregated networks may explicitly discard temporality by representing a link exists if there is some contact [156, 185], snapshots and sliding windows incorporate some temporality by aggregating networks on smaller, sequential periods [89], while fully temporal networks represent individual *events* instead of edges [185, 90, 130]. Throughout the Publications of this thesis we use several different approaches to modelling temporal data. Largely, we focus on reconstructing *static* networks, yet we address social behaviour by analyzing temporal dynamics. We'll briefly cover different approaches to handling temporal data, and discuss how temporality affects static data in the following section (3.2.1).

Many methods include different forms of temporality and staticity [130, 87]. Temporal snapshots, for instance, are ordered series of static networks, where a typically long observation window is divided into smaller periods [185, 130, 87]. Temporal snapshots have been used for analyzing the persistence of social ties [149, 142] and signatures [184, 80], as well as for evaluating generative models in, e.g., link prediction [40, 207]. While temporal snapshots represent discrete instances, sliding windows represent continuous changes [87]. A sliding window reconstructs static networks within a smaller observation period Δt that moves through time. A link is considered active within the window as long as new events fall within Δt, and inactive otherwise.

Researchers may also capture overlapping temporal structures –such as the length of relationships – using a wide array of methods. Simple temporal criteria for the birth and death of links within an observation period can capture important social behaviour, such as the activity patterns of transient relationships [81]. Focusing on spreading dynamics, Holme and Liljeros showed that simple statistics such as the birth, death and contact volume of links could characterise major aspects of spreading dynamics [88].

Temporal network theory is a sub-field that has emerged in the last decades by incorporating different approaches centered on events rather than links [130, 90]. Shifting the focus from static links to events results in a rich portfolio of temporal structures [185, 87, 90], including contact cascades [105], temporally-correlated motifs [117], or heavy-tailed inter-contact distributions [112]. Event-based networks require re-defining concepts such as centrality in a way that includes time, which may also result in different research questions [90, 130]. The paradigmatic example is, perhaps, that in temporal networks paths need to be time-respecting. This means well-established concepts such as path-based centrality need to be reformulated as reachability, or the number of nodes that can be reached in time-respecting paths [90].

Last, network reconstruction may also include formal statistical models that focus on some dynamic or topological property. Examples include models of dynamic tie strength [67], probabilistic models of uncertainty over topological features such as node degrees [170, 37], or models that use temporal dynamics such as event correlations between edges to infer community structures [163], or capture oscillator networks [172].

Our publications include a wide array of approaches to dealing with temporal data. The publications included in Chapter 4 focus on communication networks where event temporality is reflected in the structure; in Chapter 5 we use sliding windows to capture the extent to which network evolution mechanisms are present in the data. In all of these cases, empirical temporal structures that range from the inter-contact times [112, 185] to the lifetime of a relationship [142, 81] might overlap with the observation period. In the next section, we'll briefly cover how temporality in the data can impact the network reconstruction process – notions of central importance for Publications I, III and V, where we reconstruct static networks from event data.

3.2.1 Representing communication events as static networks

A first approach to dealing with event data is to represent it as a static graph by creating a link if some conditions are met, such as a minimum number of contacts [156, 40], or reciprocity [142, 185]. Such preliminary restrictions may already affect the overall network structure. DeChoudhury *et al.* found that thresholds on contact frequency resulted in rather distinct structures [40]; more recently Chowdhary *et al.* focused on temporal patterns of reciprocity that capture distinct social behaviours, such on-to-one exchange and information broadcast [41].

The choice to represent contacts as a link is not independent of the observation period, as it amounts to taking dynamic contacts and representing them as an aggregated structure, which provides an overview of social connections but may obscure structures only found in time [123, 185, 90]. This does not mean that mapping temporal events onto an aggregate network is wrong in itself: such networks have provided core insights into the structure of social systems [156, 143, 3], and continue to have a myriad of real-world applications [8, 124]. Observation windows can be used to gain insights into the social systems that produce them. Krings, *et al.* found that major structures in phone call networks stabilise at around a month —although the network overall increases in size for larger aggregation periods [118]. Using 19 months of call data, Miritello *et al.* found dormant ties can be characterised by behavioural differences within a few months i.e., inactive (dormant) ties in the future can be identified within a short temporal timeframe using boundaries from observation windows [142, 144]. In a similar line, Vergara Hidd *et al.* found that stable relationships tend to contact each other early in the observation period [81] and during weekends [206].

In Publication I, we analyse and introduce temporal features constructed from observation windows that can serve as proxies for tie strength, with the theoretical backing that strong relations tend to be relatively stable over longer periods [81, 142, 149], whereas weaker ties have higher turnover [144, 184]. Recently, Roy *et al.* analysed turnover in close friendships, which occurs at around 3% per year [180].

3.3 Inference methods for network science

Inference allows us to gain insights about whether some property, mechanism, or theory helps explain some observed phenomenon [208, 78, 152, 130, 89, 162]. Inferential methods range from null models that reject explanations that are trivial in some sense [114, 66, 3, 113], to validating models designed to capture some phenomena of interest [44, 162, 163, 164]. We can illustrate these differences under small-world networks. Considering that short paths connecting most people are prevalent in empirical networks, and also characteristic in random ER graphs, we may ask if social networks are inherently random. Taking the ER model as our baseline, called the *null* model, we can assess whether other network properties can be attributed to independently-placed links. The high clustering coefficients of empirical networks are extremely unlikely in random graphs – which helps us elucidate that other mechanisms drive the high clustering coefficients, and not random link assignments [209, 3, 114]. Formal network models can propose different mechanisms that either explain or replicate the empirical clustering coefficients [209, 162]. Watts and Strogatz famously proposed a model that rewired links from a highly clustered ring lattice, showing that a small rewiring probability preserved the local density and global connectivity of small-world networks [209]. Other models can account for clustering by modelling hierarchies [44], different mechanisms of triangle creation [210], or include combinations of social behaviour like triadic closure and homophily [164].

We use several inference frameworks throughout the publications of this thesis, which include various null models, predictive tasks or formal generative models. In Sec. 3.3.1, we discuss null models and predictive tasks which we use in Publications I, III and V. In Sec. 3.3.2 we briefly introduce generative and statistical network models, which we use in Publications II, IV and V, and describe inference methods from generative models in Sec. 3.3.4.

3.3.1 Inference methods without generative models

As we illustrated with the example of clustering coefficients in random graphs, in many cases the goal of null models is to show that it *does not* explain the phenomena, as they are regarded as simplistic in some way [208, 114, 152, 89]. Null models do not necessarily require an analytical form, as we can also assess whether a property is resilient to cleverly designed randomization processes [3, 114, 89, 66]. For example, instead of using the analytical form of the ER model, one can randomise the locations of edges or randomise edges while maintaining node degrees –a process that can be analytically described using the configuration model [151, 3, 152, 113]. Randomization is particularly fruitful for temporal data [185, 130, 89], which displays rich temporal-topological correlations. The state-of-the-art is perhaps the work of Gauvin *et. al.*, who recently proposed a generalised randomization framework for temporal data, including equivalent forms of randomization and the structures they preserve [66]. We use different forms of randomization in Publication V to show that temporal multiplexity (the use of social times) can’t be trivially attributed to high contact volumes, or that social times are transitive in the sense that triads use them at rates that are higher than expected under randomised conditions. In Publication III we use randomization methods to show that social signatures are persistent in epistolar datasets, showing that such social behaviour is also present in historical contexts and persistent over different observation periods.

Other inferential frameworks that don’t explicitly account for generative models include predictive tasks [208, 78], percolation methods [3, 114] or stochastic processes on graphs, such as random walks or spreading phenomena – for the latter, we refer the reader to the works of Masuda *et al.*[131, 130] or Newman [152], as we do not cover them in this thesis. Predictive inference refers to the process of associating one or more features known as independent variables $\mathbf{x}$ with another dependent variable $\mathbf{y}$ [208, 78]. It can be performed using statistical machine learning techniques that do involve a formal model $f(\mathbf{y}|\mathbf{x}, \theta)$, with a functional relationship f that captures general mathematical association instead of an explicit mechanism. For example, in linear regression, the features are related via weighted sum $f(\mathbf{y}|\mathbf{x}, \theta) = \sum_i \theta_i x_i$ [78]. Separately, Navarro *et al.* and Raeder *et al.* used predictive inference to find temporal features associated with tie stability [149, 169]. In Publication I, we use predictive inference under a range of machine learning methods to evaluate which tie-level communication features can serve as proxies for tie strength. Assuming that Granovetter's weak tie hypothesis is correct, we define strong ties as those that have overlapping circles of friends around them, and measure overlap in a temporal manner that includes full communication around ties (see 4.1)

Percolation is a method that assesses global network connectivity when links or nodes are removed or added [3], that can be used along with null models to perform inference on network connectivity. The emergence of a giant connected component (GCC) in the ER model (Sec. 2.2.2) can be analysed through a lens of percolation, where edges are added by increasing their probability p [3, 152]. Percolation can link local social behaviour to global network stability. Analyzing phone call networks, Onnela *et al.* showed that low-contact links behaved as weak ties in the Granovetter sense, i.e., networks became disconnected faster when removing low- than high-contact ties [156]. We extend this result in Publication V by showing that ties that are temporally monoplex –or active during a single social time during the week– are more important for network stability than low-contact ties. In other words, social behaviour expressed through the use of different societal times has structural effects on global connectivity, see Sec. 4.3.

3.3.2 Random network models

Network models $G(\theta)$ capture classes of networks that are random in some way but structured in another [3, 152, 114]. This means that they can serve as null models that account for some trivial behaviour [113], as baselines for generating data with some known property [210], or as theoretical explanations for observed patterns [162, 163, 164, 93], which is the approach we take in Publications II and IV. Common network models include the ER model that has a fixed number of nodes and link probabilities but random edge locations [55, 3, 152], the configuration model that has a fixed degree sequence and random link placement [151, 152, 130], stochastic block models (SBMs) that have groups that behave as block-level ER models [122], and exponential random graphs (ERGs) are a general framework to fix network features such as triangles [176]. Network models can also capture stochastic evolution processes such as preferential attachment [16, 52, 167], triadic closure in growing [94, 1] or rewiring scenarios [147, 109]. Naturally, many of these models can be extended or combined [103, 199, 122, 216, 152, 158].

We denote random network models as $G(\theta)$, where both mechanistic and stochastic constraints are induced by the model structure and parameters θ [152, 114]. Some network models place constraints on the connection probabilities between two nodes p_{ij} –whether directly as the ER model where $p_{ij} = p$, or indirectly as the configuration model, where a degree sequence $\{k_1, \dots, k_n\}$ induces correlations in the connection probabilities $p_{ij} \propto k_i k_j$ [151]. SBMs, which model groups defined as blocks, assign connection probabilities based on group memberships, so that $p_{ij} = p_{ab}$ if links i and j belong to groups a and b [103, 122]. Under SBMs, the expected group densities of Eq. 3.1 are the connection probabilities, i.e., $E[\rho_{ab}] = p_{ab}$, so that SBMs can serve to model homophilous communities [122], or core-periphery structures when one group is more likely to connect to the other than to itself [64].

Other group-level structures include the hierarchical model of Clauset *et al.* where connection probabilities are induced by a dendorgram –a tree-like framework where small groups can merge into larger groups in multi-scale mixing patterns [44]. Extensions of SBMs include heterogeneous degrees [103], or two-step generative models where, given a realization of homophilous blocks, triangles are created by connecting neighbors around egos [164].

By establishing a functional relationship between network features, such as connection probabilities between groups, random network models can help untangle a wide array of structures, social or otherwise [162, 87]. Such functional forms place assumptions on the underlying network mechanisms, and can thus induce biases by themselves [32, 25, 193], yet they can also be necessary for obtaining meaningful theoretical representations [162, 208]. Chang *et al.* showed that it is not possible to obtain principled estimates of network density from a single observation without assumptions on the error rates [37]. In general, it's also not possible to untangle homophily and triadic closure without some modelling assumptions [189, 210, 164]. Having explicit mechanistic representations, however, can help us untangle different effects that we can then evaluate under the lenses of empirical observations or competing models [162].

3.3.3 Modelling a sample evolution step

Although not always used in network contexts, we'd like to mention two models that we use later in this dissertation: the multinomial distribution and Markov processes. The multinomial distribution helps count the occurrences of successes across k categories when performing n independent trials. [208, 68]. For example, it counts the number of times each side of a dice $k = 6$ faces upward after n rolls. The parameters of the multinomial are the number of trials n, and a vector of probabilities of success for all categories $p = (p_1, \ldots, p_k)$, where $\sum_i p_i = 1$. If $X \sim$ Multinomial (n, p), the probability function is

$$f_X(x_1, \ldots, x_k | n, \{p_i\}) = \frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \times \ldots \times p_k^{x_k}. \tag{3.2}$$

We use the multinomial distribution in Publications V and IV in both static and dynamic contexts. In the former, we classify calling patterns according to social times by parametrising the probability vector as a tie-specific combination of population activity during the week, see 4.2. In the latter, we use the multinomial to fit network evolution mechanisms between interacting groups.

Markov processes refer to a class of stochastic models that capture memoryless temporal correlations [208, 69, 130]. This means that if G_t is an evolving network, we state the probability distribution of the graph at time t in terms of how it can evolve from the previous time-step $t-1$, independently of initial steps or the overall trajectory the network followed [1]. In Markov processes we describe transition probabilities in the form $P(G_t | G_{t-1} = g_{t-1})$, where g_{t-1} is the non-stochastic status of the network at time $t-1$ [208]. We can use Markov processes to model network evolution. These models may converge to a stationary state [3, 199, 158], which means that we can model both the evolution dynamics and the stationary distributions –fitting a preferential attachment kernel that drives dynamics [167] differs from fitting a power-law degree distribution [45, 194, 87]. We largely focus on modelling dynamics as a way of untangling social behaviour –mechanisms of homophily, triadic closure or preferential attachment– from observed structures in networks (see Sec. 5.4).

Following the framework of Pham *et al.* [167], we showcase how the multinomial distribution and Markov processes can model network transitions under preferential attachment [16, 52]. At each time step of the BA model, a new node arrives and chooses m nodes from the existing network to which to connect. Existing nodes are selected with probability proportional to their degree. So, for g_t the observed graph at time t, we choose node i of degree $k_i(t)$ with probability $\frac{k_i(t)}{\sum_j k_j(t)}$. Assuming that we don't care which specific node gains connections, but that the degree itself is what matters, for $n_k(t)$ nodes of degree k, the probability of choosing *some* node of degree k is $p_k(t) = \frac{k n_k(t)}{\sum_j j n_j(t)}$. The Markov property implies that g_t suffices to specify the probability distribution of the graph at time G_{t+1}, i.e., most of the structure is already given, and the focus is on how the structure changes

[1] We'd like to remark that many network dynamics are Non-Markovian, such as heavy-tailed inter-contact times [14, 112] or correlated bursty cascades [104]. An overview of Non-Markovian temporal network processes can be found in [130].

on a single time step, which is

$$P(G_{t+1}|\theta, G_t = g_t) \sim \text{Multinomial}\left(m, (p_k(t))\right), \tag{3.3}$$

for the observed degree sequence $\{p_k(t)\}$ in g_t. For more general attachment kernels, i.e., factors that impact the probabilities $p_k(t)$, see [167]. In Publication IV, we extend this model by including combinations of preferential attachment, homophily, and random rewiring affected by group, see Sec. 5.4.

3.3.4 Statistical inference with generative models

Statistical inference involves the formal process of evaluating whether a probabilistic model $G(\theta)$ accurately represents data for certain parameter values $\hat{\theta}$ [208, 78, 68]. Inference encompasses both determining $\hat{\theta}$ through parameter estimation, and hypothesis testing where the goal is to compare which parameter values θ_0 or θ_1 best represents data [208, 114]. Null models are a form of hypothesis tests where one of the parameters values doesn't capture any meaningful information. However, hypothesis tests can serve the more general purpose of model selection, or finding the $\hat{\theta}$ value that best represents the data. [208, 69].

The likelihood function is the central structure through which data and probability distributions interact [208, 68, 130]. The term *likelihood* refers to the inherent uncertainty about which parameter values of a random model could produce the observed data. For a collection of observations $x = \{x_i\}$, the likelihood function is the joint probability distribution of all the observations x [208, 68]. This can be easily exemplified when the observations are independent. If each observation x is described by a probability function $f_X|\theta$ the likelihood is the product of individual likelihoods $L(\theta) = \prod_i f_X(x_i|\theta)$. The *maximum likelihood* (ML) estimator refers to the parameter $\hat{\theta}_{ML}$ that optimises the likelihood function, i.e., the value that is more likely to produce the observed data under a model [208, 68].

A simple example of statistical inference is to assess whether a coin is fair when we have gotten 7 "heads" out of 10 trials. The binomial distribution $X \sim \text{Bin}(n, p)$ counts the number of successes when performing $n = 10$ trials, where each one has probability of success p [208]. In this case, the ML estimator corresponds to the fraction of successes out of trials $\hat{p}_{ML} = 7/10 = 0.7$, which could suggest that our coin is not fair. The strength of using statistical methods, however, is that we can compare different scenarios and quantify their differences. Under a null hypothesis of fair conditions $p_0 = 0.5$, we can get 7 or more "heads" with probability $P(X \geq 7|p_0 = 0.5) = 0.17$. Instead of explicitly stating whether our coin is fair, null hypotheses can serve as a baseline for how uncommon our observations are under controlled conditions.

Computational methods for inference

The two major schools of thought within probability and statistics are frequentist and Bayesian [208]. Both frameworks use similar probability distributions and functional forms; however, Bayesian statistics models uncertainty on model parameters as well [208, 68]. We'll focus on some practical implications when performing inference, but for a thorough discussion these different approaches see Gelman [68]. In broad terms, frequentist ML estimators obtain point estimates for $\hat{\theta}_{ML}$ and assess intervals of possible parameter values around those points [208]. On the other hand, Bayesian methods account for uncertainty when inferring θ by using full probability distributions $P(\theta|g)$, called *posteriors* [68, 47]. Bayesian and frequentist approaches use different inference methods as their frameworks are stated in different terms: posterior distributions are functions $P(\theta|g)$ whereas point or interval estimators $\hat{\theta}_{ML}$ points in the domain of likelihood functions.

Given a frequentist network model $G(\theta)$ and an observation g, it may be possible to find the ML estimator $\hat{\theta}_{ML}$ analytically. However, in cases when it's not possible, several methods of numerical optimization can approximate the solution [208]. These methods approximate $\hat{\theta}_{ML}$ using information from the likelihood function, such as the gradient or Hessian [114, 208, 153]. For a comprehensive overview of numerical optimization methods, see Nocedal and Wright [153]. In Bayesian frameworks, it is also possible to find the posterior $P(\theta|g)$ analytically, but not in general. Methods such as Markov Chain Monte Carlo [69] obtain sequential samples from the posterior by assessing different θ values under different sampling strategies [47]. Variational Inference methods use Gaussians or other known probabilistic distributions to find bounds on the posterior [20]. Last, likelihood-free inference methods can *approximate* likelihood functions for models $G(\theta)$ that can be expressed as simulators and where the likelihood function is somewhat difficult to construct [51, 77].

In Publications IV and V we propose statistical models for network evolution and for tie-level temporal multiplexity under frequentist frameworks. We derive their respective likelihood functions analytically and obtain ML estimators using methods of numerical optimization. In Publication II, however, we propose a network evolution model for triadic closure and choice homophily and fit to data using likelihood-free inference. We found this approach to be particularly suitable since having an explicit formulation for the likelihood function could place additional constraints on both the data and our model, which was stated in terms of simulation rules. We discuss our model and fitting in more detail on Sec. 5.4.1, but for now briefly cover likelihood-free inference.

Approximate Bayesian Computation (ABC) methods approximate the *likelihood* that a parameter produced data by measuring the similarity between observed g and simulated networks $\hat{g}$ [51]. Given an evolution model $G(\theta)$ that produces sequences of graphs for known θ, ABC methods compare samples of parameter values $\{\theta_i\}$, and accept or reject them depending on similarities between the simulated $\hat{g}_i$ and observed g networks [77]. These methods gauge similarity using summary statistics $S(g)$ of the graphs, and a distance function $\rho(S(g), S(\hat{g}))$ between the statistics of the observed g and simulated $\hat{g}$ graphs. Researchers play a role in this process as they specify the particular network structures expected to encapsulate the parameter effects in the model. Summary statistics can be, e.g., triangles, intra- and inter-group motifs or meso- and macro-scale structures. ABC methods yield unbiased approximations of the posteriors if the statistics S are sufficient, i.e., if they fully capture the effect of the parameter θ, although sufficiency is hard to determine in lack of likelihood functions [51].

3.4 Analyzing models of network dynamics

Network evolution models are useful as they help us untangle how dynamic processes, such as homophilous social behaviour, can cumulatively lead to emergent phenomena [3, 158, 98, 216]. The analysis of dynamic network processes have largely borrowed the toolset of statistical physics, which uses probabilistic modelling of microscopic interactions between components to understand the macroscopic behaviour [53]. Techniques like mean-field equations (MFEs) offer insights into the long-term effects of microscopic interactions by relying on certain key assumptions.

First, they treat particular sets of elements as equivalent [53, 3]. For example, instead of modelling each node individually, we model nodes of a certain degree, or nodes that belong to a group [199, 158]. The step-wise network evolution with preferential attachment of Eq. 3.3 accounted for this by modelling the probability that nodes of degree k will gain connections, as opposed to each node individually. These groupings can even be coarser: in our models of homophilic interactions of Publications II and IV, we focus on tracking the fraction of links within and between different groups. The second crucial assumption is to approximate changes of discrete network elements as changes of continuous averages [53, 152]. Instead of explicitly counting the number of nodes $n_k(t)$ of degree k at time t, we model the way our evolution rules affects $n_k(t)$ on average, in infinitesimal time steps. A way to assess the validity of these assumptions is to compare the evolution of our continuous approximations with simulated data using different parameter combinations.

We derive a sample rate equation for $\frac{dn_k(t)}{dt}$ the number of nodes of degree k. On average, $n_k(t)$ increases when nodes of degree $k-1$ gain connections, and decreases if its own nodes become of degree $k+1$. Just as in our multinomial model of m trials, nodes of degree k gain connections with probability $p_k(t) = \frac{kn_k(t)}{\sum_j jn_j(t)}$, with an expected $mp_k(t)$ added links. The rate of change corresponds to the difference between the expected gained and lost connections [2]

$$\frac{dn_k(t)}{dt} = mp_{k-1}(t) - mp_k(t) = m\frac{(k-1)n_{k-1}(t) - kn_k(t)}{\sum_j jn_j(t)}. \qquad (3.4)$$

MFEs track the averaged changes of system properties, and under some circumstances, can reach an equilibrium where the system dynamics are independent of time [53, 150, 3]. As an example, equation 3.4 can be integrated to derive a system of recurrent equations that correspond to a power-law degree distribution (see [3]). Overall, however, MFEs can both help us track changes and find stationary states. In Publication II, we show how the interaction of homophily and triadic closure initially leads to homophily amplification, but in the long run can lead to core-periphery structures. These different states can be visualised through a phase-space analysis, which tells the combinations of parameters that lead to one or more macroscopic effects, as well as their stability.

[2] Our model includes a simplified version. For complete derivation including alternative definitions see [3].

4. Social behaviour in communication networks

In this Chapter, we summarise some of our results where we assess how social behaviour defined from sociological perspectives can be incorporated into the reconstruction of static networks. We grounded Publication I on Granovetter's theory of the strength of weak ties [76], where our central question was to identify temporal features that could serve as proxies for tie strength. In Publication V we analysed the use of social times during the week, proposing a methodology for reconstructing multilayer networks of social times and introduced the concept of *temporal multiplexity*: the idea that some social roles are expressed through the usage of social times, so that a "work" tie will likely be active during working times, but a close relationship may be active during several social times. We framed our results in terms of Feld's theory of social foci [59], where we drew parallels between temporal and focal multiplexity. Last, in Publication III, we framed a historical dataset of epistolary metadata as a communication network, and assessed the extent to which communication patterns were also reflected in these incomplete footprints of historical communication. All these publications rely extensively on data. For this reason, we start this Chapter by describing some of the main features of the datasets we used, as well as potential sources of bias when analyzing them.

4.1 Data, preprocessing and methodological choices

Mobile dataset

The main dataset we used in Publications I and V is *Mobile*, a CDR dataset of phone call metadata from a European country during seven months in 2007. The dataset includes ~ 6.5 million nodes and ~ 26.4 million ties, which accounted for 20% of the country's market share at the time. We used a smaller version of this dataset for III. *Mobile* was provided by a particular communications operator, which could induce some biases when estimating triangles or other local structures: we can't ensure that ties are missing in a systematic manner. We account for potential biases when estimating overlap by using an extended version of the dataset that includes non-company users, i.e., people that were not clients of our operator but still contacted their users. This extended dataset thus provides complete communication ego networks of the company users, and includes ~ 76 million nodes and ~ 530 million ties.

Letters dataset
Publication III focuses on a large epistolary dataset called *Letters*, which contains metadata from $\sim 135K$ letters sent between 1510 and 1900, compiled by historians and digitised on large databases such as the Early Modern Letters Online catalogue [91]. Most of these letters were part of the so-called “Republic of Letters”, a community of mostly European scholars and ecclesiastical figures that exchanged knowledge, news and books [203, 202]. This dataset includes letters from historically-relevant thinkers such as Reneé Descartes and Galileo Galilei. In this sense, it’s a biased sample where historically prominent figures are over-represented, and where many letters are presumed lost [91]. We found that topological and large-scale features were hard to estimate, particularly as there were no known representation patters of individuals and letters, which would necessarily result in biased structures [37]. However, we evaluated whether epistolary communication replicated known patters at dyadic and ego levels.

Additional contact datasets
Throughout Publications I and III we use other datasets to compare communication patterns across different channels:

- *Forum* consists of thread postings on an online movie forum, where a contact represents a comment on a thread posted by another user over a period of seven years. There are $\sim 6K$ users and $\sim 138K$ edges.

- *Email*, called *Enron* in Publication I, consists of emails between employees of the Enron Corporation, with a particular focus on 150 users who were investigated as part of the company’s collapse [179]. There are $\sim 20K$ users and $\sim 297K$ edges, and the observation period spans 6 years.

- *Platform*, called *FB* in Publication I, consists of wall-postings on Facebook in New Orleans. Data was obtained by crawling public profiles, with links tracking one year of activity after an initial posting, resulting in an asynchronous network. There are $\sim 41K$ users and $\sim 184K$ edges.

- *Copenhagen* consists of text messages from the Copenhagen Network Study [183], which tracked the activity of around 700 university students during a period of four weeks.

These six datasets include a wide array of one-one-one communication sources. Following the realist-nominalist framework of the boundary specification problem [120], *Email*, and *Forum* adhere to the realist framework as they capture particular social foci: people working at a company, and people using a movie forum, respectively. In both cases, we don't necessarily expect ties to reflect interpersonal relationships such as friendships of family ties. *Mobile* and *Platform* follow a nominative approach. In *Mobile* we examine a subset of a large population whose only common denominator is their service provider, while in *Platform* the region and sampling strategies were determined by the researchers. *Copenhagen* and *Letters* follow a mixture of both. The former is a subset of a larger dataset that includes phone usage, online friendships and proximity data, thus capturing a rich source of ego-level behavioural data across communication channels. On the other hand, the actors of *Letters* were loosely self-organised, as the Republic of Letters included gatekeeping and brokerage [202]; however, it was also preserved by the efforts of historians based on availability and perceived importance [91].

4.2 Tie-level dynamics of communication data

Communication metadata captures behaviour at a high level of granularity, which means that it allows for a wide array of approaches to understanding temporal communication patters. Figure 4.1 depicts different ways to characterise contact patterns, highlighting ways in they differ for weak and strong ties: two people may spread their communication over long periods, favour intense *bursts* or communicate consistently for a relatively short period of time. In this section, we'll cover some of the major ways in which we can characterise such tie-level dynamics (i) as sequences of events that display bursty behaviour and are sensitive to observation periods, and (ii) as cyclical events that adhere to natural and societal rhythms [6, 4, 206].

4.2.1 Sequential dynamics of ties

A first approach to characterising the sequence of contact times $\{t_1, \ldots, t_w\}$ of a tie is to count them [156, 185, 21]. The total contact count w captures aggregate communication intensity, and was one of the first measures known to capture tie strength as defined by Granovetter: ties with higher w tend to contain more triangles around them [156, 184, 185]. Now, bursty communication patterns have an impact on contact counts: twenty calls might be concentrated in a day or spread over a month [105]. We can diminish the impact of burstiness by counting shorter activity periods, such

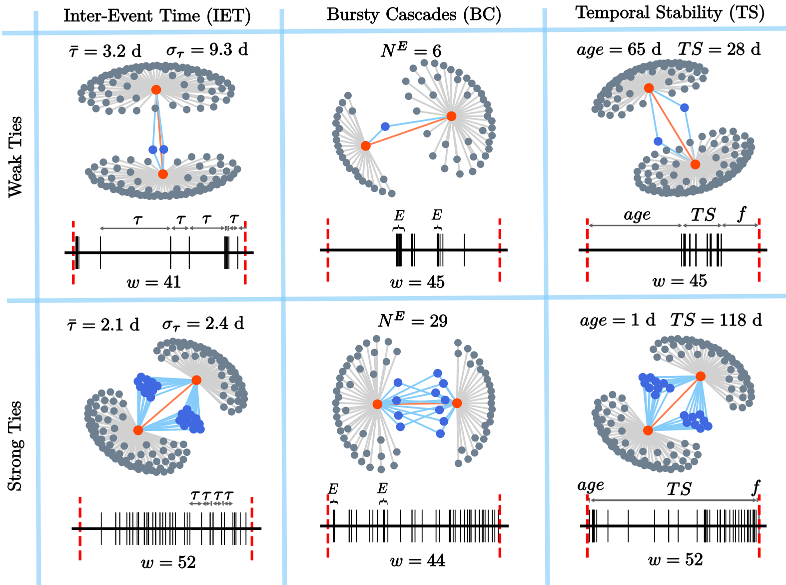


Figure 4.1. Communication patterns reflect latent tie strength through network topology. Weak ties that are topological bridges have different communication patterns from strong ties embedded within circles of friends. Columns represent measures of activity patterns. From I, licensed under CC 4.0.

as the number of active hours a_h or the number of active days a_d, features that not only capture contact intensity, but frequency of interaction [149].

Bursty dynamics
We characterise the time between contacts using the inter-event distribution (IET) [71, 204, 112], which captures the distribution of times between every pair of sequential contacts $\tau_k = t_{k+1} - t_k$ [71]. Burstiness is partially reflected on the IET distribution, as it is known to be heavy-tailed [33, 71, 112]. That is, most τ_k values are relatively small, but some are extremely large. The burstiness coefficient B helps assess these disparities by comparing how the mean IET $\bar{\tau}$ differs from the standard deviation σ_τ [71]. More formally, $B = \frac{\sigma_\tau - \bar{\tau}}{\sigma_\tau + \bar{\tau}}$ captures how IETs deviate from Poisson signals where events occur independently and the mean and standard deviation are the same [71, 204]. So, $B = 0$ for Poissonian signals, positive for bursty behaviour, and negative for regular contacts (which have low standard deviation σ_τ). Now, measures of the IET distribution are affected by the finite-size effects of the observation window [112]. These effects can partly be accounted for with statistical techniques, such as the Kaplan-Meier estimator used for survival analysis [112]. Our results show that bursty patterns and heavy IET tails are consistent even for epistolary communication, where inter-contact times tend to span weeks or years (results on Publication III).

Now, the IET distribution assumes that tie inter-contact times are independent, i.e., it's a method that discards temporal correlations across events. Such temporal correlations were accounted for by Karsai *et al.*, who defined bursty events as collections of contact sequences where each pair of contacts fall within a specified time window [105]. That is, we define a maximum inter-contact time Δt, and an event size E as the number of sequential contacts where any two consecutive calls fall within Δt. The distribution of event sizes $P(E)$ is also heavy-tailed for different definitions of Δt [105]. From bursty events we can derive metrics such as N^E the count of bursty events or the average event size $\bar{E}$.

Measures of temporal stability

Both the observation window of a dataset and the timescales of ties may range from weeks to several months or years. Measures of temporal stability attempt to characterise behavioural patterns that are consistent over fixed observation windows. Given the initial and final observation times T_{start} and T_{end}, we define a tie's age as the time between elapsed before the first contact $t_1 - T_{\text{start}}$ [142], a tie's temporal stability as the time when the time was active during the observation period, $TS = t_w - t_1$ [142] for t_w last observed contact, and the tie's freshness as the time between the last contact and the end of the observation period $\frac{T_{\text{end}} - t_w}{\bar{\tau}}$ [149], where we normalise to the average inter-event time. Miritello *et al.* used these measures to assess for link activation and dormancy in phone calls during long observation periods [142, 144].

In Publication I we introduced additional measures that account for biases in the distribution of contact events across the observation period. The core idea is in periods that range a few months it might not be trivial to assess whether ties are persistent or have decayed [142, 185, 149, 81]. However, we can compare whether communication is consistent across the observation period, or biased towards the beginning or end. We defined the average interaction time $\bar{t} = \frac{1}{w}\sum_i t_i^*$ as the average normalised timestamp value, i.e., we normalised each interaction timestamp t_i^* on $[0, 1]$ interval determined by the observation window. In turn, σ_t captured the spread of interactions across the observation period. We found that when interaction times were highly skewed towards limits of the observation window, ties tended to be topological bridges. The statistic for bias in interaction times $T = \frac{|\bar{t} - 0.5|}{\sigma_t \sqrt{N^E}}$ accounted for such imbalances.

Concepts such as temporal stability can be revealing about how links relate to a system and an observation period. Two of the datasets we examined contained links that we knew had decayed: the epistolary dataset *Letters*, and *Email*, which tracked the communication within a large company until its collapse [179]. Figure 4.2 depicts the temporal stability TS for five datasets, where each color represents quartiles of communication activity w. Both cases displayed a similar coupling between temporal stability and number of contacts, although on vastly different timescales. For low-contact

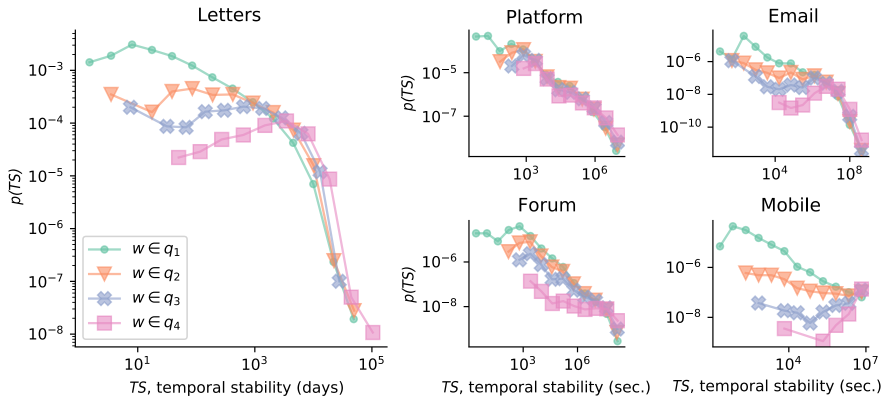


Figure 4.2. Temporal Stability for links within social systems. *Letters* and *Email* capture systems that end, and ties of all weights face sharp decay: ties of longer temporal stability become increasingly uncommon. Adapted from Publication III by removing additional subfigures.

links, TS tends to be short and consistently decreases until the distribution shows a sharp decay; for high–contact ties a short TS is less-common, consistently increasing until it also faces a sharp decay. While these results are merely descriptive, the qualitatively similarities across contact levels w could suggest that decay is induced by the systems own boundaries, or regions after which long ties become increasingly uncommon. For *Letters*, decay starts at around 10 years and lasts to up to 75 years. For *Email* decay starts at about less than a year, and continues until the end of the observation period of seven years, when the Enron Corporation collapsed [179]. In contrast, *Mobile* includes a similar initial pattern where low-contact ties tend to be shorter and high-contact ties tend to last longer; however, even transient relationships have been empirically found to last longer than our observation period of four months [81, 149]. In *Platform* all links lasted at most a year as per the collection methodology.

4.2.2 Time usage during the week

Human behaviour is also reflected in cyclical patterns following daily rhythms and weekly social times [97, 65, 6]. Daily rhythms have been useful in characterising differences in chronotype [5] and sleep health [4, 165]. Now, while chronotypes are associated with individuals instead of ties, previous work had suggested that similar time usage during the day can reflect homophilous patterns expressed through chronotypes, such as teenagers communicating with their peers [5]. Following the work of Aledavood *et al.* [6], we characterised daily activity patterns of node i via the fraction of contacts per hour $p_i = (p_i^0, \ldots, p_i^{23})$ over the whole observation period. For each tie, we defined the similarity of time usage via the Jensen-Shannon Divergence (JSD), a measure commonly used for comparing probability distributions [6]. This allowed us to measure whether two nodes behaved similarly during the day, potentially capturing underlying homophilic behaviour.

Our second approach focused on identifying weekly social times, such as whether links were largely active during times associated with work or leisure. We used two different methodologies for characterising weekly behaviour in Publications I and V. In both cases, we started by creating weekly activity profiles for each tie: the number of contacts that each tie placed during each of the $7 \times 24 = 168$ hours of the week. In Publication I, we obtained Ci clusters of correlated activity over the week, classifying each tie according to their profile of weekly activity C^* that contained the fraction of contacts placed during each cluster (details in I).

Latent social times from population-level data

In Publication V, we expanded this notion and introduced a methodology for reconstructing multilayer networks based on social times during the week. We used a matrix of full activity profiles X that captured population-wide time usage and applied Non-Negative Matrix Factorization (NNMF) [110, 4] with orthogonality constraints to represent activity in terms of J latent components. NNMF can approximate X as the product of two low-rank and non-negative matrices: one that associates the hours of the week to the latent social times H, while matrix G associates the social times to the activity of each tie. Figure 4.3 captures the overall network reconstruction process, including a visualization of the latent signals over the week (the columns of H). We focused on $J = 4$ signals which we named according to the time during the week that captured the most activity; a combination of daytime and evenings, and working days and weekends. It is these signals that we refer to as *social times*.

Using the matrix of latent social times H, we framed the reconstruction problem as one of statistical inference, as it would allow us to perform model selection on layers as opposed to using the linear approximations G of the original NNMF problem. We modeled the weekly activity profile X_i of tie i as a multinomial random variable,

$$X_i \sim \text{Multinomial}\left(w_i, h(\alpha_i)\right), \tag{4.1}$$

where w_i is the total contact count and $h(\alpha_i) = \hat{H}\alpha_i$ is a linear combination of the latent social times $\hat{H}$ [1]. Multinationals model the distribution of w_i trials over $k = 168$ categories with different probabilities. This way, coefficients α_i parameterise the *probability* that a tie is active at each hour, depending on how it balances different social times. We implemented a model selection procedure to obtain the combination of layers α_i that could best describe the usage of social times. We reconstructed multilayer

[1]The matrix of social times $\hat{H}$ is scaled to meet probability constraints.

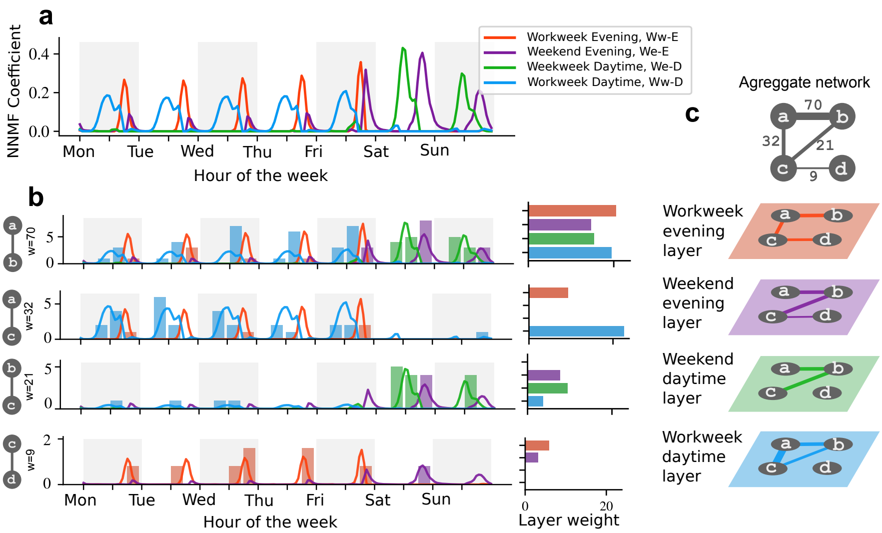


Figure 4.3. Multiplexity is temporally expressed through social times. **a** Latent activity patterns H correspond to social times. **b** The histograms of weekly activity of ties is modeled by combinations of latent social times αH. **c** We reconstruct networks where social times are represented as layers, and define a tie's temporal multiplexity through the usage of layers α. Figure from Publication V.

networks of temporal usage by defining the weight of layer l as $w_l = w\alpha_l$.

We defined *temporal multiplexity* as the entropy of the layer coefficients $S(\alpha_i) = -\sum_j^J \alpha_{ij} \log(\alpha_{ij})$. Entropy is an information-theoretic measure of uncertainty for probability distributions, where no entropy $S(\alpha_i) = 0$ implies monoplexity —one layer coefficient is 1 and the rest are zero. On the other hand, we refer to high-entropy ties as multiplex, as they are equally active during all social times.

The usage of social times is tie-specific

We tested the extent to which multiplexity could be induced by a large contact volume w or bursty behaviour. Initial descriptive statistics of the reconstructed network, depicted on Figure 4.4, suggested that this might be the case, which seems natural as ties with higher number of contacts tend to be present in more layers. We tested whether the balance of layer weights could be trivially explained by contact volumes by randomising contact times according to the calling patterns of the population. In other words, we developed a null model where a tie's weekly behaviour was determined by the number of contacts randomised over the population-level activity. Our results show that our randomised model consistently induced high multiplexity values for high-contact ties, while low-contact ties may have heterogeneous multiplexity values based on random chance.

However, removing the average effect of induced multiplexity given w showed that ties display rich time usage regardless of contact counts: ties

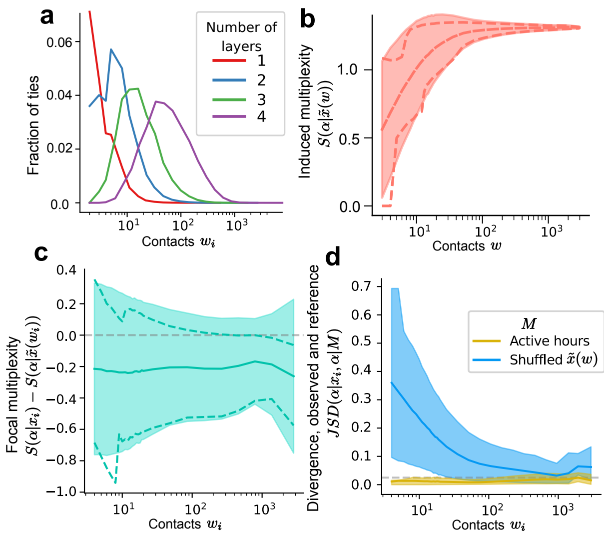


Figure 4.4. Ties favour their own social times. a The number of layers increases along with the contact counts w. **b** We characterise the multiplicity induced by contact counts using a null model of randomised contacts. **c** Focal multiplexity is consistently heterogeneous across contact counts, and **d** classifying coarse-grained activity shows resilience to bursty behaviour. From V.

have consistently heterogeneous multiplexity values that are also smaller than the multiplexity trivially induced by contact volume w. This means that ties tend to favour their own social times regardless of the induced effects, so that even ties with hundreds of calls can have low multiplexity values. Last, to account for the effect of bursty behaviour we performed inference on layer coefficients using coarse-grained weekly activity. That is, we transformed the weekly activity profiles X into binary profiles of active hours that only retain information on whether calls where placed at an hour of the week, but not the amount of calls. We found that the coarse-grained usage of social times was largely consistent with the full activity profiles across contact counts w. This strongly suggests that the usage of social times is robust to burstiness, in the sense that simply knowing that two people have called each other at some hour can give almost as much information as knowing how much they called at that hour.

4.3 The many faces of tie strength

The notion that tie strength was related to a topological role that had network-wide implications led to several efforts to characterise tie strength [128, 75, 127, 212]. Many such efforts were based on the definition proposed by Granovetter: a "possibly linear" combination of reciprocity, time, emotional intensity and intimacy [76], and largely relied on self-reported surveys that focused on finding proxies to measure such dimensions [215, 63]. Early on, Marsden and Campbell suggested that tie strength could be multidimensional [128], while Feld's characterization of ties in terms of social foci reinforced that notion, as it suggested that the topological patterns of clustering or bridging would be associated to combinations of different social spaces [59, 61].

The use of auto-recorded data such as CDRs and social media interactions led to a wide array of definitions and proxies for tie strength. Onnela *et al.* were among the first to characterise tie strength in terms of number of contacts w and assess its effect on global connectivity [156]. For social media, many efforts have helped characterise tie strength in terms of platform-specific mechanisms [99, 12]; a recent review of some of these methods can be found in [166]. These measures are not necessarily divorced from perceived tie strength, e.g., Wuchty and Uzzi found agreement between, e.g., contact counts and self-reported data [213].

In Publication I, we performed a systematic review of temporal communication patterns that could serve as proxies for tie strength, comparing different communication channels from different datasets. Instead of only using the aggregated overlap over the whole observation period, we defined the temporal overlap of a tie $\hat{O}^t$ as the average overlap over a sliding window of $\Delta T = 1$ month. This measure allowed us to capture temporally consistent triangles[2]. Our choice of Δt was motivated by previous research on alter-rotation rates [142, 141], and stability of network features [118].

We used four machine learning models to identify the features of temporal communication that best served as proxies for tie strength, using a range of cutoff values for temporal overlap that determined binary levels of weak-strong ties. We evaluated the prediction task using Matthews Correlation Coefficient (MCC), a metric that can handle imbalanced and asymmetric data in classification tasks [27].

We derived our main conclusions from the *Mobile* dataset as it contained full ego networks around our focal ties on a major communication channel, and was not constrained by particular social foci. Our results, on Figure 4.5, showed that several features captured topological information that outperformed the contact counts w in the predictive task. Notably, measures derived from contact counts that accounted for bursty behaviour tended to be the best-performing features: active days a_h and hours a_h and the number of bursty events N^E. These were followed by measures dependent on the observation window: temporal stability, the spread of interaction times σ_t, the tie age and the bias of interaction times $T = \frac{|\bar{t}-0.5|}{\sigma_t\sqrt{N^E}}$. Measures dependent on the observation period likely capture transient relationships [81], showing an association between the stability of a relationship and

[2] As a reference, we compared temporal overlap with the aggregate over the whole observation period, but found an overall stronger coupling with the temporal version.

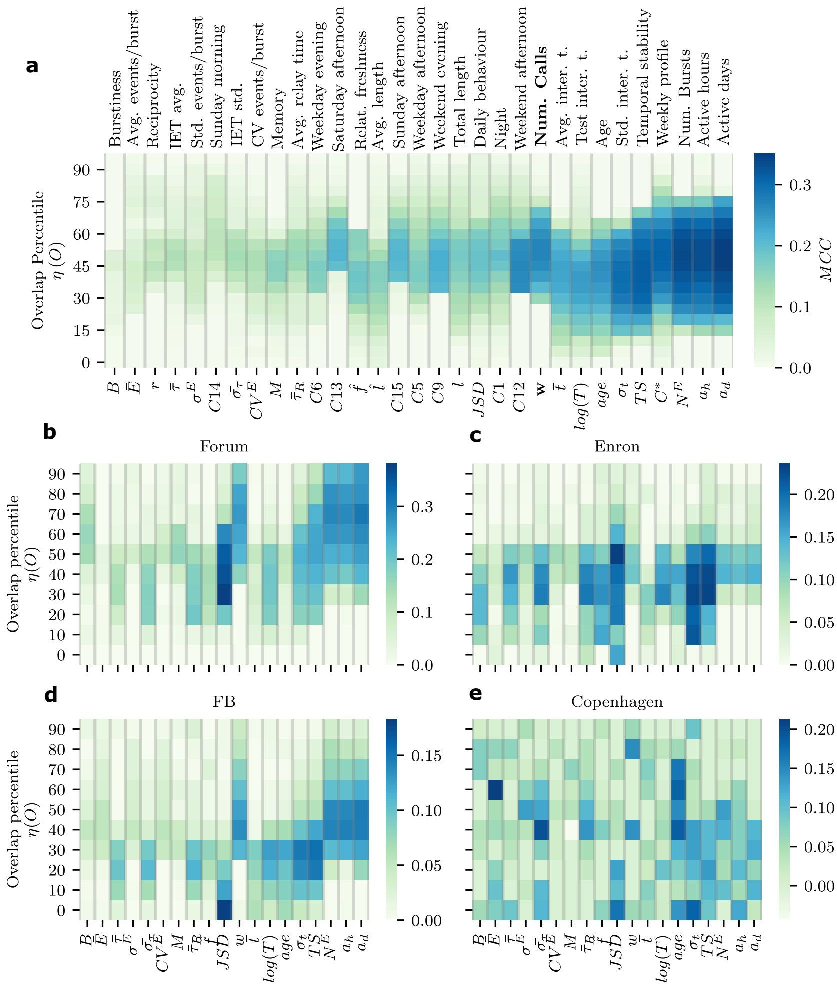


Figure 4.5. Temporal features of dyadic communication predict tie strength. We assess the predictive capacity of temporal communication features on five datasets (**a-e**). Intensity features that account for burstiness, and features of temporal stability tend to perform well, with results varying per dataset. Figures adapted from I by combining two figures and removing additional subfigures. Licensed under CC 4.0.

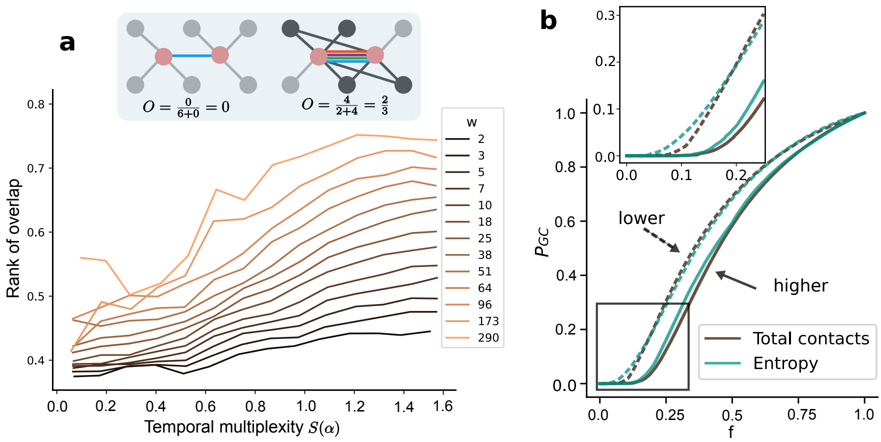


Figure 4.6. Temporally monoplex ties are topological bridges. **a** Overlap increases along with temporal multiplexity across activity levels w. **b** Monoplex ties are more important than low-contact ties for global network stability, leading to a GCC faster than w. Figure from Publication V.

broader social contexts such as overlapping circles of friends.

For the other datasets we found that the relative strength of the results varied according to the social system at study. Burstiness-adjusted contact counts were still good predictors of tie strength on social media sites. Overall, most datasets showed that measures dependent on the observation window captured tie strength, with temporal stability, age and skewness of interaction times showing different degrees of predictive capacity. Strikingly, for most of these datasets the similarity in daily contact patterns JSD showed high association with topological overlap, which suggests context-specific behaviour: similarity in daily patterns was particularly strong in social media sites and emails within a company. While JSD was not highly indicative of tie strength on mobile phone usage by itself, we found that it was highly informative of tie strength *when combined* with measures of communication intensity such as active days a_h and event counts N^E (results on Publication I). We found the weekly profile of time usage C^* to be highly predictive of tie strength –only behind the burstiness-adjusted contacts for the *Mobile* dataset. However, we also found that particular clusters were predictive of tie strength, such as weekend afternoons and evenings.

The strength of multiplex ties

In Publication V, we expand this notion by exploring the association between temporal multiplexity and tie strength from local and global perspectives, depicted on Figure 4.6. We framed our results in terms of Feld's focus theory, where we argued that despite the lack of explicit social foci, the use of social times could reflect underlying social foci. At a local level, our results confirmed that balanced time usage (higher entropy of layer coefficients

$S(\alpha)$) was associated with higher topological overlap across diverse contact levels w. Even extremely low-contact ties, from 2 to 5 contacts over seven months, have higher overlap if they display temporal multiplexity; while high-contact ties ($w \geq 120$) would be more likely to behave as bridges if they were temporally monoplex. These results also hold for global network connectivity. When performing a percolation analysis, where we add ties according to their contact volume or multiplexity, we found giant connected component (GCC) appears faster when adding monoplex ties than low-contact ties. This means that ties that are associated to either one or few social times are important for global connectivity than ties with few contacts at any time (results in Publication V).

4.4 Ego-centric time usage

Social signatures refer to the particular ways in which individuals divide their contacts among their alters [184]. For a person's sequence of outgoing contacts $\{w_j\}$ during an observation period, a social signature captures the fraction of contacts associated to the *ranks* of the alters. Usually, people tend to have a small fraction of high-contact ties and a wide array of lower-contact alters. Social signatures are revealing of a node's social behaviour as they capture the particular shapes of contact heterogeneity, which has been found to be stable across time and communication channels, even as alters do tend to change [184, 80].

Overall, we can test for the stability of social signatures using randomised null models. Given different temporal snapshots, we computed d_{self}, the JSD between the social signatures of an individual across three temporal snapshots, and compared them with d_{ref}, the social signatures of other individuals with a similar number of alters. We analysed the stability of social signatures on the *Letters* dataset for a wide array of snapshots sizes ranging from 1 to 10 years, thus analyzing the stability of social signatures over periods of 3 to 30 years. The stability of social signatures was consistent across window sizes and minimum number of out-going alters (details in III).

Close friends call at similar times

The individualised way in which egos use time is also reflected on the way they distribute their alters over social times. In Publication V, where we reconstructed multilayer networks of social times, we analysed the behaviour of egos across layers by focusing on how the alters were topologically distributed over social times. We framed usage of social time in terms of focus theory, which states that small and constraining social foci result in more triangles around an ego. Within this context, we argued that if an ego devoted a small fraction of their alters to that layer, the usage of

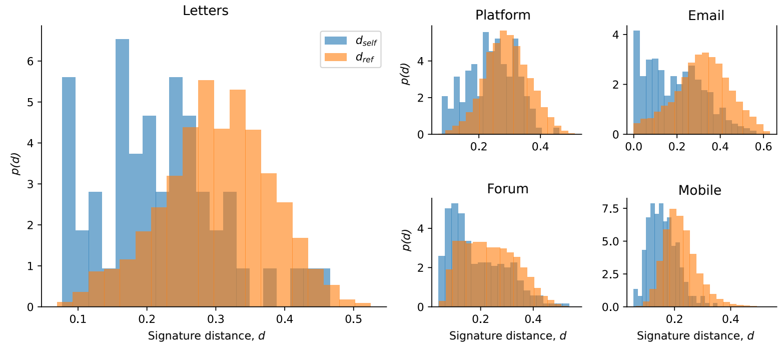


Figure 4.7. Social signatures are persistent in historical epistolar communication. The distance between social signatures across time d_{self} is consistently smaller than the distance to other egos d_{ref}. Figure from III, licensed under CC 4.0.

the social time would be transitive, so that alters would be connected in that layer as well. Conversely, for egos that communicated with *most* of their alters during a social time we could expect to see less evidence of focalised activity, i.e., less transitivity. Then, we tested whether “important” alters tended to be over- or under-represented in particular layers, where importance refers to the embeddedness of the alter in the personal network of the ego.

We followed a two-step approach, where we first showed that alter allocation to layers was ego-specific. In other words, just as ties have their own levels of temporal multiplexity, egos devote different social times to different subsets of alters (details in Publication V).

We developed a null model of independent alter sampling that allowed us to assess whether the observed transitivity in a layer could stem alters communicating with the ego independently of each other, and not because the triad uses the social time together. For an ego with n_l alters in layer l, we took independent alter samples of size n_l without replacement, where each alter was selected with probability proportional to their log-weight $\log(w_a)$. If two sampled alters were connected in the aggregate, then we assumed maximum transitivity, and their connection would also exist in a layer. We compared the empirical number of triangles around alters and the expected number of triangles under independent sampling.

Figure 4.8 depicts the results of our analysis. When the fraction of alters in a layer is small, we can expect that both the ego and the alters use the social time in a transitive manner, suggesting transitive use of social times. When egos allocate most of their alters to a layer, social times tend to be less transitive than expected, which does not suggest focalised activity. Those triangles still exist, but they are divided across different layers. We also found that weekend layers tend to be more transitive. These results align with those of Vergara Hidd *et al.* [206], who found that transient relationships —which likely have few triangles—, don’t tend

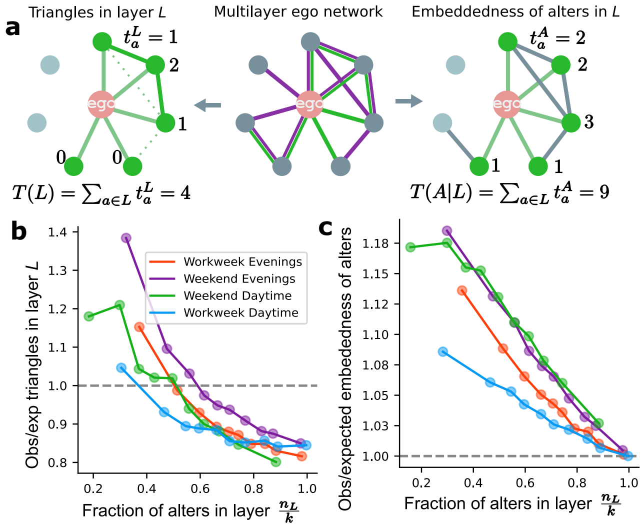


Figure 4.8. Small social times behave as constraining foci. **a** We measure the transitivity of social times by counting triangles around egos, and the aggregate embeddedness of alters selected **b-c** Social times are transitive –alters also use the time–, only when egos use a social time for a small subset of alters.

to communicate during the weekend. Last, we used the same sampling procedure to assess the embeddedness of the alters in a layer. That is, whether the alters allocated to those layers were overall embedded in the ego network, regardless of the social time. We found that such embedded alters tended to be over-represented in all layers, particularly if the fraction of alters in the layers was small. However, weekend layers over-represented such alters the most, i.e., egos tend to communicate more during the weekend with those who are well-connected within their social circle.

5. The dynamic interplay between social structures and mechanisms

In Publications II and IV we model group interactions through *choice homophily*, where social identities can favour connections within a social group [137, 138], but the available set of candidates for making new connections is constrained by the network structure [116, 191]. We examine how choice homophily interacts with two structural mechanisms, modelling its dynamic interplay with triadic closure in Publication II, and with preferential attachment in Publication IV. Our analytical results show how structural constraints can interact with choice homophily: triadic closure can induce more homophily in the network and result in increased patterns of homophily, while preferential attachment can lead to one group becoming dominant in a core-periphery structure, even when both groups as homophilous. We estimate the amount of homophily amplification in empirical datasets, as well as empirical core-periphery networks. We find that that in most empirical networks the periphery is heterophilous, with preferential attachment entrenching the structural dominance of the core.

5.1 Homophily as an interaction mechanism

Network evolution models tend to be relatively simple as a means for isolating the purported effects of the mechanisms themselves [3]. Our work was inspired by Schelling's model of residential segregation [186] and the empirical analysis of homophily in universities by Kossinets and Watts [116]. Both studies posited that the *observed* homophily in empirical systems could be explained by a combination of *choice* homophily that corresponds to the importance that a person places on their group identity, and homophily that is *induced* by structural effects of collective behaviour –such as triadic or focal closure [116]. Translated to networks, this means that we needed our model to untangle homophily as a mechanistic choice, and as an observed effect in the network.

This notion of choice is not the only way to model dynamic homophily, which can reflect patterns of similarity that respond to different contexts.

Based on Axelrod's model of dissemination of culture, Murase *et al.* modeled homophily as a multidimensional vector of choices to assess the effect of the number of homophilic features on network structure [9, 147]. In their model, homophily captures structural biases, where new connections are selected from the subset of same-group nodes. Karimi *et al.* focused on structural asymmetries, and incorporated homophily onto a preferential attachment kernel, so that the probability of creating a link to an existing node was proportional to its degree and its group assignment [100, 122]. Abebe *et al.* analysed a growing model where homophily interacted with second-order search: arriving nodes made new connections to N_S random nodes within the same group, and N_D connections to random nodes in a different group, with homophily enforced by setting $N_S > N_D$. Their model incorporated *triadic search*, with the arriving nodes connecting to the neighbors of their newly-made connections[1], finding that it could work along with the homophilic constraints and lead to more inter-group connections [1].

Modelling choice homophily

We analyse the interaction of choice homophily with triadic closure in a rewiring model in Publication II, and with preferential attachment in growing and rewiring models in Publication IV. Given group a, we focus on untangling homophily as a mechanism and as an observed pattern in the network o_a by defining a parameter of choice homophily s_a in terms of network dynamics: at each step, a node is faced with a candidate to which to connect to, and choice homophily is the probability to create a link to the candidate based on group attributes. This approach allows us to independently model homophily and the networked mechanisms that affect the candidate selection processes. We also parameterised the relative amount of times that candidates were found via triadic closure or preferential attachment with a parameter of probability c. Otherwise, a candidate was selected via random mixing, where all nodes have the same probability of being selected.

Our three models share central steps of candidate selection and acceptance via choice homophily, which we iterated over large periods of time. For our rewiring models, the simulation algorithms were:

1. Select a random focal node for rewiring.

[1] I use the term *triadic search* to emphasise that triangles are created after creating a link to a random node, and not by closing an existing open triad. This form of triadic closure is commonly used in growing models, such as the Jackson-Rodgers [94] model or the general growing model of Overgoor *et al.* [158]. In this context, however, triadic search exploits second-order homophily effects by adding connections to random nodes and then connecting to some of their neighbors. Such second-order effects include, e.g., following random same-group links that lead to more homophilous nodes [57].

2. With probability c select a random candidate via triadic closure or preferential attachment, for Publications II and IV, respectively. Otherwise, the candidate is randomly selected.

3. Accept the candidate with probability choice homophily s_{ab}, which differs for equal and different groups.

4. If the candidate is accepted, remove a random link from the focal node.

The growing model of Publication IV shares steps 2-4, and only differs in terms of node addition. In other words, the focal node is not random, but arrives to the network and finds m candidates.

Our definition of choice homophily can be expressed as a matrix of probabilities, where entry s_{ab} denotes the probability that a node of group a accepts a candidate node of group b. If the candidate is selected, then we create a link. We simplified this definition by setting $s_{ab} = 1 - s_{aa}$ as a means of capturing the actual acceptance rate of groups: $s_{aa} > 0.5$ for homophilous groups, $s_{aa} < 0.5$ for heterophilous groups, and the equality for neither. This definition affects the rate at which the network changes, but not the overall dynamics from an analytical perspective (see Publication IV).

5.2 Assessing group interactions in networks

5.2.1 Observed homophily

In terms of topological patterns, observed homophily —also known as assortative mixing or simply homophily— is commonly measured using the assortativity coefficient, which compares observed patterns of group mixing with the expected patterns under random rewiring [152, 160, 101]. Patterns of observed homophily are, however, difficult to assess as group mixing is subject to structural asymmetries and heterogeneous mixing [57, 101, 160]. The friendship paradox (Sec. 2.2.1) is a good example of how structural asymmetries can result in biased estimates. Eom and Jo proposed a generalised version of the friendship paradox that is of the form “your friends have more of this attribute than you” and observed it for empirical networks [95]; the limits of this claim were recently analysed by [58]. Lee *et al.* applied this notion under unequal groups, using a generative model to assess how differences in homophilic patterns can lead to distorted perception of the prevalence of minority groups in social systems [122]. More recently, Evtushenko and Kleinberg analytically showed local measurements of homophily are affected by second-order effects, i.e., homophily is higher if you follow a same group link than if you follow a different-group link [57]. Such local differences in mixing patterns were also analysed by Peel *et al.*, who described measures of local assortativity as "multiscale-mixing": the distribution of local assortativity values can be multimodal and skewed, with some parts of the network being more or less assortative than others. They proposed a measure that incorporates larger neighborhoods for assessing mixing patterns [160]. Last, Karimi and Oliveira showed that the assortativity coefficient is subject to group-size biases, and proposed a measure that adjusts for size imbalances in the assortativity coefficient [101].

We used similar notions of structural biases to measure observed patterns of homophily from our network models in Publication II. Our approached focused on two groups a and b, where N_a is the number of nodes in group a, $N = N_a + N_b$ is the total number of nodes, and relative group sizes are $n_a = \frac{N_a}{N}$ and $n_b = 1 - n_a$. We tracked three network statistics that played different roles in both the network evolution and the analysis. First, we defined a group mixing matrix P as the fraction of links between groups as $P_{ab} = \frac{L_{ab}}{L}$. Here, L is the total number of links and L_{ab} the links between groups a and b. Naturally, the entries of P follow the condition $P_{aa} + P_{ab} + P_{bb} = 1$. Next, we defined a matrix of transition probabilities T, with entries defined as the probability that if you follow a link from group a, you end up in group b. For same-group transitions, these probabilities are

$$T_{aa} = \frac{2P_{aa}}{2P_{aa} + P_{ab}}. \tag{5.1}$$

Unlike the group-mixing P matrix, the transition probability T matrix is not symmetric, with inter-group transitions defined by the complement $T_{ab} = 1 - T_{aa}$. In Publication II. The transition matrix already captures a notion of observed homophily as it measures the observed rate at which you will reach the same or a different group. However, this rate is subject to group imbalances. We can see this in the extreme case where almost all members of a minority group are connected among themselves and weakly connected to the other group (e.g., a SBM where a minority has intra-group connection probability close to one, and low inter-group connection probability). In the scenario we're describing, the minority is homophilous as it is fully dense, and more likely to connect to itself than to the other group. Following a random link might still lead to the majority group if inter-group connections are not if the majority is large enough and thus provides more opportunities to connect. To account for these imbalances, we measured observed homophily o_a of group a as a group-size correction of the transition matrix (see II),

$$o_a = \frac{n_b T_{aa}}{n_a T_{ab} + n_b T_{aa}} \tag{5.2}$$

5.2.2 Measuring core-peripheries

In Publication IV we focused on mechanisms that lead to core-periphery networks, where one group dominates connections over the other. The first network-scientific definition of core-periphery came from Borgatti and Everett [26], who used block-models to define an idealised version of core-periphery, where the core is fully connected to itself, the periphery is disconnected from itself, while inter-group connections laid on a spectrum between fully-disconnected and fully-connected. They introduced a measure of core-periphery as the matrix-level correlation between the observed network and the idealised block matrix. Their work was followed by a wide array of efforts to both infer and measure core-peripheries [85, 177, 49, 48]. In the case of inference, the task is to find which nodes belong to which group, while in measurement the goal is to assess the degree of structural dominance. Our goal, however, is to quantify the core-peripheriness of networks with pre-defined groups.

Gallagher *et al.* formalised two notions of core-peripheries based on group densities. They referred to *hub-and-spoke* as the case when group densities are ordered, so that if $\rho_{aa} > \rho_{ab} > \rho_{bb}$ then group a is a core and group b is the periphery, more likely to connect to the core than to itself. In the case when $\rho_{ab} = \rho_{bb}$, the network is called *layered*, as the periphery is just as likely to connect to the core as to itself.

In Publication IV we used Gallagher's notions of core-periphery to quantify the degree of structural dominance when a group was *density-dominant*, i.e., where $\rho_{aa} > \max\{\rho_{ab}, \rho_{bb}\}$. We assessed the core-peripheriness of group a via the inter-group relations,

$$\Omega_a = \frac{\rho_{ab} - \rho_{bb}}{\rho_{ab} + \rho_{bb}}, \tag{5.3}$$

which is positive when the groups are a hub-and-spoke, zero when it's layered, and negative when the groups form separate communities. We can write the densities in terms of the group-mixing matrix P as $\rho_{aa} \approx \frac{\rho P_{aa}}{n_a^2}$, where ρ is the overall network density.

While Ω_a is informative of how the periphery is related to the core, it is not informative of how the core behaves. Given the wide array of definitions of core-periphery, we also used the original measure from Borgatti and Everett [26], who defined a correlation to the " idea" structure where the core is fully connected to itself $\rho_{aa} = 1$, the periphery is disconnected from itself $\rho_{bb} = 0$ and inter-group connectivity exists on a spectrum $\rho_{ab} = \alpha$ with $\alpha \in [0, 1]$ that can be estimated. We adapted their discrete correlation measure to our continuous context as r_a. This measure takes values in the range $[-1, 1]$, and can be interpreted as a correlation to the ideal.

5.3 From structural polarization to structural dominance

5.3.1 Mean-field equations

We analysed our models using mean-field approximations, where we focused on tracking changes in the group-mixing matrix P. We expressed the evolution rules on a matrix M that captured the probabilities of creating inter- or intra-group links at each timestep. For example, under preferential attachment, a candidate node of group b is selected with probability $\frac{K_b}{K_a+K_b}$, where K_b is the sum of all the degrees of group b. We expressed the degree ratio in terms of the group mixing matrix $\frac{K_b}{K_a+K_b} = \frac{1}{2}P_{ab}+P_{bb}$. Under random node selection, a node from group b is selected with probability n_b. In a single time step, the probability of creating a link from a to b is

$$M_{ab} = \left[c\left(\frac{1}{2}P_{ab} + P_{bb}\right) + (1-c)n_b\right] s_{ab}. \tag{5.4}$$

Changes on the group-mixing matrix P depend on link addition P^{+}_{aa} and deletion P^{-}_{aa}. In the first case, this meant selecting a random node from a and then creating a new connection to a, so $P^{+}_{aa} = n_a M_{aa}$. The probability of deleting a random neighbor from group a is determined by the transition probability T_{aa}. Link deletion occurs if any link is created from a, i.e., with probability $M_{aa} + M_{ab}$. Thus, $P^{-}_{aa} = n_a T_{aa}(M_{aa} + M_{ab})$, and the master equation for the group-mixing of a is

$$\frac{dP_{aa}}{dt} = n_a M_{aa} + n_a T_{aa}(M_{aa} + M_{ab}). \tag{5.5}$$

We derived MFEs in terms of group mixing for our three network evolution models. Major differences stemmed from particular mechanisms: for triadic closure, the probability of creating a new $a-a$ link was given by the squared transition matrix $\left(T^2\right)_{aa}$. For the growing model, the MFEs were dependent on the overall number of edges L, which became negligible at the fixed points (see Publication IV).

5.3.2 The networked effects of choice homophily

Homophily amplification

Our results from Publication II confirmed the empirical results of Kossinets and Watts, that hypothesised that triadic closure could dynamically amplify choice homophily (Figure 5.1). Our baseline model with no triadic closure ($c = 0$) showed that choice and observed homophily aligned on average, yet including even small amounts of triadic closure lead to observed homophily becoming larger than the choice homophily ($o_a > s_a$). We named this gap induced homophily, and found that it became increasingly larger for higher

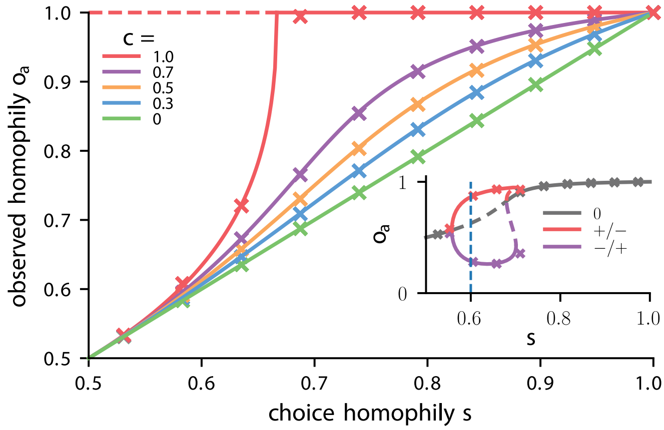


Figure 5.1. Triadic closure induces more homophily. Increasing the amounts of triadic closure c leads to more observed homophily. Inset: memory effects dependent on initial conditions can lead to core-peripheries. Lines represent fixed points of MFEs, and crosses simulations. Adapted from II by removing subfigures. Licensed under CC 4.0.

amounts of triadic closure. For networks with no random mixing ($c = 1$), the groups became disjoint for relatively small amounts of homophily $s_a \geq 2/3$.

Our model also showed that the dynamic interplay between triadic closure could lead to two competing states: homophily amplification and core-periphery structures. Depending on the parameter configuration, networks may reach any of these stationary points, where in many cases homophily amplification captures an unstable fixed point, with core-periphery appearing at the latter stages of network evolution.

Preferential attachment leads to core-peripheries

In Publication IV, we focused on the interplay between choice homophily and preferential attachment, finding that it can lead to core-periphery structures even when both groups are homophilous (Figure 5.2). When new connections relied exclusively on preferential attachment, small amounts of homophily ($s \geq 0.5$) sufficed to lead to sharp hub-and-spokes. In the case when both groups were of equal size ($n_a = n_b$) and equal homophily ($s_a = s_b$), any of the groups could become a core or a periphery, with results dependent on initial conditions and stochastic fluctuations. Our results held in general for different parameter configurations. Notably, differences in the choice homophily of the groups, or in group sizes affected the degree of homophily amplification or degree of core-peripheriness.

We evaluated both a growing and a rewiring model in as a means of comparing the effect of the model parameters and the overall evolution frameworks; namely, node addition or link rewiring. In this sense, we don't

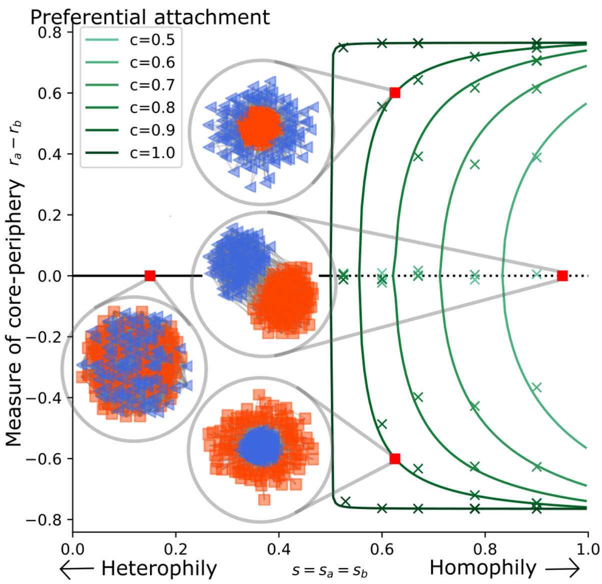


Figure 5.2. Preferential attachment leads to core-periphery. Preferential attachment c can limit the options to connect within the same group, breaking the effect of homophily and leading one group to become dominant over the other as a core-periphery. For equal group sizes and homophilies, both groups can be cores according to initial conditions. Adapted from IV by adding textual labels. Licensed under CC 4.0.

expect empirical networks to be neither fully growing nor fully rewiring in the way our model specifies. However, using two evolution frameworks allowed us to assess baseline scenarios, expecting empirical networks to be a combination of growing, rewiring, and other frameworks. This way, in the growing model links are rigid and capture cumulative advantage. In the rewiring higher degrees are more likely to both gain and lose connections.

We qualified the main differences as the growing model more commonly resulting in layered networks (the periphery connecting as much to itself as to the core), and the rewiring model more commonly producing hub-and-spoke networks, with a sharp core and a periphery connected mostly to the core. Overall, we found that cores tend to display a higher correlation to Borgatti and Everett's idealised notion under the rewiring framework, with higher correlation requiring group size disparities and high levels of preferential attachment.

The results of Publications II and IV suggest that this effect may be a byproduct of implicit preferential attachment in the triadic closure mechanism. A plausible explanation is that core-periphery is a result of connecting to more popular links via triadic closure. This could be tested by analyzing alternative definitions of triadic closure, such as selecting neighbors with probability inversely proportional to their degree.

5.4 Uncoupling dynamics and structural effects in empirical networks

We analysed a broad array of datasets to test whether these mechanisms could also explain observed patterns of homophily and core-periphery in empirical social systems. Our since our goal was focused on dynamics, we analysed temporal datasets where the "static" observed topology corresponded to sequences of networks obtained from sliding windows. For Publication II, our datasets included *Facebook*, a collection of 100 universities in the United States divided by classes during 2005 [200]; *PolBlogs* which contained linked political blogs divided by affiliations [2]; *School* contained proximity data for two days, with categories divided by gender [92]; *Last.fm* included self-reported friendships and gender; while *Directors*, was a network of boards of directors in Norway, divided by gender [187].

We expanded our analysis of the latter dataset in Publication IV under the name *Boards*, where we examined behavioural changes after the implementation of an affirmative action policy. In the same publication, we also analysed *Twitter*, where the minority group represented political actors on discussions about climate politics in Finland for 11 months [39]. We used two citation networks to analyse our growing model. In *Cit-geo*, groups corresponded to geographical regions, and in *APS* to subfields in physics [115]. Since homophily can be more generally understood as assortative attachment [173, 152], we also analysed *Airport*, a dataset of airport use in the United States, with groups determined by regions [154]. We systematically analysed group combinations to determine whether groups were related as core-peripheries under thresholds for Ω_g.

5.4.1 Induced homophily in empirical networks

Data-fitting methods

In Publication II we used two data-fitting methods to address different modelling shortcomings when deriving an analytical likelihood function. In particular, while our simulations were well-approximated by the MFEs, the squared transition matrix $(T^2)_{aa}$ would likely be a bad local proxy for empirical triadic closure. In our model, choice homophily works as a rejection mechanism with a clear algorithmic interpretation, but that requires to account for "rejected" candidates that do not exist in empirical data. In the literature, the problem of disentangling homophilic effects from network structure has been considered in different scenarios that may include an underlying generative model or not. In general, both triadic closure and homophily can lead to community-like structures [210], and so disentangling underlying mechanisms generally requires strong modelling assumptions [164, 189]. To be clear, it is possible to derive an analytical likelihood function of our generative model, which does meet such strong assumptions. For example, Peixoto recently introduced a variant of SBM that models homophily and triadic closure as part of a two-step generative process, successfully recovering links produced by either mechanism [164]. More recently, Sajjadi *et al.* defined a measure of choice homophily that uses the configuration model as the null [182]. We opted to use a dual fitting approach that would not impose restrictions on the empirical form of triadic closure and on the unobserved "rejected" candidates. The first fitting method estimated the difference between observed and generating homophily from a static perspective, while the second fitting method focused on dynamics via likelihood-free inference.

The first method was based on the MFEs at the fixed points. In this case, the core assumption was that the empirical networks had evolved enough to reach stationary states, allowing us to compute a functional form for choice homophily s_a and s_b in terms of the observed network structure (empirical transition matrices T and relative group sizes) and a parameter of triadic closure c. In short, this approach allowed us to estimate the relative amount of induced homophily for different values of triadic closure via an amplification measure $A_a(c) = \frac{o_a - s_a}{o_a}$, showing that in networks with *any* amounts of triadic closure, patterns of observed homophily can stem from structural constraints. Our results on Fig. 5.3 (*top*) showed that in the presence of triadic closure it's not trivial to assess the extent to which observed static homophilic patterns stem from underlying homophilic preferences or local structural constraints.

For the second method we used BOLFI, a framework for Bayesian likelihood-free inference [77]. As we mentioned in Sec. 3.3.4, the family of likelihood-free ABC methods rely on comparing statistics between simulated and empirical networks. We selected a set of network statistics by systematically evaluating their inferential performance. In other words, we analysed sets of statistics that could correctly recover the parameters that produced synthetic networks. Given that network dynamics could lead to several stable and unstable solutions, our parameter space did not only depend on evolution parameters (c, s_a, s_b), but also on initial network configurations, which we stated in terms of initial transition probabilities T^0_{aa} and T^0_{bb}. We evaluated sets of statistics across the five-dimensional parameter space, settling on statistics that captured elements of intra- and inter-group dynamics, including colored triangles, clustering coefficients for nodes of different degrees and colors and the group mixing matrix P. In our evaluations, the largest estimated mean absolute errors between seed θ and estimated parameters $\hat{\theta}$ were present in regions with unstable solutions.

Results

Our results under both fitting methods on Fig. 5.3 were qualitatively consistent, although the levels of induced homophily varied depending on

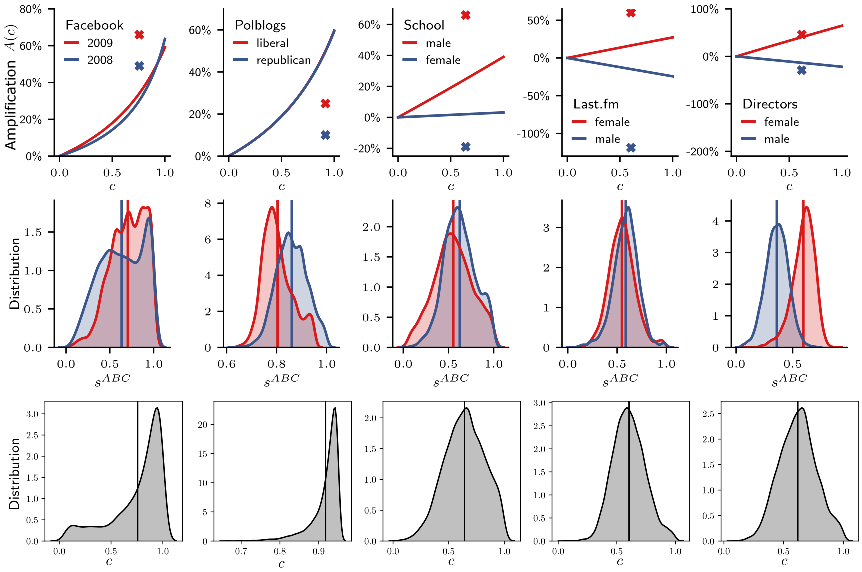


Figure 5.3. Homophily amplification and parameter estimates.. Adapted from II by removing estimates for the Facebook 100 dataset, and including posteriors c^{ABC} from the supplementary material. Licensed under CC 4.0.

the datasets and the fitting method. Particularly for the *Facebook* and *Polblogs* data we found high levels of triadic closure that likely amplified the underlying homophilic tendencies. In other datasets we found negative amplification, and weaker evidence of triadic closure (on average $c \in [.6, .7]$, with broad posterior distributions). In *School* and *Last.fm*, group homophilies were centered around random mixing and also included broad distributions, while for *Boards* the distributions were not as broad, but included a homophilic group and one heterophilic group. We further analysed *Boards* in Publication IV, finding that their particular dynamics were also largely affected by group imbalances.

5.4.2 Preferential attachment entrenches structural dominance

Data-fitting method

In Publication IV we derived the explicit likelihood functions by modelling network evolution as Markov processes, explicitly addressing the rejected candidates by marginalising over unobserved values in our likelihood function. We focused on performing inference on the amount of preferential attachment c and homophily s_a and s_b. We modeled the step-wise network evolution as a multinomial random variable. In a single time step, a focal node would find m candidates from either group to link to. For a focal node of group a, the probability of connecting to a node of group b is given by M_{ab} (see Eq. 5.4). This formulation in terms of the M matrix is, however, not well-defined for determining the amount of preferential attachment c. We used a degree-corrected distribution: at each timestep, we obtained p_k^{ab} the probability that a node from group a would connect to a node of degree k in group b,

$$p_k^{ab} = \left[c \frac{n_k^b k}{\sum_j j(n_j^a + n_j^b)} + (1-c) \frac{n_k^b}{\sum_j (n_j^a + n_j^b)} \right] s_{ab}, \tag{5.6}$$

For a focal node of group $i(t) = a$, we parameterised the evolution from one timestep to the other by a vector of probabilities over groups and degrees $p = (p_k^{aa}, p_k^{ab})_k$ [2],

$$P(G_t | G_{t-1}, \theta, i(t) = a) = \text{Multinomial}\left(m, (p_k^{aa}, p_k^{ab})_k\right) \tag{5.7}$$

We defined choice homophily as a rejection mechanism: a candidate node is either accepted or not. Evaluating the performance of our likelihood function when recovering seed parameters revealed that unobserved rejected candidates particularly affected the growing model. While both rewiring and growing models include rejections that are unobserved in empirical data, we attributed the disparity to the number of trials of each model. When rewiring $m = 1$ by definition, and s_a and s_b were well-approximated via observed successes. In contrast, for the growing case $m \geq 1$ was not directly observed from the data and resulted in highly-biased estimates of choice homophily. We explicitly modeled $p^{an} = 1 - \sum_k \left[p_k^{aa} + p_k^{ab}\right]$ the probability of rejecting a candidate of any degree, and obtained the marginal probability distribution over all possible rejected values (details in SM of Publication IV).

Results
Our systematic analysis of the phase space revealed that preferential attachment could amplify the exposure of one group over the other and lead to structural dominance even when both groups were homophilous. Analyzing empirical datasets revealed a starker picture: peripheral groups display heterophilous behaviour that is then amplified by preferential attachment, leading to reinforced structural disadvantages. Figure 5.4 depicts group densities and the estimated effect of preferential attachment on group mixing P. The *Twitter* dataset was a particularly interesting example, as the minority group was composed by political actors while the periphery by non-politicians. Our results show that politicians tend to share the content of fellow political figures, while other participants shared the messages of politicians at higher rates than of non-politicians. Our networks of political discussions display partisan sorting on a political spectrum [39]. However, this effect is not explicitly captured by the heterophilous

[2] Here the tuple notation emphasises that there are two probabilities per degree, but p is a column vector.

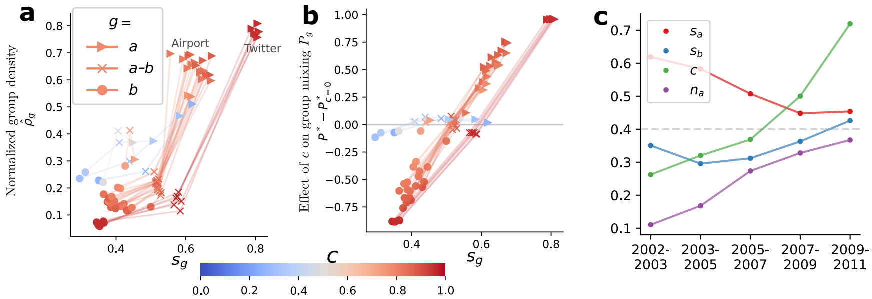


Figure 5.4. Homophily and preferential attachment in empirical datasets entrench structural disparities. **a** Empirical group densities as a function of choice homophily and preferential attachment (color). **b** Effect of preferential attachment in group mixing. *c* Evolution of boards dataset. Adapted from IV by removing growing networks and including parameter estimates for *Boards*. Licensed under CC 4.0.

behaviour of the peripheries, which more likely reflects that politicians are stakeholders in public policy, a status that grants relevance to their messages. Preferential attachment, however, has an augmentative effect on the homophilous-heterophilous pairing, so that retweeting by non-politicians cumulatively amplifies the messages of political figures, who end up becoming a dense core. These effects are strong enough despite both the core and the periphery displaying polarised structures on the discussion networks [39].

Last, we'd like to discuss *Boards*, a dataset that tracked the gendered composition of Boards of Directors in Norway during 9 years on a monthly basis. This dataset focused on an affirmative action policy by the government which began as an informal trial to achieve gender parity between 2004-2006. Due to insufficient results, the government established a two-year period to achieve 40% representation of each gender in each board, which was achieved by 2009. Figure 5.4 **c** tracks the evolution of parameters during the full observation period. By focusing on gender parity per board, this policy reduced the homophily gap between women a and men b. However, initially small preferential attachment increased to around $c \approx 0.7$ in the last observation period, while women still made up less than 40% of the overall. In practice, this policy increased gender representation, but also led to "superstar" women directors who were present in several boards. We did not assess the effect of preferential attachment on men, but for women these results were first observed by [187].

6. Discussion

6.1 Summary and answers to the research questions

The networked nature of social behaviour affects interpersonal relationships, individuals, groups, institutions, communities and societies. Social ties are in constant flux, following patterns of creation, maintenance and decay that depend on the traits of individuals, but also on social patterns such as group identities and social times. Throughout this dissertation, we explored how such patterns of interactions can be represented and analysed as networks that reflect social behaviour through static and dynamic components. Building from large bodies of work that range from sociology, behavioural studies, and network analysis, we approached sociality as reflected and affected by network structure and dynamics. We focused on two related but conceptually different perspectives. The first one prioritised high-granularity communication data, analyzing how sequences of contact contain rich temporal information about ties. The second one emphasised tunable models inspired by sociological and network studies, which allowed us to analyse the macroscopic effects of simple social behaviours. To this end, we'll now focus on how the main insights from this thesis relate to our overarching research questions.

Q1. How do contact patterns reflect sociological concepts of tie strength and multiplexity?

In our work, we extensively analysed known communication features that ranged from contact intensity and frequency, bursty behaviour, stability across an observation period, and the roles of social times. Many such features had been studied before, some others we proposed, and a perhaps larger number of features we missed, such as temporal reciprocity [41]. In Publication I we assessed how such features reflected tie strength by turning Granovetter's weak link hypothesis upside down: we assumed that tie strength was reflected topologically through overlapping circles of friends, and used it as a baseline to measure the extent to which different features could serve as proxies for tie strength. Based on previous research about relationship turnover and stability, we used a temporal measure of overlap that captured the average overlap on month-wide windows.

Our results highlighted the rich array of temporal features that capture tie strength, ranging from communication frequency through the counts of active days and hours, to stability of communication in the observation window, to the use of social times over the week. Our results highlighted the known roles of communication frequency, but also the importance of the observation period to assess relationships that are likely transient [149, 81], as well as the importance of social times during the week, which have been devoted less attention [97, 206].

We then expanded the latter concept in Publication V by proposing a methodology to capture multiplexity as the use of different social times reconstructed as multilayer networks, which we compared with Feld's focal multiplexity [59]. Temporal multiplexity provides a new framework through which we can analyse contact patterns that also reflects topological tie strength, helping us characterise even low-contact ties as weaker or stronger based on their multiplexity. We showed that although temporal multiplexity had a component induced by contact counts, much of the empirical multiplexity reflected heterogeneous tie-specific behaviour that was resilient to bursty dynamics. Individuals and small clusters used social times in their own ways as well. Looking at egos, we showed that the relative amount of alters in a layer was indicative of group behaviour, so that when egos devoted a social time to a small subset of alters then the alters would be much more likely to use the time as well, particularly on weekends.

Our framework provides a novel and intuitive approach to understanding widely-studied temporal contact patterns, including burstiness, individual communication strategies, and tie creation and decay. For example, since the usage of social times seems to be resilient to coarse-graining weekly activity, our method could potentially help explain patterns of both correlated and uncorrelated bursty behaviour [105, 96], which were previously only analysed in terms of either daily or aggregate weekly cycles [97], not social times.

Q2. How consistent is such social behaviour on different communication channels?

Temporal patterns of human communication have been widely studied during the last couple of decades, providing fruitful advances in our understandings of phenomena such as burstiness, social signatures, or patterns in tie creation, persistence, and decay. However, many such patterns in social behaviour can be affected by differences in communication channels, adherence to different social systems stemming from particular contexts, or be obscured by different time windows.

In Publications I and III we examined how temporal features of communication reflected on different communication channels besides mobile calls, such as social media platforms, text messages, emails, and historical letters. We focused on how the features themselves were reflected on different mediums, and not how individuals use different channels – which is of interest as well [80, 7]. Contact frequency measures such as active days, and biases in the observation windows are consistent proxies of tie strength on different channels –including the historical letters of III–. Such commonalities of interactions systems suggest deep layers of human behaviour. However, their relative importance is dependent on intrinsic elements of the social system and communication channel, and extrinsic elements such as the observation period or design methodologies. In this sense, while such layers of social behaviour strongly suggest commonalities, their particular configuration is system-dependent. For example, measures of similarity in daily rhythms were particularly important when looking at social media platforms or work emails.

By analyzing historical letters as communication networks we could uncover persistent behavioural features in dyadic communication on Publication III. Epistolary tie-level dynamics reflect qualitatively similar features to contemporary mediums. Most importantly, our results showed that social signatures were also a persistent feature of egos across time ranges, strongly suggesting that such communication strategies reflect underlying human behaviour that are not only a byproduct of our time, but perhaps related to one-on-one communication mediums. These results, however, were highly constrained by data availability, as there was no way to know the rates at which letters or people were missing, although we did know that some people were over-represented. This meant that some ties and egos were more likely to produce consistent results, unlike large-scale topological features that were subject to irregular sampling biases.

Q3. How can we untangle the behavioural patterns of group identities from their effects on networked interactions?

In Publications II and IV we proposed models for the micro-level interaction between group identities and networked processes, such as triadic closure and preferential attachment. Modelling homophily as a choice, where a node is faced with a candidate to which to connect and evaluates such candidate based on their group, allowed us to disentangle the effects of homophily as a social behaviour and the network mechanisms that may lead to over- or under-exposure of some group.

Our results from II showed that triadic closure could amplify choice homophily under a wide array of conditions, including group size imbalances and different amounts of random rewiring. Our results provide a mechanistic understanding of previously-held notions of how homophily can be induced by triadic closure, which constrains the pool of potential friends by searching within local neighborhoods. In our model, such neighborhoods can be initially heterogeneous, but become more homogeneous as the network evolves. This suggests that iterated triadic closure and choice homophily alone can lead to polarised social networks, yet more realistic extensions of the model could account for other mechanisms that lead to the homogenization of local neighborhoods, such as social foci.

Our model of preferential attachment in IV, on the other hand, sheds light on mechanisms that lead to group-level inequalities. By exposing nodes to high-degree nodes irrespective of groups, preferential attachment can amplify same-group connections for one group, and diminish them for the others. Our model of triadic closure could also result in core-periphery networks; after analyzing preferential attachment by itself, our results suggest that this effect is a byproduct of preferential attachment induced by triadic closure, where following links leads to more popular nodes – future work could help untangle these two mechanisms.

We found evidence of homophily amplification on empirical datasets using two fitting methods –one that assumed the networks were at a fixed state, and one where we dynamically fitted parameters. Using maximum-likelihood estimates for the dynamics on empirical core-periphery networks, we found most periphery groups to be heterophilous, cores to be homophilous and evidence of preferential attachment. These results have direct applications on, e.g., social media analysis, as they show that dominant positions in the networks can become entrenched through the interplay of preferential attachment through algorithmic exposure or “sharing content”, as well as homophilous dynamics in the core.

6.2 Future work

Our contributions to understanding how social behaviour is expressed in networked systems could be extended in a wide array of ways, both as a basis for future research lines or to test limitations. Our work on network reconstruction methods has immediate applications, such as the use of proxies for tie strength or networks based on social times. In particular, our framework for reconstructing multilayer networks of weekly times could provide fruitful research opportunities. Social times can affect spreading dynamics, reflect changing roles in relationships, or shed light on how egos balance their social interactions. Our results concerning the second research question could be improved by a more thorough examination of how particular elements of social systems such as size, system boundaries and communication channels, relate to the relative elements shaping human behaviour. This could help better understand how, e.g., information spreads in phone call networks as opposed to social media platforms.

Other possible extensions that combine both of our methodological approaches could include generative models of social times. This could take many forms, such as parameterising them in terms that capture notions of social foci, or assessing their effect on the dynamic composition of tie strengths around ego networks, in line with the recent work of Íñiguez *et al.* [93].

Our minimal models of choice homophily, triadic closure and preferential attachment could be more formally extended to analyse polarization. In this regard, effective understanding of polarization should be best approached by taking several of these mechanisms into account, and highlighting the different roles of local and global mechanisms. Classical models such as Axelrod's or the more recent work of Murase *et al.* have shown that binary segregation can follow regime changes in multidimensional scenarios: multidimensional identities can become segregated via mechanisms such as triadic closure if the number of identities is small enough, but networks remain mixed if the possible number of identities is large enough [9, 147]. In this sense, while triadic closure can amplify homophily, it is perhaps not the central issue that leads to segregation —meeting friends of friends is a basic form of exploring local social connections. A more interesting question is perhaps to understand the mechanisms that push the multidimensional landscape of local identities and opinions beyond a threshold small enough to become globally polarised. In other words, perhaps the problem is not that *local* mechanisms such as triadic closure lead to polarization in scenarios with binary groups, but that the *global* landscape of opinions is small enough to lead to binary groups that are then locally reinforced.

Our analysis of Twitter networks, for instance, showed that homophily could operate in terms of *status*, with politicians preferring to connect among themselves and the general population preferring to connect to politicians. Status homophily, along with preferential attachment, lead to core-periphery in discussion networks. Given the prominent role of political actors in both broadcast and social media, one approach to understanding polarization would be to assess the extent to which homophily operates both in terms of status –political figures and opinion leaders communicate with each other, reacting to their own context–, and in terms of political identities and opinions, which may be globally determined by the pool of choices offered by opinion leaders and political figures through preferential attachment and their status. Philosopher O. Táíwò recently examined such mechanisms from sociological perspectives, arguing that political actors have an outsized role in shaping the narratives around identity politics [196]. Following this framework, an interesting extension of our work would be to analyse whether shrinking landscapes of political identities are structurally favoured through combinations of these local and global mechanisms.

Social behaviour is a truly multifaceted phenomena that drives and is impacted by the networked nature of social systems, and future work will surely extend, expand and even refute our current understanding of societal dynamics.

References

[1] Rediet Abebe, Nicole Immorlica, Jon Kleinberg, Brendan Lucier, and Ali Shirali. On the effect of triadic closure on network segregation. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, EC '22. ACM, July 2022.

[2] Lada A. Adamic and Natalie Glance. The political blogosphere and the 2004 u.s. election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, LinkKDD '05, pages 36–43, New York, NY, USA, 2005. ACM.

[3] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, January 2002.

[4] Talayeh Aledavood, Ilkka Kivimäki, Sune Lehmann, and Jari Saramäki. Quantifying daily rhythms with non-negative matrix factorization applied to mobile phone data. *Scientific Reports*, 12(1), April 2022.

[5] Talayeh Aledavood, Sune Lehmann, and Jari Saramäki. Social network differences of chronotypes identified from mobile phone data. *EPJ Data Science*, 7(1), October 2018.

[6] Talayeh Aledavood, Eduardo López, Sam G. B. Roberts, Felix Reed-Tsochas, Esteban Moro, Robin I. M. Dunbar, and Jari Saramäki. Daily rhythms in mobile telephone communication. *PLOS ONE*, 10(9):e0138098, September 2015.

[7] Talayeh Aledavood, Eduardo López, Sam G. B. Roberts, Felix Reed-Tsochas, Esteban Moro, Robin I. M. Dunbar, and Jari Saramäki. *Channel-Specific Daily Patterns in Mobile Phone Communication*, page 209–218. Springer International Publishing, 2016.

[8] M. Tarik Altuncu, Ayse Seyyide Kaptaner, and Nur Sevencan. Optimizing the access to healthcare services in dense refugee hosting urban areas: A case for istanbul, 2019.

[9] Robert Axelrod. The dissemination of culture: A model with local convergence and global polarization. *The Journal of Conflict Resolution*, 41(2):203–226, 1997.

[10] Robert L. Axtell, Omar A. Guerrero, and Eduardo López. Frictional unemployment on labor flow networks. *Journal of Economic Behavior &; Organization*, 160:184–201, April 2019.

[11] Joe Ayres. Strategies to maintain relationships: Their identification and perceived usage. *Communication Quarterly*, 31(1):62–67, January 1983.

[12] Lars Backstrom and Jon Kleinberg. Romantic partnerships and the dispersion of social ties. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW 14*, 2014.

[13] Abhijit Banerjee, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. The diffusion of microfinance. *Science*, 341(6144), July 2013.

[14] Albert-László Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, May 2005.

[15] Albert-László Barabási. Scale-free networks: A decade and beyond. *Science*, 325(5939):412–413, July 2009.

[16] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.

[17] Louise Barrett, S. Peter Henzi, and David Lusseau. Taking sociality seriously: the structure of multi-dimensional social networks as a source of information for individuals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599):2108–2118, August 2012.

[18] Marco Bastos, Carlo Piccardi, Michael Levy, Neil McRoberts, and Mark Lubell. Core-periphery or decentralized? topological shifts of specialized information on twitter. *Social Networks*, 52:282–293, January 2018.

[19] Christian Bick, Elizabeth Gross, Heather A. Harrington, and Michael T. Schaub. What are higher-order networks? *SIAM Review*, 65(3):686–731, August 2023.

[20] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, April 2017.

[21] Vincent D Blondel, Adeline Decuyper, and Gautier Krings. A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1), August 2015.

[22] Eszter Bokányi, Eelke M. Heemskerk, and Frank W. Takes. The anatomy of a population-scale social network. *Scientific Reports*, 13(1), June 2023.

[23] S. Borgatti and Daniel Halgin. On j. clyde mitchell's the concept and use of social networks. In Mario L. Small and Brea L. Perry, editors, *Personal Networks: Classic Readings and New Directions in Egocentric Analysis*, chapter 3. Cambridge University Press, September 2021.

[24] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca. Network analysis in the social sciences. *Science*, 323(5916):892–895, feb 2009.

[25] Stephen P. Borgatti, Daniel J. Brass, and Daniel S. Halgin. *Social Network Research: Confusions, Criticisms, and Controversies*, page 1–29. Emerald Group Publishing Limited, July 2014.

[26] Stephen P Borgatti and Martin G Everett. Models of core/periphery structures. *Social Networks*, 21(4):375–395, October 2000.

[27] Sabri Boughorbel, Fethi Jarray, and Mohammed El-Anbari. Optimal classifier for imbalanced data using matthews correlation coefficient metric. *PLOS ONE*, 12(6):e0177678, June 2017.

[28] Anna D. Broido and Aaron Clauset. Scale-free networks are rare. *Nature Communications*, 10(1), March 2019.

[29] Ronald S Burt. *Structural holes*. Harvard University Press, London, England, October 1995.

[30] Ronald S. Burt. Bridge decay. *Social Networks*, 24(4):333–363, October 2002.

[31] Ronald S. Burt. Structural holes and good ideas. *American Journal of Sociology*, 110(2):349–399, September 2004.

[32] Carter T. Butts. Revisiting the foundations of network analysis. *Science*, 325(5939):414–416, July 2009.

[33] Julián Candia, Marta C González, Pu Wang, Timothy Schoenharl, Greg Madey, and Albert-László Barabási. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224015, 2008.

[34] Simone Centellegher, Eduardo López, Jari Saramäki, and Bruno Lepri. Personality traits and ego-network dynamics. *PLOS ONE*, 12(3):e0173110, March 2017.

[35] D. Centola. Influencers, backfire effects, and the power of the periphery. In Mario L. Small and Brea L. Perry, editors, *Personal Networks: Classic Readings and New Directions in Egocentric Analysis*, chapter 2. Cambridge University Press, September 2021.

[36] Damon Centola and Michael Macy. Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3):702–734, November 2007.

[37] Jinyuan Chang, Eric D. Kolaczyk, and Qiwei Yao. Estimation of subgraph densities in noisy networks. *Journal of the American Statistical Association*, 117(537):361–374, July 2020.

[38] F Stuart Chapin. *Human activity patterns in the city*. Wiley series in urban research. John Wiley & Sons, Nashville, TN, November 1974.

[39] Ted Hsuan Yun Chen, Ali Salloum, Antti Gronow, Tuomas Ylä-Anttila, and Mikko Kivelä. Polarization of climate politics results from partisan sorting: Evidence from finnish twittersphere. *arXiv preprint arXiv:2007.02706*, 2020.

[40] Munmun De Choudhury, Winter A. Mason, Jake M. Hofman, and Duncan J. Watts. Inferring relevant social networks from interpersonal communication. In *Proceedings of the 19th international conference on World wide web*. ACM Press, 2010.

[41] Sandeep Chowdhary, Elsa Andres, Adriana Manna, Luka Blagojević, Leonardo Di Gaetano, and Gerardo Iñiguez. Temporal patterns of reciprocity in communication networks. *EPJ Data Science*, 12(1), March 2023.

[42] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), February 2021.

[43] Aaron Clauset and Nathan Eagle. Persistence and periodicity in a dynamic proximity network, 2012.

[44] Aaron Clauset, Cristopher Moore, and M. E. J. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, May 2008.

[45] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[46] Janice Compton and Robert A. Pollak. Family proximity, childcare, and women's labor force attachment. *Journal of Urban Economics*, 79:72–90, January 2014.

[47] Mary Kathryn Cowles and Bradley P. Carlin. Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91(434):883–904, June 1996.

[48] Peter Csermely, András London, Ling-Yun Wu, and Brian Uzzi. Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, 1(2):93–123, 10 2013.

[49] M.R. da Silva, Hongwu Ma, and An-Ping Zeng. Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks. *Proceedings of the IEEE*, 96(8):1411–1420, August 2008.

[50] Sebastian Deri, Jeremie Rappaz, Luca Maria Aiello, and Daniele Quercia. Coloring in the links: Capturing social ties as they are perceived. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–18, November 2018.

[51] Xavier Didelot, Richard G. Everitt, Adam M. Johansen, and Daniel J. Lawson. Likelihood-free estimation of model evidence. *Bayesian Analysis*, 6(1), March 2011.

[52] Thomas A. DiPrete and Gregory M. Eirich. Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology*, 32(1):271–297, August 2006.

[53] Teunis C. Dorlas. *Statistical Mechanics: Fundamentals and Model Solutions*. CRC Press, April 2021.

[54] Stefan Dumont. correspsearch –– connecting scholarly editions of letters. *Journal of the Text Encoding Initiative*, 6(10), 2016.

[55] P. Erdös and A. Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290, 1959.

[56] Tucker Evans and Feng Fu. Opinion formation on dynamic networks: identifying conditions for the emergence of partisan echo chambers. *Royal Society Open Science*, 5(10):181122, October 2018.

[57] Anna Evtushenko and Jon Kleinberg. The paradox of second-order homophily in networks. *Scientific Reports*, 11(1), June 2021.

[58] Anna Evtushenko and Jon Kleinberg. Node-based generalized friendship paradox fails. *Scientific Reports*, 13(1), February 2023.

[59] Scott L. Feld. The focused organization of social ties. *American Journal of Sociology*, 86(5):1015–1035, March 1981.

[60] Scott L. Feld. Why your friends have more friends than you do. *American Journal of Sociology*, 96(6):1464–1477, May 1991.

[61] Scott L. Feld, Devin T. Knighton, and Alec McGail. Reflections on the focused organization of social ties and its implications for bonding and bridging. In Mario L. Small and Brea L. Perry, editors, *Personal Networks: Classic Readings and New Directions in Egocentric Analysis*, chapter 13. Cambridge University Press, September 2021.

[62] Santo Fortunato and Darko Hric. Community detection in networks: A user guide. *Physics Reports*, 659:1–44, November 2016.

[63] Noah E. Friedkin. A guttman scale for the strength of an interpersonal tie. *Social Networks*, 12(3):239–252, September 1990.

[64] Ryan J. Gallagher, Jean-Gabriel Young, and Brooke Foucault Welles. A clarified typology of core-periphery structure in networks. *Science Advances*, 7(12):eabc9800, 2021.

[65] Yerali Gandica, João Carvalho, Fernando Sampaio dos Aidos, Renaud Lambiotte, and Timoteo Carletti. Stationarity of the inter-event power-law distributions. *PLOS ONE*, 12(3):e0174509, mar 2017.

[66] Laetitia Gauvin, Mathieu Génois, Márton Karsai, Mikko Kivelä, Taro Takaguchi, Eugenio Valdano, and Christian L. Vestergaard. Randomized reference models for temporal networks. *SIAM Review*, 64(4):763–830, November 2022.

[67] Valeria Gelardi, Didier Le Bail, Alain Barrat, and Nicolas Claidiere. From temporal network data to the dynamics of social relationships. *Proceedings of the Royal Society B: Biological Sciences*, 288(1959), September 2021.

[68] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, November 2013.

[69] Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), November 1992.

[70] E. N. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, December 1959.

[71] K.-I. Goh and A.-L. Barabási. Burstiness and memory in complex systems. *EPL (Europhysics Letters)*, 81(4):48002, jan 2008.

[72] Amit Goldenberg, Joseph M. Abruzzo, Zi Huang, Jonas Schöne, David Bailey, Robb Willer, Eran Halperin, and James J. Gross. Homophily and acrophily as drivers of political segregation. *Nature Human Behaviour*, 7(2):219–230, November 2022.

[73] Sandra González-Bailón, David Lazer, Pablo Barberá, Meiqing Zhang, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Deen Freelon, Matthew Gentzkow, Andrew M. Guess, Shanto Iyengar, Young Mie Kim, Neil Malhotra, Devra Moehler, Brendan Nyhan, Jennifer Pan, Carlos Velasco Rivera, Jaime Settle, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud, and Joshua A. Tucker. Asymmetric ideological segregation in exposure to political news on facebook. *Science*, 381(6656):392–398, July 2023.

[74] John M. Gottman and Paul R. Peluso. *Dynamic Models of Social Interaction*, page 17–29. Springer International Publishing, 2018.

[75] Mark Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510, November 1985.

[76] Mark S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[77] Michael U. Gutmann, Jukka Cor, and er. Bayesian optimization for likelihood-free inference of simulator-based statistical models. *Journal of Machine Learning Research*, 17(125):1–47, 2016.

[78] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., 2001.

[79] Fritz Heider. *The psychology of interpersonal relations.* John Wiley & Sons Inc, 1958.

[80] Sara Heydari, Sam G. Roberts, Robin I. M. Dunbar, and Jari Saramäki. Multichannel social signatures and persistent features of ego networks. *Applied Network Science*, 3(1), May 2018.

[81] Valentín Vergara Hidd, Eduardo López, Simone Centellegher, Sam G. B. Roberts, Bruno Lepri, and Robin I. M. Dunbar. The stability of transient relationships. *Scientific Reports*, 13(1), April 2023.

[82] Takayuki Hiraoka, Abbas K. Rizi, Mikko Kivelä, and Jari Saramäki. Herd immunity and epidemic size in networks with vaccination homophily. *Physical Review E*, 105(5), May 2022.

[83] Paul W. Holland and Samuel Leinhardt. Local structure in social networks. *Sociological Methodology*, 7:1, 1976.

[84] Betina Hollstein. Georg simmel's contribution to social network research. In Mario L. Small and Brea L. Perry, editors, *Personal Networks: Classic Readings and New Directions in Egocentric Analysis*, chapter 1. Cambridge University Press, September 2021.

[85] Petter Holme. Core-periphery organization of complex networks. *Phys. Rev. E*, 72:046111, Oct 2005.

[86] Petter Holme. Modern temporal network theory: a colloquium. *The European Physical Journal B*, 88(9), September 2015.

[87] Petter Holme. Rare and everywhere: Perspectives on scale-free networks. *Nature Communications*, 10(1), March 2019.

[88] Petter Holme and Fredrik Liljeros. Birth and death of links control disease spreading in empirical contact networks. *Scientific Reports*, 4:4999, 2015.

[89] Petter Holme and Jari Saramäki. *A Map of Approaches to Temporal Networks*, page 1–24. Springer International Publishing, 2019.

[90] Petter Holme and Jari Saramäki, editors. *Temporal Network Theory*. Springer International Publishing, 2019.

[91] Howard Hotson and Thomas Wallnig, editors. *Reassembling the Republic of Letters in the Digital Age: Standards, Systems, Scholarship*. Göttingen University Press, 2019.

[92] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, 271(1):166–180, February 2011.

[93] Gerardo Iñiguez, Sara Heydari, János Kertész, and Jari Saramäki. Universal patterns in egocentric communication networks. *Nature Communications*, 14(1), August 2023.

[94] Matthew O Jackson and Brian W Rogers. Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915, May 2007.

[95] Hang-Hyun Jo and Young-Ho Eom. Generalized friendship paradox in networks with tunable degree-attribute correlation. *Physical Review E*, 90(2), August 2014.

[96] Hang-Hyun Jo, Takayuki Hiraoka, and Mikko Kivelä. Burst-tree decomposition of time series reveals the structure of temporal correlations. *arXiv:1907.13556 [physics.data-an]*, 2019.

[97] Hang-Hyun Jo, Márton Karsai, János Kertész, and Kimmo Kaski. Circadian pattern and burstiness in mobile phone communication. *New Journal of Physics*, 14(1):013055, jan 2012.

[98] Marko Jusup, Petter Holme, Kiyoshi Kanazawa, Misako Takayasu, Ivan Romić, Zhen Wang, Sunčana Geček, Tomislav Lipić, Boris Podobnik, Lin Wang, Wei Luo, Tin Klanjšček, Jingfang Fan, Stefano Boccaletti, and Matjaž Perc. Social physics. *Physics Reports*, 948:1–148, February 2022.

[99] Indika Kahanda and Jennifer Neville. Using transactional information to predict link strength in online social networks. In *Third International AAAI Conference on Weblogs and Social Media*, 2009.

[100] Fariba Karimi, Mathieu Génois, Claudia Wagner, Philipp Singer, and Markus Strohmaier. Homophily influences ranking of minorities in social networks. *Scientific Reports*, 8(1), July 2018.

[101] Fariba Karimi and Marcos Oliveira. On the inadequacy of nominal assortativity for assessing homophily in networks. *Scientific Reports*, 13(1), November 2023.

[102] Fariba Karimi, Verónica C. Ramenzoni, and Petter Holme. Structural differences between open and direct communication in an online community. *Physica A: Statistical Mechanics and its Applications*, 414:263–273, November 2014.

[103] Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83:016107, Jan 2011.

[104] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki. Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E*, 83(2), February 2011.

[105] Márton Karsai, Kimmo Kaski, Albert-László Barabási, and János Kertész. Universal features of correlated bursty behaviour. *Scientific Reports*, 2(1), May 2012.

[106] Márton Karsai, Kimmo Kaski, and János Kertész. Correlated dynamics in egocentric communication networks. *PLoS ONE*, 7(7):e40612, jul 2012.

[107] Márton Karsai, Nicola Perra, and Alessandro Vespignani. Time varying networks and the weakness of strong ties. *Scientific Reports*, 4(1), February 2014.

[108] János Kertész, János Török, Yohsuke Murase, Hang-Hyun Jo, and Kimmo Kaski. *Multiplex Modeling of Society*, page 84–100. Oxford University Press, November 2018.

[109] János Kertész, János Török, Yohsuke Murase, Hang-Hyun Jo, and Kimmo Kaski. *Modeling the Complex Network of Social Interactions*, page 3–19. Springer International Publishing, August 2020.

[110] Keigo Kimura, Yuzuru Tanaka, and Mineichi Kudo. A fast hierarchical alternating least squares algorithm for orthogonal nonnegative matrix factorization. In Dinh Phung and Hang Li, editors, *Proceedings of the Sixth Asian Conference on Machine Learning*, volume 39 of *Proceedings of Machine Learning Research*, pages 129–141, Nha Trang City, Vietnam, 26–28 Nov 2015. PMLR.

[111] M. Kivela, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, July 2014.

[112] Mikko Kivelä and Mason A. Porter. Estimating interevent time distributions from finite observation periods in communication networks. *Physical Review E*, 92(5), nov 2015.

[113] Sadamori Kojaku and Naoki Masuda. Core-periphery structure requires something else in the network. *New Journal of Physics*, 20(4):043012, apr 2018.

[114] Eric D. Kolaczyk. *Statistical Analysis of Network Data: Methods and Models*. Springer New York, 2009.

[115] Hyunsik Kong, Samuel Martin-Gutierrez, and Fariba Karimi. Influence of the first-mover advantage on the gender disparities in physics citations. *Communications Physics*, 5(1):1–11, 2022.

[116] Gueorgi Kossinets and Duncan J. Watts. Origins of homophily in an evolving social network. *American Journal of Sociology*, 115:405–450, 2009.

[117] L. Kovanen, K. Kaski, J. Kertesz, and J. Saramaki. Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. *Proceedings of the National Academy of Sciences*, 110(45):18070–18075, October 2013.

[118] Gautier Krings, Márton Karsai, Sebastian Bernhardsson, Vincent D Blondel, and Jari Saramäki. Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science*, 1(1), May 2012.

[119] Jussi M. Kumpula, Jukka-Pekka Onnela, Jari Saramäki, Kimmo Kaski, and János Kertész. Emergence of communities in weighted networks. *Physical Review Letters*, 99(22), November 2007.

[120] Edward O Laumann, Peter V Marsden, and David Prensky. The boundary specification problem in network analysis. *Research methods in social network analysis*, 61(8), 1989.

[121] David Lazer, Eszter Hargittai, Deen Freelon, Sandra Gonzalez-Bailon, Kevin Munger, Katherine Ognyanova, and Jason Radford. Meaningful measures of human society in the twenty-first century. *Nature*, 595(7866):189–196, June 2021.

[122] Eun Lee, Fariba Karimi, Claudia Wagner, Hang-Hyun Jo, Markus Strohmaier, and Mirta Galesic. Homophily and minority-group size explain perception biases in social networks. *Nature Human Behaviour*, 3(10):1078–1087, August 2019.

[123] Sune Lehmann. *Fundamental Structures in Temporal Communication Networks*, page 25–48. Springer International Publishing, 2019.

[124] Tracey Li, Jesper Dejby, Maximilian Albert, Linus Bengtsson, and Veronique Lefebvre. Estimating the resilience to natural disasters by using call detail records to analyse the mobility of internally displaced persons, 2019.

[125] Nicole Liebers and Holger Schramm. Parasocial interactions and relationships with media characters - an inventory of 60 years of research, 06 2019.

[126] Eduardo López, Omar A. Guerrero, and Robert L. Axtell. A network theory of inter-firm labor flows. *EPJ Data Science*, 9(1), November 2020.

[127] P. V. Marsden and K. E. Campbell. Reflections on conceptualizing and measuring tie strength. *Social Forces*, 91(1):17–23, August 2012.

[128] Peter V. Marsden and Karen E. Campbell. Measuring tie strength. *Social Forces*, 63(2):482, December 1984.

[129] Naoki Masuda and Petter Holme. Predicting and controlling infectious disease epidemics using temporal networks. *F1000Prime Reports*, 5, March 2013.

[130] Naoki Masuda and Renaud Lambiotte. *A Guide to Temporal Networks*. WORLD SCIENTIFIC (EUROPE), June 2020.

[131] Naoki Masuda, Mason A. Porter, and Renaud Lambiotte. Random walks and diffusion on networks. *Physics Reports*, 716–717:1–58, November 2017.

[132] Antonis Matakos, Sijing Tu, and Aristides Gionis. Tell me something my friends do not know: diversity maximization in social networks. *Knowledge and Information Systems*, 62(9):3697–3726, April 2020.

[133] Heather Mattie and Jukka-Pekka Onnela. Edge overlap in weighted and directed social networks. *Network Science*, 9(2):179–193, February 2021.

[134] Cassie McMillan. Worth the weight: Conceptualizing and measuring strong versus weak tie homophily. *Social Networks*, 68:139–147, January 2022.

[135] J. Miller McPherson and Lynn Smith-Lovin. Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *American Sociological Review*, 52(3):370, June 1987.

[136] Miller McPherson and Jeffrey A. Smith. Network effects in blau space: Imputing social context from survey data. *Socius: Sociological Research for a Dynamic World*, 5:237802311986859, January 2019.

[137] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[138] Miller McPherson, Lynn Smith-Lovin, and Craig Rawlings. The enormous flock of homophily researchers: Assessing and promoting a research agenda. In Mario L. Small and Brea L. Perry, editors, *Personal Networks: Classic Readings and New Directions in Egocentric Analysis*, chapter 16. Cambridge University Press, September 2021.

[139] Angelo Mele. A structural model of homophily and clustering in social networks. *SSRN Electronic Journal*, 2017.

[140] Stanley Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.

[141] Giovanna Miritello, Rubén Lara, and Esteban Moro. Time allocation in social networks: Correlation between social structure and human communication dynamics. In *Understanding Complex Systems*, pages 175–190. Springer Berlin Heidelberg, 2013.

[142] Giovanna Miritello, Rubén Lara, Manuel Cebrian, and Esteban Moro. Limited communication capacity unveils strategies for human interaction. *Scientific Reports*, 3(1), June 2013.

[143] Giovanna Miritello, Esteban Moro, and Rubén Lara. Dynamical strength of social ties in information spreading. *Physical Review E*, 83(4), apr 2011.

[144] Giovanna Miritello, Esteban Moro, Rubén Lara, Rocío Martínez-López, John Belchamber, Sam G.B. Roberts, and Robin I.M. Dunbar. Time as a limited resource: Communication strategy in mobile phone networks. *Social Networks*, 35(1):89–95, jan 2013.

[145] Esteban Moro, Dan Calacci, Xiaowen Dong, and Alex Pentland. Mobility patterns are associated with experienced income segregation in large us cities. *Nature Communications*, 12(1), July 2021.

[146] Yohsuke Murase, Hang-Hyun Jo, János Török, János Kertész, and Kimmo Kaski. Modeling the role of relationship fading and breakup in social network formation. *PLOS ONE*, 10(7):e0133005, July 2015.

[147] Yohsuke Murase, Hang-Hyun Jo, János Török, János Kertész, and Kimmo Kaski. Structural transition in social networks: The role of homophily. *Scientific Reports*, 9(1), March 2019.

[148] Yohsuke Murase, János Török, Hang-Hyun Jo, Kimmo Kaski, and János Kertész. Multilayer weighted social network model. *Physical Review E*, 90(5), November 2014.

[149] Henry Navarro, Giovanna Miritello, Arturo Canales, and Esteban Moro. Temporal patterns behind the strength of persistent ties. *EPJ Data Science*, 6(1), dec 2017.

[150] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Physical Review E*, 64(2), July 2001.

[151] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2), July 2001.

[152] Mark Newman. *Networks*. Oxford University Press, London, England, 2 edition, July 2018.

[153] Jorge Nocedal and Stephen Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, NY, 2 edition, July 2006.

[154] Department of Transportation. Airline origin and destination survey (db1b), 2022. Available at BTS: `https://www.transtats.bts.gov/DatabaseInfo.asp?QO_VQ=EFI&YvOx=D`. Accessed February 10, 2022 [Organization as author].

[155] Marcos Oliveira, Fariba Karimi, Maria Zens, Johann Schaible, Mathieu Génois, and Markus Strohmaier. Group mixing drives inequality in face-to-face gatherings. *Communications Physics*, 5(1), May 2022.

[156] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, April 2007.

[157] J-P. Onnela, J. Saramäki, Jörkki Hyvönen, G. Szabó, M. Argollo de Menezes, K. Kaski, A-L. Barabási, and J. Kertész. Analysis of a large-scale weighted network of one-to-one human communications. *New Journal of Physics*, 9(179), 2007.

[158] Jan Overgoor, Austin Benson, and Johan Ugander. Choosing to grow a graph: Modeling network formation as discrete choice. In *The World Wide Web Conference*, WWW ’19, page 1409–1420, New York, NY, USA, 2019. Association for Computing Machinery.

[159] Patrick S. Park, Joshua E. Blumenstock, and Michael W. Macy. The strength of long-range ties in population-scale social networks. *Science*, 362(6421):1410–1413, December 2018.

[160] Leto Peel, Jean-Charles Delvenne, and Renaud Lambiotte. Multiscale mixing patterns in networks. *Proceedings of the National Academy of Sciences*, 115(16):4057–4062, April 2018.

[161] Leto Peel, Daniel B. Larremore, and Aaron Clauset. The ground truth about metadata and community detection in networks. *Science Advances*, 3(5):e1602548, 2017.

[162] Leto Peel, Tiago P. Peixoto, and Manlio De Domenico. Statistical inference links data and theory in network science. *Nature Communications*, 13(1), November 2022.

[163] Tiago P. Peixoto. Network reconstruction and community detection from dynamics. *Physical Review Letters*, 123(12), September 2019.

[164] Tiago P. Peixoto. Disentangling homophily, community structure, and triadic closure in networks. *Physical Review X*, 12(1), January 2022.

[165] Ignacio Perez-Pozuelo, Bing Zhai, Joao Palotti, Raghvendra Mall, Michaël Aupetit, Juan M. Garcia-Gomez, Shahrad Taheri, Yu Guan, and Luis Fernandez-Luque. The future of sleep health: a data-driven revolution in sleep science and medicine. *npj Digital Medicine*, 3(1), March 2020.

[166] Isidoros Perikos and Loizos Michael. A survey on tie strength estimation methods in online social networks. In *Proceedings of the 14th International Conference on Agents and Artificial Intelligence*. SCITEPRESS - Science and Technology Publications, 2022.

[167] Thong Pham, Paul Sheridan, and Hidetoshi Shimodaira. Pafit: A statistical method for measuring preferential attachment in temporal complex networks. *PLOS ONE*, 10(9):e0137796, September 2015.

[168] Derek De Solla Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5):292–306, September 1976.

[169] Troy Raeder, Omar Lizardo, David Hachen, and Nitesh V. Chawla. Predictors of short-term decay of cell phone contacts in a large scale communication network. *Social Networks*, 33(4):245–257, October 2011.

[170] Sebastian Raimondo and Manlio De Domenico. Measuring topological descriptors of complex networks under uncertainty. *Physical Review E*, 103(2), February 2021.

[171] Daniel Redhead and Eleanor A. Power. Social hierarchies and social networks in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1845), January 2022.

[172] Thorsten Rings, Timo Bröhl, and Klaus Lehnertz. Network structure from a characterization of interactions in complex systems. *Scientific Reports*, 12(1), July 2022.

[173] Mark T. Rivera, Sara B. Soderstrom, and Brian Uzzi. Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annual Review of Sociology*, 36(1):91–115, June 2010.

[174] Abbas K. Rizi, Ali Faqeeh, Arash Badie-Modiri, and Mikko Kivelä. Epidemic spreading and digital contact tracing: Effects of heterogeneous mixing and quarantine failures. *Physical Review E*, 105(4), April 2022.

[175] Sam G.B. Roberts and Robin I.M. Dunbar. The costs of family and friends: an 18-month longitudinal study of relationship maintenance and decay. *Evolution and Human Behavior*, 32(3):186–197, May 2011.

[176] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173–191, May 2007.

[177] Michaela Puck Rombach, Mason Alexander Porter, James H. Fowler, and Peter J. Mucha. Core-periphery structure in networks. *SSRN Electronic Journal*, 2013.

[178] Puck Rombach, Mason A. Porter, James H. Fowler, and Peter J. Mucha. Core-periphery structure in networks (revisited). *SIAM Review*, 59(3):619–646, January 2017.

[179] Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI*, 2015.

[180] Chandreyee Roy, Kunal Bhattacharya, Robin I. M. Dunbar, and Kimmo Kaski. Turnover in close friendships. *Scientific Reports*, 12(1), June 2022.

[181] Miguel Ruiz-García, Juan Ozaita, María Pereda, Antonio Alfonso, Pablo Brañas-Garza, José A. Cuesta, and Angel Sánchez. Triadic influence as a proxy for compatibility in social relationships. *Proceedings of the National Academy of Sciences*, 120(13), March 2023.

[182] Sina Sajjadi, Samuel Martin-Gutierrez, and Fariba Karimi. Unveiling homophily beyond the pool of opportunities, 2024.

[183] Piotr Sapiezynski, Arkadiusz Stopczynski, David Dreyer Lassen, and Sune Lehmann. Interaction data from the copenhagen networks study. *Scientific Data*, 6(1), December 2019.

[184] J. Saramäki, E. A. Leicht, E. Lopez, S. G. B. Roberts, F. Reed-Tsochas, and R. I. M. Dunbar. Persistence of social signatures in human communication. *Proceedings of the National Academy of Sciences*, 111(3):942–947, jan 2014.

[185] Jari Saramäki and Esteban Moro. From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. *The European Physical Journal B*, 88(6), 2015.

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

[187] Cathrine Seierstad and Tore Opsahl. For the few not the many? the effects of affirmative action on presence, prominence, and social capital of women directors in norway. *Scandinavian Journal of Management*, 27(1):44 – 54, 2011. Part Special Topic Forum on “Architectural Competitions”.

[188] Indira Sen, Fabian Flöck, and Claudia Wagner. On the reliability and validity of detecting approval of political actors in tweets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2020.

[189] Cosma Rohilla Shalizi and Andrew C. Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 40(2):211–239, May 2011.

[190] Mario L. Small and Brea L. Perry, editors. *Personal Networks: Classic Readings and New Directions in Egocentric Analysis*. Cambridge University Press, September 2021.

[191] Mario Luis Small. *Unanticipated Gains: Origins of Network Inequality in Everyday Life*. Oxford University PressNew York, August 2009.

[192] Marc A. Smith, Itai Himelboim, Lee Rainie, and Ben Shneiderman. *The Structures of Twitter Crowds and Conversations*, page 67–108. Springer International Publishing, 2015.

[193] Urszula Strawinska-Zanko and Larry S. Liebovitch. *Introduction to the Mathematical Modeling of Social Relationships*, page 1–15. Springer International Publishing, 2018.

[194] Michael P. H. Stumpf and Mason A. Porter. Critical truths about power laws. *Science*, 335(6069):665–666, February 2012.

[195] Kaiyuan Sun, Enrico Ubaldi, Jie Zhang, Márton Karsai, and Nicola Perra. *The Effects of Local and Global Link Creation Mechanisms on Contagion Processes Unfolding on Time-Varying Networks*, page 305–324. Springer International Publishing, 2019.

[196] Olúfémi O Táíwo. *Elite capture*. Haymarket Books, May 2022.

[197] Taro Takaguchi, Naoki Masuda, and Petter Holme. Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics. *PLoS ONE*, 8(7):e68629, July 2013.

[198] William Hedley Thompson, Per Brantefors, and Peter Fransson. From static to temporal network theory: Applications to functional brain connectivity. *Network Neuroscience*, 1(2):69–99, June 2017.

[199] Riitta Toivonen, Lauri Kovanen, Mikko Kivelä, Jukka-Pekka Onnela, Jari Saramäki, and Kimmo K. Kaski. A comparative study of social network models: Network evolution models and nodal attribute models. *Soc. Networks*, 31:240–254, 2008.

[200] Amanda L. Traud, Peter J. Mucha, and Mason A. Porter. Social structure of facebook networks. *Physica A*, 391(16):4165 – 4180, 2012.

[201] Gergő Tóth, Johannes Wachs, Riccardo Di Clemente, Ákos Jakobi, Bence Ságvári, János Kertész, and Balázs Lengyel. Inequality is rising where social network segregation interacts with urban topology. *Nature Communications*, 12(1), February 2021.

[202] Charles van den Heuvel, Scott B. Weingart, Nils Spelt, and Henk Nellen. Circles of confidence in correspondence. *Nuncius*, 31(1):78–106, 2016.

[203] Dirk van Miert. What was the Republic of Letters? A brief introduction to a long history (1417–2008). *Groniek*, 204/205:269–287, 2016.

[204] Alexei Vazquez, Balázs Rácz, András Lukács, and Albert-László Barabási. Impact of non-poissonian activity patterns on spreading processes. *Physical Review Letters*, 98(15), April 2007.

[205] Lois M. Verbrugge. Multiplexity in adult friendships. *Social Forces*, 57(4):1286, June 1979.

[206] Valentín Vergara Hidd, Mailun Zhang, Simone Centellegher, Sam G. B. Roberts, Bruno Lepri, and Eduardo López. The rhythms of transient relationships: allocating time between weekdays and weekends. *Royal Society Open Science*, 10(10), October 2023.

[207] Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD*, 2011.

[208] Larry Wasserman. *All of statistics : a concise course in statistical inference*. Springer, New York, 2010.

[209] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, June 1998.

[210] Sophie Wharrie, Lamiae Azizi, and Eduardo G. Altmann. Micro-, meso-, macroscales: The effect of triangles on communities in networks. *Physical Review E*, 100(2), August 2019.

[211] William Wilmot. *Dyadic Communication*. Addison-Wesley Pub. Co., 1987. ISBN: 0394358260.

[212] S. Wuchty. What is a social tie? *Proceedings of the National Academy of Sciences*, 106(36):15099–15100, sep 2009.

[213] Stefan Wuchty and Brian Uzzi. Human communication dynamics in digital footsteps: A study of the agreement between self-reported ties and email networks. *PLoS ONE*, 6(11):e26972, nov 2011.

[214] Yanni Yang, Alex Pentland, and Esteban Moro. Identifying latent activity behaviors and lifestyles using mobility data to describe urban dynamics. *EPJ Data Science*, 12(1), May 2023.

[215] Richard A. Zeller, Steven L. Nock, and Edward G. Carmines. Measurement in the social sciences: The link between theory and data. *Contemporary Sociology*, 11(1):79, January 1982.

[216] Bin Zhou, Petter Holme, Zaiwu Gong, Choujun Zhan, Yao Huang, Xin Lu, and Xiangyi Meng. The nature and nurture of network evolution. *Nature Communications*, 14(1), November 2023.

Aalto-DT 188/2024

The combination of network science, sociology, and high-granularity data sources has provided unique insights into social systems, enabling us to understand how small-scale human behavior scales up to shape society. Our lives are networked and embedded in an ever-evolving tapestry of social interactions, where individual behavior and dynamic relationships interact with and within social spaces, times, and identities.

This doctoral thesis builds on a large body of work in network science and sociology, offering new insights into two major subfields of social network analysis. First, it explores how social behavior—ranging from tie strength to overlapping social roles—is reflected in communication mediums. Second, it examines how networked interactions can have large-scale effects on social systems, including increased polarization and structural inequality.

9 789526 420127

ISBN 978-952-64-2012-7 (printed)
ISBN 978-952-64-2013-4 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science

www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
THESES