UNIVERSITY OF BIRMINGHAM

Note: Please sign this form and present it to the Bindery with your thesis, so that it can be bound at the front of the Library deposit copy

RESEARCH THESIS
Author's Declaration

Full name (block capitals, surname first): STICH CHRISTOPH

Full title of thesis/dissertation (block capitals): HUMAN MOBILITY AND SOCIAL TIES IN CONTEXT: FROM PLACES TO PERSONALITY

College/School/Department (block capitals): GEOGRAPHY, EARTH AND ENVIRONMENTAL SCIENCES

Date of award of degree (leave blank):

1. I understand that one printed and one electronic copy of my thesis/dissertation (the Work) will be deposited in the University Library (the Library) and in a suitable electronic repository, for permanent retention.
2. Without changing the content, the Library or any third party with whom it has an agreement to do so, may convert either copy into a format or medium for the purpose of long-term retention, accessibility and preservation.
3. The Library will publish, and/or arrange with appropriate third parties for the non-exclusive publication of, a bibliographic description of the thesis/dissertation, and the author's abstract.
4. Unless arrangements are made to the contrary, (see paragraph 6. below), the Library is authorised to make the Work available for consultation in the Library, and via a recognised inter library loans system. The Library is authorised to make the electronic copy of the Work freely accessible to individuals and institutions - including automated agents - via the Internet.
5. Rights granted to the University of Birmingham through this agreement are entirely non-exclusive. I retain all my rights in the Work in its present version or future derivative works.
6. I understand that I may apply to the University to retain the right to withhold access to the content of my thesis/dissertation. Access to the paper version may be withheld for a period which shall not normally exceed four calendar years from the congregation at which the degree is conferred. The actual length of the period will be specified in the application, together with the precise reasons for making that application. The electronic copy may be withheld from dissemination via the web or other networks for any period.
7. I have obtained permission for any use made of substantial amounts of published or unpublished copyright material (text, illustrations, etc) where the rights are owned by a third party, other than as permitted under either The Copyright Designs and Patents Act 1988 (as modified by any related or successor legislation) or the Terms and Conditions of any Licence governing its use.
8. The content of the copies I shall deposit with the Library will be the final version of my thesis, as approved by the Examiners.
9. I understand that the Library and administrators of any electronic theses repository do not hold any obligation to take legal action on behalf of myself, or other rights holders, in the event of a breach of intellectual property rights, or any other right, in the material deposited.
10. I understand that, in the event of my thesis/dissertation being not approved by the Examiners, this declaration will become null and void.

Signature: ……………………………………… **Date:** 04.03.2020

For Library use (please leave blank):

Classmark: **Accession number:** **Control number:**

eTheses Repository url:

Human Mobility and Social Ties in Context: From Places to Personality

by

Christoph Stich

A thesis submitted to
The University of Birmingham
for the degree of
Doctor of Philosophy

School of Geography, Earth, and
Environmental Sciences
The University of Birmingham
January 2020

Abstract

Recent years saw an increasing proliferation of the use of digitally generated traces of data for understanding human behaviour. The quantitative understanding of social networks as well as patterns of human mobility benefited tremendously from these new sources of data. The main dynamics of both social networks and human mobility such as a propensity of humans for heterogeneous behaviour, how humans choose to explore new places, or the fact that both spheres are intrinsically linked are now fairly well understood.

However, how various other factors mediate the observed dynamics is still relatively unknown, not least due to the difficulty in obtaining adequate data. Thus, for my thesis I focus on how a variety of factors—places, longer-term dynamics, the personality of individuals, or neighbourhoods—might be a driver of various aspects of social and mobility behaviour.

I used data from the Copenhagen network study that tracked 847 students with smartphones and measured their social encounters as well as the locations they visited for a whole academic year. I further utilised a variety of methods for analysing the data ranging from applied machine learning over inferential statistics to social network analysis. Using this dataset, I found that the qualities of places were very informative for understanding future encounters between students, that the longer-term dynamics shaped both social

and mobility behaviour, and that while personality had a significant effect on the observed regularity of behaviour, its effect was rather small.

Acknowledgements

I would like to thank everyone. I am especially thankful to my first supervisor Emmanouil Tranos for supporting me every step of the way; I could not have wished for a better supervisor and I am sure that without him the journey to becoming an independent researcher would have been much harder and a lot less enjoyable. I am also very grateful for the insightful discussions with my second supervisor Mirco Musolesi. The support I received from Sune Lehmann was instrumental for not only accessing the dataset used in this thesis but also for shaping the ideas that became this thesis.

I would like to express my gratitude to Vedran Sekara for helping me derive the social cores for gatherings of students,[1] Piotr Sapiezynski for providing access to the improved location traces,[2] Lenka Mudrova for very helpful insights and discussions that benefited this thesis, and Katy Mawhood for proof reading the thesis.

I would also like to thank both the Economic and Social Research Council (Grant ES/J50001X/1) and the School of Geography, Earth, and Environmental Sciences at the University of Birmingham for providing financial support during my PhD study.

Furthermore, my thesis would not have been possible without the following open source

[1]As described in Sekara, Stopczynski, and Lehmann (2016)

[2]Sapiezynski, Stopczynski, Gatej, and Lehmann, 2015

software: *Python*,[1] several, external *Python* libraries,[2] *R*,[3] and *Latex*.[4]

Finally, greatest thanks go to my partner Lenka, who was always there for me, to my friends, who stuck with me all those years, and to my family, who made me who I am today. Thank you little sister, mum, and dad as well as all the others.

[1]van Rossum, 1995

[2]Hunter, 2007; Jones, Oliphant, and Peterson, 2001; McKinney, 2010; Pedregosa and Varoquaux, 2011; Perez and Granger, 2007; Seabold and Perktold, 2010; van der Walt and Aivazis, 2011; Waskom et al., 2018

[3]R Core Team, 2019

[4]https://www.latex-project.org/

Contents

List of Figures

List of Tables

"Begin at the beginning", the King said gravely, "and go on till you come to the end: then stop"

Lewis Carroll

1 Introduction

Every day countless people commute to work, go out to bars and restaurants in the evening, visit and message their friends, and carry with them phones that can track their movement and social interactions. Moving from A to B and socialising with others are so ingrained into our lives that it stands to wager that most of us do not think twice about the infrastructure that enables us to move about so freely, about the digital systems that allow communication between far flung places almost instantaneously, and about the huge amount of time we dedicate on grooming and maintaining social relationships. Humans spend an inordinate amount travelling and socialising with others, and not without good reason: both social relationships and mobility are key aspects of human life.

1.1 A Short History of Social Network Analysis & Mobility Studies

Social networks, with whom we form social ties, are central to who we are as humans. Social relationships shaped the development of children (Cochran & Brassard, 1979), were related to a variety of health outcomes (Perkins, Subramanian, & Christakis, 2015), and potentially gave rise to our unusually large brains (Dunbar, 1993; Dunbar, 2009).

Given the salient role social ties play for humans, social network analysis became a prolific field of study in recent decades. The modern field of social network analysis consists of four defining axioms (Freeman, 2011):

1. the notion that ties between actors are salient,
2. collection of data related to links between actors,
3. a heavy reliance on graphs to study patterns, and
4. models to describe and explain the observed patterns.

It was not until the 1930s that researchers used all those four properties together and thus the modern field of network analysis emerged (Freeman, 2011). In the 1970s social network analysis eventually became widely recognised as an independent field of research. (Freeman, 2011). According to Borgatti, Mehra, Brass, and Labianca (2009) among the two most influential works in this period were: the work of Lorrain and White (1971) on structural equivalence and Granovetter's (1973) work on the strength of weak ties. Lorrain and White (1971) were among the first to start to take network structure rather than individuals into account. They proposed that nodes that were structurally equivalent (i.e. nodes that have exactly the same set of ties to others) occupy equivalent social positions in the network and should thus be treated similarly. Their work proved highly influential, even though in practice the criterion of identical sets of ties was soon relaxed by other researchers (Freeman, 2011). Granovetter (1973) argued that in “strongly” knit communities there was a high degree of transitivity between nodes; this being the defining feature of a socially tightly integrated group. However, this could be detrimental for spreading new information. The information circulating in a group was to some degree redundant as almost everyone talked to everyone else. “Weak” ties to members outside of the group on the other hand were by definition not well connected to the group itself. They could thus act as bridges to other communities and access novel sources of information. Slightly later Feld (1981) also published his influential work on the focused organization of ties. He discovered that ties between individuals did not form randomly but were often organised around external focal points such as sport clubs or shared workplaces.

In the late 1990s physicists started to publish studies on social networks and thereby revolutionised the field (Borgatti et al., 2009; Freeman, 2011). Watts and Strogatz (1998) wrote about the small-world property of social networks and showed that distances between nodes were significantly shorter than those expected by chance. Barabasi and Albert (1999) analysed the degree distribution of nodes and found that they were heavily skewed towards the nodes that already had the majority of ties. While Watts and Strogatz (1998) and Barabasi and Albert (1999) were the first physicists to publish on social networks, a deluge of publications by other physicists on social networks soon followed (Freeman, 2011). While physicists often rediscovered findings of earlier social network researchers such as the small world property of networks or the inequal distribution of ties (Freeman, 2011), they also contributed to new findings such as that social gatherings were composed of a stable core of members (Sekara et al., 2016) or that mobility patterns and social behaviour were intrinsically linked (Alessandretti, 2018). As time went on the two communities, traditional social network researchers and physicists, seemed to have found a common language and now formed a more coherent research field (Borgatti et al., 2009; Freeman, 2011).

In the same vein that social ties are of utmost importance to humans, mobility or the propensity of humans to travel from one place to another, is another key aspect of everyday life, from hunter-gatherers (Kelly, 1983) to modern day commuters (Lima et al., 2016). The amount of time spent on travelling stayed remarkably stable over human history (Marchetti, 1994), which suggests that mobility is indeed an innate quality of being human.

With a rapidly urbanising landscape in the later half of the 19th century came an increasing interest to not only study life in cities (Simmel, 1969) but also the flows of people (Greenwood & Hunt, 2003). While lifetime migration data were available as early as the 1850s for the US and the UK (Greenwood & Hunt, 2003), arguably it took several decades for the first quantitative study of migration (Ravenstein, 1885) to be published in the 1880s. Beginning with the 1940s census data began to include information about individuals' prior residence and in 1960 the first study utilising microdata on migration was conducted (Greenwood & Hunt, 2003).

It was, however, not until Torsten Hägerstraand's seminal article "What about people in regional science?" in 1970 that modern studies of human mobility emerged specifically accounting for individual behaviour (Hägerstrand, 1970). Before then researchers studied mobility mostly on an aggregate level and focused on the "modal man" and "modal woman" (Hägerstrand, 1970) and applied aggregate level models such as the Gravity Model (Greenwood & Hunt, 2003). Hägerstrand was one of the first to suggest that, conversely, there were fundamental links between the micro-situation of the individual and the aggregate outcome.

The approach became later to be known as "time-geography" or "time-space geography". Time-geography in short provides a framework for the study of socio-environmental mechanisms and how events occurred in both time and space (Ellegard & Svedin, 2012). It built upon the insight that concurrently one needs a "way of finding out the workings of large socio-environmental mechanisms" and that "the individual is indivisible" (Hägerstrand, 1970, p. 20f). At the behavioural level time geography was distinct in its attempt to capture the movements and activity patterns of people instead of aggregates.

While in the 1980s time-geographic methods had fallen out of favour due to a lack of tools and individual level data, in the 1990s new GIS-based research led to a resurgence of the field (Neutens, Witlox, & Demaeyer, 2007). Eventually time-geography approaches led to important contributions such as assessing accessibility on an individual level, new models of travel behaviour, improved spatial decision making, and understanding network-related travel possibilities (Neutens et al., 2007; Neutens, Schwanen, & Witlox, 2011).

It is noteworthy that around the same time Hägerstrand proposed his new framework for understanding mobility, social network analysis gained traction as a research field. Both approaches tried to understand complex phenomena from the bottom up. As Alessandretti (2018) noted, Thomas Schelling developed one of the earliest agent-based models (Schelling, 1971) around the same time as well. Schelling again studied how individual actions can lead to emergent properties of the whole. In his example, he showed how only a slight preference for similar neighbours could lead to segregated communities over time.

The 1970s, thus, marked a shift from the study of aggregate statistics about human behaviour to the quantitative study of individual human behaviour. Interestingly this epistemological shift coincided with the proliferation of computing power and the widespread adoption of the micro-processor in the 1970s (Ceruzzi, 2003) as well as the beginnings of the study of explicitly complex systems (Vemuri, 1978). Although, until very recently the lack of suitable individual level data hampered development of both understanding individual mobility and social relationships on a quantitative level (Watts, 2007; Lazer, Brewer, Christakis, Fowler, & King, 2009). After all both, social network analysis and studies of human mobility, heavily depend on data about individual human behaviour.

1.2 Contribution & Chapters

Traditionally data about both social networks and human mobility was expensive to collect and thus comparatively less well studied on a larger scale (Lazer et al., 2009). However, as everyday behaviour became increasingly mediated by digital technologies, humans left ever more digital traces of their behaviour behind. Consequently, both fields have seen a torrent of new quantitative studies dealing with various aspects of both social and mobility behaviour. For Chapter 2, I thus review the main strains of both fields with an emphasis on studies using digital traces of behaviour. In general, the main drivers and dynamics of both social networks and human mobility patterns are now relatively well understood. Various studies (Alessandretti, 2018; De Domenico, Lima, & Musolesi, 2013; Larsen, Urry, & Axhausen, 2006; Cho, Myers, & Leskovec, 2011; Scellato, Noulas, & Mascolo, 2011) furthermore showed that both mobility behaviour and social networks were intrinsically linked. What remains less well studied is how various other factors might (or might not) shape both social as well as mobility behaviour. That is not to say that other factors were completely disregarded by previous work. The most obvious factors such as time, age, socio-economic status, gender, race, personality, etc. and their effect on both social networks and mobility behaviour have indeed been studied (Section 2.5). Nevertheless, a lack of suitable multi-dimensional data greatly hindered the analysis of potential, other mediating factors such as the role of places for predicting future encounters, how longer term dynamics shape the interplay between mobility and social ties, and how personality traits shape regularities of behaviour.

Thus, for my thesis I aimed to evaluate light on how a variety of contextual and mediating factors influenced observed patterns of social and mobility behaviour. Given access to data collected via the Copenhagen Network Study (CNS), which I describe in Chapter 3, I considered how the following various factors shaped social and mobility behaviour and the relationships between the two, using a variety of methods ranging from social network analysis over time series analysis to applied machine learning. I also briefly review the methods and metrics that are common to several chapters in Chapter 3. The three mediating factors that I researched for my thesis were:

First, I tried to understand the role of places for predicting future co-occurrences. There is evidence that the type of places hold discriminatory power for predicting future co-occurrences. However, current studies (Scellato, Noulas, & Mascolo, 2011; Yang, Chawla, Basu, Prabhala, & La Porta, 2013; Sekara et al., 2016) failed to account for potential confounding factors. In Chapter 4, I consequently analysed the role the type of places play for predicting future encounters, while simultaneously accounting for social, spatial, and temporal variables.

Second, aggregated measures of both social ties and mobility behaviour seemed to have an influence on how much people travel and how much they socialise (Viry, Kaufmann, & Widmer, 2009; Stanley et al., 2011). Nonetheless, the current literature focused mainly on more traditionally acquired data and was ambiguous to whether more social ties increase or actually decrease mobility. Hence, in Chapter 5, I analysed the interaction between social and mobility behaviour at longer time-scales.

Third, while it was established that human behaviour with regards to both social ties and travel is to a large extent regular (Clauset & Eagle, 2007; Song, Qu, Blumm, & Barabasi, 2010), what effect personality might have on the regularity of behaviour was largely unknown. In Chapter 6, I looked to what extent personality traits are a driver of regularity of social and travel behaviour.

In short, my contribution consisted of shedding light on the role various mediating factors play for both social ties and patterns of human mobility such as place, long-term temporal dynamics, and personality. Chapter 7 then summarises the findings and lays out further avenues for research.

There are two motives for reading a book; one, that you enjoy it;
the other, that you can boast about it.

Bertrand Russell

2
Literature review

As this thesis was very much situated at the intersection of human mobility and social networks, I first give an overview of how digital traces of behaviour are re-shaping research into human behaviour. I then briefly review the existing literature of both fields with a particular focus on digital traces of human behaviour. As my research for this thesis addressed how mediating and contextual factors shaped the observed dynamics, I also briefly discuss previous studies dealing with various other factors and their effects on behaviour.

2.1 Digital Traces of Behaviour

Recent years saw an increasing proliferation of the use of digitally generated data for understanding human behaviour. The invention of the internet and the widespread adoption of mobile devices created a platform to digitally capture most aspects of everyday life with a high spatial and temporal resolution (Arribas-Bel, 2014). Digitally generated data allowed the study of human behaviour in a variety of contexts on very different spatial and temporal scales ranging from patterns of mobility and its predictability (De Domenico et al., 2013; Eagle & Pentland, 2006; González, Hidalgo, & Barabási, 2008a; Sadilek & Krumm, 2012; Schneider et al., 2013; Song, Qu, et al., 2010) over human behaviour in economic arenas (Radicchi, Baronchelli, & Amaral, 2012; Preis, Moat, & Stanley, 2013) to the analysis of political trends (Adamic & Glance, 2005; Carpenter, Esterling, & Lazer, 2004). Human mobility itself was governed by processes at various spatial scales ranging from intra-urban (Noulas, Scellato, Lambiotte, Pontil, & Mascolo, 2012; Reades, Calabrese, & Ratti, 2009; Wu, Zhi, Sui, & Liu, 2014), over inter-urban (Liu, Sui, Kang, & Gao, 2014; De Montis, Barthelemy, Chessa, & Vespignani, 2007), and national (Brockmann, Hufnagel, & Geisel, 2006; Sobolevsky et al., 2014), to global (Barrat, Barthelemy, Pastor-Satorras, & Vespignani, 2004; Balcan et al., 2009; Hawelka et al., 2014) dynamics. Data for these studies was predominantly collected via online social networks, financial transactions, GPS enabled devices, and mobile phones.

Furthermore, social relationships and interactions between humans were studied in a variety of settings. Researchers used online social media (Grabowicz, Ramasco, Gonçalves, & Eguíluz, 2014; Goncalves, Perra, & Vespignani, 2011; Takhteyev, Gruzd, & Wellman, 2012; Volkovich, Scellato, Laniado, Mascolo, & Kaltenbrunner, 2012), online communications (Burke & Kraut, 2014; Johansen, 2004; Kossinets & Watts, 2006), call data records (Blondel, Decuyper, & Krings, 2015; Onnela et al., 2007) to study social networks and the dynamics unfolding within them. Another strain of research utilised Bluetooth, WiFi and RFID enabled devices to study the proximity of individuals (Stopczynski et al., 2014; Sekara et al., 2016; Clauset & Eagle, 2007; Eagle & Pentland, 2006) and their face-to-face interactions (Isella et al., 2011; Cattuto et al., 2010; Starnini, Baronchelli, & Pastor-Satorras, 2013; Zhang, Li, Xu, & Vasilakos, 2015).

The proliferation of digitally recorded interactions and behaviour was beginning to lift the traditional constraints on research into individual human behaviour as each online interaction and increasingly also our movements through space leave behind digital bread crumbs of human behaviour. No longer were the only methods available to send out surveys or to interview people. To generate data about social phenomena and behaviour via surveys or observation was a rather costly and laborious process; moreover, studies that relied on self-reports to, for example, study social ties or travel behaviour suffered from cognitive bias, errors of perception and framing ambiguities (Marsden, 1990; Watts, 2007). As a result historically studies about individual human behaviour were relatively small in terms of sample size (Lazer et al., 2009; Watts, 2007).

These new studies of behaviour improved our quantitative understanding of phenomena that had previously often only been considered from a qualitative point of view (Barabasi, 2005; Eagle & Pentland, 2009; Alessandretti, 2018; Williams, 2013). Thus, they potentially bridged the gap between a very quantitative world-view and the social sciences (Lazer et al., 2009); a development that had and still has tremendous implications for research concerned with the social world. Especially since more and more of our communications and interactivity happen online or leave behind digital traces of behaviour. For example, the temples of consumerism, shopping malls and high-streets, were increasingly faced with consumers, who ordered their goods online and as a consequence become ever more often deserted (Zhang, Zhu, & Ye, 2016). Facebook's dominance in shaping our interactions online, led to calls to safeguard democratic elections from *undue* interference in how ideas and information were spread via the platform (D'Ancona, 2018). Not only had smartphone usage become near ubiquitous in the Western world, but nomophobia—smartphone addiction—was proposed to be included in the Diagnostic and Statistical Manual of Mental Disorders (Bragazzi & Del Puente, 2014).

For the first time, one could begin to study interactions of a multitude of humans and their movements through space at a resolution that was sensitive to individual level effects. As a result a new type of science into human behaviour was emerging that leveraged the capacity to collect and analyse data both at an unprecedented scale and scope (Lazer et al., 2009). As data had unequivocally transformed fields such as biology and physics, the emerging field of computational social science held the promise of potentially transforming our understanding of our lives, organizations, and societies in a fashion that was barely conceivable just a few years ago (Lazer et al., 2009). In this view Big Data might provide a new lens with which to look at the social realm.

Yet, it was not the sheer size of those new data sources alone, that made it such a valuable tool for researching human behaviour. Often sources of Big Data enabled researchers to apply filters to find the salient pieces of information or to aggregate the data into the right temporal or spatial resolution (González-Bailõn, 2013). They further enabled longitudinal research into patterns of behaviour and allow the capture the dynamics of complex systems (Holme, 2015). Traditionally longitudinal data about social networks and individual mobility was relatively scarce due to the required effort in collecting the data (Xu et al., 2015; Zhao, Stehlé, Bianconi, & Barrat, 2011; Watts, 2007).

Additionally the deluge of digital traces of behaviour might lead to a dramatic change regarding what it means to do research in the humanities and social sciences. Several authors (Berry, 2011; Kitchin, 2013, 2014; Rabari & Storper, 2015) argued that digital technologies and the increased availability of data were affecting both the epistemologies as well as the ontologies of social scientific research. Many of the statistical techniques currently in use in Geography were created in a context of relatively sparse datasets and new methods might be required for the analysis of spatial data (Arribas-Bel, 2014; Kitchin, 2013). Furthermore, historically, the data used by social scientists was aggregated into pre-determined categories such as "city", "census tract", "industry", "age", "occupation", and so on (Rabari & Storper, 2015). While those data categories were built upon theoretical assumptions about the social world, they also hid a considerable amount of heterogeneity (Rabari & Storper, 2015) Computer science, thus, could very well start to play a foundational role for the humanities, "supporting and directing their development and issuing lucid directives for their inquiry" (Berry, 2011, p. 9) by both providing new detailed sources of data as well as new tools for analysing said data.

As data and computational models became more important for doing social scientific research, the traditional social sciences could be losing their dominant position as gatekeepers for understanding the social world (Savage & Burrows, 2007) and social scientists might be cut off from doing potentially interesting studies altogether (Kitchin, 2013). Big Data posed a fundamental challenge to the authority of the social sciences to define knowledge about the social as "[i]t permits a dramatically increased range of other agents to claim the social for their own" (Burrows & Savage, 2014, p. 5). This development was particularly worrisome as corporations whose operations and customers generate a lot of digital traces in the first place, became de facto gatekeepers of knowledge about the social world (Lazer et al., 2009; Savage & Burrows, 2007). Not even democratic institutions could easily supervise who holds what data about whom. For example, the House of Commons needed to resort to an archaic procedure to force a plaintiff in a US american court case against Facebook to hand over documents that would have otherwise been unavailable to parliament but deemed crucial for investigating Facebooks's practices (Cadwalladr, 2018). And as algorithms that used these data as input shaped ever more aspects of our lives, closed off collections of data had implications on daily life not just for the way social scientific research was conducted.

Notwithstanding the challenges digital traces posed to social scientific research and to society as a whole, studies using them clearly shed light on a variety of novel behavioural insights. In particular, the area of human mobility as well as social networks saw a plethora of studies using digital traces. The fields of mobility studies and social networks were easily among the most prolific areas of study concerning individual human behaviour using digital traces. Especially, data generated by mobile phones was shown to be useful for both fields and allowed the study of the link between patterns of human mobility and dynamic social networks. Recent examples include an upper limit to the predictability of human mobility (Song, Qu, et al., 2010), evidence for a link between social ties and human mobility (Alessandretti, Sapiezynski, Sekara, Lehmann, & Baronchelli, 2018), and how relatively simple community structures seem to govern the evolution of networks (Sekara et al., 2016).

2.2 Social Ties

While the power of a network approach lies in its ability to simplify complex systems (Holme, 2015), acquiring data was historically a key limitation. Until relatively recently collecting data about human networks was costly and prone to error (Watts, 2007). Not least due to the difficulty of collecting data, the standard approach to analysing networks was to treat associations between nodes as temporally invariant and permanent (Blonder, Wey, Dornhaus, James, & Sih, 2012). Researchers shifted towards studying temporal dynamics of social networks only relatively recently (Blonder et al., 2012; Holme & Saramaki, 2012; Holme, 2015).

Consequently, studies utilising temporal data to study networks became more popular (for an overview see Holme and Saramaki (2012) and Holme (2015)); of particular interest were how networks themselves evolved (Krings, Karsai, Bernhardsson, Blondel,

& Saramäki, 2012; Kossinets & Watts, 2006, 2009; Kovanen, Karsai, Kaski, Kertesz, & Saramaki, 2011) and how dynamic processes propagated through networks as connections between individuals could serve to spread infectious diseases, ideas, information, and gossip (Lima, 2016; Karsai, Iniguez, Kaski, & Kertész, 2014; Kramer, Guillory, & Hancock, 2014; Lu, Roberts, Lio, Dunbar, & Crowcroft, 2009; Meo, Ferrara, Fiumara, & Provetti, 2014; Iribarren & Moro, 2009; Tizzoni, Sun, Benusiglio, Karsai, & Perra, 2015; Holme & Liljeros, 2014; Perra, Gonçalves, Pastor-Satorras, & Vespignani, 2012; Rocha & Blondel, 2013; Romero, Meeder, & Kleinberg, 2011; Williams & Musolesi, 2016). Furthermore, in recent years studies of networks began to include several layers of analysis concurrently (Boccaletti et al., 2014). Examples include the simultaneous analysis of calls, texts, and physical interactions to account for different dynamics in other layers of the multiplex network (Eagle & Pentland, 2006; Stopczynski et al., 2014) or the analysis of spatio-temporal networks that considerd both space and time simultaneously (Williams & Musolesi, 2016).

The expansion in scope and complexity of the study of human social networks very heavily relied on broader and more detailed sources of data. Typical sources of data nowadays include call data records of mobile phone users (Karsai, Kaski, Barabasi, & Kertesz, 2012; Kovanen, Kaski, Kertesz, & Saramaki, 2013; Miritello, Lara, Cebrian, & Moro, 2013), emails (Bird, Gourley, Devanbu, Gertz, & Swaminathan, 2006; Holme & Liljeros, 2014), online social networks (Kanai, Bahrami, Roylance, & Rees, 2011; Romero et al., 2011; Nguyen & Szymanski, 2012; Takhteyev et al., 2012), and scientific studies that track their users either via RFID badges (Cattuto et al., 2010) or mobile phones (Eagle & Pentland, 2006; Stopczynski et al., 2014; Wang et al., 2017).

Using the above mentioned new sources of data several characteristic features of social networks had either been identified in recent years or confirmed on a much larger scale. In particular, I will briefly discuss the following nine characteristics:

1. the propensity of "heterogeneous" (i.e. diverse) activity in networks,

2. differences in tie strength (most often between weak and strong ties and differences in how individuals chose to maintain their ties),

3. preferential attachment of nodes to others that already have a high degree of connections,

4. the tendency for friends-of-friends to become friends themselves,

5. the emergence of community structures in networks,

6. small-world properties of real life networks,

7. similarity of individuals that form ties, and

8. the bursty nature of temporal interactions in networks,

9. the regularity of social interactions.

For a graphical overview of how these concepts related to each other see Figure 2.1.

2.2.1 Heterogeneities

Observed social networks were characterised by a propensity of individuals for heterogeneous activity in the network (Barrat et al., 2004; Cattuto et al., 2010; Candia et al., 2008; Ghoshal & Holme, 2006; Lima, 2016; Kim & Altmann, 2017; Karsai, Perra, & Vespignani, 2014; Perra et al., 2012; Onnela et al., 2007; Starnini et al., 2013; Stehle et al., 2011; Ubaldi et al., 2016), which significantly impacted social dynamics playing out within the network (Min, Goh, & Vazquez, 2013; Rocha & Blondel, 2013). While the behaviour of individuals itself was fairly static, the distribution of behaviour that could be observed was rather wide; in other words more heterogeneous than one would initially expect. The size of individuals' networks differed considerably (Miritello et al., 2013; Roberts, Dunbar, Pollet, & Kuppens, 2009), even though the overall size of an

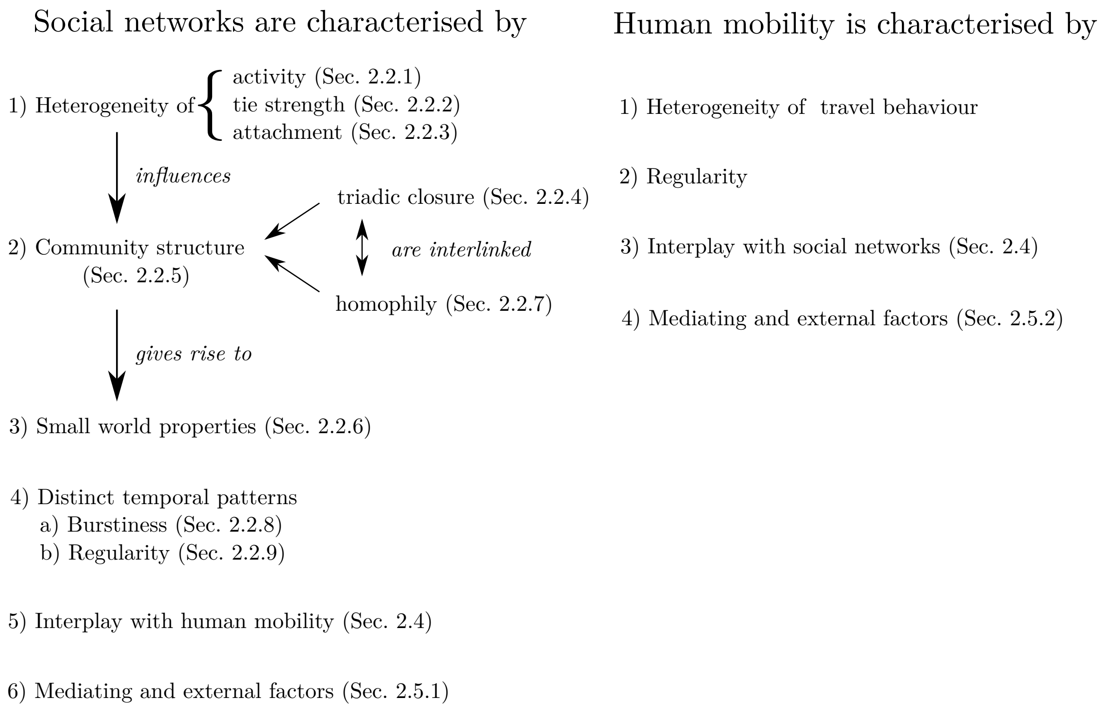

Figure 2.1: Main Discussed Concepts and How They Are Related

The figure provides an overview of the main concepts discussed in this literature review. On the left side are the main characteristics of social networks as discussed in this chapter; on the right side, are the main features of human mobility as presented in this chapter. Note, that several factors such as heterogeneity, regularity, interplay of social networks and mobility, and the existence of mediating factors shaped both social networks and human mobility.

individual's social network seemed to be limited to around 150 (Dunbar, 1993; Hill & Dunbar, 2003). Furthermore, individuals showed differences in the network structure of their ego-network and the turnover of ties (Centellegher, Lopez, Saramaeki, & Lepri, 2017), where ego-networks are the subset of ties that are immediately adjacent to each individual in the graph. In general, social relationships followed a skewed distributions (in the case of social networks the right tail of the distribution was usually long and heavy) and frequently a power-law, $P(k) \sim k^{-\beta}$, or exponential function, $P(k) \sim e^{-\lambda k}$ (Arnaboldi, Conti, Passarella, & Pezzoni, 2013; Song, Wang, & Barabasi, 2012; Hossmann, Spyropoulos, & Legendre, 2011; Tizzoni et al., 2015), where the shape of the distribution usually indicates an inequal distribution of ties.

It is noteworthy here that whether a distribution was scale-free or exponential with a heavy-tail was an active debate (Stumpf & Porter, 2012). Stumpf and Porter (2012, p. 665) even went so far as to argue that "[m]ost reported power-laws lack statistical support and mechanistic backing" and that the knowledge that the distribution was heavy tailed was far more important than whether it follows a power-law or not.

These heterogeneities could partially be explained by various factors external to the network itself. Gender (Dunbar & Spoors, 1995; Igarashi, Takai, & Yoshida, 2005), age (Ajrouch, Blandon, & Antonucci, 2005; Sander, Schupp, & Richter, 2017; Wrzus, Hänel, Wagner, & Neyer, 2013; Zhaoyang, Sliwinski, Martire, & Smyth, 2018), socio-economic status (Ajrouch et al., 2005; Campbell, Marsden, & Hurlbert, 1986), physical attractiveness or fitness (Ali, Amialchuk, & Rizzo, 2012; Reis, Nezlek, & Wheeler, 1980), personality characteristics (Alessandretti, 2018; Kalish & Robins, 2006; Pollet, Roberts, & Dunbar, 2011; Staiano et al., 2012; Lu et al., 2009) and the built-environment (Boessen, Hipp, Butts, Nagle, & Smith, 2017; Xu, Belyi, Bojic, & Ratti, 2017) all influenced how individuals socialised in networks. Moreover, there is evidence that personality traits (i.e. fundamental traits) shaped how individuals formed ties (Wehrli, 2008) and exchanged alters in their network (Centellegher et al., 2017).

Notwithstanding the external factors, Ubaldi et al. (2016) suggested two main dynamics that led to the observed heterogeneity: different levels of activity as well as a different strategies for allocating resources.

Firstly, not everyone was equally socially active in a given network and participated at a different rate (Cattuto et al., 2010; Iribarren & Moro, 2009; Perra et al., 2012; Muchnik et al., 2013; Starnini et al., 2013; Ubaldi et al., 2016). As activity and degree distribution of an individual were highly correlated (Muchnik et al., 2013) variations in activity resulted in a diverse number of observed interactions. Furthermore, several analytical models were proposed that linked the heterogeneity of activity to the heterogeneity of degrees (Perra et al., 2012; Starnini et al., 2013; Moinet, Starnini, & Pastor-Satorras, 2015; Ubaldi et al., 2016) and spreading processes (Iribarren & Moro, 2009; Tizzoni et al., 2015). This finding held for a variety of different types of networks ranging from dating sites (Ghoshal & Holme, 2006) over call data records (Onnela et al., 2007; Saramäki & Moro, 2015) to face-to-face networks (Starnini et al., 2013).

Secondly, individuals varied in how they allocated their resources to form ties; in general they had a choice of “bonding” within their group or ”bridging” communities (Putnam, 2000). For Putnam (2000) the trade-off between inclusive and exclusive tie formation was one of the most essential features of social life. Some might have favoured re-visiting a limited number of ties frequently, thus bonding and strengthening the connection between them and the other person. Others might have chosen to explore a range of weaker social contacts, potentially leading to more diverse but weaker ties (Granovetter, 1973; Levin & Cross, 2004; Putnam, 2000; Karsai, Perra, & Vespignani, 2014).

2.2.2 Tie Strength

In general, the heterogeneity of tie strengths between nodes of a network is a typical feature of human networks and especially the dichotomy between strong and weak ties was a key ingredient of empirically observed social networks (Granovetter, 1973; Meo et al., 2014; Karsai, Perra, & Vespignani, 2014; Kumpula, Onnela, Saramäki, Kaski, & Kertész, 2007; Kumpula, Onnela, Saramäki, Kertész, & Kaski, 2009; Onnela et al., 2007; Song et al., 2012). Among others, it could be found in networks in the workplace (Gee, Jones, & Burke, 2017; Levin & Cross, 2004; Zenou, 2015), in online social networks (Goncalves et al., 2011; Meo et al., 2014), in collaboration networks (Karsai, Iniguez, et al., 2014; Newman, 2001), and communication networks (Ara & Van Alstyne, 2011; Onnela et al., 2007; Saramäki & Moro, 2015).

The source of the observed heterogeneities in tie strengths could again be traced to a trade off between competing factors and how individuals tried to resolve it. Social relationships generally imbued individuals with various advantages, while imposing a maintenance cost (Takano & Fukuda, 2017). Moreover, there seemed to be an upper limit to the overall complexity of social network individuals could maintain (Dunbar, 2009; Miritello et al., 2013). In other words, there was an upper limit to the total number of ties and the emotional intensity of each relationship they could maintain. Consequently, humans needed to balance the following three competing objectives (Takano & Fukuda, 2017):

First, humans were unsurprisingly inherently social and had a need for close relationships. They further received various benefits from maintaining social ties. Previous researched showed that social ties played an important role for mental health (Kawachi & Berkman, 2001), physical well-being (Helliwell & Putnam, 2004) and even morbidity (Holt-Lunstad, Smith, & Layton, 2010).

Second, it took effort to maintain social ties, which limited the capacity of an individual to socialise (Dunbar, 1998). Especially, since the time costs of maintaining social ties were significant. In line with other primates, humans across various cultures spent about 20% of their waking time on grooming relationships with others (Dunbar, 1998). And there was a clear limit to how much time individuals could and were willing to invest in ties (Saramaki et al., 2014). And the deeper the relationship the more time was needed to maintain it. Not only was network size negatively related to emotional closeness of the ties of an individuals (Roberts et al., 2009) but also contact frequency was linked to emotional closeness (Hill & Dunbar, 2003). Takano and Fukuda (2017) found that the total number of ties N was inversely proportional to the mean tie strength m^a, where $a > 1$, which indicated the higher cost of social grooming for stronger ties.

Third, maintaining social ties in a coherent group was a relatively expensive cognitive task (Dunbar, 1993; Dunbar, 2009). The size and complexity of human social networks correlated with various factors of mental aptitude such as memory and processing constraints (Powell, Lewis, Roberts, García-Fiñana, & Dunbar, 2012; Stiller & Dunbar, 2007). Humans had thus also to make a trade-off between the cognitive costs of establishing new ties and the benefits they might accrue from their network (Dunbar, 1993; Dunbar, 2009; Powell et al., 2012; Stiller & Dunbar, 2007).

However, social networks were not static. Social networks, including tie strengths, were changing to various degrees (Centellegher et al., 2017; Kossinets & Watts, 2006). For example, Miritello et al. (2013), using call data records (CDR[1]), found that on average about 75% of links stayed active over a seven month period; a significantly larger percentage than would be expected by chance alone. Arnaboldi et al. (2013) by using Twitter data, found a much higher turnover rate of around 75% over just five months. The different rates of turnover between Twitter and mobile phone based ties might also be due to differences in tie strength itself.

[1]Phone companies record the calls and texts customers make within their network for billing purposes. Those records usually included a time stamp and the location to which base station a phone was connected (Lima et al., 2016).

Interestingly, even though there was a significant turnover of ties, how humans balance the competing interests off quantity and quality was highly unique and hardly varied with respect to both time and changes in network composition (Saramaki et al., 2014). Some, “social explorers”, allocated their efforts more uniformly, cast a wider net, and exhibited a frequent turnover of ties (Centellegher et al., 2017; Miritello et al., 2013). Others, “social keepers”, showed a high persistence in the ties they keep and a focus on their closest relationships (Centellegher et al., 2017; Miritello et al., 2013).

This apparent trade-off between quantity and quality was also apparent in how individuals chose the medium for social grooming and in the resulting network structure; higher grooming costs generally led to smaller but deeper networks (Takano & Fukuda, 2017). While using text communications via the internet, people developed rather wide and shallow networks (Arnaboldi et al., 2013), whereas more elaborate forms of interactions such as telephone calls and face-to-face were associated with stronger relationships (Burke & Kraut, 2014). Unsurprisingly more costly forms of grooming such as face-to-face interactions led to higher satisfaction than less intensive forms of grooming such as texting (Vlahovic, Roberts, & Dunbar, 2012).

Different types of ties also played different roles for the network. While weak ties appeared to be necessary to maintain a network’s structure and overall connectedness (Kumpula et al., 2009), strong ties played a crucial role for maintaining local communities (Onnela et al., 2007). Furthermore, weaker ties provided access to a wider range of non-redundant information (Granovetter, 1973; Putnam, 2000; Meo et al., 2014; Levin & Cross, 2004) and enabled a significant portion of salient discussions (Small, 2013). By definition in tightly knit groups there was a high degree of transitivity and strong ties hence may had a negative role for information spreading as strong ties constrain information in clumps of strongly connected social groups (Karsai, Perra, & Vespignani, 2014).

2.2.3 Preferential Attachment

Individuals, however, did not form new connections to alters randomly. The network structure, or who was connected to whom, influenced how new ties were established. One of the most well known and earliest proposed mechanisms was preferential attachment (Barabasi & Albert, 1999; Jeong, Néda, & Barabási, 2003; Newman, 2001) and recent research (Pham, Sheridan, & Shimodaira, 2015; House, Read, Danon, & Keeling, 2015; Klimek & Thurner, 2013) further corroborated the process of preferential attachment. Preferential attachment, in short, was the idea that the probability of a node to acquire a new link scales with the degree of that node (Barabasi & Albert, 1999). In other words, already popular nodes received more new ties than less popular ones as nodes preferentially attach to those that were already well connected. It could explain the observed scale-free or heavy tailed distributions observed in many networks (Pham et al., 2015). While Barabasi (2012) argued that preferential attachment was one of the most profuse concepts of network science and its impact was hard to miss, there was still some debate whether preferential attachment applies to social networks as well (Newman, 2008).

Preferential attachment had important implications for inequality that might be exacerbated by the network structure of unequally distributed connections. As social ties had a bearing on labour markets (Calvo-Armegol & Jackson, 2004; Finneran & Kelly, 2003), income (Dawid & Gemkow, 2013), health (Perkins et al., 2015), and quality of life (Cattell, 2001), an unequal distribution of ties could lead to unequal outcomes in a variety of contexts.

2.2.4 Closure

Yet, the process of preferential attachment alone was not sufficient to generate realistic networks. In particular, networks generated with just preferential attachment as a generative principle were not very realistic. They exhibited a very low clustering coefficient (Bianconi, Darst, Iacovacci, & Fortunato, 2014), a salient feature of real social networks (Watts & Strogatz, 1998), and no community structure (Bianconi et al., 2014), where communities are roughly defined as sub-graphs that have high internal density so that between group connections are relatively sparse.

Clustering in real life networks implied a high proportion of loops of short length (Newman, 2003). Triads, loops of length three, have been found to be especially prevalent in social networks (Klimek & Thurner, 2013; Kumpula et al., 2007). The intuition is that triads represent transitive relationships between nodes, or in other words friends-of-friends are also friends. The idea that friends-of-friends are more likely to become friends (also referred to as triadic closure or balanced triads) seemed to be a fundamental principle of how social networks were organised and was a fairly well established process in social networks (Bramoullé, Currarini, Jackson, Pin, & Rogers, 2012; Bianconi et al., 2014; Laurent, Saramäki, & Karsai, 2015; Klimek & Thurner, 2013; Kossinets & Watts, 2006, 2009).

Triadic closure was, however, not the only empirically observed process that could give rise to community structures in the network. Focal closure, by contrast, followed from an alternative theory of tie formation, that of "social interaction foci" (Feld, 1981). The idea was that an individual's life was structured around foci, which were defined as social, psychological, legal, or physical entities around which joint activities were organised. In essence, all the different groups, settings, institutions and places such as workplaces, voluntary organizations, sport clubs, families, etc. around which social life was often structured. These foci, that were external to the network itself, provided repeated opportunities for ties to form between individuals (Kossinets & Watts, 2006, 2009) and shaped the resulting network structure (Doreian & Conti, 2012). Foci as well as triadic clustering could give rise to communities in networks (Feld, 1981).

Whereas triadic closure implied a local attachment bias for creating new links, focal closure was external to the network and could thus be seen as attachment bias outside the immediate neighbourhood of the node (Kumpula et al., 2007; Kumpula et al., 2009; Laurent et al., 2015); notwithstanding the fact that triadic and focal closure could overlap in real life social networks and potentially shape the evolution of networks concurrently (Laurent et al., 2015; Kumpula et al., 2007; Kumpula et al., 2009).

2.2.5 Community Structure

As mentioned above social networks exhibited a pronounced community structure, which not only shaped with whom individuals connected but also how dynamic processes spread within real life social networks (Newman, 2003). In networks certain isomorphic subgraphs appeared more commonly than due to chance alone and were often referred to as network motifs (Milo et al., 2002). Nonetheless, not all significant and recurring subgraphs represented communities. The defining features of communities were that ties within the group were highly transitive within the group but hardly with respect to the rest of the network (Meo et al., 2014; Watts & Strogatz, 1998; Newman, 2001). Analogously in a weighted social network the ties within a community were much stronger than globally between individuals (Hossmann et al., 2011). Clustering was thus the result of nodes organising into more or less distinct groups or communities, where communities varied in size, connectedness, and distribution of tie strengths (Ravasz & Barabási, 2003; Hossmann et al., 2011).

Those communities were themselves then in turn organised hierarchically (Clauset, Moore, & Newman, 2008; Palla, Derényi, Farkas, & Vicsek, 2005; Ravasz & Barabási, 2003). For example, it is easily imaginable that a community of students attending the same class is embedded in community of a particular school which in itself is embedded within the group of the wider university (Palla et al., 2005). Further complicating matters was that individuals could and usually were part of several overlapping communities at the same time (Palla et al., 2005). Most people were members of different groups centred around various foci such as the workplace, school, and family.

And communities exhibited varying degrees of turn over. Smaller groups, corresponding to stronger ties between their members, were usually fairly static (Palla, Barabási, & Vicsek, 2007; Sekara et al., 2016), whereas bigger groups, such as institutions, showed a relatively high rate of change (Palla et al., 2007). Sekara et al. (2016) showed that there exists a remarkably stable core of group members for smaller groups. Those social cores were in fact so stable and regular that they could be used to predict the arrival of missing group members. In a sense, the core of these communities represented the various social settings and foci each of us move through in our daily lives. Uncovering the communities in social networks received considerable attention and a variety of algorithms was proposed for both static and temporal networks (Girvan & Newman, 2002; Fortunato, 2010; Fortunato & Hric, 2016; Scott & Carrington, 2011).

2.2.6 Small World

The community structure of networks gives rise to one of the most famous properties of social networks, the small world property of networks (Milgram, 1967; Watts & Strogatz, 1998). Notwithstanding the fact that social networks were highly clustered, the average path length between any two nodes was usually relatively short, at least it was much shorter than for regular lattices (Milgram, 1967; Watts & Strogatz, 1998). Connections between communities beyond the local neighbourhood of nodes dramatically decreased the average path length (Watts & Strogatz, 1998). For example, it took a relatively short number of hops in a social network to connect two people from opposite sides of a country (Milgram, 1967). Often small-word networks had a high degree of hubs that were very well connected within the network (Karsai et al., 2011). While not all social networks necessarily exhibited small world properties, they were relatively common for social networks (Karsai et al., 2011; Hossmann et al., 2011).

2.2.7 Homophily

In general, homophily (also referred to as assortative mixing), is the tendency of individuals in networks to interact with others that were similar to them with respect to different individual characteristics. It is fairly well documented phenomenon for social networks. Both in the real world as in virtual communities variables along lines such as emotion, gender, age, education, religion, proximity, network position, and physical characteristics were linked to increased interaction and tie formation between individuals (Bollen, 2011; Centola & van de Rijt, 2015; Huang, Shen, & Contractor, 2013; Holme, 2003; Marsden, 1988; Putzke, Gloor, Fischbach, & Schoder, 2010; McPherson, Smith-lovin, & Cook, 2001; Skopek, Schulz, & Blossfeld, 2011). Homophily could even lead to positive network effects for those that as a consequence of homophily were now better connected (DiMaggio & Garip, 2011; Vigier, 2014)

Researchers identified several different mechanisms that could lead to the observed homophily. Centola and van de Rijt (2015) proposed two individual level explanations for why we observe homophily in social ties, *choice homophily* and *social contagion*. The most obvious one is *choice homophily*, the propensity of individuals to preferentially form connections to those that were similar to themselves (McPherson & Smith-Lovin, 1987). The second individual level mechanism, *social contagion*, acts on the preferences of the individuals themselves. Friends could influence the views and behaviours of their friends (Bond et al., 2012; Christakis & Fowler, 2007; Centola & van de Rijt, 2015; McPherson

& Smith-Lovin, 1987; Karsai, Iniguez, et al., 2014), where social influence could both act at a global and local level (Pan, Hou, & Liu, 2017) and controversial ideas could have unexpectedly large effects on adoption with repeated exposure (Romero et al., 2011). While some innate characteristics could not be influenced by friends such as race and gender, a wide array of other variables were susceptible to peer influence: ranging from political views and partisanship (Bond et al., 2012; Campos, Heap, & de Leon, 2017; Peng, Liu, Wu, & Liu, 2016) over emotions (Bollen, 2011; Kramer et al., 2014) to health related behaviour such as smoking and obesity (Centola & van de Rijt, 2015; Christakis & Fowler, 2007).

In addition, there were several structural mechanisms that might lead to homophily even though individuals did not necessarily have a bias towards interacting with similar others: *induced homophily*, *triadic closure*, and *preferential attachment* (Centola & van de Rijt, 2015). *Induced homophily* occured when the groups and institutions, within which individuals were embedded, were themselves homogeneous (McPherson & Smith-Lovin, 1987). Organizational and institutional sorting processes inherently limited the heterogeneity of others an individual was exposed to in settings like schools (Currarini, Jackson, & Pin, 2010; Elman & O'Rand, 2007; Moody, 2001), in academia (Wang & Zhu, 2014), at workplaces (Bertrand & Mullainathan, 2004; Feld, 1982; Ruef, Aldrich, & Carter, 2003) and at voluntary organizations (McPherson & Smith-Lovin, 1987). Institutions were often focal points for forming new ties (Feld, 1981) and repeated exposure to others increases one's opinion about them (Swap, 1977). Hence, homogeneous institutions inhibited tie formation to heterogeneous others by excluding them a priori (Centola & van de Rijt, 2015). They then acted as an implicit filter of individuals into homogeneous groups along race, gender, education, and other characteristics. In general, homogeneity within groups and institutions played a significant role for determining overall homophily, even when the individuals did not necessarily have a strong preference for homogeneous ties (McPherson

& Smith-Lovin, 1987). For example, Asian students at American high schools did not seem to have a strong preference for friends of the same race but were still highly racially segregated with whom they interact (Currarini et al., 2010).

Furthermore, *triadic closure*, could also lead to homophilic outcomes without the individuals necessarily seeking out similar others (Centola & van de Rijt, 2015; Kossinets & Watts, 2009; Wimmer & Lewis, 2010). If A and B are friends and are similar to each other and B and C are friends and similar to each other, then B and C are likely to become friends as well and also be similar to each other without either B or C actively seeking out a connection that is similar to them (Centola & van de Rijt, 2015).

Last, individuals in networks often had a preference to form ties to others they perceived as desirable (Centola & van de Rijt, 2015). Examples include preferential attachment to popular, fit or attractive nodes in the network (Ali et al., 2012; Barabasi, 2012; Wang & Zhu, 2014). The less desirable nodes then were not only at a disadvantage from forming new ties, because they were less popular (i.e. they had a lower probability of being selected by chance simply because they were less popular; Barabasi, 2012) but also because they were perceived as less desirable alters (Centola & van de Rijt, 2015). The less desirable nodes in the network were then forced to form ties to each other as they were effectively excluded from forming ties to the overall population (Centola & van de Rijt, 2015). As the less desirable nodes shared certain characteristics by their nature of being less popular, this systematic exclusion could then re-enforce homophily in a variety of settings ranging from health (Ali et al., 2012) over dating (Skopek et al., 2011) to scientific collaboration (Wang & Zhu, 2014).

The above described processes, a tendency for individuals to select similar individuals as friends, biased opportunities to meet new individuals, as well as social contagion could act simultaneously. For example, (Kossinets & Watts, 2009) showed that the dynamic interplay between *choice* and *induced homophily* could over generations amplify the bias for homophily. Although certain types of networks and settings might be more amenable to certain types of biases (McPherson et al., 2001).

Homophily had a clear impact on the overall network structure leading to tighter communities and an overall difference in the degree distribution of nodes. Bramoullé et al. (2012) showed, theoretically, that homophily led to more integrated groups that shared a common property than for groups that did not. The bigger the homophily bias the bigger the effect was, no matter the group size. Furthermore, homophily could exacerbate preferential attachment (Vigier, 2014) as those that already had a lot of connections were further favoured by homophily as the nodes rich in ties became even richer (Kim & Altmann, 2017). In particular, the overall degree distribution in a log-log plot changed from concave to convex the greater the level of homophily (Kim & Altmann, 2017), where a convex shape indicates a higher level of inequality.

2.2.8 Temporal Auto-Correlation of Behaviour

Another very common aspect of human interactions in networks was that the interactions were not randomly distributed in time. If the frequency distribution of the time between events were uniformly sampled, one would expect the resulting data to follow a poisson process (Barabasi, 2005). However, human activities in networks were highly correlated in time (Barabasi, 2005; Candia et al., 2008; Backlund, Saramaeki, & Pan, 2014; Goh & Barabasi, 2008; Karsai et al., 2011; Karsai et al., 2012; Karsai, Perra, & Vespignani, 2014; Laurent et al., 2015; Moinet et al., 2015; Ubaldi et al., 2016) and usually were heavy-tailed (Barabasi, 2005; Holme, 2003; Johansen, 2004).

In other words, interactions of humans in networks were not ahistoric but exhibited patterns of a memory effect (Karsai et al., 2012; Karsai, Perra, & Vespignani, 2014) and reinforcement processes (Zhao et al., 2011). Once an individual had interacted with another there was then a higher chance that more interactions occurred in a relatively short period of time—"the memory effect" or burst. There was, however, some debate about the exact strength of the memory effect in social networks (Goh & Barabasi, 2008). Furthermore, past interactions increased the probability of future interactions—the reinforcement process—while the longer an individual went without interactions the lower the probability of future interactions.

As Karsai et al. (2012) pointed out the observed bursty patterns of behaviour aligned well with the Decision Field Theory of psychology (Busemeyer & Townsend, 1993), where each decision is a threshold phenomenon and a stimulus needs to reach a certain level to trigger the individual to choose an action from the large number of possible actions. This makes sense as an interaction initiated by another person such as a phone call or text messages acts as a trigger for a reciprocal interaction. As human social actions were strongly reciprocal (Fehr, Fischbacher, & Gaechter, 2002) and friendships needed regular investment in maintenance or otherwise they risked breaking down (Duck, 1999). Unrequited social interactions potentially led to the end of the social relationship, thus explaining the bursty pattern of observed interactions.

Furthermore, the non-Poissonian waiting times between events affected the dynamics unfolding on the network. The burstiness of interactions slowed down dynamical processes spreading via the network as the high variance increased the expected waiting times between events (Saramäki & Moro, 2015), even though initially heterogeneous contact patterns could enable a relatively fast initial growth (Rocha, Liljeros, & Holme, 2011; Rocha & Blondel, 2013). The slower rate of spreading then could also lead to longer prevalence times of epidemics (Min et al., 2013). While burstiness generally hindered the dissemination of information at large scales by increasing the duration of the fastest temporal path (Saramäki & Moro, 2015), if the networks were dense enough global information cascades were still possible (Backlund et al., 2014). On a local level the community structure of the network could lead to rapid dissemination of information even in bursty networks

(Miritello, Moro, & Lara, 2011).

What was more, temporal patterns of interactions could also be found for neighbourhoods within networks, which to an extent reflected social behaviour in groups, and not just dyads. At the smallest level one could find temporal correlations between interactions to and from one individual's acquaintances and between them (Karsai et al., 2011; Miritello et al., 2011). However, temporal correlations existed for mesoscopic structures as well.

Similarly to network motifs in static network, one may distil all observed configurations of temporal subgraphs from a temporal network given a choice of time interval (Kovanen et al., 2011). Temporal motifs were usually further restricted to the classes of isomorphic valid subgraphs, where the isomorphism was taken to include the temporal order of events (Kovanen et al., 2011). As a result, two temporal subgraphs were only considered to be isomorphic if they were both topologically equivalent and the order of their constituting events was identical. The most common temporal motifs involved only two nodes (Kovanen et al., 2011) but even higher order motifs usually followed just one temporal mesoscopic patterns (Kovanen et al., 2011; Kovanen et al., 2013; Zhang et al., 2015). Those motifs themselves exhibited bursty temporal dynamics (Zhang et al., 2015), indicating that even interactive group behaviour was bursty in nature.

Nodes in similar temporal motives furthermore tended to have similar properties—temporal homophily—with respect to node attributes (age, gender) and links (intra- or inter-communities) beyond that was predicted by the aggregate structure of the network (Kovanen et al., 2013). There were gender related differences in communication motifs as well and motifs within groups were more complicated than motifs between different groups (Kovanen et al., 2013).

2.2.9 Regularity

While temporal events in networks exhibited a bursty pattern, this did not mean the absence of regularities. External cycles shaped the dynamics of interactions considerably as Clauset and Eagle (2007) reported that topology of social networks evolved at a broad range of time scales. Different calendar cycles were observable as periodicities in the time series of the network. Daily as well as weekly rhythms in human behaviour could be observed (Clauset & Eagle, 2007; Jo, Karsai, Kertesz, & Kaski, 2012; Krings et al., 2012; Scellato, Musolesi, Mascolo, & Latora, 2010) as human behaviour was phase-locked to the day night cycle (Saramäki & Moro, 2015) and human life was generally structured around a seven day week. Underlining the importance of the home-work-home daily schedule, social behaviour during the week was significantly more regular than on weekends (Sekara et al., 2016)

2.3 Human Mobility

It was, however, not just social networks that had received high levels of academic attention. Ever since the seminal work of Hägerstrand (1970), how individuals moved around space had been a core problem of understanding human behaviour. Traditional studies relied heavily on self-reported travel behaviour in order to analyse people's movements and activities (Xu et al., 2015). Yet conducting such studies was costly as they included detailed information not only about the respondents but also about the trips they undertook (Xu et al., 2015).

Newer studies increasingly utilised digital traces of travel behaviour often leading to better data quality (Wan et al., 2013). Among the new sources of data were mobile and call data records (Becker et al., 2013; Barbosa, Lima-Neto, Evsukoff, and Menezes, 2015; Csaji et al., 2013; Krings, Calabrese, Ratti, and Blondel, 2009; Sevtsuk and Ratti, 2010; Williams, Thomas, Dunbar, Eagle, and Dobra, 2014), online location based social networks (Bapierre, Jesdabodi, & Groh, 2015; Cho et al., 2011; Cheng, Caverlee, Lee, & Sui, 2011; Jurdak, Zhao, Liu, Jaoude, & Cameron, 2015; Noulas et al., 2012; Wu et al., 2014), GPS traces of vehicles (Bazzani, Giorgini, Rambaldi, Gallotti, & Giovannini, 2010; Jiang, Yin, & Zhao, 2009; Krause & Zhang, 2018; Lima et al., 2016; Luo, Cao, Mulligan, & Li, 2016), transportation cards (Hasan, Schneider, Ukkusuri, & González, 2013; Roth, Kang, Batty, & Barthélemy, 2011; Seaborn, Attanucci, & Wilson, 2009; Sun, Axhausen, Lee, & Huang, 2013), digital monetary transaction (Lenormand et al., 2015; Sobolevsky et al., 2014) and digital traces of individuals (Eagle & Pentland, 2006; Stopczynski et al., 2014; Wang et al., 2017). Recent studies found two main characteristics of human mobility behaviour: travel was characterised by heterogeneities in behaviour and a surprising regularity as well as predictability of travel, which was not necessarily a new finding. However, the scope and scale of studies using digital traces of mobility behaviour was unprecedented allowing for a much better quantitative understanding of patterns of human mobility.

For a graphical overview of the main concepts related to human mobility discussed in this chapter see also Figure 2.1.

2.3.1 Heterogeneities

Similarly as in social networks human mobility was characterised by heterogeneities. Individuals exhibited a broad spectrum of travel ranges (Alessandretti, 2018) and their particular travel patterns were highly unique (De Montjoye, Hidalgo, Verleysen, & Blondel, 2013). Human mobility, as activity in social networks, was often heavy tailed and indeed a wide variety of empirical studies found evidence supporting this using a variety of data sources such as bank notes, CDR, and GPS traces (González et al., 2008a; Song, Koren, Wang, & Barabasi, 2010; Rhee et al., 2011; Baronchelli, Ferrer-i-Cancho, Pastor-

Satorras, Chater, & Christiansen, 2013; Lu, Wetter, Bharti, Tatem, & Bengtsson, 2013; Wang et al., 2011; Wang, Han, & Wang, 2014; Zhao, Musolesi, Hui, Rao, & Tarkoma, 2015; Jurdak et al., 2015; Beiro, Panisson, Tizzoni, & Cattuto, 2016; Cheng et al., 2011; Hawelka et al., 2014; Brockmann et al., 2006; Yan, Han, Wang, & Zhou, 2013).

The distribution of spatial displacements (i.e. offsets) of individuals $\triangle r$ usually followed a power-law, $P(\triangle r) \sim \triangle r^{-\beta}$), while, especially at shorter distances, exponential distributions, $P(\triangle r) \sim e^{-\lambda \triangle r}$, were also observed (Alessandretti, 2018). There is some evidence that human mobility might follow other heavy tailed distributions such as a log-normal distribution, $P(\triangle r) \sim (1/\triangle r) * e^{-(log \triangle r - \mu)^2/\sigma^2}$(Alessandretti, 2018). For a comprehensive review of mobility models and the observed distributions, I refer the reader to Alessandretti (2018).

Often the observed distribution could be described as a Levy flight as $1 < \beta \leq 3$ (Baronchelli et al., 2013; Rhee et al., 2011; Zhang et al., 2018), where a Levy walk is a random walk, whose step lengths follow a heavy tailed distribution. Levy walks were highly efficient for exploring unknown areas (Baronchelli et al., 2013). It is noteworthy here that while human mobility was clearly not a random walk, it nevertheless often shared a lot of statistical similarities with a Levy flight.

One possible reason for the observed variety in empirical distributions was the possibility that human mobility was inherently multi-modal (Seaborn et al., 2009; Jurdak et al., 2015; Zhao et al., 2015; Yan et al., 2013). Jurdak et al. (2015) reported that the distribution of $\triangle r$ was best described as a mixture function differentiating between movements within the same site, within the same city, and between different cities. Another possible factor influencing the observed variety in distributions of $\triangle r$ is the difference in spatio-temporal resolution of the underlying data. Tuhin, Kevin, Nathaniel, Scott, and Nazeem (2016) found that most common metrics of human mobility including $\triangle r$ were sensitive to the spatio-temporal resolution and direct comparisons of data with different resolutions might not be meaningful.

In general, however, all studies confirmed that $\triangle r$ was skewed and the important take away was that "[m]ost people go a short distance; few people go a long distance" (Stouffer, 1940). Even when varying the resolution of the data, $\triangle r$ was found to be consistently heavy tailed (Tuhin et al., 2016). As a result human mobility patterns overall were highly heterogeneous and the tail of $\triangle r$ for the individuals with the most displacements was often "heavier" than for exponential distributions.

Fewer studies explored the distribution of $\triangle t$ as sampling was often uneven (Alessandretti, 2018). Often data sets only recorded the position or the movement of an individual after a specific action by that individual. For example, on Foursquare, an online location based social network, users needed to actively check-in at a venue for their location to be recorded (Noulas et al., 2012). Call data records usually only indicated the location of a user when a call was placed or a text sent (Lima et al., 2016). They most often did not include the handshake data between the phone and cell phone tower. However, studies that had access to more evenly sampled data confirmed that $P(\triangle t)$ was heavy tailed as well (Song, Koren, et al., 2010; Schneider et al., 2013; Alessandretti, 2018; Bazzani et al., 2010; Rhee et al., 2011; Wang et al., 2014).

There were several possible explanations for the observed heterogeneities proposed in the literature:

– the distribution of anchor locations of individuals, that dominate individual travel patterns,

– different transportation modes,

– the built environment,

– differences in how people allocate their resources when travelling, and

– mediating factors such as age or ethnicity.

First, travel between the top n locations dominated the everyday mobility of most people. Analogous to preferential attachment in social networks, humans showed a significant propensity to return to locations frequently visited before, that meant the probability of visiting a place in the future was directly related to how often an individual had visited that place in the past (Jurdak et al., 2015; Song, Koren, et al., 2010). Consequently, the majority of people frequently only visited six or fewer locations (Csaji et al., 2013; Isaacman et al., 2011) and travelled between the top two locations for most people, was the most significant observed travel behaviour (Pappalardo, Simini, et al., 2015; Xu et al., 2015). Unsurprisingly those two anchor points of everyday mobility consisted for most people of the work place and the home (Xu et al., 2015). As the most frequented locations formed areas that people only infrequently left (Bagrow & Lin, 2012), the mobility between the top n locations was hence responsible for a salient part of the observed heterogeneities (Pappalardo, 2016). Compounding the effects of visiting frequent locations, was the tendency of humans to prefer to revisit locations they had recently visited (Barbosa et al., 2015).

In particular, as people tended to visit locations that were in the same spatial neighbourhood, the locations visited by individuals tended to cluster in a small number of mobility cores. This suggested two different modes of mobility for intra-urban displacement (Jurdak et al., 2015). While inter-core mobility was strongly correlated with the overall radius of gyration, once individuals were constrained to only intra-core mobility their travel behaviour became much more homogeneous (Pappalardo, 2016), even if their frequency of their trips might still be heterogeneous (Roth et al., 2011). In other words, inter-core travel was responsible for much of the observed heterogeneities.

Second, another factor contributing to the observed heterogeneities in travel behaviour is the different travel modes individuals employ. Zhao et al. (2015) found four distinct transportation modes, such as walk/run, bike, train/subway, and car/taxi/bus. The distribution of displacements of each mode separately could be well approximated by a log-normal distribution. However, when taken together the four transportation modes produced a characteristic power-law distribution of displacements (Zhao et al., 2015).

Third, travel behaviour showed different characteristics at different scales. Both intra-urban travel (Noulas et al., 2012) and inter-urban travel (Liu et al., 2014) appeared to not necessarily follow a power-law and seemed to be guided by the built environment (Zignani, Papandrea, Gaito, Giordano, & Rossi, 2014). The variability in behaviour could thus partly be explained by differences in the distribution of places in each city (Noulas et al., 2012) and how the built environment shaped traffic behaviour (Jiang et al., 2009) rather than differences in behaviour per se. Furthermore, the organization of space itself could be an important reason for the observed heterogeneities as streets themselves exhibited scaling behaviour (Huang, Zhu, Ye, Guo, & Wang, 2016). The most popular roads saw a disproportionate amount of traffic as well as the 80% of roads were connected below average whereas 20% were connected above average (Huang et al., 2016; Jiang, 2007; Jiang et al., 2009). Moreover, the origin/destination pairs of trips were unevenly distributed in space, whereas 80% of locations were scattered but 20% of locations were densely clustered. Jiang et al. (2009) showed that these two scaling properties of the built environment could help explain scaling properties of human mobility.

What was more, the density of the urban environment might have been a decisive factor in how travel distances were distributed within cities. As a result, travel time and not spatial distance might be a more appropriate distance metric (Williams & Musolesi, 2016; Zignani et al., 2014; Phithakkitnukoon, Smoreda, & Olivier, 2012). For example, people in New York travelled a significantly shorter distance on a given day than people in LA (Becker et al., 2013), where New York was generally a much more dense urban environment than LA. Similarly to earlier findings about migration (Stouffer, 1940) as the number of intervening opportunities between possible destinations increased, the travelled distances of individuals decreased (Noulas et al., 2012). The probability of a transition to a destination place was thus inversely proportional to the relative rank (i.e. how many other places were closer than the destination) of it raised to a power α, where α varied slightly from city to city (Noulas et al., 2012).

Fourth, people employed different strategies for allocating their resources. Some individuals chose to spend most of their time at just a relative low number of locations, whereas others chose to distribute their time more broadly between their top n locations. Individuals could be broadly categorised into two classes based on their travel behaviour: "explorers" and "returners" (Pappalardo, Simini, et al., 2015). For "returners" the largest part of their travel behaviour consisted of travel to the top n locations, whereas the travel pattern of "explorers" was not dominated by travel to the top n locations. However as n increased ever more "explorers" also turned into "returners" and at around $n = 8$ most "explorers" became "returners" as well. This dynamic was similar to how individuals chose to allocate resources in their social network, either investing heavily in just a few connections or cultivating more but shallower social ties (Miritello et al., 2013).

Fifth, mediating factors such as age (Yuan, Zheng, & Xie, 2012), the built environment (Clark, Scott, & Yiannakoulias, 2013), culture (Amini, Kung, Kang, Sobolevsky, & Ratti, 2014; Wu, Wang, & Dai, 2016), ethnicity (Silm & Ahas, 2014), income (Pappalardo, Pedreschi, Smoreda, & Giannotti, 2015), season (Isaacman et al., 2011), weather (Clark et al., 2013) and personality (Alessandretti, 2018) were also proposed to explain the heterogeneities in observed travel behaviour.

2.3.2 Predictability & Regularity

The average, low cardinality of the set of visited locations was consistent with the overall high predictability of location traces. Historical patterns of travel were highly informative for predicting future travel; and the more historical data were available the better the prediction (Lu et al., 2013). Up to around 90% of an individual's mobility was predictable (Lu et al., 2013; Song, Qu, et al., 2010) and only 17 network motifs were sufficient to describe 90% of all observed mobility for various different countries (Schneider et al., 2013). Remarkably the predictability of mobility behaviour was fairly independent of the distance an individual travelled on a regular basis (Song, Qu, et al., 2010).

Moreover, the regularity in observed travel behaviour held for different time scales. Regularities existed both at the daily (Bagrow & Lin, 2012; Schneider et al., 2013; Sevtsuk & Ratti, 2010), weekly (Cheng et al., 2011; Csaji et al., 2013; Sevtsuk & Ratti, 2010), and aggregate level (Ahas, Aasa, Silm, & Tiru, 2010; Sevtsuk & Ratti, 2010). And while individual travel routines were somewhat unstable in the long term, the tendency to explore new locations decreased with the observed time (Schönfelder & Axhausen, 2003; Song, Koren, et al., 2010).

Everyday human activity spaces were also usually spatially constrained around the most important locations shrinking the space of probable next locations significantly. Individuals hardly deviate from the confidence ellipsoid spanned by their most visited locations (Cho et al., 2011; Schönfelder & Axhausen, 2003). Furthermore, when travelling between locations individuals rarely travelled outside an ellipsoid with the origin and destination as foci (Lima et al., 2016). In general, the space of actively visited locations stayed fairly constant at around 25 locations as well as the distribution of how time was allocated to different locations (Alessandretti et al., 2018).

In short, individuals usually did not select their next destination at random and their movements were most often spatially confined. This greatly contributed to their observed regularity as the possible space of likely next locations was relatively small compared to an unbounded space of possibilities. Furthermore, this also allowed for the predictions of mobility traces; sometimes far into the future. By extrapolating robust patterns of mobility behaviour (Sadilek & Krumm, 2012), as the activity space of users only gradually evolved (Alessandretti et al., 2018), one could fairly accurately predict individual mobility behaviour months or sometimes even years in advance.

Driving the regularities were several dynamics. Unsurprisingly, the commute between work and home played an important role for a significant portion of the population as both home and work were the two most common locations (Csaji et al., 2013; Ranjan, Zang, Zhang, & Bolot, 2012). Indeed, several studies found peaks in the distribution of travel behaviour and waiting times that corresponded to the typical workday schedule. Periodicities that occurred at various frequencies correspond to a typical schedule on a workday (Alessandretti, 2018). Peaks at around 4 hours, at around 8 hours, at around 12, and at around 24 hours mapped to part-time work, full-time employment, spending a day at home, and to daily routines respectively (Alessandretti, 2018; Hasan et al., 2013; Schneider et al., 2013; Sun et al., 2013). Given the observed regularities of individual patterns of mobility, unsurprisingly aggregate mobility patterns followed a similar distribution with peaks at 8, 12, and 24 hour intervals (Sevtsuk & Ratti, 2010).

Another possible driver of travel regularity were meetings with pre-existing friends and acquaintances (Cho et al., 2011). Social activities were after all responsible for a significant portion of all trips by individuals (van den Berg, Arentze, & Timmermans, 2009; van den Berg, Kemperman, & Timmermans, 2014). About half of all face-to-face interactions happened outside the work or the home and thus required travel to other places (van den Berg et al., 2014) While social interactions exhibited clear periodic patterns, it was unclear from the existing literature how much exactly of the observed social regularity was due to underlying periodic mobility behaviour and how much pre-existing social relations shaped the regularity of travel behaviour.

Last, leisure activities might affect periodic travel patterns. While leisure travel was arguably less stable than commuting to work, there still existed a high level of repetition for leisure travel (Schlich, Schoenfelder, Axhausen, & Hanson, 2004); especially on a weekly level leisure travel might have significantly contributed to the periodicity of travel behaviour.

2.4 Interplay of Social Ties & Human Mobility

As both social networks and mobility are key aspects of human life, unsurprisingly they are not independent of each other. Social networks were embedded in geographic space and travel behaviour was shaped by social networks. Around 80% of an individual's trips were within just 20km of a peer's home location and this decreased to just 10km for dense, urban areas (Phithakkitnukoon et al., 2012). Moreover, the size of a person's geographic activity space on the one hand and the amount of interactions and size of network on the other hand were correlated (Alessandretti et al., 2018; Puura, Silm, & Ahas, 2017; Scellato, Noulas, Lambiotte, & Mascolo, 2011; Berg, Arentze, & Timmermans, 2012; Carrasco, Hogan, Wellman, & Miller, 2008; Yuan, Raubal, & Liu, 2012). Even in online communities, physical distance between individuals was negatively associated with the likelihood of a tie between them (Huang et al., 2013; Takhteyev et al., 2012).

Furthermore, people travelled to meet their social others leading to the observed pattern of mobile homophily, where friends were having highly similar mobility behaviour. The more similar two people's trajectories were the higher the likelihood that they were close in the social network (Bapierre et al., 2015; Toole et al., 2015). Not only were people that call each other more likely to be co-located in space (Calabrese, Smoreda, Blondel, & Ratti, 2011), but network proximity, friendship, and tie strength were all influenced by distance and by the similarity of one's trajectory (Cho et al., 2011; Grabowicz et al., 2014; Shi, Wu, Chi, & Liu, 2016; Toole et al., 2015; Wang et al., 2011). And social ties and mobility behaviour co-evolved over time (Dong, Lepri, & Pentland, 2011; Alessandretti, 2018).

While the interplay between social networks and mobility was influenced by mediating variables such as age, gender, and the built environment (Puura et al., 2017; Yuan, Raubal, & Liu, 2012), the two main mechanisms that facilitated the connection between social networks and human mobility were mobility shaping friendships and simultaneously friendships shaping mobility.

On the one hand, face-to-face contact was still highly valued and essential for maintaining social ties (Burke & Kraut, 2014; Larsen et al., 2006; Vlahovic et al., 2012). The chance of interacting and meeting with others was highest for those physically close to us (Preciado, Snijders, Burk, Stattin, & Kerr, 2012; van den Berg et al., 2009; van den Berg et al., 2014). Recent advances in communication technologies seemed not to have decreased the need for corporeal interactions (Mok, Wellman, & Carrasco, 2010). In other words, our mobility shaped our social network. Spatial distance in general played an important role for friendship formation, because individuals were more likely to form ties with those who lived close by (Liben-Nowell, Novak, Kumar, Raghavan, & Tomkins, 2005; Preciado et al., 2012). The probability of being friends decreased dramatically with distance; about two thirds of all friendships were contingent on geography (Liben-Nowell et al., 2005), where medium to longer distances played a more discriminatory role for friendships than short distances (Backstrom, Sun, & Marlow, 2010). Studies confirmed this effect for CDR (Krings et al., 2009; Lambiotte et al., 2008), online social networks (Liben-Nowell et al., 2005; Nguyen & Szymanski, 2012; Takhteyev et al., 2012; Volkovich et al., 2012), and simulation studies (Shi et al., 2016).

However, in dense urban environments distance played less of a pronounced role. The denser the urban environment was the shorter the average distance travelled due to social ties were (Phithakkitnukoon et al., 2012). What was more salient in dense environments were shared sets of locations as there usually existed a variety of destinations for each travel need (Noulas et al., 2012; Shi et al., 2016; Toole et al., 2015; Wang, Kang, Bettencourt, Liu, & Andris, 2015); especially locations other than the home or work place were frequently associated with social interactions (Picornell et al., 2015). Consequently the more similar users were with respect to their travel history, the more likely they were to be friends and the stronger their connection (Toole et al., 2015). Shared locations most likely increased the likelihood of "bumping" into each other and the opportunities provided by overlapping routines could then foster friendship formation (Verbrugge, 1977). It is noteworthy to point out that co-occurrence alone did not automatically lead to friendships. Individuals might co-occur many times but not become friends (Sun et al., 2013) and co-occurrences might be more important for strengthening existing friendships than forming new ones (Shi et al., 2016).

Distance further mediated how people maintained their social relationships and the structure of their social network. The closer people were, the more likely they were to see or contact each other (Carrasco, Miller, & Wellman, 2008; Mok et al., 2010), a fact that was useful for predicting missing links and/or future interactions (Crandall et al., 2010; Wang et al., 2011). Geographic closeness also increased the probability that users belonged to the same tightly connected community (Volkovich et al., 2012) and the probability for triadic closure decreased with distance (Lambiotte et al., 2008).

On the other hand, a considerable amount of a person's mobility was motivated by social ties; that was the social dimension was a key reason for travel (Belot & Ermisch, 2009; Carrasco, Miller, & Wellman, 2008; Cho et al., 2011; Larsen et al., 2006; Eagle & Pentland, 2009; Nguyen & Szymanski, 2012). Once ties were established, spatial distance played less of a pronounced role for maintaining friendships than for forming friendships (Cho et al., 2011; Preciado et al., 2012). In fact, our social network actively altered our mobility behaviour. People dedicated a significant amount of their mobility behaviour to maintaining existing ties. Even long trips were not uncommon to meet existing ties (Larsen et al., 2006). Roughly 15% to 30% of all trips could be attributed to meeting others (Cho et al., 2011; Grabowicz et al., 2014; Toole et al., 2015) and having close friends living close by significantly reduces the observed travel distance (Belot & Ermisch, 2009).

Furthermore, a significant factor for intra-urban moves of households was proximity to local ties, or in other words people moved to where there friends were already living (Metcalf, 2013). As a salient part of human mobility was motivated by social interactions, it was unsurprising that knowing the mobility patterns of your friends could improve the predictions of where you will be next (Cho et al., 2011; Grabowicz et al., 2014; De Domenico et al., 2013; Beiro et al., 2016). And distinct social groups exhibited distinct, shared patterns of mobility (Eagle & Pentland, 2009); it seemed as if, social groups imposed their own pattern of shared mobility onto their members.

2.5 Mediating Factors Shaping Mobility & Social Ties

New sources of data enabled the study of human behaviour on a scale hitherto not imaginable. The statistical patterns of how individuals socialised and moved around space are now fairly well established (Section 2.2 and 2.3). In short, humans were inherently social beings that were faced with a trade off to balance limited resources and followed routines operating on different time scales. Human mobility and social ties were also intrinsically linked as social relationships determined a big part of human mobility and geographic space shaped who becomes friends with whom.

What is more, human mobility and social behaviour were not a closed nor mono-causal system. A variety of mediating and contextual factors could influence the observed outcomes. I would like to point out that while for certain attributes such as age, gender or attractiveness, there was a clear causal direction, for other variables such as economic behaviour it was much harder to attribute a definitive causal direction. What is more, these mediating factors might interact and influence the observed behaviour simultaneously making it even harder to assign causality to observational data alone. Hence, it is important to account for the "context", ranging from the attributes of the individuals to the environment they were in, within which we can observe both mobility and social behaviour.

2.5.1 Social Networks

Studies found evidence for how gender, age, economic resources, attractiveness, personality, geography, the type of places individuals socialise, and temporal factors shaped social networks.

Gender

Males and females tended to have different social network structures. Dunbar and Spoors (1995) found that both genders had significantly more ties to the same gender than the opposite gender, or in other words gender was a driver of homophily. In particular, women had a larger amount of ties to female friends and relatives, whereas men had more ties to male friends and kinship relationships play less of a role for them.

What was more, females tended to invest significantly more resources in their personal networks both qualitatively and quantitatively. Females usually had not only larger personal networks (Moore, 1990) but formed larger groups within their networks (Igarashi et al., 2005), interacted more with their alters in their networks, gave and received more support through their network (Hays & Oxley, 1986), had more face-to-face contacts (van den Berg et al., 2009), and used more media mediated messaging to groom their ties than men (Igarashi et al., 2005).

Age

Age was another mediating factor shaping an individual's network. There is evidence that with age the size of personal social networks and the frequency of interaction with non-family members decreased (Carstensen, 1992; Wrzus et al., 2013; Miritello et al., 2013; Sander et al., 2017; Shaw, Krause, Liang, & Bennett, 2007; Zhaoyang et al., 2018). Moreover, personal networks tended to become older with age (Ajrouch et al., 2005). In contrast, familial ties did not seem to be negatively influenced by age, even though familial ties were usually geographically more spread out than friendships (Carrasco, Miller, & Wellman, 2008). One group of researchers (Sander et al., 2017; Shaw et al., 2007; Wrzus et al., 2013) found stable levels of interactions with family members, whereas Zhaoyang et al. (2018) found increased levels of interactions. Interestingly life course events such as moving residence and the birth of a child did not necessarily influence the life span trajectory of contact frequency (Sander et al., 2017).

Economic Resources

Economic success and behaviour were also correlated with network structure. At the individual level job opportunities were linked to the social network. Whom one knew significantly altered salary negotiations (Seidel, Polzer, & Stewart, 2000) and changes in CDR data could predict layoffs as well as who was affected by them (Toole et al., 2015). Furthermore, economic inequality was mirrored in the network position of individuals (Campbell et al., 1986; Decuyper et al., 2014; Yannick, Eric, Alvarez-Hamelin, Carlos, & Karsai, 2016; Luo, Morone, Sarraute, Travizano, & Makse, 2017) and individuals in professional occupation had less proximial personal networks (Ajrouch et al., 2005). On a broader level, economic development of a region was linked to the amount and diversity of communication ties of its inhabitants (Eagle, Macy, & Claxton, 2010; Mao, Shuai, Ahn, & Bollen, 2015).

Attractiveness

As humans preferentially formed ties to nodes they viewed as attractive, health, fitness and attractiveness were also mediating factors shaping social networks. For example, obese adolescents had fewer friends and were less well socially integrated (Ali et al., 2012) and obese people were regularly mistreated by others (Carr, Jaffe, & Friedman, 2008). Conversely, attractiveness was positively correlated with satisfaction with interactions (Reis et al., 1980).

Personality

Personality traits were also shown to shape social network dynamics (Wehrli, 2008). Montjoye, Quoidbach, Robic, and Pentland (2013) found that openness was related to the diversity with whom an individual calls and texts. Individuals that scored high on openness also tended to have higher levels of network turnover and larger variations among their peers (Centellegher et al., 2017). Extraversion was strongly linked to the number of friendships, although extroverts were not necessarily emotionally closer to others in their network (Kalish & Robins, 2006; Quercia, Bodaghi, & Crowcroft, 2012; Pollet et al., 2011). Moreover, Selfhout et al. (2010) discovered that individuals that scored high on agreeableness were more likely to be selected as friends by others. Personality also appeared to be related to the network position individuals inhabit (Oliveira, 2011; Kalish & Robins, 2006). Individuals, who felt vulnerable to external forces tended to favour closed networks of weak connections, whereas individualists tended to seek stronger ties to others that were themselves not connected (Kalish & Robins, 2006).

Geography

Geography was a consequential factor for social networks as well. The spatial context, within which social networks were embedded in, could significantly affect social networks themselves (Adams, Faust, & Lovasi, 2012). Previous research suggested that spatial factors such as population density, diversity of land uses, and design were important for the number of an individual's social ties and where they were located (Boessen et al., 2017). Population density acted as an important driver of local network characteristics and spatial heterogeneity, especially at smaller scales, was reflected in heterogeneous network characteristics (Butts, Acton, Hipp, & Nagle, 2012). Similarly, Wang et al. (2015) showed that dense downtown areas act as a hub for many heterogeneous social groups and many high-degree individuals.

In addition, the distribution of POIs within an area played a role for the spatial distribution of ties (Xu et al., 2017). Commercial buildings, shopping malls, education, and community centres were associated with an increased bonding potential for existing ties (Xu et al., 2017). What is more, the characteristics of individual places seemed to influence social networks as well. The diversity of individuals visiting a certain location was a strong indicator of potential future connections between individuals with less popular and diverse locations being especially conducive to tie formation (Scellato, Noulas, & Mascolo, 2011). Categories of venues such as food, night life, and residence on Foursquare also had a far greater probability for co-located individuals to be friends than all other categories (Brown, Noulas, Mascolo, & Blondel, 2013). Overall this suggested that certain types of places and areas played an important role for fostering social ties.

Temporal Factors

While the temporal evolution of social networks (for a review see Holme and Saramaki, 2012 and Holme, 2015) and other dynamics such as burstiness (Section 2.2.8) were extensively studied there is also evidence that other temporal factors play a role for social networks as well. For one, the size of an individual social network stayed remarkably stable over time (Alessandretti, 2018) even though there was a constant change in network composition (Centellegher et al., 2017; Miritello et al., 2013; Arnaboldi et al., 2013), suggesting that while individuals adapted their networks over time they had a limited capacity for the amount of social interactions (Dunbar, 1998). Furthermore, with whom people socialised unsurprisingly changed over the course of the week (Sekara et al., 2016).

2.5.2 Mobility

Mobility was shaped by a similar set of mediating factors than social networks; in particular gender, age, socio-economic status, culture, ethnicity, the built environment, temporal factors, personality all seemed to influence observed patterns of mobility.

Gender

Gender was a significant variable forming travel behaviour. Females were significantly less likely to be multi modal (Dill et al., 2015) and poorer women faced significant barriers to mobility (Salon & Gulyani, 2010). Interestingly the movement radius and the travelled distances of females could be larger than (Yuan, Raubal, & Liu, 2012), smaller than (Lenormand et al., 2015), or equal to (Kang et al., 2010) that of men. Notwithstanding the contradictory results for travel radii, women generally encountered more fixed out-of-home activities in their everyday lives irrespective of their employment status and travelled further to work than men (Kwan, 2000).

Age

In general mobility was heavily tied to the commute (Xu et al., 2015), and thus unsurprisingly, young employed, and active people were the most mobile group (Puura et al., 2017; Yuan, Raubal, & Liu, 2012; Lenormand et al., 2015). Whereas for children and teenagers as the grew older independent mobility increased (Fyhri, Hjorthol, Mackett, Fotel, & Kyttä, 2011; Kang et al., 2010), the older adults became the less they travelled, not only due to retirement but also to an increasing share of individuals with impairments that make travel harder (Tacken, 1998).

Socio-Economic Status

There exists also a large body of literature linking socio-economic indicators and travel behaviour. Socio-economic indicators of both individuals (Carrasco, Miller, & Wellman, 2008; Frias-Martinez & Virseda, 2012; Frias-Martinez, Soto, Virseda, & Frias-Martinez, 2013; Soto, Frias-Martinez, Virseda, & Frias-Martinez, 2011; Pappalardo, Pedreschi, et al., 2015) and regions (Marchetti et al., 2015) were correlated to travel behaviour. In short, the richer a person the further and more frequently they travelled (Etminani-Ghasrodashti & Ardeshiri, 2015; Murakami & Jennifer, 1997; Prendergast & Williams, 1981). Available income also shaped mode choices as poorer individuals walk much more frequently than richer individuals (Klinger & Lanzendorf, 2016; Murakami & Jennifer, 1997).

Culture & Ethnicity

Another factor shaping mobility behaviour were cultural and ethnic differences. While there was no clear consensus on how culture might shape mobility, what was clear was that different cultural or ethnic groups had different patterns of mobility. For example intra-tribal mobility was much more pronounced than inter-tribal mobility (Amini et al., 2014), the heterogeneities of mobility patterns in China could be explained by dialect communities (Wu et al., 2016), Russian speakers in Estonia had significantly altered travel patterns, especially with respect to longer distances, compared to the majority population (Puura et al., 2017; Silm & Ahas, 2014), and race influenced travel behaviour of individuals in Chicago (Luo et al., 2016). Cultural travel choices were, however, not fixed but amenable by the local context an individual lived in (Klinger & Lanzendorf, 2016).

Built Environment

Unsurprisingly the built environment also shaped people's mobility, but it was not just density (Clark et al., 2013; Noulas et al., 2012) or the scaling properties of the street network (Jiang et al., 2009) that affected travel behaviour. Density, diversity, and design of the built environment were all found to have an impact on a person's mobility. In a meta-analysis, Ewing and Cervero (2010) found that the denser an area, the closer amenities, and the more mixed areas the more people walked. Denser and more diverse areas also had a positive effect on transit usage. Conversely, less accessible neighbourhoods with fewer transit stops and further away from the city centre were related to an increase in automobile usage. However, Ewing and Cervero (2010) cautioned to take their findings at face value due the often small sample sizes and the lack of control for attitudes and residential preferences. Nevertheless, more methodologically sound research still corroborated the finding that denser and more mixed neighbourhood led to more walking and less driving (Handy, Cao, & Mokhtarian, 2005; Hong, Shen, & Zhang, 2014). For an overview of qualitative reviews I refer the reader to Ewing and Cervero (2001).

Furthermore, the type of places people visited and thus the purpose of a trip informed behaviour as well. In fact, the observed mobility behaviour and the purpose of a trip were so interwoven that passively collected trip data were often enough to successfully infer a trip's purpose (Lu & Zhang, 2015). As trips to work places and homes were fairly regular (Section 2.3.2), these trips were more predictable than trips undertaken to meet with others or to go shopping (Krause & Zhang, 2018). As commuting was for most people the dominant part of their everyday mobility (Xu et al., 2015), the home and work locations were of particular importance for determining travel behaviour. Whereas other frequently visited places that were not an individual's home or work, were associated with travel induced due to social ties (Picornell et al., 2015).

Temporal Factors

External rhythms such as the weekday/weekend schedule and business hours (Section 2.3.2), seasonal changes (Isaacman et al., 2011), or travel around holidays (Wu et al., 2016) also significantly impacted travel behaviour. People unsurprisingly travelled less in colder months (Isaacman et al., 2011) as adverse weather had a negative impact on travel (Clark et al., 2013).

Personality

Personality traits were also proposed to explain the heterogeneities in observed travel behaviour (Alessandretti, 2018; Montjoye et al., 2013; Chorley, Whitaker, & Allen, 2015). In particular, there is evidence that both openness to experience (Montjoye et al., 2013) and conscientiousness (Chorley et al., 2015) were associated with a larger variety of visited locations. Conversely, neurotic users appeared to visit significantly fewer locations (Noe, Whitaker, Chorley, & Pollet, 2016).

2.6 Summary

A burgeoning number of studies used new sources of digital traces of behaviour to look at both social networks and mobility behaviour. They found that human behaviour was characterised by a large degree of variety and shaped by various trade-offs humans have to make between re-visiting previous locations and ties and exploring new places and relationships (Section 2.2 and 2.3). Furthermore, there is clear evidence that a variety of mediating and contextual factors played an important role for the observed dynamics for both social networks and human mobility. Nevertheless, studies using digital traces of behaviour that looked at mediating factors were still relatively rare, at least compared to the vast trove of studies that dealt with social network analysis or human mobility in general. While the most obvious mediating factors such as age, gender, socio-economic status, race, ethnicity, cultural background, and personality were researched (Section 2.5), relatively little was known about how other factors influence the observed statistical patterns.

Several potential mediating factors are currently not well studied. A non-exhaustive list of not well studied topics includes the role places play for future co-occurrences, how longer term dynamics shape the interplay between mobility and social ties, how personality traits shape the previously established regularities of behaviour in the social realm and in geographic space, how neighbourhoods might facilitate certain kinds of social and travel behaviour, how parenting styles predispose the trade-off between exploration and exploitation strategies, how cliques and communities are both influenced by homophily, social contagion as well spatial and temporal constraints simultaneously, and how mobile homophily co-evolves and shapes the personal social network.

As discussed in Section 1.2, for this thesis I have decided to focus on the role of places for future co-occurrences, how longer term dynamics affect the interplay between mobility and social networks, how personality shapes regularity of spatial and social behaviour, and what role neighbourhood effects play for digitally observed behaviour.

Errors using inadequate data are much less than those using no data at all.

Charles Babbage

3
Data & Methods

This chapter briefly reviews the data I used for my PhD thesis as well as several methodological concepts utilised throughout the thesis.

3.1 Data

3.1.1 Description

The data I used for my thesis consisted of the dataset collected by the Copenhagen Network Study (Stopczynski et al., 2014). The dataset tracked 847 students at the Danmarks Tekniske Universitet (Technical University of Denmark, hereafter DTU) for a couple of years using smartphones provided by the researchers. Around 22% of the students in the study were female and around 78% male. The research subjects were typically between 19 and 21 years old.

The dataset contained call and text logs, GPS traces, scans of WiFi access points, as well as scans of nearby Bluetooth devices of the students. The scale of the dataset provided an unprecedented level of detail and at the same time breath of the daily life of a cohort of students. For the first time a significant portion of participants' "everyday" peers was covered by a study.

While data were collected for 24 months from September 2013 to September 2015, the study was initially designed to only collect data for one year. Consequently the first academic year provided the highest sample rate of behaviour and I focused my analysis on the first academic year. As can be seen in Figure 3.1 there was a noticeable drop off in the amount of location as well Bluetooth traces collected after the first academic year.

Furthermore, additional information about the students was collected via questionnaire including the Big Five personality traits (Goldberg, John, Kaiser, Lanning, & Peabody, 1990). The data collection was approved by Datatilsynet (the Danish Data Protection Agency) and all participants provided informed consent to the data collection.

I would like to point out that the dataset had a clear bias as many other sources of digital data about human behaviour (Arribas-Bel, 2014). First, the dataset was heavily imbalanced with respect to both the socio-demographic variables age and gender as most

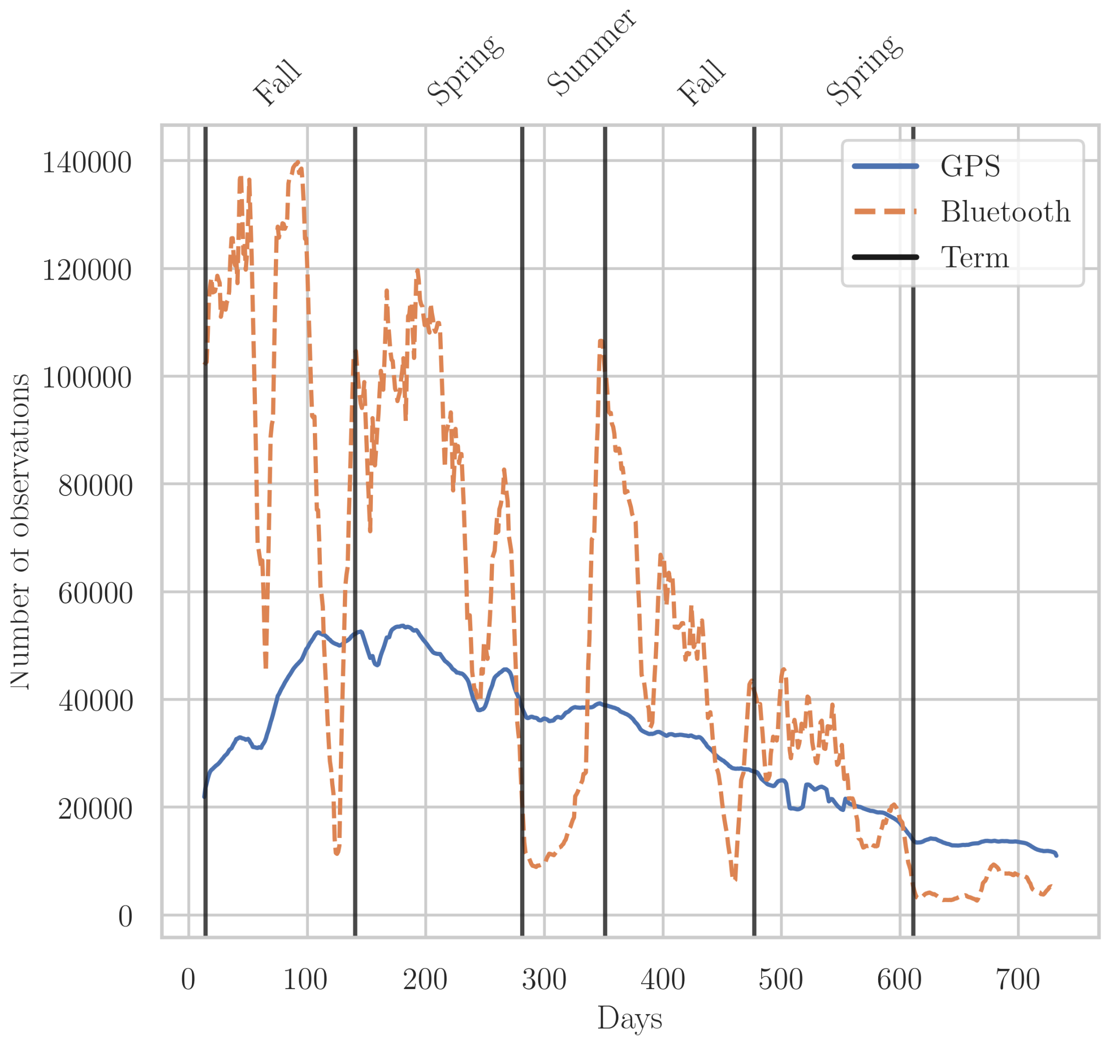

Figure 3.1: Number of Observations

The figure shows the two-week moving average of the total number of observations for the whole study period. There was a clear and pronounced drop off after the second term of the first academic term in the number of observations for term three and four, which I therefore excluded.

participants in the CNS were male and between 19 and 21 years old. This bias clearly limited my ability to accurately account for both the effects of age and gender within the dataset (Section 2.5). Second, the dataset only included students at one particular university, the DTU, in Copenhagen. Students in fact might have a very different social and mobility behaviour than other social groups. Their day to day schedule is typically organised around lectures on campus; arguably heavily influencing both with whom they socialise and where they travel to. Thus, extrapolating my findings to a wider population should either be only undertaken with caution or might very well not be possible at all. However, the relatively unique background and life situation of the students on the other hand provided a relatively large sample of relatively homogeneous individuals to study the effects of various mediating factors.

3.1.2 Pre-Processing

I used the Bluetooth traces of the students to infer, which other students they encountered over the course of the study. As I had the MAC addresses of each of the student's smartphones I could easily map the Bluetooth scans to individual students. However, I could not reliably map the MAC address of other devices to individuals external to the study, as I did not know what type of device the other addresses represents. Thus, I only considered the Bluetooth data of the other participants of the study and further only traces that had a signal strength of -80 dBm or stronger. Sekara and Lehmann (2014) showed this to be a reliable cut-off value for close and unobstructed physical proximity for this dataset.

For location traces I focused my analysis on stop locations and discarded any GPS traces for which I detected movement (that included movement on buses or trains and with bikes or cars). I further adopted the convention of (Cuttone, Lehmann, & Larsen, 2014) that a user must spend at least ten minutes at a location to be a meaningful location. I also enhanced the spatial and temporal resolution and fidelity of my location traces by adding information about a user's proximity to geo-located WiFi hotspots as described in in (Sapiezynski et al., 2015)[1] and merged locations that were very close together via DBSCAN (Pedregosa & Varoquaux, 2011) with $\epsilon = 5$ and the minimum number of points in a cluster set to five to avoid classifying noise as a shared location, and using the Haversine distance (see Figure 3.2 for choosing an appropriate value of ϵ).

For chapter 6 I also used the *social context* a student resided in as a dependent variable. Sekara et al. (2016) noted that there was a distinctive difference in how a certain set of individuals participated in a physical gathering. The core individuals of a gathering participated significantly longer in the gathering than other individuals. To derive the *social context* a students was in, I first hierarchically clustered all physical encounters between students to find all gatherings in the data. In a second step, I found the maximum gap in the distribution of participation rates for each gathering. If the gap was significantly larger, then I could expect to occur by chance, all individuals before the gap formed the social core, or in my case the *social context*.

As the meetings of the core individuals were also fairly regular and consistent (Sekara et al., 2016), they allowed me to label each gathering an individual is part of via its core members. Thus, I could observe which social cores or groups a student encountered over the course of the study. For the details of how I inferred the social cores, I refer the reader to Sekara et al. (2016).

[1] I however only enhanced the data for Chapter 5 and 6 as the method was not published when I worked on Chapter 4. As Chapter 4 deals with the comparative performance of different prediction setups this should not affect the results.

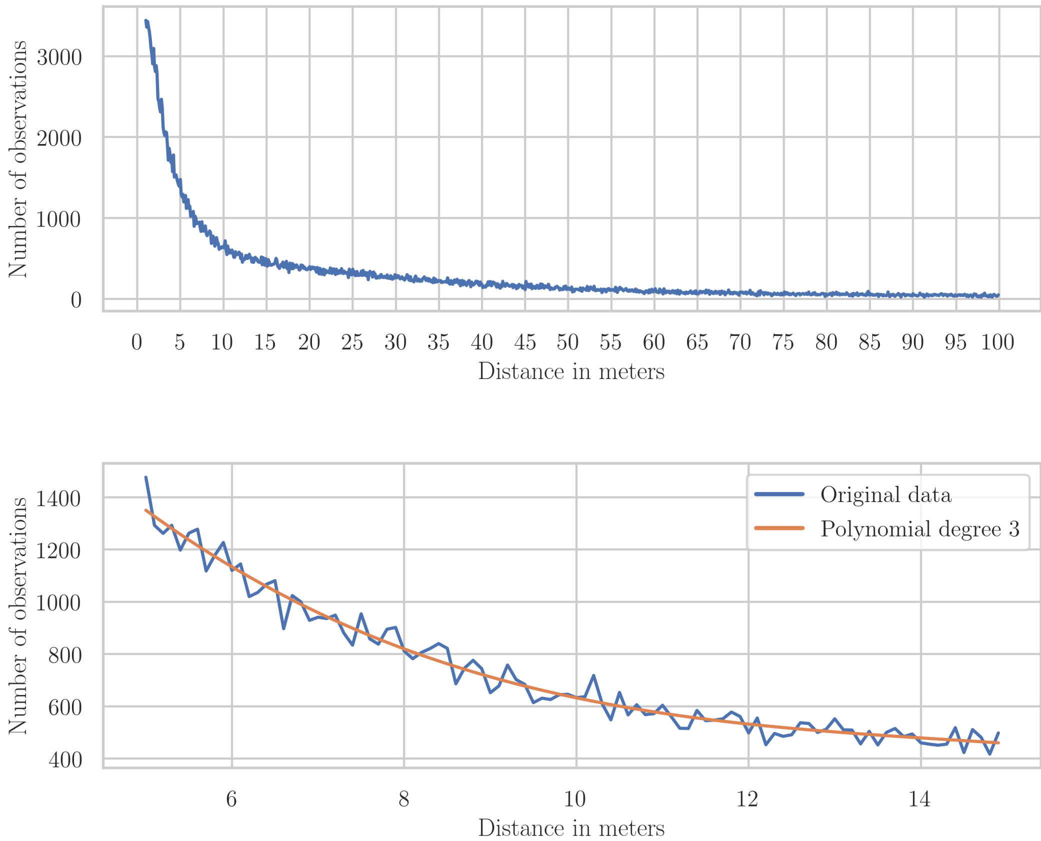

Figure 3.2: Distance to the Four Nearest Neighbours

I used the data for the first term to determine the appropriate value for DBSCAN. It is generally recommended that ϵ is chosen where the above plot shows an "elbow" (Schubert, Sander, Ester, Kriegel, & Xu, 2017). While the most suitable range for ϵ was between 5 and 15 meters, when zooming in there was no clear "elbow" visible any more. As smaller values of ϵ are generally preferable to larger ones (Schubert, Sander, Ester, Kriegel, & Xu, 2017), I thus decided to set ϵ to five meters.

3.2 Methods

I used a wide variety of methods ranging from applied machine learning over time series analysis to local indicators of spatial association for each chapter. The relevant methodology was thus only discussed when needed in each chapter. However, several key metrics were used in more than one chapter. I briefly review these in this chapter.

3.2.1 General

The Shannon entropy of a random variable X is defined as (Mezard & Montanari, 2009):

$$H(X) = -\sum_{i=0}^{N-1} p_i log_2 p_i \tag{3.1}$$

where p_i in my case is the empirical probability of observing state i of X. In case, I am dealing with time series data, $p(i)$ is the temporary uncorrelated probability that the time series is in state i (Song, Qu, et al., 2010).

The important properties of H that make it useful as a measure of information content are (Mezard & Montanari, 2009):

1. $H(X) \geq 0$.

2. $H(X) = 0$ if and only if the random variable X is certain. This means that X only has one state with probability one.

3. $H(X)$ is maximum when all M events i of X are equi-probable with $p(i) = 1/M$. The entropy is then $H(X) = log_2 M$.

4. If X and Y are two independent random variables, then $H(X,Y) = H(X) + H(Y)$.

5. For any pair X, Y of random variables in general $H(X,Y) \leq H(X) + H(Y)$.

6. Additivity for composite events. $H(X,Y) = H(Y) + H(X|Y) = H(X) + H(Y|X)$.

Intuitively, the Shannon entropy gives a measure for the uncertainty of a random variable, or how much information one could learn by knowing the particular state of a random variable. The larger the entropy the less information one has a priori about the random variable (Property 2 and 3). Entropy, however, also lends itself as a measure of the diversity of a discrete random variable. The higher the entropy the higher the diversity; in other words the less certain one is about the state of a random variable a priori the more diverse it is. For the rest of the thesis, I used entropy as a measure of diversity of a discrete random variable, where I used the empirical probability distribution to calculate $H(X)$.

Directly related to $H(X)$, is the concept of mutual information (I). At its core $I(X,Y)$ measures how much one random variable Y can tell us about another random variable X (Latham & Roudi, 2009). Mutual information is defined as (Latham & Roudi, 2009):

$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \tag{3.2}$$

where $H(X|Y)$ is the conditional entropy, that means the average uncertainty about X after observing a second random variable Y (Latham & Roudi, 2009). Thus, $I(X,Y)$ is the reduction of $H(X)$ after observing Y, or how uncertain I am about X after knowing the state of Y.

3.2.2 Social Network

Since its inception modern Social Network Analysis made heavy use of graphs to represent ties between actors (Freeman, 2001). In this framework social relationships between individuals are viewed as an edge in a graph $G = (V,E)$, where V represents the set of vertices or individuals and the set of edges E. For each pair of individuals $u, v \in V$ the graph G has an edge $e \in E$ if and only if one observes a relationship between u and v. Analogously the connections can also be represented as an adjacency matrix, where its elements are non-zero for each corresponding edge. For example, given the adjacency matrix $\mathbf{A}$ the element $\mathbf{A}_{u,v}$ represents the relationship between u, v.

In a time-invariant graph a common metric to measure the connectedness of vertices is degree centrality defined as

$$C_D = deg(v) = |E_v| \tag{3.3}$$

where E_v is the set of all edges incident to v.

For my thesis I was however dealing with temporal data. This meant that instead of having only one G, I had a time-ordered set of several G or one time-varying graph G_t. A time-varying graph $G_t = (V_t, E_t)$ has a different set of vertices and edges at each time point t. Alternatively G_t can be represented as a three dimensional tensor $\mathbb{A}$, where $\mathbb{A}_{u,v,t}$ represents the relationship between u and v at t.

In a time-varying graph G_t, one can still calculate C_D for each time slice

$$C_{D_t} = deg(v_t) \tag{3.4}$$

to derive a time series of C_D and then analyse the time series.

However, C_D does not account for the variety of social relationships an individual might have. To account for the diversity of the set of peers of an individual, I calculated $H(\phi_v)$, where ϕ are all observed relationships with alters for v over a set of time slices $\{t_{i+1}, t_{i+2}, ..., t_{i+n}\}$.

3.2.3 Mobility

To assess the average travel distance of individuals I calculated the radius of gyration (González, Hidalgo, & Barabási, 2008b). It is defined as follows:

$$r_g(t) = \sqrt{\frac{1}{n_c(t)} \sum_{i=1}^{n_c} (\mathbf{r}_i - \mathbf{r}_{cm})^2}, \tag{3.5}$$

where $\mathbf{r}_i$ represents the $i = 1, ..., n_c(t)$ position on the Earth's surface recorded for an individual and $\mathbf{r}_{cm} = 1/n_c(t) \sum_{i=1}^{n_c} r_i$ is the centre of mass of that individual's trajectory.

In other words the radius of gyration is the square root of the average squared distance between each visited location and the centre of mass of a user's trajectory, or the linear size occupied by a user's mobility pattern. The radius of gyration is conceptually very similar to the standard deviation (Agresti & Barbara, 2009):

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{3.6}$$

where $\{x_1, x_2, ..., x_n\}$ are the sampled values of a distribution. Intuitively both describe the deviation of sampled values from a centre; the Radius of Gyration then indicates how far an individual deviates from their centre of mass (i.e. they size of their typical geographic activity space). Note that similarly to the sample mean the Radius of Gyration does not take multi-modality of the data generating process into account.

However, the Radius of Gyration does not account for the diversity of the visited locations. I again calculated $H(\rho_v)$ where ρ_v is the set of all visited locations of the individual v for a given time period to account for the variety of locations.

'Somewhere' is always critically tied to the 'somewhere' of a moment earlier.

Torsten Hägerstrand

4

Predicting Future Co-occurrences

4.1 Introduction

Few would doubt that space, time and the social realm are intrinsically linked. Geography has always been interested in the role spatial, temporal, and social factors play in shaping human behaviour. However, it can be rather difficult to separate the effect an individual factor has on human behaviour from other dynamics. After all, human behaviour is inherently interwoven with space and time. As Hägerstrand (1970, p. 10) emblematically stated "'somewhere' is always critically tied to the 'somewhere' of a moment earlier".

Several studies tried to disentangle space, time, and social factors in recent years. Backstrom et al. (2010) showed that the probability of friendship between people decreases with distance. Scellato, Noulas, Lambiotte, and Mascolo (2011) studied the properties of location-based social networks and found that about 40 percent of all links in location-based social networks were shorter than 100km. Lambiotte et al. (2008) concluded that the likelihood of a tie in a mobile communications network followed a gravity model (i.e. the likelihood of a tie between two users decreased exponentially with distance). Toole et al. (2015) employed the coupling of social ties and mobility behaviour to build a mobility model that included choices based on social contacts. They showed that the ratio of acquaintances, co-workers, and friends/family in a person's ego network shaped their mobility behaviour. Studying the mobility patterns and virtual interactions of people, Larsen et al. (2006) argued that nearby strong ties were crucial for a individual's network as they found that phone calls, texting, and face-to-face meetings became less regular with distance.

Recently researchers also called attention to how space itself could influence personal relationships (Adams et al., 2012). Boessen et al. (2017) discovered that the built environment had a significant effect on how people socialised. They highlighted the potential role the built environment could have for fostering the formation of social ties. Both Doreian and Conti (2012) and Butts et al. (2012) showed that the structure of social networks could be partly explained by spatial factors.

Noulas, Shaw, Lambiotte, and Mascolo (2015) and Scellato, Noulas, and Mascolo (2011) both utilised the social and spatial properties of location-based social networks to propose a link-prediction model. Brown et al. (2013) developed a model for the evolution of city-wide location-based social networks, which demonstrated that friends tended to meet at specific—more "social"—places. De Domenico et al. (2013) used the mobility data of friends to improve user movement prediction. Last, Cho et al. (2011) built a mobility model incorporating both periodic movement of individuals as well as corporeal travel induced by social ties.

An extensive amount of research has already been conducted on the interplay between the social realm, place, and time. However, studies so far were either limited to a very specific type of network or did not jointly deal with all three factors. On the one hand, several studies that accounted for spatial and temporal features focused on a narrow set of social interactions such as online social networks or encounters in face-to-face networks. One group of research projects studied very topical online social networks such as the Foursquare network (Scellato, Noulas, & Mascolo, 2011) or the Flickr network (Crandall et al., 2010), while another group focused on studies of face-to-face encounters solely in highly structured and defined settings such as a museum, a conference, or a primary school (Isella et al., 2011; Stehle et al., 2011; Zhao et al., 2011). Whereas Noulas et al. (2015) analysed spatial, temporal, and social features but focused on networks of places instead of individuals.

On the other hand, studies that analysed more broadly defined social networks did not assess spatial and temporal features at the same time. Although Yang et al. (2013) used information about when and in which network configuration people have met as features for their link-prediction algorithm, they did not incorporate spatial features. Sekara et al. (2016) utilised the regularity of social group structures to predict missing members of the group. However, place did not play a role in their subsequent prediction task. While Wang et al. (2011) successfully employed the similarity of trajectories of users for predicting phone calls between users, they did not take any other temporal or spatial features into account.

In short, I believed that a joint assessment of spatial, temporal, and social features is crucial for understanding the true dynamics behind social encounters as human interactions might be spatially, temporally, and/or socially confounded with each other.

Consequently, my contribution consisted of three parts:

1. ascertaining whether geographic places themselves hold discriminatory power,

2. assessing the "simultaneous" predictive information of geographic,temporal, and social features for a changing network of encounters, and

3. understanding if different types of social encounters networks influence the overall predictability.

Overall, I tried to better understand what factors drive the evolution of a human social encounter network, and how I could use salient features for predicting future encounters.

4.2 Problem Definition

A common way of dealing with social relations within populations is to view social ties—in my case social encounters—as edges (hereafter also links and ties) in a graph. Conceptualising social relations as edges in a graph had the advantage that analysing social relations as graphs was fairly well studied problem and allowed me to rely on state-of-the-art methods for predicting future encounters (Peng, Baowen, Yurong, & Xiaoyu, 2015). Furthermore, viewing the problem as a time-varying graph enabled me to account for social network dynamics. In particular, I phrased the problem of predicting an encounter as a link prediction problem in a time-varying graph G_t that represents encounters.

4.2.1 Encounter

For my study I defined an encounter as physical proximity as measured by a smartphone via a Bluetooth measurement. I used a Bluetooth signal of -80 dBm or stronger to indicate encounters as Sekara and Lehmann (2014) showed this to be a reliable cut-off value for close and unobstructed physical proximity for this dataset. Given that I was only interested in time spent at stop locations, this meant an encounter in my study represented either the physical co-location of two students in the same room or in close proximity outdoors. Sekara et al. (2016) used this definition of face-to-face encounter to study the evolution and structure of dynamic social networks.

However, I was not interested in predicting short encounters that are only due to chance but rather in more meaningful, longer encounters. Thus, I adopted the convention of the *Rochester Interaction Record (RIR)* for meaningful encounters, where they were defined to last at least ten minutes (Reis & Wheeler, 1991).

4.2.2 Social Encounter Graph

To construct the time-varying, undirected social encounter graph $G_t = (V_t,\ E_t)$, where V_{t} are the set of students at t and E_{t} the set of all meaningful encounters between them, I first discretised my data into intervals of 30 minutes. I chose an interval of 30 minutes to be able to account for the irregularity of the Bluetooth measurements and still be able to find meaningful encounters between students. In case, a meaningful encounter of at least ten minutes was not represented in the resulting graph due how I discretised the time steps, I assigned it to the period t with which it had the biggest overlap; I broke ties between intervals randomly. As the majority of interactions in the dataset were either shorter than ten minutes or significantly longer than ten minutes, this did not significantly alter the resulting graph (Figure 4.1). To summarise, any edge $e \in G_t$ represents a meaningful encounter between students that was at least ten minutes long as observed by at least one student.

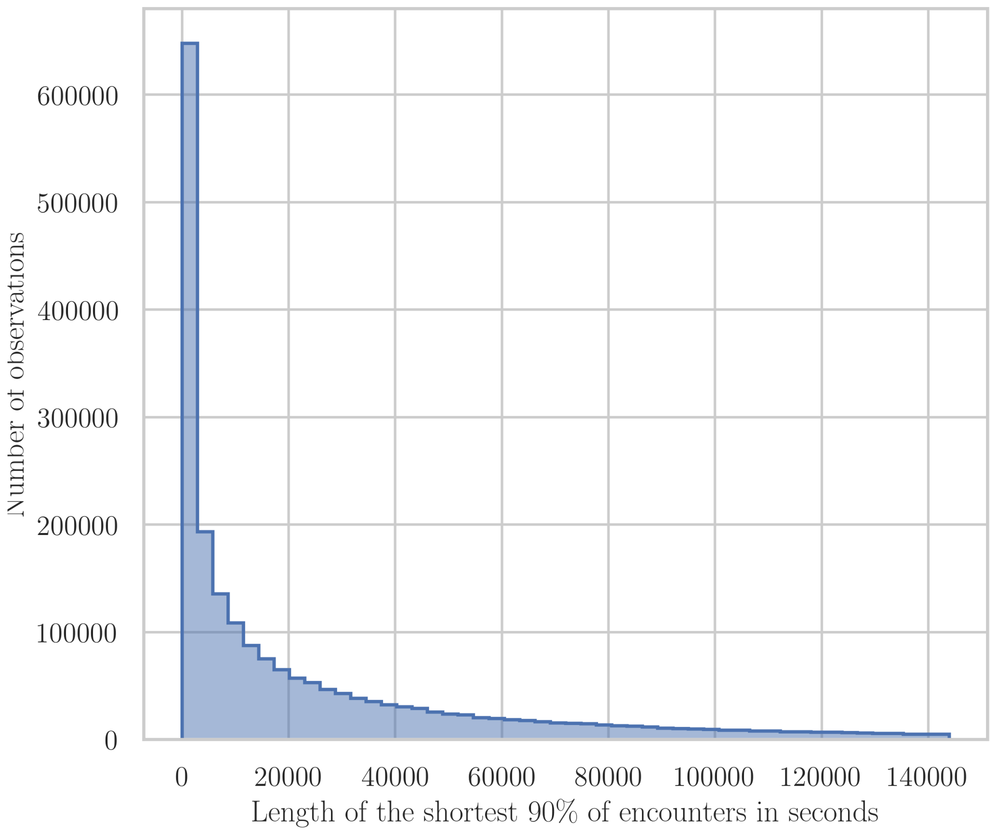

Figure 4.1: Distribution of the Encounter Lengths Between Students

The figure shows the distribution of inferred lengths of for the shortest 90% of encounters in seconds Note that most encounters were either shorter than 10 minutes or considerably longer and that the distribution was heavily skewed with a heavy tail.

4.2.3 Link Prediction

As I conceptualised social encounters as edges in a graph, the problem of predicting future encounters between any two students became equivalent to predicting whether an edge between nodes in the graph exists. More formally, in a human encounter network G_t, the link prediction task is to predict whether e at time $t + n$ exists for the vertices $u, v \in V_t$. Effectively, I was trying to predict who will meet whom for ten minutes or more during period $t + n$. This is equivalent to predicting all the new ties that form, the ties that do not change, and all the ties that will dissolve from time period to the next, or in other words predicting the network structure of G_{t+n}. Formulating the problem this way had the advantage of including link dissolution—a not well studied problem in link-prediction (Peng et al., 2015)—quite naturally in the problem definition.

4.3 Predicting Future Encounters

After defining my problem in the previous section, I specify how I implement my approach for predicting links between nodes. In particular, I describe which algorithm I used for prediction, which features I used for predicting future encounters, and how I built my models.

4.3.1 Random Forests

Random forests consistently performed well in link-prediction tasks (Peng et al., 2015). I thus opted to use them for my prediction task as well (Pedregosa & Varoquaux, 2011). At its core, random forests are an ensemble learning algorithm for classification built upon decision trees. However, decision tress are sensitive to initial conditions (Altmann, Tol Si, Sander, & Lengauer, 2010) and can easily over-fit the data (Ho, 2002). To deal with these problems Breiman (2001) proposed to use a set of decision trees. He defined a random forest as a classifier that consists of a collection of tree-structured classifiers $\{h(x, \Theta_k), k = 1, ...\}$, where $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a vote for the most popular class at input x. To protect against over fitting each split of a decision tree only considers a random subset of all features.

Recall that I was trying to predict future social encounters between $u, v \in V_t$. Thus, for each individual u, I trained a separate random forest classifier R. R can be understood to be a mapping from my input space (the features I used for prediction) to the output space (encounters of students at G_{t+n}). Thus, each R tried to learn for each user u their individual function of whether u and v would encounter each other in the next time period. Conditional probabilities can be estimated by simply counting the fraction of trees in the forest that vote for a certain class, which usually delivers good probability estimates (Olson & Wyner, 2018). The probability of an edge e between u and v could then be seen as the average fraction of trees that voted for e between u and v. Note as each user u had its own R the estimated probability of the edge e from u to v, might be different from the edge e from v to u.

4.3.2 Features

I generally used features that had been used in the literature for my link-prediction task. All my features accounted for the general likelihood of an encounter occurring, for the various contexts an encounter could take place in, or were derived from the encounter graph of the students. The three contexts I was particularly interested in understanding their role for encounters were time, space, and social factors and thus most of my features were related to them. In order to assess the predictive information of each of those contexts, I created the following five sets of features:

Baseline Features

Baseline features accounted for the idea that two students that met each other often and frequently were more likely to meet each other in the future than two students who hardly ever met. I constructed as baseline features for all my models whether the two nodes met in the previous time period or in other words whether I could observe a tie between them (*edge*), the amount of elapsed time since the last meeting (*recency*), and the total amount of time I observed two nodes together (*time spent together*) as described in Yang et al. (2013).

Temporal Features

The time related features captured variations in temporal behavioural patterns as when two students met could in itself be an important clue for the type of relationship between two students. For example, if two students only ever meet during normal working hours then they are most likely just colleagues at university, but if they also meet after work or on the weekend then their relationship should be qualitatively different. Let M now be the set of all meetings between two nodes u, v in the training period. I built a feature vector (*hour-of-day(M)*) of length 24, that counted the total amount of the encounters between u and v at every hour of the day as well as feature vector (*day-of-the-week(M)*) of length 7, that counted the total amount of encounters between students at every day of the week. If an encounter occurred in more than one bucket, I distributed it proportionally for both *hour-of-the-day* as well as *day-of-the-week*. I also included the current *hour of the day* as well as the current *day of the week* as a feature, so that each R could keep track of when and where the current encounter occurred.

Spatial Features

I observed that there was a difference in whether two people meet at a place a lot of people visit and thus with high place entropy or at "quieter" place with low place entropy. Or in other words, if two student met at the university, a very popular place for students, the information content of that meeting was relatively low, but if two people met at their respective homes then this was a much more unlikely and more noteworthy event. I thus derived the minimum *place entropy* of the set of all observed locations of meetings between any u, as as feature as well (Scellato, Noulas, & Mascolo, 2011).

I also inferred the *relative importance* of each venue for each user u by measuring the amount of time a user spent there. I then ranked the venues by the *relative importance* for each user. Arguably the more time a student spent at a location the more important that location was for that student; encounters at more important locations as measured by the time students spent there could thus signify a more important social relationship as well. I thus also included the *maxRank(relativeImportance(u, v))* of any meeting between u, v.

Based on Oldenburg's seminal paper (Oldenburg & Brissett, 1982), I derived geographic contexts in which encounters occurred as features as well. For a graphical overview of the derived aggregated behaviour of the students in each setting, see Figure 4.2. Oldenburg argued that in order for communities to thrive they needed places away from the home ("first place") and the workplace ("second place"); hence they needed "third places". Examples of third places were cafes, clubs, and parks. Several studies used Oldenburg's concept of "third places" to highlight the importance they played for social encounters (for examples see among others Glover and Parry, 2009; Mair, 2009 and Rosenbaum, Ward, Walker, and Ostrom, 2007). Others used a classification similar to Oldenburg's to understand and predict human mobility on a larger scale (Cho et al., 2011; Eagle & Pentland, 2009).

Analogous to Oldenburg I distinguished between several different geographic settings a student could be in: the *home*, the *university*, a *third place*, and *other*. I inferred the locations as follows:

First, I found the home location for each student by clustering all his or her location measurements between 11PM and 4AM using DBSCAN (Ester, Kriegel, Sander, & Xu, 1996)[1] into the set of spatial clusters C. I used DBSCAN as I did not have to specify the amount of clusters beforehand as I do not know how many clusters each individual might have. Each cluster $c \in C$ then represented an area where a lot of locational measurements were taken for that user. I then selected $max(|c|)$ as a student's home location.

Second, for assigning students to the *university* context I mapped the campus of their university and checked whether students were within 50 meters of the campus. As some students lived in dormitories on campus I gave precedence to the *home* location when assigning location measurements to their respective contexts.

Third, to infer *third places* I adopted the approach of Sekara et al. (2016) for inferring significantly more important contexts given a distribution of observed times in a given context. For each student, I constructed the set of all the stop locations S a student visits. For each $s \in S$, I could also observe the amount of time *t(s)* a student spent there and rank the resulting distribution of stop times in descending order, giving one $T(s)$. I observed that for most students there was a clear gap in $T(s)$; this implied that students visited some locations very often and some locations very rarely. I defined as *third place* any location s that appeared before the biggest gap in $T(s)$, where the biggest gap in $T(s)$ was significantly larger than I would expect by random sampling of stop times from a uniform distribution, that was neither *home* nor *university*. This way, I could ensure that *third places* were only places where students spent significantly more time than at all other locations they visit.

Fourth, any other $s \in S$ was classified as *other*.

[1]I used the implementation of DBSCAN from Pedregosa and Varoquaux (2011).

Let $context(u, v)$ now be the function that counts the amount of time two nodes $u, v \in G$ have spent together at the different geographic contexts: *Home, work/university, third place,* and *other*. I included the amount of time spent at each spatial context as a feature. The reasoning was that the amount of time two nodes spent together in different geographic settings should contain information about the type of their relationship. I used the current spatial context—home, university, third place, or other—of the encounter as a feature as well. It seemed reasonable to expect that two students, who met regularly in a certain setting were more likely to meet should one of them currently be in that setting.

Last, I included the *Jaccard similarity*, $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$, where A and B are the set of visited locations (Ranjan et al., 2012). There is evidence that the more similar two individuals were with respect to their mobility the more likely they were to be friends as well (Bapierre et al., 2015; Toole et al., 2015) and thus might be indicative of future encounters.

Social Features

I also accounted for the social setting an encounter occurs in. If two students met at the university during a course this was nothing extraordinary in my dataset, but if two students met alone on the campus there was a higher likelihood that they were socialising. Let now $P_{u,v}$ be the distribution of the number of other people from the study that are present when two nodes $u, v \in G$ meet. I then used $avg(P_{u,v})$ as a feature.

What is more, the social configuration two students met in could also play an important role for predicting future encounters. Building upon the concept of triadic-closure, that is the phenomenon in social network that friends of friends are likely to become friends themselves, Yang et al. (2013) proposed to use triadic periods as a feature for predicting encounters. The main idea was to count the different possible arrangements of triads in the encounter graph, or in other words the different possible configurations of co-locations

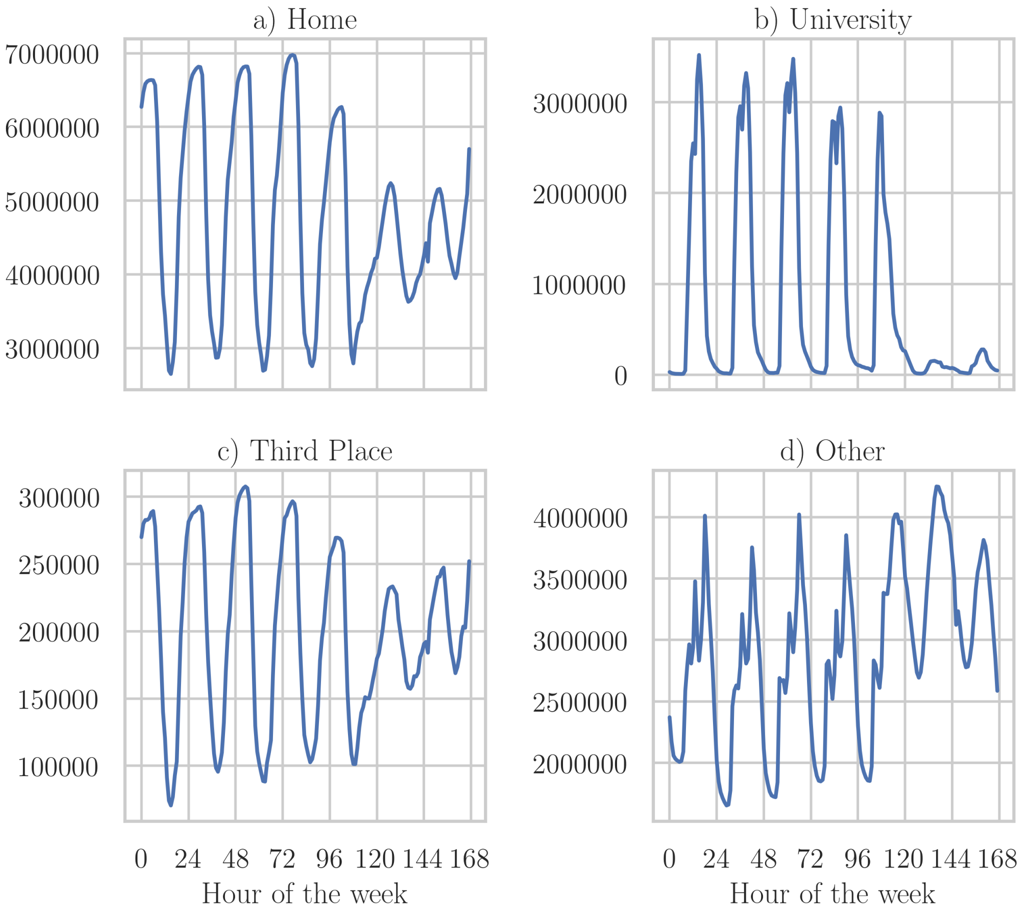

Figure 4.2: Weekly Aggregated Observations by Geographic Context

The figure counts the total amount of observations at each hour of the week for each of the inferred geographic contexts: *Home, work/university, third place,* and *other* and thus provides an aggregate view of the activity of the students. I could observe that students in aggregate mostly followed a diurnal pattern. During weekdays students left their *home* in the morning, then attended *university* before going to an *other* location. The observed pattern for *third places* mirrored that for the *home* setting indicating that *third places* in my dataset were places were students would spend the night (possibly at a partner's home location). On the weekends, the sutdents could not be observed at the *university* often but could regularly be found at *other* locations. Overall, the aggregated patterns of activity of the students for each context were in line with what I expected. Interestingly, students seemed to spend nights at *third places*, thus further highlighting the importance of those places.

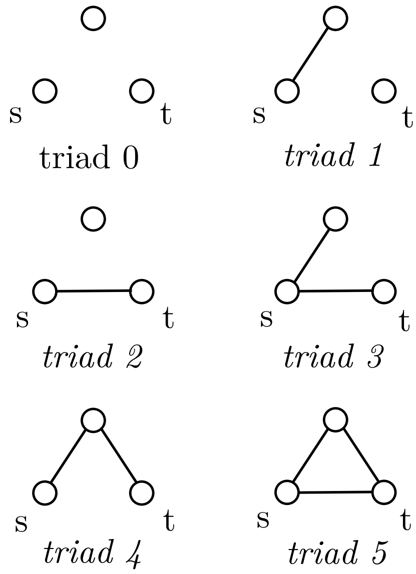

Figure 4.3: Triadic Periods

The figure shows all possible configuration of triadic periods or in other words all possible configuration of edges in a (sub)graph G with only three nodes.

at a particular location (for all possible configurations of triadic periods see Figure 4.3). This is equivalent to accounting for the immediate neighbourhood of every u in G_t.

Interestingly Bianconi et al. (2014) showed that triadic closure was a leading driver in how social networks evolve. And triadic periods likely accounted for the dynamic of triadic closure in the encounter graph as well.

Network Topology Features

In previous studies on link-prediction features derived from the wider network topology of the social graph were used extensively. The core idea of all network metrics is that friends of friends are likely to become friends themselves. However, they differ in how they formulate this idea mathematically. In particular, I included *preferential attachment (PA)*, *weighted prop flow (weighted PF)* and *Adamic-Adar (AA)* (Peng et al., 2015) after seeing favourable performance for those three metrics when designing my experiments. The *PA* metric indicates that new nodes will more likely attach to nodes that already have a high degree. It is defined as $PA(u, v) = |\Gamma(u)| \cdot |\Gamma(v)|$, where $\Gamma(v)$ is the set of neighbours of node v and $|\Gamma(v)|$ be the number of neighbours of node v. PF is the probability that a restricted random walk starts at node u and ends at node v with no more than s steps. Weighted PF uses the weights of links (in my case how much time two students spent together in the previous time period) as transition probabilities. AA is defined as the inverted sum of the logarithmic degrees of neighbours shared by the two nodes $A(u, v) = \sum \frac{1}{\log|N(u)|}$, where $N(u)$ is the set of nodes adjacent to u.

4.3.3 Evaluating Temporal Prediction Models

I used the first academic term for building and validating my model, whereas I tested my hypotheses on the second academic term of the dataset, where each term consisted of 13 weeks. As I was dealing with time series data, I used one-step forecasts with re-estimation as described in Hyndman and Athanasopoulos (2013) to make sure my models did not have access to training data from the future, where a step was 12.5% of the data and I used at least 50% of the available data to train each model. In other words, I evaluated my model at four different time points for the second half of the available data, where I retrained my model for each time point with all available data at that time point.

4.3.4 Search Space

Every $u \in G_t$ has N potential candidates for encounters at G_{t+n} as every node can meet every other node. Thus, the unrestricted search space is $N * (N - 1)$. This was impractically large as in my data I would need to predict more than twelve billion potential edges for each term. A common strategy to deal with the huge search space is to only consider as potential candidates for a new tie nodes that are thought of to be more likely to become connected in the first place. It is known that in social networks friends of friends are more likely to become friends than by chance alone and this property could be exploited for a prediction task (Scellato, Noulas, Lambiotte, & Mascolo, 2011). To limit the computational complexity, I adopted the convention of Scellato, Noulas, Lambiotte, and Mascolo (2011) for my work, where I restricted the prediction space to alters that a student had either encountered before or whom a student's alters had themselves encountered before (i.e. friends of friends).

4.3.5 Feature Preparation Interval

I had to decide on how many temporal slices of G_t I used to construct my features. However, several of the features I was interested in representing longer term dynamics between students such as the places they usually met and how similar their trajectories were, whereas several other features such as the other people present at a current meeting represented shorter term dynamics. I thus opted to introduce a longer term feature preparation interval $\triangle\tau$ and a shorter feature preparation interval $\triangle T$ that I used to generate the appropriate features.

Yang et al. (2013) showed that the length of the feature preparation interval has an impact on the performance of the resulting link prediction. To determine the most appropriate hyper-parameters for my model, I tested the performance of my model with various values of $\triangle T$ and $\triangle\tau$ for the first academic term (Figure 4.4 and Table 4.1). In particular, I was interested if values of $\triangle\tau$ that corresponded to longer periodicities such as two and four weeks and longer intervals for $\triangle T$ might improve the performance of my models. I found that a $\triangle T$ interval of 30 minutes and a $\triangle\tau$ interval of one week respectively had the best performance and I used those values for training and evaluating the remaining models for the second academic term.

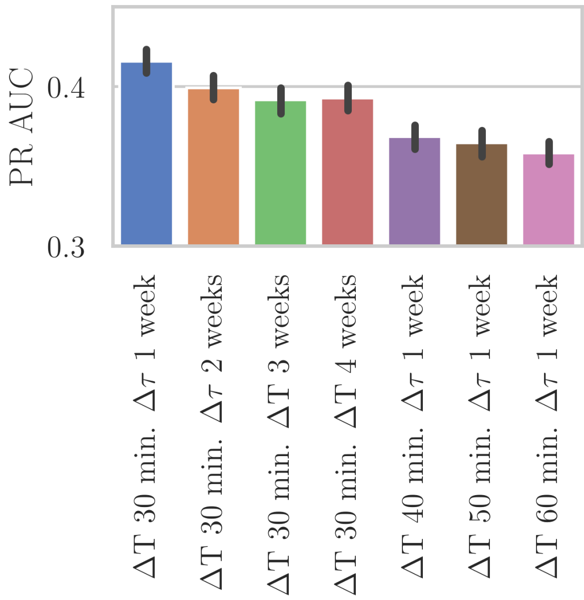

Figure 4.4: Performance Hyper-Parameters

The plot shows the effect of various values of $\triangle\tau$ and $\triangle T$ had on the performance of my link-prediction task, where the error bars represent the 95% confidence intervals. While the overall differences between the models were relatively small, the model with $\triangle\tau$ of 30 minutes and $\triangle T$ of one week clearly performed best. Thus, I have used those values for building and evaluating my models for the second term.

	Mean	CI 95%
ΔT 30 min. $\Delta\tau$ 1 week	0.42	(0.40,0.42)
ΔT 30 min. $\Delta\tau$ 2 weeks	0.40	(0.39,0.41)
ΔT 30 min. ΔT 3 weeks	0.39	(0.38,0.40)
ΔT 30 min. ΔT 4 weeks	0.39	(0.38,0.40)
ΔT 40 min. $\Delta\tau$ 1 week	0.37	(0.36,0.38)
ΔT 50 min. $\Delta\tau$ 1 week	0.36	(0.36,0.37)
ΔT 60 min. $\Delta\tau$ 1 week	0.36	(0.35,0.37)

Table 4.1: Cross-Validation Precision-Recall AUC Scores

The tables lists the effect of various values of $\triangle\tau$ and $\triangle T$ had on the performance of my link-prediction task, where the 95% confidence intervals are reported in the column to the right of the scores.

4.3.6 Model Construction

In order to test the importance of each domain for predicting future encounters, I constructed several different models. Each of the models I built has access to a different set of features. Should the context of an encounter have played a role than my contextual features should have also been relevant for predicting future encounters. Table 4.2 lists each model and its corresponding features and also indicates, if the feature was derived using $\triangle T$ or $\triangle \tau$ as a feature preparation interval.

As a benchmark to test my predictions against I first developed a *null* model for a time-evolving weighted encounter graph with dissolving ties. My *null* model was adapted from Newman and Girvan (2004), where the edges of the graph were randomly rewired under the constraint that the expected degree matches the original degree distribution. In my case, this meant that the expected amount of encounter of each node $u \in G_t$ followed the original distribution of meetings, but the encounters between any two nodes $u, v \in G_t$ were chosen at random.

Besides the null model, I constructed a *base* model that only contained the baseline features. I further built a *temporal* model, a *social* model, a *spatial* model and a *network topology* model by adding to the base models the feature set that pertains to that domain. The *context* model consisted of the baseline features as well as the temporal, spatial, and social features. The *full* model consisted of all features. I also, after my experiments, constructed a *refactored* model based on top five features of the *full* model.

Sometimes one however might not have access to the whole network and might only be in possession of node level data. Hence, one is unable to calculate or reliably estimate the features that utilise the wider network topology I described above. I simulated such a scenario by building *node* model that only incorporated features that could be obtained from the ego-network of a node. In particular, the features for the node model were: The baseline features, and all the spatial, temporal, and social features with the limitation that “triad 4” could not be distinguished from “triad 1” and “triad 5” not from “triad 3”.

Feature	Base	Node	Soc.	Spat.	Temp.	Con.	Net.	Full	Ref.	$\triangle\tau$	$\triangle T$
Recency	x	x	x	x	x	x	x	x		x	
Activeness	x	x	x	x	x	x	x	x		x	
Overall time tog.	x	x	x	x	x	x	x	x		x	
Current hour		x			x	x		x			x
Hour-of-the-day vector		x			x	x		x		x	
Current weekday		x			x	x		x			x
Day-of-week vector		x			x	x		x		x	
Place entropy		x		x		x		x			x
Min(place entropy)		x		x		x		x		x	
Relative importance		x		x		x		x			x
MaxRank(relative importance)		x		x		x		x	x	x	
Current spatial context		x		x		x		x			x
Time at home tog.		x		x		x		x		x	
Time at university tog.		x		x		x		x		x	
Time at third places tog.		x		x		x		x		x	
Time at other places tog.		x		x		x		x		x	
Jaccard trajectories		x		x		x		x		x	
Avg. amount of people		x	x			x		x	x		x
Triadic periods		0,1,2,3	x			x		x	0,3		x
Preferential attachment							x	x			x
Weighted Prop Flow							x	x	x		x
Adamic Adar							x	x			x

Table 4.2: Model Features

The table depicts the various models and the set of features that was used for training, where the rows represent the features and the columns the models as described in Section 4.3.6.
Numbers for *triadic periods* indicate that I have only used that particular *triadic period* as a feature.

4.4 Findings

To compare the performance of my different models I chose to report the precision, the recall, the precision-recall curve, and the area under the precision-recall curve ($PR\ AUC$), where precision is plotted on the y-axis and recall on the x-axis (Davis & Goadrich, 2006). Precision and recall are defined as follows (Davis & Goadrich, 2006): In a binary classification task, true positives (TP) are instances correctly labelled as positives, whereas false positives (FP) are incorrectly labelled as positives. Conversely, true negatives (TN) are examples correctly labelled as negatives and false negatives (FN) refer to positive examples erroneously labelled as negatives. Recall is then $\frac{TP}{TP+FN}$ and precision $\frac{TP}{TP+FP}$.

The area under the PR curve can then be directly used to compare the performance of different models (i.e. the bigger the area the better the model) and is suited to evaluate the performance of an algorithm if there is a large class imbalance as in my data (Davis & Goadrich, 2006).

4.4.1 Performance of the Link-Prediction Algorithm

First, all models performed significantly better than the *null* model, however only the *network* model had a higher PR AUC score than the *base* model (Figure 4.5 and 4.6 as well as Table 4.3). Unsurprisingly, nodes are not randomly interacting with other nodes but exhibit learnable patterns (at least to a certain degree).

Second, the models that had the highest PR AUC score were the *network* and the *base* model, even though they have access to a lot fewer features than other models. Thus, initially it looked like only a sub-set of features seemed to be important for the prediction task and some features appeared to even be detrimental for predicting future encounters. The network topology of the social encounter graph G appeared to be very discriminative on its own.

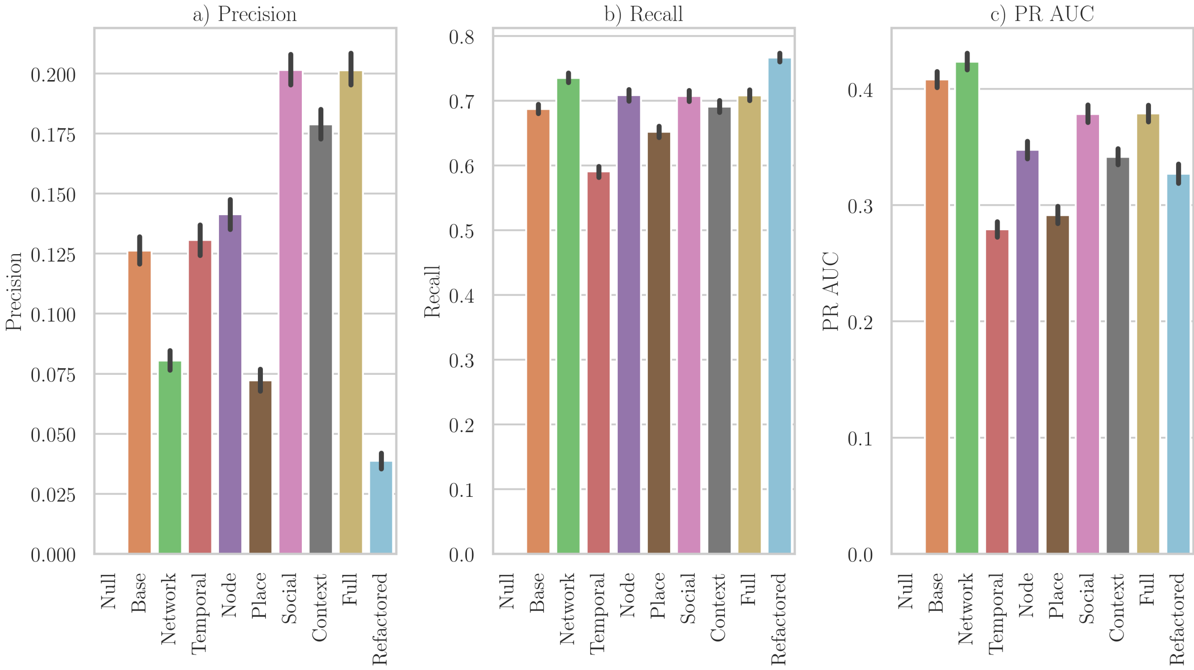

Figure 4.5: Model Scores

The figure depicts the precision, the recall, and the area-under-the-curve scores for the precision-recall curves of the different models. While the *base* and *network* model had the highest PR AUC score, both the *social* and *full* model had the highest precision. The recall scores were relatively high in comparison to the precision score for all models.

	$\overline{x}_{Precision}$	CI 95%	$\overline{x}_{Recall}$	CI 95%	$\overline{x}_{PR}$	CI 95%
Null	0.00	(0.00,0.00)	0.00	(0.00,0.00)	0.00	(0.00,0.00)
Base	0.13	(0.12,0.13)	0.69	(0.68,0.70)	0.41	(0.40,0.42)
Network	0.20	(0.20,0.21)	0.71	(0.70,0.72)	0.38	(0.37,0.39)
Time	0.08	(0.08,0.08)	0.74	(0.73,0.74)	0.42	(0.42,0.43)
Node	0.14	(0.14,0.15)	0.71	(0.70,0.72)	0.36	(0.34,0.36)
Place	0.07	(0.07,0.08)	0.65	(0.64,0.66)	0.30	(0.28,0.30)
Social	0.03	(0.04,0.04)	0.77	(0.76,0.77)	0.32	(0.32,0.34)
Context	0.20	(0.20,0.21)	0.71	(0.70,0.72)	0.38	(0.37,0.39)
Full	0.13	(0.12,0.14)	0.60	(0.58,0.60)	0.27	(0.27,0.29)
Refactored	0.13	(0.17,0.19)	0.69	(0.68,0.70)	0.34	(0.33,0.35)

Table 4.3: Model Scores

The tables lists the precision, the recall, and the area-under-the-curve scores for the precision-recall curves of the different models with the 95% confidence interval always in the column to the right of reported scores. While the *base* and *network* model had the highest PR AUC score, both the *social* and *full* model had the highest precision. The recall scores were relatively high in comparison to the precision score for all models.

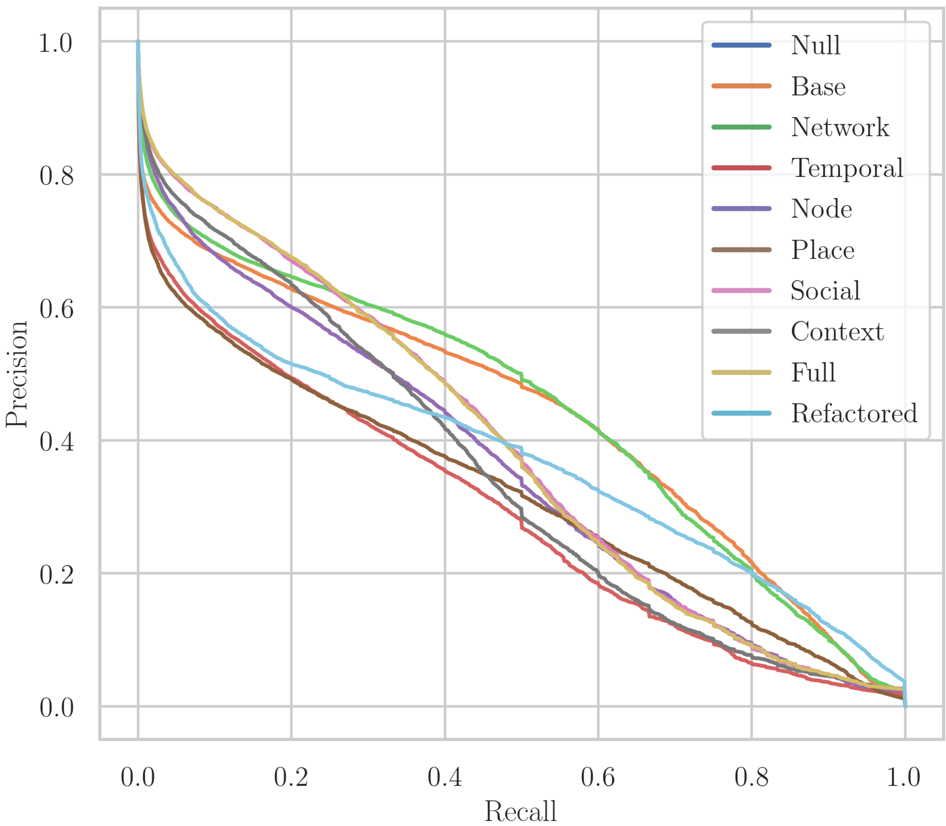

Figure 4.6: Precision Recall Curves

The figure depicts the precision-recall curves of the different models. As I fitted a separate tree R for each student, these curves were built by averaging the individual PR curves of each R. The *network* model performed best, while the *base* model was only slightly worse overall; in particular those two models managed to keep a relatively high precision score for higher recall values. The *social* and *full* model had relatively high precision scores as well.

Third, the *social* and *full* model had a significantly higher PR AUC score than all models except the *base* and *network* model. In particular with respect to precision both the *social* and the *full* model performed much better than any other model, which became apparent not only in the model scores (Figure 4.5) but also in the PRC curves (Figure 4.6). This indicated that while social features might not have been overall as important as the network topological features, they were still relatively important for correctly predicting whether an encounter occurred.

However, all models had a low precision score compared to the recall scores. This indicated that all models suffered from a relatively large amount of false positives. Given the relatively sparse nature of the social encounter graph G, this finding was not unexpected as there were many more opportunities for false positives than for false positives.

Feature importance

I also investigated the relative importance of the features for predicting future encounters for the *full* model (Figure 4.7). Interestingly the top five features—*average amount of people, weighted prop flow, triadic closure 0, triadic closure 3* and *max(relative importance)* accounted for roughly 50% of the expected contribution to the final prediction.

The relative importance of the features was consistent with the low scores for the models that did not include network topological features. Interestingly the social features *triadic closure 0* and *triadic closure 3* were also important highlighting the process of triadic closure in my dataset and partly explained the comparatively good performance of the *social* and *full* model. Triadic closure was consistently shown to be a driving feature of tie formation in networks (Bianconi et al., 2014). This makes sense as when triadic closure occurred, students were already spatially close to each other and thus more likely to encounter each other.

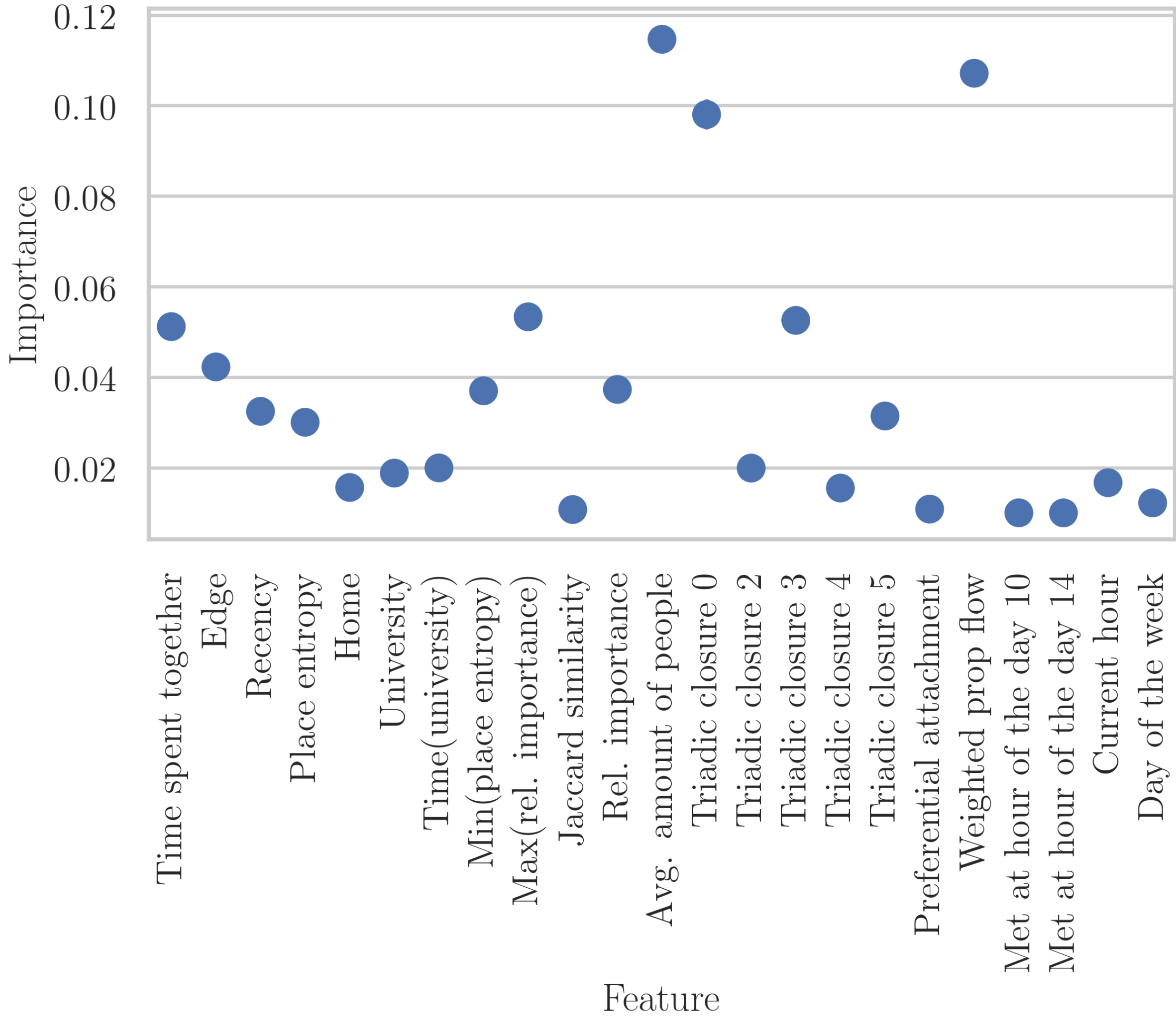

Figure 4.7: The Importance of the Different Features for the *Full* Model

The plot shows how important each feature of the *full* model was for predicting e at time $t + n$. It only depicts features whose importance was bigger than 0.01. Both *triadic closure 0* and *number of people* were among the most important features indicating the importance of knowing the social context of where encounters took place. Furthermore, *weighted prop flow* was important as well, highlighting the role the wider social encounter graph played for predicting encounters. In total the top five features accounted for about 50% of the expected contribution to the final prediction.

	Mean	CI 95%
G^{all}	0.38	(0.37,0.39)
G^{social}	0.34	(0.33,0.35)
G^{uni}	0.49	(0.48,0.49)

Table 4.4: Precision-Recall AUC Scores for Different Link Types

The tables lists the performance difference for the various definitions of G, where the 95% confidence intervals is reported in the column to the right of the scores.

4.4.2 Predicting Different Types of Links

I was also interested in whether the type of relationship (i.e. whether the students were just colleagues, or also socialised outside of university) between nodes affected the predictability of encounters. In order to explore this question, I constructed two new encounter graphs. Recall that G_t was based on all spatial encounters between students regardless of *where* and *when* these encounters took place (hereafter G_t^{all}). I constructed G_t^{social} based on all the encounters that took place between nodes $u, v \in G_t^{social}$ before 9AM or after 6PM local time on weekdays, on the weekend, or in a spatial context other than university. In other words, I was trying to capture the non-university/work related encounters only that happened either after the normal "working" hours, or in a different place than the university. I, furthermore, constructed G_t^{uni} that was derived only from encounters between nodes $u, v \in G_t^{uni}$ that happened between 9AM and 6PM on weekdays and whose spatial context was university.

As I can see in Figure 4.8 and Table 4.4, the performance for the models based on G_t^{social} was worse than for the models based on G_t^{all}. An explanation could be that "social" encounter are less regular than other encounters; meetings between friends are usually varied in time and place.

Unsurprisingly, the performance for the models based on G_t^{uni} was significantly better than for those based on G_t^{all}. Unsurprisingly students were interacting and meeting regularly; quite likely at the university itself as students from the same year had a similar

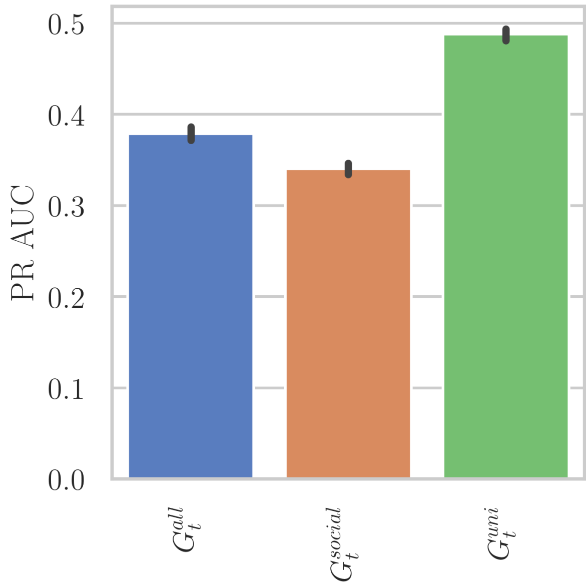

Figure 4.8: Model Scores

The plot visualises the performance difference for the various definitions of G, where the error bars represent the 95% confidence intervals. Unsurprisingly the prediction for G_{uni} had a higher PC AUC score reflecting the potential higher regularity of meeting one's "colleagues" at university.

schedule for lectures.

4.5 Discussion

The main finding of my research was that features about *whom* one meets and the wider network topology of the social encounters significantly improved my predictions, while information about *when* and *where* one meets did not seem to play an as important role for my prediction task. Furthermore, and in contrast to previous research that information about *where* individuals meet did not seem to play a pronounced role for predicting future encounters between individuals in my dataset of students (Scellato, Noulas, & Mascolo, 2011; Yang et al., 2013). It appeared that almost all information was already contained in the network topology of G and the social context rather than in the spatial and temporal setting.

One possible explanation for the relatively low importance of spatial and temporal features could be that as people move through their daily lives, the information of where they are is already embedded in who else is physically close. For example, one is with their partner there is a high chance that one is either at home or at the partner's home; if one is with their friends from university then there is a high chance that one is meeting them at university. In a sense the social contexts individuals (Sekara et al., 2016) inhabit might intrinsically be linked to spatial places.

Thus, and while out of scope for this work, one interesting route to explore would be to not only map but also conceptualise human behaviour not in the traditional dimensions of time and space as in time-geography (Hägerstrand 1970) but in a reference frame of time, social, and spatial dimensions.

Last, the performance of my link-prediction algorithm was significantly better when considering all ties rather than just social ties but worse than when considering just university ties. I believe that a better understanding the role different types of relationships play for encounters could be a fruitful avenue for future research and whether potential other factors can help improve the prediction of social ties.

However, one has to be careful when generalising from my sample of students to the whole population. While I was not aware of any reason my findings should not also hold for a wider population, the dataset in my study represented after all just one sample of a network. Furthermore, my classification of geographic places was rather broad and did not allow for a detailed analysis of those factors. I believe that a more fine-grained analysis of the role of geographic place is an interesting prospect for future research, especially in conjunction with an expanded analysis of the predictability of different types of ties.

I couldn't tell you why, because I never had any why.

Marcel Duchamp

5 The Interplay of Long-Term Social & Mobility Behaviour

5.1 Introduction

As discussed in Section 2.4, travel behaviour is intrinsically linked to social networks. People regularly visit their friends or undertake joint activities at various places. Human mobility behaviour and social networks were so intertwined that Wang et al. (2011) were able to use mobility networks as proxy for social networks and vice versa. Urry (2002) reminded us that co-presence is obligatory for many forms of social life. Corporeal travel constitutes one of the cornerstones of modern societies. In our modern life "[m]obility has become a central aspect of social integration" (Viry et al., 2009). However, maintaining social ties to far flung peers is not without its costs. One not only has to have the resources to travel but one also needs to be able to spare the time.

As social networks and mobility behaviour are so co-dependent, a lack of access to either has clear implications for the other domain. If one cannot visit their friends, it is hard for them to sustain their social ties. Conversely, if one does not have a lot of social ties, there are fewer reasons to travel. Several empirical studies looked at the relationship between social relationships and travel behaviour and what role inequality in one domain played for the other domain. However, most modern quantitative studies did not used longer time frames as units of analysis. Previous studies either used a static snapshot (see for example Carrasco and Cid-Aguayo 2012; Brown et al. 2013; Viry et al. 2009 and Frei and Axhausen 2007) or focused on shorter time frames for analysis (see for example Cho et al. 2011; De Domenico et al. 2013 and Nguyen and Szymanski 2012), although some few studies (Alessandretti, 2018; Scellato, Noulas, & Mascolo, 2011) accounted for longer term dynamics. This is noteworthy as long-term dynamics might be considerably different than short term interactions between social ties and mobility behaviour as social networks and mobility behaviour might evolve at different timescales (Chapter 2).

And while previous researchers found ample evidence for potential relationships between social factors and mobility, previous results did not necessarily paint a coherent picture. For example, an individual's high mobility might be at the same time beneficial and detrimental for maintaining their social ties. It is thus unclear from theory alone which potential relationships between social variables and mobility behaviour to study. I dealt with this problem by using data driven models for uncovering potential relationships in my data. In particular, I adopted the concept of Granger-causality (Granger, 1969) to test for relationships between various social network and mobility variables with a vector autoregressive (VAR) model. While Granger-causality was originally developed in the context of economic time series data, several recent studies dealing with transportation issues also successfully adopted the concept of Granger-causality. They studied a variety of transport related topics ranging from the substitution effect of transport modes (Castillo-Manzano, Pozo-Barajas, & Trapero, 2015) over the relationship between air transport and economic growth (Hakim & Merkert, 2016) to trade and air travel (Hakim & Merkert, 2016).

Furthermore, previous studies often did not explicitly try to simultaneously model relationships between travel behaviour and social variables. Last, I did not simply presuppose a best estimator for my system of equations but empirically tested several possible ways to infer causal pathways for my specific problem. My main contribution, thus, consisted of studying at the same time possible the causal relationship between social ties and the propensity to travel in a data driven way.

5.2 Background

While there is evidence that mobility as well as social ties co-evolve over longer time periods (Alessandretti, 2018), most studies did not try to account for longer-term interactions between the two spheres. On the one hand, studies that were purely interested in predicting future behaviour usually did not account for changing temporal patterns over a longer timescale (Cho et al., 2011; Wang et al., 2011; De Domenico et al., 2013). On the other hand, studies that linked mobility to social networks usually did not study the co-evolution of the two spheres as they did not deploy a longitudinal design (Berg et al., 2012; Grabowicz et al., 2014; Shi et al., 2016).

What is more, the relationship between social ties and mobility was not clear cut.

In certain circumstances mobility could have a positive effect on a person's social network, while in other the need to travel a lot might be detrimental for maintaining social ties. While better access to mobility allowed people to maintain a stronger social network, a long and solitary commutes or geographically spread out social networks reduced opportunities for socialising.

At the same time social ties allowed access to more "social" resources and life opportunities and could allow people to travel more. Being able to gain more resources from one's social network should have facilitated access to better means of transportation, especially ridesharing, and freed up resources to travel in the first place.

Schwanen et al. (2015) posited that mobility and social ties were entwined processes. "[I]n any given locality, for any person, community and/or social group, there likely exists a range of overlaps and pathways through which these processes affect, and are affected by, each other" (Schwanen et al., 2015). There are several possible causal pathways for how mobility could affect social networks and for how social networks might affect mobility behaviour. While Schwanen et al. (2015) focused on social exclusion and transportation disadvantage, they nevertheless provided an exhaustive overview of how social ties might influence mobility and how mobility might shape an individual's social network. I thus followed their overall classification of causal pathways from social networks to mobility and vice versa but abridged and adapted them for my particular research questions.

5.2.1 Causal Pathways From Mobility to Social Networks

As face-to-face interactions were still an essential part of building and maintaining social ties, it was necessary for individuals to travel to maintain old social ties or form new ties. Even today in our internet permeated lives, one could observe that the likelihood of a friendship between two people decreased with distance (Preciado et al., 2012). Urry (2002) even considered corporeal travel to be "necessary and appropriate for a rich and densely networked social life for different social groups". A higher capacity for mobility might be beneficial for social ties in various ways and there were several ways how an individual's capacity for mobility might help form new and maintain old ties.

First, mobility might directly affect the amount of practical, material, informational, and emotional support a person might receive from their social network (Schwanen et al., 2015). For example, Carrasco and Cid-Aguayo (2012) studied two neighbourhoods with different income levels in a Chilean city and showed that individuals with a car receive more emotional, monetary and informational resources from their social networks. Although they cautioned that car ownership and income levels might be confounded and overall contact frequency between individuals was not affected by car ownership.

Second, one's capacity for mobility might affect a persons level of travel, activities and interaction with others, and through that their social networks (Schwanen et al., 2015). The number of trips a person could undertake was found to significantly, negatively correlate with the risk of being socially excluded (Stanley et al., 2011). Being able to move from A to B generally allowed individuals to maintain more spread out and larger social networks. Viry et al. (2009) used data about the social networks of commuters in the largest Swiss agglomeration—Zurich, Geneva, and Basel—to study the relationship between social ties and the potential for mobility. They found that a high mobility allowed individuals to maintain or even widen their social network, as commuters could benefit from being at the intersection of spatially separated social circles. In another study, Frei and Axhausen (2007) noted that car ownership and with it the increased potential for automotive travel had a positive effect on the geography of a person's social network. Being less anchored to a particular locality was positively associated with an increased geographic size of a person's social network. Moreover, the ability to use public transportation seemed to be have wider positive social impacts. Green, Jones, and Roberts (2014) reported that a free bus pass for elderly citizens in London not only increased their physical well-being but also improved their social life and opened up a new sphere for socialization, the shared space of the bus. Utsunomiya (2016) showed that local bus services were positively correlated with various positive outcomes in an neighbourhood such as trust, amount of social ties, and participation in local activities.

Nonetheless, a more geographically spread out social network could increase the cost for maintaining such a network and hence the need for more mobility might lead to a sparser social network overall. Putnam (2000) suggested that travelling alone by car had negative consequences for the formation of social relationships as this time was essentially lost to form social ties via corporeal interactions. And indeed Mattisson, Håkansson, and Jakobsson (2014) found that the longer people were commuting by car in the south of Sweden the lower their participation in social activities. Furthermore, for already disadvantaged groups of people such as single women with children, migrants, less educated people, and people with special needs an increased expected mobility made it more difficult for them to maintain significant social ties. Having no access to a car, living far away from accessible transport or social meeting places such as shops and bars, and lacking the organizational and temporal resources to maintain social ties over a distance all contributed to this (Jiron, 2007; Viry et al., 2009).

Third, the knowledge and skills a person had regarding transport options and opportunities had implications for the total knowledge regarding transportation that was available within social networks (Schwanen et al., 2015). While the empirical evidence for this particular pathway between mobility and social networks was rather limited (Schwanen et al., 2015), various other empirical studies existed that highlighted the importance of social networks for knowledge sharing (for an overview see Nieves and Osorio, 2013).

5.2.2 Causal Pathways From Social Networks to Mobility

Social ties were an important factor for determining access to mobility. Access to and opportunities for travel might all be influenced by the amount of resources and support one is able to get from one's social network.

First, social networks could have an effect on the level of practical, material, informational, and emotional support an individual receives (Schwanen et al., 2015). This in turn might influence their access to and their knowledge about transportation resources and opportunities. Larger and wider personal networks had clear benefits in how many social connections an individual could access for support. For example, Lovejoy and Handy (2011) described how Mexican immigrants relied on their social networks for enabling access to automotive transportation. Individuals in the study frequently got a lift or borrowed the car of someone else. Consequently, having larger and wider networks made it easier for individuals to access cars. However, there is evidence that not just the size of an individuals social network but also the diversity of (Agneessens, Waege, & Lievens, 2006) and activity within (Silvis & Niemeier, 2009) a person's personal network determined access to inter-personal resources. Silvis and Niemeier (2009) found that they more active seniors are in their social network the more likely they were to ride share regularly, while the size of an individual's personal network did not play an important role. Moreover, social ties did not just allow better access to automotive transport. A larger social network allowed individuals to determine which destinations are worthwhile to visit in the first place (Agosto & Hughes-Hassell, 2005) and to gain better access to formal and informal markets relevant for transportation (Oviedo Hernandez & Titheridge, 2015).

Second, social networks might impose various space-time constraints on an individual's activity and travel pattern (Schwanen et al., 2015). Di Ciommo, Comendador, López-Lambas, Cherchi, and Ortúzar (2014) studied travel behaviour among residents in a suburb of Madrid for one working week. Besides traditional variables such as age, gender, income, and car ownership, the authors also included receiving help and participation in voluntary activities as explanatory variables for modelling travel behaviour. They reported that individuals that received help with various domestic tasks such as childcare or housekeeping were more likely to walk or use public transport. Conversely, individuals that participated in voluntary activities were more likely to use automotive transport. In the authors' view people participating in voluntary commitments had less uncommitted time available to them than people receiving help with various tasks. Thus, people who participated in voluntary activities tried to save time by using individual modes of transport. Schwanen et al. (2015) pointed out that these findings concured with the idea that resources gained from social networks are conducive to more *social* modes of transport.

5.2.3 Research Questions

The above mentioned studies all highlighted potential causal pathways between an individual's social network and an individual's mobility; nonetheless, it was far from clear which causal pathways were not only generalizable but applicable to my specific dataset. For example, high mobility might enable individuals to foster more diverse social ties. At the same time, a high mobility might lead to a lack of time to maintain those ties (Section 5.2.1). While a wider social network was generally associated with a higher potential for mobility (Section 5.2.2), it was unclear whether a higher potential for mobility not only led to more corporeal travel but also enabled individuals to explore more varied places.

Moreover, the benefits individuals could accrue from their networks or through their mobility were not just dependent on the total amount of social ties nor the total amount of an individual's mobility. The variety of social ties or the variety of places one visits constituted an important role. Granovetter (1973) highlighted the importance of having a varied social network and the role weaker ties can play for information transfer. Similarly

Seibert, Kraimer, and Liden (2001) showed that network structure and not just the total number of ties was directly related to social resources. Furthermore, Kaufmann, Bergman, and Joye (2004) stressed the importance of how options and conditions regulated the access to means of mobility, thus highlighting the importance of variety for mobility. Consequently, I was also interested in the variety of a person's social network.

Last, most empirical studies cited above in Section 5.2.1 and 5.2.2 did not explore the drivers and effects of both social networks and mobility at the same time. They tended to focus on either social networks or mobility in their analysis. However, given the ample empirical evidence for different causal pathways between the two spheres, there was the possibility that the two are intertwined.

Given access to temporal data of the movement patterns as well as the social encounters and virtual interactions of 847 students in Copenhagen (Chapter 3), I attempted to untangle the entwined processes of social networks and mobility. I tried to answer the questions of to what extent social networks "caused" mobility and to what extent mobility "caused" social networks with a particular focus on longer-term dynamics. While, there is ample empirical evidence for causal pathways between both social networks towards mobility and vice versa, it was unclear how previous findings exactly related to mobility and social networks.

Thus, I opted for the following overarching approach: First, instead of confirming hypotheses I explored possible causal pathways in my data. I therefore used a data-driven approaches for uncovering "causal" relationships instead of confirming theoretically derived relationships with my data. Second, I focused on long-term dynamics not only because there is relatively little research dealing with long-term interactions between social networks and mobility. Third, I attempted not only to capture the total amount of social ties and the total amount of the distance travelled by any individual, but I also tried to account for the variety in social ties as well as for the variety in places individuals visit in my analysis. Fourth, I opted to study the potential drivers and effects of both social networks on mobility patterns, and of mobility on social ties contemporaneously. Therefore, I treated both the social network as well as mobility as both independent and dependent variables in my analysis.

5.3 Methodology

5.3.1 Problem Definition

I argued in Section 5.2.3 that both social networks and mobility were more than just the sum of a person's social ties or respectively the total distance they travelled. Hence, to assess the social network of an individual, I measured not only the total amount of *physical encounter* between peers but also their overall *virtual interactions* and the entropy of the set of peers they had interacted with (hereafter *peer entropy*). Similarly, I captured different aspects of an individual's mobility by not only measuring an individual's *radius of gyration* but also the entropy the places they had visited (hereafter *location entropy*).

As I had time ordered data I also wanted to preserve the information about the ordering of events in time as temporal information can hold important clues for understanding complex systems. Consequently, I was interested in studying the dynamics of a multivariate time series composed of my variables *physical encounter*, *virtual interactions*, *peer entropy*, *radius of gyration* and *location entropy*. As I excluded contemporaneous dependencies between the variables, Vector Autoregressive Models (VAR) models were a natural fit for my multi-variate time series data (Luetkepohl, 2005). Let $X(t) \in \mathbb{R}^d \times 1$ for $t = 1, ..., T$ be a d-dimensional multivariate time series and L the maximum included lag in the model. I could then fit the following VAR model with L lags:

$$\mathbf{X}(t) = \sum_{\tau=1}^{L} \mathbf{A}_\tau \mathbf{X}(t-\tau) + \epsilon(t) \tag{5.1}$$

where A_τ is a matrix of coefficients for every τ and $\epsilon(t)$ are a white Gaussian random vector (i.e. the residuals). Given two time series X_i and X_j the coefficient $A(j,i)_\tau$ can be seen as the lagged effect of X_i on X_j at τ.

Formally, I wanted to minimise my prediction error as well as to maximise the *sparseness* of the matrix $\mathbf{A}$, representing all coefficients for all time points τ. In words, I attempted to find the coefficients of $\mathbf{A}$ that are salient for estimating future states of the system. One straightforward way to define this notion of predictive causality is Granger causality.

5.3.2 Granger Causality

Granger causality, named after Granger (1969), as a concept of causality is based on prediction and conditional dependence. The idea is that if a signal X_i "granger-causes" a signal X_j, then knowledge of the past values of X_i should help to improve the prediction of X_j (Seth, 2007). Underlying this definition is the notion that the cause should precede the effect as well as that the knowledge of the cause should measurable improve one's predictions.

More formally, let $\mathcal{I}(t)$ be the information set that includes all information at point t and $\mathcal{I}_{-X_i}(t)$ the information set that includes all information except information about the signal X_i. Then X_i does not granger-cause X_j if

$$(X_j(t+1) \perp\!\!\!\perp \mathcal{I}(t))|\mathcal{I}_{-X_i}(t) \tag{5.2}$$

which is equivalent to

$$P(X_j(t+1)|\mathcal{I}(t) \cap \mathcal{I}_{-X_i}(t) = P(X_j(t+1)|\mathcal{I}_{-X_i}(t)) \tag{5.3}$$

In words, if $X_j(t+1)$ is conditionally independent of the information about signal X_i, then X_i does not granger-cause $X_j(t+1)$ (Eichler, 2012). Conversely, if I am estimating $X_j(t+1)$ and if one of the coefficients $A(j,i)$ for $\tau = 1, ..., L$ is significantly different from zero, then X_i is said to granger-causes X_j (Luetkepohl, 2005).

For the rest of the chapter I followed Granger's definition of causality. I would like to highlight here that Granger-causality is not irrefutable proof that a causal relationship exists between two variables as it only uses observational data. Furthermore, I want to point out that the set of possible confounding variables is for all practical purposes always bigger than what I could physically observe if there is not a randomised control trial. Last, Stokes and Purdon (2017) showed that using finite order VAR, noting that infinite VAR processes are not practical to compute nor to obtain sufficient data for—to estimate "causal" relationships could introduce significant bias in the "causal" estimates.

5.3.3 Estimating the VAR process

While empirical evidence suggested a causal relationship between social networks and mobility, it was not clear how well previous findings generalised (especially since some previous findings were contradicting each other).

Traditionally VAR models to test Granger-causality were estimated using the ordinary least squares estimator (OLS) (for empirical examples see among others Narayan and Smyth, 2005, 2009 and Tselios, 2014). While the Gauss-Markov theorem states that in linear regression models in which the errors have expectation zero, are uncorrelated and have equal variance, the OLS estimator is the best linear unbiased estimator (BLUE), other estimators such as the James-Stein estimator, LASSO, or ridge regression with lower variance exist (Chipman, 2014). All the alternative estimators rely on some form of finding a parsimonious set of coefficients of **A** that are not zero, while at the same time minimising the MSE on the training set. Prominent choices are regularization of the OLS cost function or shrinkage to learn significant causal relationships (see among others Burge, Lane, Link, Qiu, and Clark 2009; Chen, Resnick, Davatzikos, and Herskovits 2012; Prinzie and Van Den Poel 2011).

It thus not a priori clear which estimator for my VAR system of equation was best suited for my problem at hand. Consequently, I split my data into a training and a test set and evaluated the performance of the various models on the test set. In particular, I used a rolling-window forecast on the last three months of my data. The idea was that the true causal structure should be better for predicting future values than models that included more false positive or false negative "causal" relationships. I furthermore restricted my model space to linear models as I could then use the within transformation to account for individual time invariant covariates in my panel data without introducing additional covariates (Bruederl & Ludwig, 2015).

In the end I decided to compare the following five estimators for my VAR process:

1. The standard estimator for Granger-causality—OLS with Wald test (Luetkepohl, 2005),
2. S3L (Rahmadi et al., 2018; Rahmadi, Groot, Heins, Knoop, & Heskes, 2017),
3. LASSO regression with cross-validation (Nagarajan, Scutari, & Lebre, 2013),
4. LASSO regression with stable specification search (Meinshausen & Bühlmann, 2010),
5. James-Stein shrinkage estimator (Opgen-Rhein & Strimmer, 2007).

While this list did not represent all possible estimators, I believed that it stroke a reasonable balance between well established estimator such as OLS as well as LASSO regression and newer estimators such as S3L and the James-Stein shrinkage estimator.[1] All of those approaches are able to infer "causal" relationships from the data alone. Furthermore, all algorithms try to find the most parsimonious set of variables that explains future observations.

5.3.4 Operationalisation

Defining the Temporal Aggregation Interval

Recall that I was particularly interested in longer-term dynamics as they had been studied relatively sparingly before. Using a longer temporal window to aggregate my data furthermore avoided fluctuations due to weekly periodicity of human behaviour (Williams, 2013) as well as smoothened breaks in the academic year such as the Christmas holidays or the term break.

To find an appropriate temporal level of aggregation that captured the stability that studying more aggregate, longer-term dynamics implied, I calculated the aggregate rate of change R for each variable over all users for time windows ranging from 1 to 90 days (Figure 5.1). However, as the plot was relatively hard to inspect visually I also calculated the mean of the five different R as well as the standard deviation. Figure 5.2 showed that while other window lengths resulted in slightly lower average rate of change, using one month (i.e. 30 days) as window length not only resulted in a very low average rate of change but also a low standard deviation indicating that all variables were relatively stable at this level of aggregation.

While other intervals such as 43 days have an overall lower mean and standard devi-

[1]For a more detailed description of the various estimators see Appendix A.1.1.

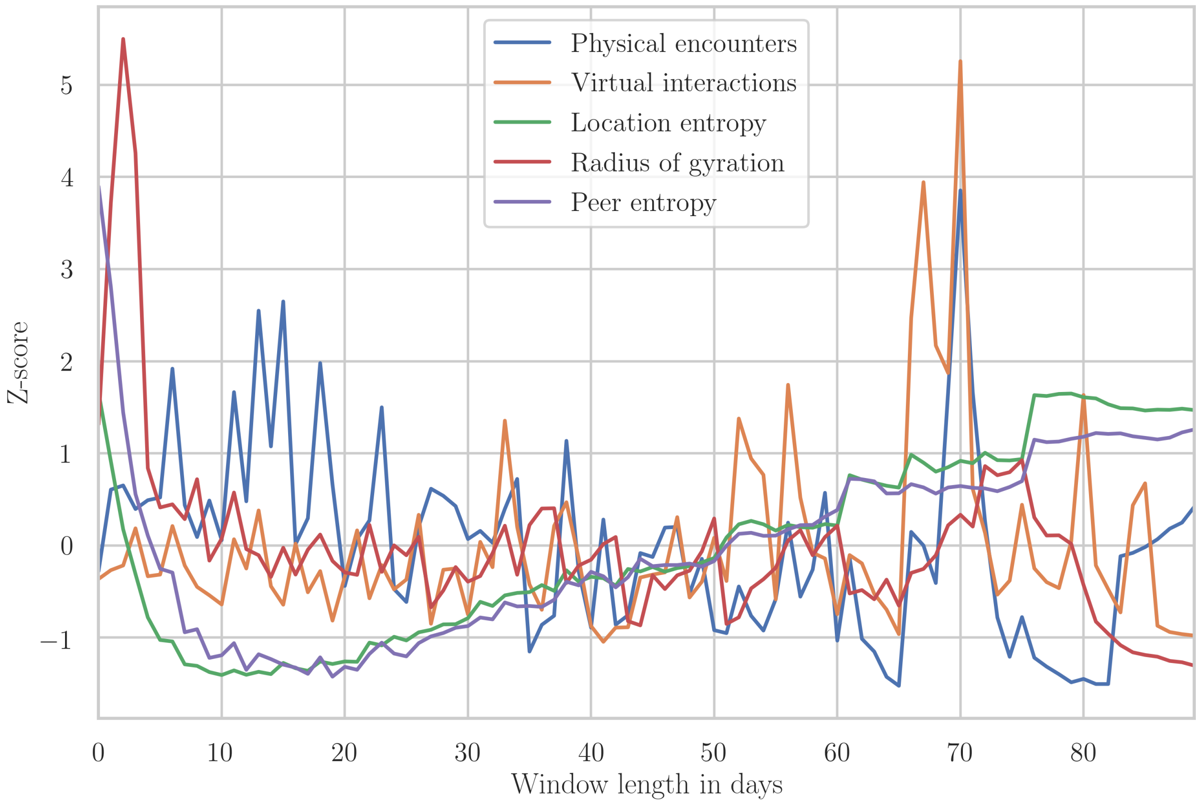

Figure 5.1: Standardised, Aggregate Rate of Change per Window Length

The figure shows the z-standardised aggregated rate of change R for each of the variables for different time windows. In other words, it shows how much the change depends on the length of the window of observations. However, the figure was relatively hard to interpret visually as five different variables are overlapping, thus I also calculated the mean and standard deviation of R in Figure 5.2.

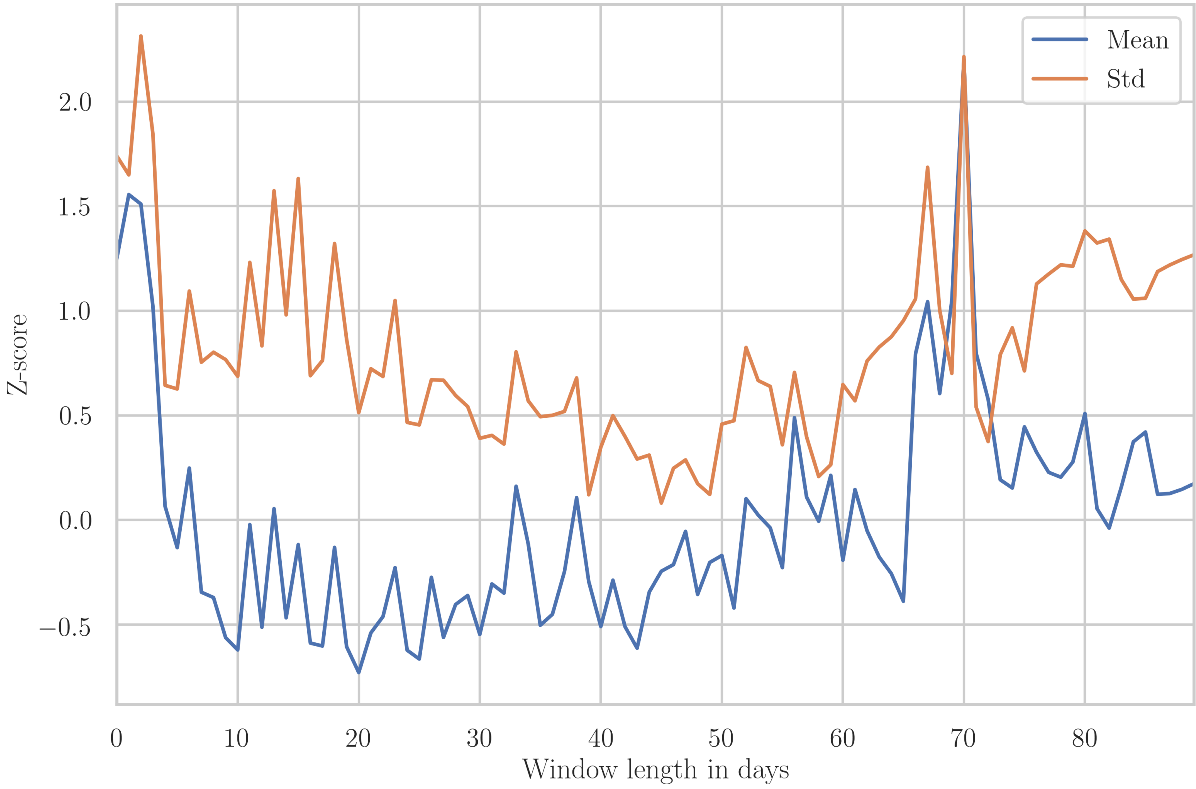

Figure 5.2: Mean and Standard Deviation of R per Window Length

The figure depicts the average as well as the mean of z-standardised rate of change R for different window lengths. One can see that at 30 days the average score of R for all variables was relatively low as indicated by a low average and standard deviation. The spike for R at around 70 days for both *physical encounters* and *virtual interactions* can be explained by the fact that at this aggregation level several users had at least one period with a very low count of observations followed by a period with thousands of observations thus increasing the average overall.

ation for R, I nevertheless decided to use 30 days for two reasons: First, the differences between the aggregation intervals were rather small and, second, an aggregation interval of 30 days was not only more easily interpretable as monthly behaviour but aligned with previous research into the stability of both social networks and mobility patterns (Alessandretti, 2018; Krings et al., 2012). Krings et al. (2012) showed that social network are most similar to themselves at the 30 day interval and Alessandretti (2018) found that the cardinality of the set of visited locations stays relatively constant at a monthly level.

Measuring the Social Network

In order to assess the social network of a student, I measured their overall physical encounters with other students, their virtual interactions with all their peers, and also the entropy of the set of peers they were meeting. I utilised the Bluetooth scans of the phones to derive the total amount *physical encounters* between students. Sekara et al. (2016) showed this to be a reliable way to estimate physical co-location of students.

I used the amount of phone calls as well as the texts as indicators for *virtual interactions* between peers. To capture the variety of social behaviour of students I calculated the *peer entropy* for each student. *Peer entropy* is defined as the Shannon entropy (Equation 3.1) for each time-period for the set of all social encounters as well as interactions of a student.

Measuring Mobility Behaviour

In order to measure the mobility of a student, I computed the *radius of gyration* for each student for each time period. I also included the *location entropy* as a measure for the variety of places a student visits. *Location entropy* is defined as in Equation 3.1, where p_i in the case of *location entropy* represents the relative probability of observing a student at location i. The *location entropy* should tell us whether a person is just going to the same places or is actually also exploring a variety of different places.

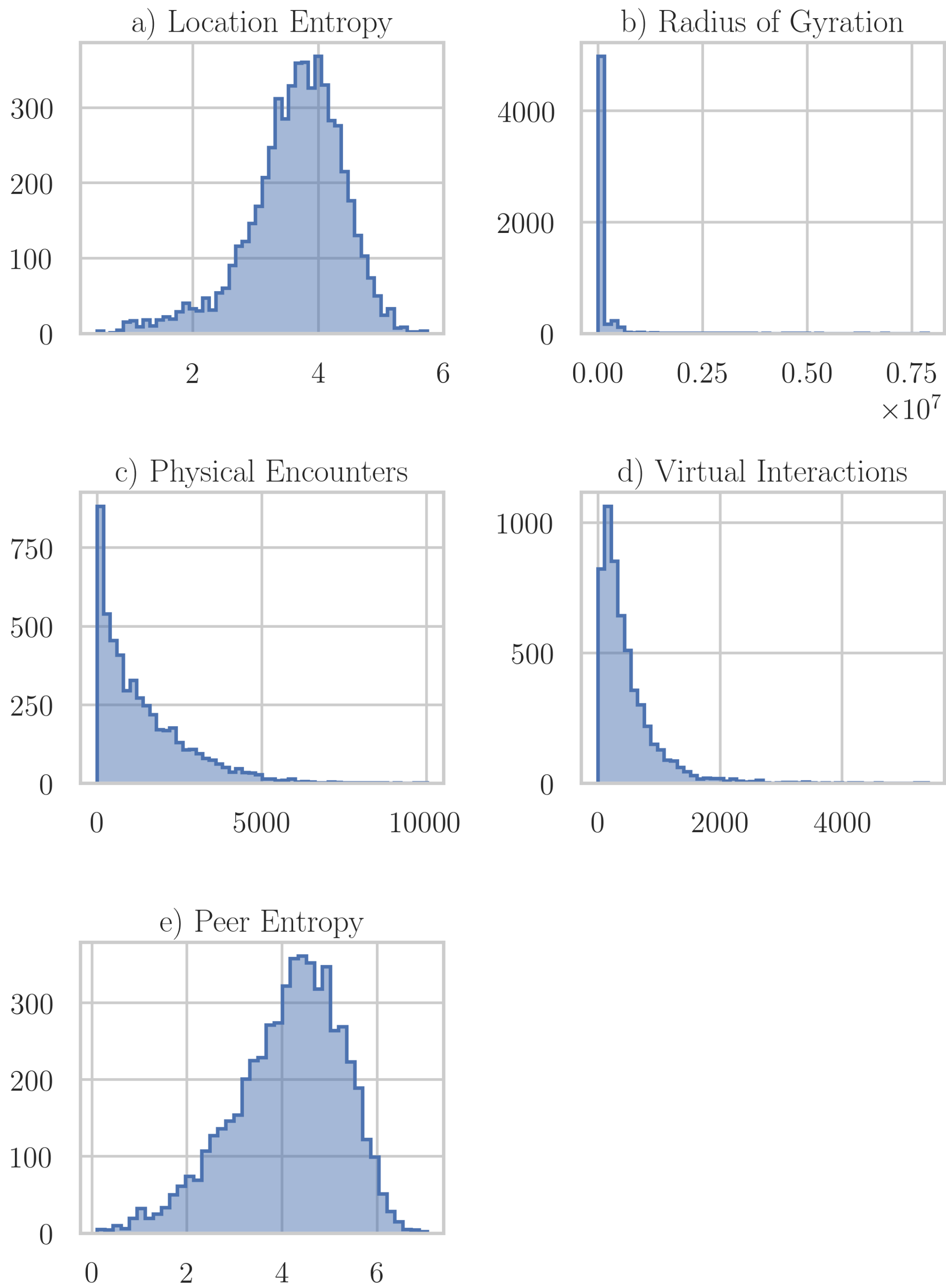

Figure 5.3: Distribution of Variables

The figure depicts the distribution of all the variables for my VAR system of equations. *Radius of gyration, physical encounters* and *virtual interactions* were all heavily skewed with long right tail. This means that for these three variables most students had relatively low values, whereas for some students I could observe values significantly larger. A heavily skewed distribution was common for both social networks (Section 2.2.1) and observed patterns of mobility (Section 2.3.1).

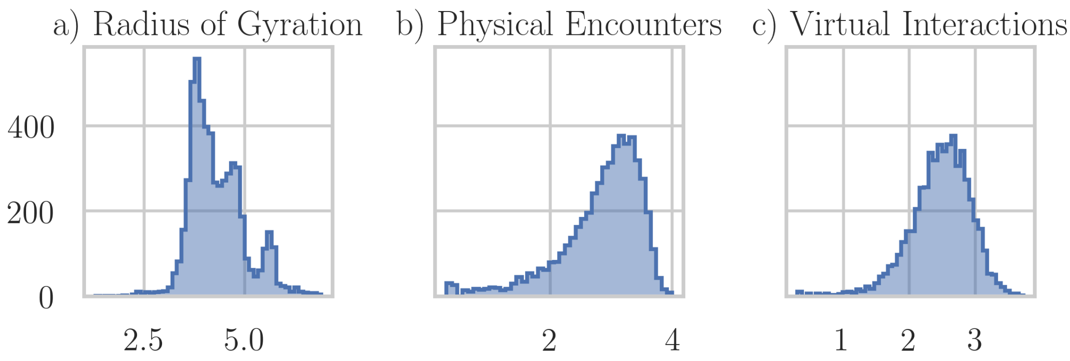

Figure 5.4: Log Transformed Variables

After log transforming the variables *radius of gyration, physical encounters* and *virtual interactions*, there distributions were much less skewed and were much better suited for an analysis with linear models.

I transformed my discrete variables *physical encounters*, *virtual interactions*, and *radius of gyration* as they are heavily skewed with $log(x+1)$ (Figure 5.3 and Figure 5.4). Furthermore, I also dropped data points for individuals, where I observed no movement at all. As those observations were most likely missing observations rather than actually measuring that a student has not moved for a month.

Confounding Variables

However, several other unobserved confounding factors might influence either my variables related to the social network, mobility behaviour, or both. For example, a more extroverted student might meet more people, while at the same time being more likely to travel in the first place. Or an overall economically less well off student might not be able to afford to travel as much and thus meet fewer people.

While the absence of latent variables is essential for learning the true causal structure (Nagarajan et al., 2013), this assumption is hard to verify in a real-world non-experimental setting as the set of all possible confounding variables is usually not known. While I knew several confounding variables for each student such as gender, age, and personality, I did not know various others such as their socio-economic background. I thus decided to demean my data for each student by subtracting the mean for each student from their respective observations (also referred to as within transformation in the literature) and thereby to account for all unobserved time-invariant covariates (Bruederl & Ludwig, 2015). I would like to caution the reader however that as in a Fixed Effects model this did not account for time varying *latent variables*.

VAR Pre-Requisites

Before I estimated my VAR process, I had to make sure my VAR process is stationary and that I picked a sufficient number of lags. First, I used the Maddala and Wu (1999)[1] test of stationarity for panel data to make sure that my multi-modal time series was stationary. The Maddala and Wu test for my time series significantly rejected the hypothesis that my time series was not stationary. Second, I used the OLS estimator and the commonly used Bayesian Information Criterion (BIC), which penalised overly complex models to determine the "optimal" amount of lags that I should include in my model (Seth, 2007). I compared BIC scores for the *full* VAR models containing up to to five lags. I found that the optimal order of the VAR process is one.

5.4 Findings

5.4.1 Evaluating the Different VAR Estimators

To evaluate the relative performance of the different ways to estimate my VAR system of equation, I combined a rolling window forecast with bootstrap sampling.

First, I created three tuples of temporally ordered training and test data

[1] I used the implementation of Croissant and Millo (2008).

Model	Loc. ent.	R. of gyr.	Phys. enc.	Virt. int.	Peer ent.
Full	0.21	0.31	0.99*	0.25	0.99*
OLS	0.22	0.28	0.99*	0.25	0.99*
LASSO CV	0.42	0.13	0.40	0.08	0.31
LASSO Stable	0.21	0.31	0.99*	0.15	0.99*
Shrinkage	0.21	0.31	0.99*	0.25	0.99*
S3L	0.40	0.20	0.67	0.00	0.99*

Table 5.1: ζ Model Scores

The table shows how often a model performs strictly better than the *null* model for predicting each dependent variable. The statistic ζ based on 10,000 bootstrapped samples of the test set. Interestingly the only two variables where the models performed better than the *null* model wehre *physical encounters* and *peer entropy.* *: $p < 0.05$

$(train_{T-1-i}, test_{T-i}), i \in \{0, 1, 2\}$, where the first $T - 1 - i$ observations were the training data and the corresponding time slice at $T - i$ formed the respective test set.

Second, for each tuple of training and test data I compared the performance of each model to a *null* model that for each variable just used the mean of that variable as its prediction. In other words, the *null* model had a bias of one and zero variance. In OLS parlance, I only fitted the intercept for each variable for the *null* model. I also defined a *full* model that includes all possible variables as independent variables.

I defined my test statistic ζ as the difference of the MSE of the prediction of the *null* model and the prediction of the model $m \in \{OLS, S3L, LASSO - CV, LASSO - Stable, Shrink\}$ for each variable of my VAR process. Next I bootstrapped ζ from 10,000 samples drawn from the current test set at $T - i, i \in \{0, 1, 2\}$.

Third, I summed ζ over all tuples of training and test data. I say that a model m performed better than the *null* for variable y if it performed so in at least 95% of all bootstrapped samples.

See Table 5.1, for the relative performance of each model in comparison to the null model and Table 5.2 for how parsimonious each model was by averaging the amount of coefficient each type of models uses for predicting the different dependent variables;

Model	Loc. ent.	R. of gyr.	Phys. enc.	Virt. int.	Peer ent.
Full	5	5	5	5	5
OLS	1.33	1	1	0	3
LASSO CV	3.33	3.66	4.33	4	5
LASSO Stable	4	4.66	4.33	5	4
Shrinkage	1.33	0.33	2.33	0.33	2.33
S3L	1.33	0.66	0.66	0	1

Table 5.2: Average Amount of Coefficients

The table shows the average amount of coefficients for each model over the three tuples of training and test data (excluding the coefficient for the intercept). Generally speaking the fewer coefficients a model used on average the more parsimonious it was.

ceteris paribus if a model used fewer coefficients on average it was more parsimonious. Interestingly, the *full* model performed at a similar level to the *OLS, LASSO-Stable* as well as the *Shrinkage* model. For predicting both variables *physical encounters* as well as *peer entropy*, the *full, OLS, LASSO-Stable* and *Shrinkage* model performed significantly better than the *null* model.

As the total ζ scores were relatively similarly for the *full*, *OLS*, *LASSO-Stable* as well as *Shrinkage* model, I wanted to ensure that the other models actually performed better than the *OLS* model. In other words I wanted to only include additional "causal" edges if those helped the models to predict the dependent variables better. I thus defined a new test statistic ω as the difference of the prediction of the *OLS* model and the prediction of the models $m \in \{Full, LASSO - Stable, Shrinkage\}$. I then repeated the testing procedure outlined above. Table 5.3 shows that none of the other models performed better than the *OLS* model in more than 95% of all bootstrapped samples. The additional coefficients did not improve the performance of the other models as compared to the *OLS* model for any of my independent variables.

Thus, I used the coefficients as estimated by the *OLS* model that had access to the most temporal information for further discussion. Figure 5.5 depicts the coefficients of the this final *OLS* model, whereas Table 5.4 lists all coefficients of the model in tabular

Model	Loc. ent.	R. of gyr.	Phys. enc.	Virt. int.	Peer ent.
Full	0.35	0.31	0.88	0.25	0.89
LASSO Stable	0.63	0.60	0.31	0.08	0.01
Shrinkage	0.37	0.31	0.88	0.25	0.63

Table 5.3: ω Model Scores

The table shows how often the other models perform better than the *OLS* model; in fact none of the model performed better than the *OLS* model more than 95% of the time. The values were again based on 10,000 bootstrap samples of the test set.

	Loc. ent.$_t$	R. of gyr.$_t$	Phys. enc.$_t$	Virt. int.$_t$	Peer ent.$_t$
Intercept	0.03	0.06	-0.02	0.05	-0.03
Loc. ent.$_{t-1}$	0.00	0.00	-0.19	0.00	-0.13
R. of gyr.$_{t-1}$	0.00	0.00	0.00	0.00	0.00
Phys. enc.$_{t-1}$	0.03	0.00	0.24	0.00	0.07
Virt. int.$_{t-1}$	0.00	0.00	0.00	0.00	-0.09
Peer ent.$_{t-1}$	0.00	-0.01	0.00	0.00	0.19

Table 5.4: OLS Estimated Coefficients

The coefficients for the final OLS model. Rows represent the independent variables at t_1, whereas the dependent variables at t are organised column wise. In contrast, to the graphical representation of the Granger-causal relationships (Figure 5.5) the table also contains the values for the intercept term.

form as well. Furthermore, the final *OLS* model seemed to be a decent fit for the data as the normal Q-Q plots of the residuals and my dependent variables revealed that for the most part the residuals followed a normal distribution (Figure A.1.1) and for the most part I could only detect weak correlations between the dependent variables and the residuals (Figure A.1.1). However, there were two noteworthy correlations between *physical encounters*$_t$ and $\epsilon_{Phys.enc.}$ and *Peer entropy*$_t$ and $\epsilon_{Peer\,ent.}$, indicating that there might possibly was an unobserved, confounding, "social" variable present.

5.4.2 Results of the OLS Model

Overall, I discovered evidence for several potential Granger-causal relationships in the data. In particular, I found relationships from *physical encounters*, *virtual interactions*,

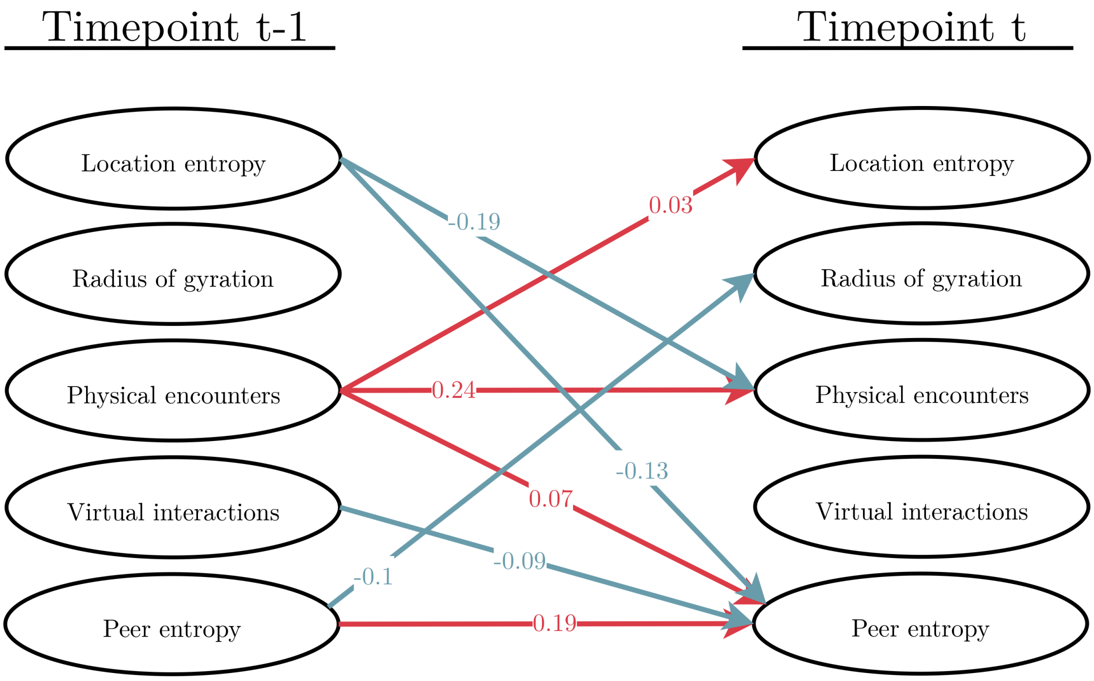

Figure 5.5: OLS Estimated Significant, Causal Coefficients

The figure provides a graphical representation of the estimated Granger-causal relationships in the dataset, where each edge represents the *OLS* model estimate for the respective coefficient. Red coefficients were positive, while blue coefficients were negative in the model. Overall the coefficients seemed to suggest that *location entropy* had a negative effect on social variables, while social variables only positively affected other social variables, but the *radius of gyration* negatively.

and *location entropy* to *peer entropy*, from *location entropy* to *physical encounters* and *peer entropy*, and from *peer entropy* to the *radius of gyration*.

Several of the other estimated, significant coefficients represented auto-regressive processes. *Physical encounters*, and *peer entropy* were all positively associated with past values of themselves; indicating that effects that might change either of those variables might persist for several months for individual students.

In particular, *physical encounters* had a positive relationship with *peer entropy*, whereas *location entropy*, *virtual interactions*, and *peer entropy* had each a negative relationship with their respective dependent variables. This seems to suggest two things: First, physical encounters led to a slight propensity to meet a variety of people in the future and the variety of peers one met decreased the observed future travel radius. Second, having a high variety of visited locations was detrimental to maintaining social ties and led to an overall decrease in the amount and variety of observed social connections. Conversely, meeting a variety of peers in the previous time period also overall decreased the observed mobility.

5.5 Discussion

My results showed that longer term dynamics might indeed be shaping both mobility and social dynamics. Both social and mobility behaviour thus might not only evolve at shorter time frames but also longer ones. This has important implications for understanding the interplay between both mobility and social behaviour as structural inequalities might influence the observed behaviour significantly over long time periods. For example, students that might have a high *location entropy* in one month might thus have a lower rate of *physical encounters* the next month. As *physical encounters* were auto-regressive the effects of the initial higher value for *location entropy* might persist for a substantial amount of time.

I would like to point out, however, that the *OLS* model only performed significantly better than the *null* model for the dependent variables *physical encounters* and *peer entropy* (see Table 5.1). Thus, even though several coefficients of the *OLS* model for the other dependent variables were significant, the results should be treated with caution as the *null* model essentially just used the mean for estimating future states. One possible explanation is that the relationship between the other variables might be better captured by a non-linear model or was not further observable with my choice of temporal aggregation.

In fact, analysing the interplay between social ties and mobility on different timescales might be an interesting further research question. In particular, it is known that social networks evolved at various short term timescales (Scellato et al., 2010). However, relatively little was known about how the interplay between social networks and mobility might evolved at various time scales and as my research shows longer term dynamics are worth investigating.

In contrast to parts of the surveyed literature (Section 5.2.2), I did not find a strong positive relationship between social variables and variables representing mobility. At least for my study being socially well connected did not consistently lead to a higher *radius of gyration* or in the case of *location entropy* only in a minor increase; or in other words for students in Denmark being well connected to their peers did not automatically seem lead to a higher mobility. Possible explanations for the lack of a positive relationship between the social network and mobility were the similar socio-economic status of all individuals in the study and the high availability of relative cheap forms of transportation (i.e. public transport and excellent cycling infrastructure). Relatively low barriers to mobility might possibly negate the added value of being socially close to someone that has a higher potential for mobility (e.g. somebody that owns a car in highly car dependent city). What is more, *peer entropy* was a significant, negative coefficient for the *radius of gyration*. Students that had a very diverse set of peers in one month significantly reduced their overall mobility in the next month. Socialising within their peer group appeared to actually discourage travel behaviour.

I also discovered a positive relationship between *physical encounters* and *peer entropy*. Not surprisingly the more people one had met in the previous month, the more varied the set of people one met was in the next month. Meeting more people in the previous month appeared to allow people to form more new ties than they could have otherwise. Interestingly, I did not find evidence for the reverse relationship: *peer entropy* did not seem to lead to more *physical encounters*. One interpretation for the lack of a reverse relationship is the limited amount of time people could use to socialise with other people. Just because one potentially had access to more people to interact with as measured by the *peer entropy* did not necessarily mean that one also had the temporal resources to spend time with more people.

Furthermore, I found evidence for a negative relationship between *location entropy* and both *physical encounters* and *peer entropy*. However, while I could not measure the *physical encounters* of the students outside the group, I did not find a similar relationship for my variable *virtual interaction* that accounted for all virtual ties of at student. As suggested by others (Jiron, 2007; Mattisson et al., 2014; Putnam, 2000; Viry et al., 2009) having to travel more and having to maintain a geographically diverse set of social ties, limited a person's ability to socialise rather than to maintain a wider set of social relationships. However, a lack of encounters within the observed social network might possibly be compensated by students with encounters outside the observed social network.

Last, *virtual interactions* did have a negative relationship with *peer entropy*. Students did not seem to be able to compensate for either a lower mobility or other social indicators with *virtual interactions*; rather *virtual interactions* led to a lower diversity of the peers of students as students possible substitute meeting in person with phone calls and/or text messages.

I see that time divided is never long, and that regularity abridges all things.

Abel Stevens

6 Personality & Regularity of Behaviour

6.1 Introduction

The last decade saw an increasing number of studies analysing digital traces of human behaviour ranging from data collected via online social networks, over Bluetooth and GPS enabled devices to call data records of millions of people; especially data collected in one way or another via phones was proven to be a fruitful avenue for research (Lazer et al., 2009). These studies dealt with a wide array of topics such as mobility (González et al., 2008a; Song, Qu, et al., 2010), the formation of social ties (Cho et al., 2011), the regularity of behaviour (Williams, 2013) and the influence of personality on human behaviour (Lambiotte & Kosinski, 2014). Unsurprisingly personality influenced a variety of behaviour ranging from sexuality (Bourdage, Lee, Ashton, & Perry, 2007) to academic achievements (Connor & Paunonen, 2007). Past studies that used digital traces to study the effects of personality on behaviour mainly focused on three areas: social media, social networks, and mobility.

There is evidence that social media usage as well as language used on social media was shaped by personality traits. Park et al. (2015) reported that the language used on social media correlated with personality traits. Schwartz et al. (2013) used differential language analysis to uncover a substantial variation of language driven by personality, age, and gender. Last, Kosinski, Stillwell, and Graepel (2013) discovered that Facebook likes alone were highly predictive of not only the personality traits of a person but also of age, gender, intelligence, political and religious views, and sexual orientation.

What is more, how people socialised and their position in their social network was shaped by their personality traits (Morelli, Ong, Makati, Jackson, & Zaki, 2017). Research by Quercia et al. (2012) demonstrated that extroverts had larger networks, while Friggeri et al. (2012) showed that introverts were part of fewer, but larger communities and that extroverts acted as bridges between communities. Personality traits were also an important driver of tie formation as people with similar personality traits were more likely to be friends (Noe et al., 2016; Whitaker, Noe, Whitaker, & Allen, 2016).

Last, personality also affected the mobility patterns of people. Oliveira (2011) were able to use features inferred from call data records and the call graph to predict an individuals personality and Montjoye et al. (2013) managed to infer personality traits from CDR alone. Furthermore, Staiano et al. (2012) showed that the network structure of social interactions alone could be used to predict personality traits. Personality traits could also explain differences in check-in behaviour of users in a location based social network (Chorley et al., 2015) and personality traits were further able to help to explain variations in socio-spatial behaviour (Alessandretti, 2018).

Previous work clearly indicated that various aspects of behaviour captured by digital traces were linked to personality traits. However, one salient feature, the regularity of human behaviour, and how it was intertwined with personality traits had not been studied before. While several studies established that human behaviour was to a relatively large degree regular (see among others González et al., 2008a; Song, Qu, et al., 2010 and Williams, 2013), to the best of my knowledge how personality could shape regularity remained unclear.

My contribution thus lay in studying in how personality might be a driver of regularity of behaviour as observed through digital traces. In particular, I tried to bridge the gap between more computational oriented studies that analysed regularity of behaviour and studies that focused on the effects of personality on behaviour. As a large degree of human behaviour was shown to be regular, understanding regularity of behaviour might have important implications for understanding behaviour in general.

6.2 Background

As I was interested in how personality traits potentially shaped the regularity of behaviour, I briefly review other studies here that found effects of personality on behaviour using digital traces and previous findings with regards to regularity of behaviour.

6.2.1 Personality

In psychology, trait theory posits that humans have relatively stable patterns of behaviour, thoughts, and emotions (Allport, 1966). Furthermore, traits differ between individuals and can influence behaviour. One of the most widespread models of personality is the five factor model (FFM). The five personality factors are openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. While how to best measure personality was still an ongoing research question (Ashton & Lee, 2008; Lee, Ogunfowora, & Ashton, 2005), studies of digital traces of behaviour predominantly used the FFM (Lambiotte & Kosinski, 2014; Chorley et al., 2015; Noe et al., 2016). I followed that convention for my work here.

Openness

People that score high for openness tend to be curious and creative. They are generally open for new experiences. Conversely, people who score low on openness tend to be conservative (Noe et al., 2016). The entropy of visited places and the entropy of contacts with whom an individual called and texted was a significant predictor for openness (Montjoye et al., 2013) Or in other words, the more diverse the set of visited places and the more diverse the set of called friends was, the more likely somebody was to score high on openness. While Chorley et al. (2015) found no correlation between the diversity of visited places and openness, people who scored high on openness tended to visit social and popular venues. Social venues might be popular, because more open people tended to enjoy socialising and meeting new people.

Conscientiousness

Conscientious people tend to be disciplined and organised as opposed to easy-going and inconsistent. They further are rather goal oriented and stay focused on their tasks (Noe et al., 2016).While Chorley et al. (2015) discovered that conscientious users had significantly more check-ins than less conscientious users, it was unclear whether more conscientious users were really visiting more locations or were just more diligent in recording check-ins. Furthermore, Amichai-Hamburger and Vinitzky (2010) found that conscientious users had more friends on Facebook.

Extraversion

Extroverts usually seek engagement with the external world. They tend to be more outgoing, talkative, and sociable; they generally have a larger circle of friends. In contrast, introverts generally have a lower amount of friends and enjoy doing things on their own (Noe et al., 2016).While extraversion was not associated with a particular check-in behaviour (Chorley et al., 2015), they tended to have a larger circle of friends (Pollet et al., 2011; Lambiotte & Kosinski, 2014). Consequently, extraversion was the strongest predictor for the number of friendships among personality traits (Quercia et al., 2012). Very extroverted individuals tended to have on average twice as many friends as very introverted individuals.

Agreeableness

Agreeable individuals value getting along well with others and are generally perceived to be friendly and likeable. Disagreeable individuals tend to place self-interest above getting along with others, are less likely to compromise, and less gullible (Noe et al., 2016).Balmaceda, Schiaffino, and Godoy (2014) reported that agreeable users of social networking sites tend to communicate with extroverted and emotionally stable communication partners. They also discovered that agreeable users were more likely to communicate between themselves than disagreeable users. Furthermore, Selfhout et al. (2010) showed that individuals that scored high on agreeableness were mere likely to be selected as friends by others. At the same time individuals that scored high on agreeableness had a higher number of check-ins Chorley et al. (2015). Yet, Chorley et al. (2015) did not find a significant relationship between agreeableness and check-in diversity.

Neuroticism

Neurotic people are sensitive and nervous, and have a tendency to experience negative emotions such as anger, anxiety, and depression. The opposite of neuroticism is often referred to as emotional stability; and emotionally stable people tend be to calmer and more self-confident (Noe et al., 2016).Conversely, neuroticism was linked to a low tolerance of stress and adverse stimuli (Norris, Larsen, & Cacioppo, 2007). Neuroticism was negatively correlated with the number of check-ins at social venues (Chorley et al., 2015). Moreover, Noe et al. (2016) discovered that neurotic users might socialise less and check-in to fewer locations than others. Last, Montjoye et al. (2013) found mobility features to be useful for predicting neuroticism. The most important features for predicting neuroticism were the daily travelled distance, the entropy of visited places, and the lagged time series of phone calls and texts. However, the paper did not report the type of relationship between the used features and neuroticism (i.e. it was unclear whether a high or low mobility was indicative of neuroticism).

6.2.2 Regularity of Behaviour

Understanding recurring patterns of behaviour has clear implication for not only understanding individual human behaviour but also for how predictable a person's behaviour is. Several studies found a rather remarkable level of regularity in both individual's spatial as well as social behaviour. However, the concept of regularity comprises not just showing up at the same place at the same time. Meeting the same people at the places one visits and having stable patterns of behaviour over time are also important dimensions. In the literature, I found three different types of regularity with respect to digital traces:

1. periodicity of events,
2. predictability, and

3. regularity of behaviour with respect to settings and contexts.

First, periodicity is not only an important characteristic of human behaviour but also a straightforward measure of how regular a person's behaviour is. In short, the more periodic behaviour is the more regular it is. Using various datasets, Williams (2013) studied the periodicity of behaviour using the irregularity of an individual's visits to a certain location per week. They found that visits to academic buildings were the most regular, while visits to outdoor locations were the least regular. Scellato et al. (2010) showed that while network dynamics of human contact networks were non-stationary (i.e. evolving over time) they were still relatively periodic with respect to the 4, 8 and 64 hour interval, where 4 and 8 hours most likely relate to the work day and 64 hours roughly to the weekday/weekend dichotomy. Furthermore, the cardinality of the set of places visited in the last 30 days stayed more or less constant for most people (Alessandretti et al., 2018). This indicated that for a given time period people were highly periodic in the places they visited.

Second, given past observations, the overall predictability of behaviour was another way to assess regularity of behaviour. An individual's behaviour might follow very regular patterns of behaviour but not necessarily be periodic. For example, an individual might very regularly visit a coffeeshop after being done with university no matter at which time their lectures finish. The more predictable a person's behaviour is, arguably, the more regular that person is, or otherwise one could not predict their future state.

González et al. (2008a) discovered that as individuals returned to a few highly frequented locations, such as home or work, they overall displayed a large degree of predictability. Song, Qu, et al. (2010) used the entropy of location traces to estimate an upper bound of the predictability of human mobility behaviour at about 93% of all mobility traces. Lima et al. (2016) reported that motorists only used a small number of possible routes for frequent trips and were thus relatively predictable in their daily routing behaviour.

Third, I looked at the regularity of a student's behaviour in different settings. The underlying idea was that even when a student might not be regular in their temporal patterns, they might still be very regular in what they did and with whom they interacted in certain places. For example, they might always meet a group of friends for the same amount of time or visit a certain place for set period of time. Or they might still be regular with respect to whom they met at a certain locations, or with whom the virtually interacted in certain social or spatial contexts. For example, Sekara et al. (2016) discovered that during periods of high unpredictability with respect to their physical location (i.e. on weekends), people nevertheless were highly predictable with respect to which other people they met.

6.2.3 Data

Previous research concerning digital traces of behaviour mainly analysed three spheres of human behaviour: spatial behaviour (Alessandretti et al., 2018; Chorley et al., 2015; Song, Qu, et al., 2010; Williams, 2013), social behaviour (Morelli et al., 2017; Sekara et al., 2016; Quercia et al., 2012) and virtually mediated interactions (Montjoye et al., 2013; Kanai et al., 2011; Kosinski et al., 2013). By having access to data from the Copenhagen network study (Chapter 3), I was able to study the effect of personality on all three spheres of behaviour as the dataset includes data of 847 students and their psychological traits as measured by the FFM. As can be seen in Figure 6.1, the distributions of the different personality traits more ore less followed a bell curve for the dataset, but students were slightly more agreeable and slightly less neurotic overall.

I could use these data to look at an individual's spatial (which places they visited), social (which social groups they interacted with), and virtual (whom they texted and called) behaviour. In particular, I was interested in how personality traits derived from

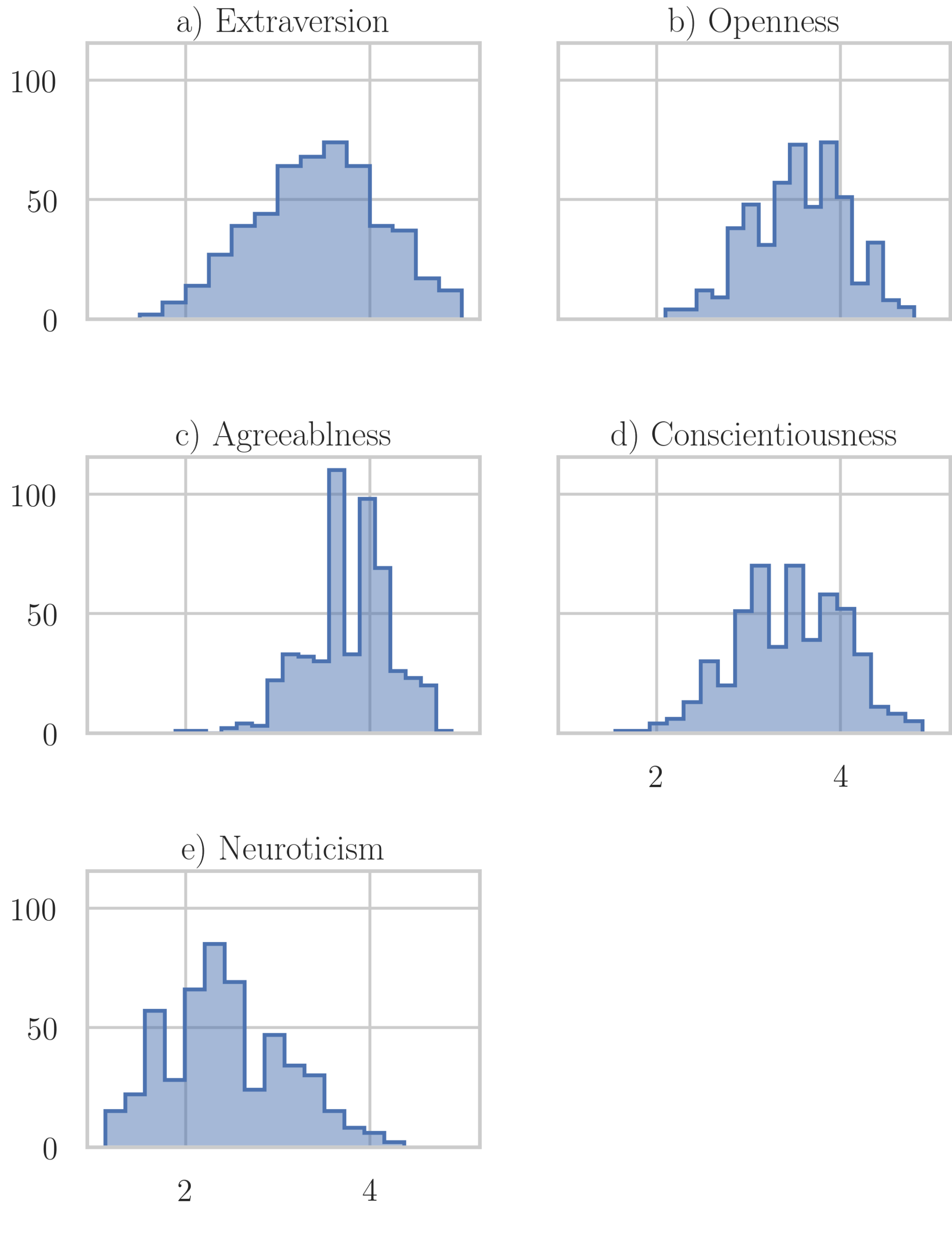

Figure 6.1: Distribution of the Big Five Personality Traits

Overall the personality traits of the students in the study were relatively equally distributed and more or less followed a bell curve. The exception being that students were slightly more agreeable and slightly less neurotic as a population.

the FFM mediated my three aspects of regularity. Thus, for each user, I constructed a time series of stop locations, a time series of the core social group that was present, or in other words in which social contexts the student was in, and a time series of calls and texts describing the virtual interactions of each user.

Recall that a gathering of students usually consisted of a stable core of participants (Section 3.1.2). In order to build the time series of social contexts, I used the social core a student was co-located with to assign unique labels to each social setting. I note that the methodology of Sekara et al. (2016) required me to bin the data into five minute buckets. Various bucket widths were tested in the study, and the authors found that to detect social groups reliably a five minute interval worked best. For consistency, I thus binned my other time series into five minute wide buckets as well.

To construct the time series of stop locations, I followed the process as described in Chapter 3 for processing location traces. In summary, an individual had to at least spend ten minutes at a stop location for it to be a meaningful location.

I also assigned each unique phone number that a student interacted with a label and binned the calls and texts into five minute wide buckets. Due to the relative sparsity of the data, there were almost no collisions due to a student interacting with more than one phone number within one bucket. Although in the handful of cases where a collision occurs, I randomly selected a label to use for that bucket. In order to be able to estimate my entropy based metrics faithfully, I created a secondary series of phone calls and texts based on the labels that preserves the ordering of all calls and texts without binning.

6.3 Hypotheses

Previous studies clearly showed that various aspects of digitally captured behaviour were linked to personality traits. Several other works also established that human behaviour was, to a relatively large degree, regular. However, not much is known about how personality traits shaped the regularity of digital traces of human behaviour. Thus, I tried to answer the following question: *What role did personality play for the regularity with which individuals visited their stop locations, met with their social groups, and called and texted each other?*

Understanding how regular individuals were with respect to their behaviour could not only inform predictions about future behaviour but also improve traffic models, business decisions, and recommendations from digital assistants. Given the reviewed findings (Section 6.2.1), I expected personality traits to play a small but statistically significant role in shaping regularity. I assumed that people who scored high on openness to be less regular and predictable in their behaviour; both with respect to their spatial, social as well as virtual behaviour. As extroverts acted as bridges between communities (Friggeri et al., 2012), extroverts potentially had to balance scheduling demands from a diverse set of social groups. Hence, I expected them to be less regular overall with respect to their social behaviour, spatial, and virtual behaviour. There is some empirical evidence that conscientiousness was correlated with a larger amount of friends, thus there was a higher a priori chance of individuals being less regular in their behaviour. However, as conscientiousness is associated with high levels of organization and discipline, I nevertheless expected high conscientiousness to be associated with a high spatial, social, and virtual regularity. Fourth, agreeableness appeared to lead to a higher probability of forming friendships, and I therefore assumed agreeable individuals to have a more varied social circle and thus a lower regularity with respect to their social interactions. Fifth, as neurotic individuals tended to interact more with other neurotic individuals and tended to have a smaller set of friends, I expected them to be more regular with respect to their social behaviour.

In summary my hypotheses regarding regularity were:

– **H1** Openness was correlated with less spatial and social regularity.

- **H2** Extraversion was correlated with less spatial and social regularity.
- **H3** Conscientiousness was correlated with more spatial and social regularity.
- **H4** Agreeableness was correlated with less social regularity.
- **H5** Neuroticism was correlated with more social regularity.

6.4 Methodology

As interactions between students were relatively sparse and discrete, I could not use traditional measures of frequency analysis such as Fourier transform or wavelet analysis. Furthermore, using auto correlation or in my case auto-mutual information was problematic as in the data short term lags dominate the auto-mutual information function (Figure 6.2), where auto-mutual information is the mutual information (Equation 3.2) of a time series with a lagged copy of itself. This was conceptually very similar to using an auto-correlation plot to estimate the auto regressive part of a time series, except auto-mutual information is more suited to discrete values. Students that simply had longer meetings or spent more time at a certain location had a higher overall auto-mutual information without necessarily being more periodic in their overall behaviour.

6.4.1 Measuring Periodicity

One common and appropriate approach for relative sparse event based data is to focus on the time between successive events. If the events happen periodically then the interval between events should also be relatively stable over time. For each distinct set of events, I can calculate the coefficient of variation (c_v) of the inter-event intervals defined as the

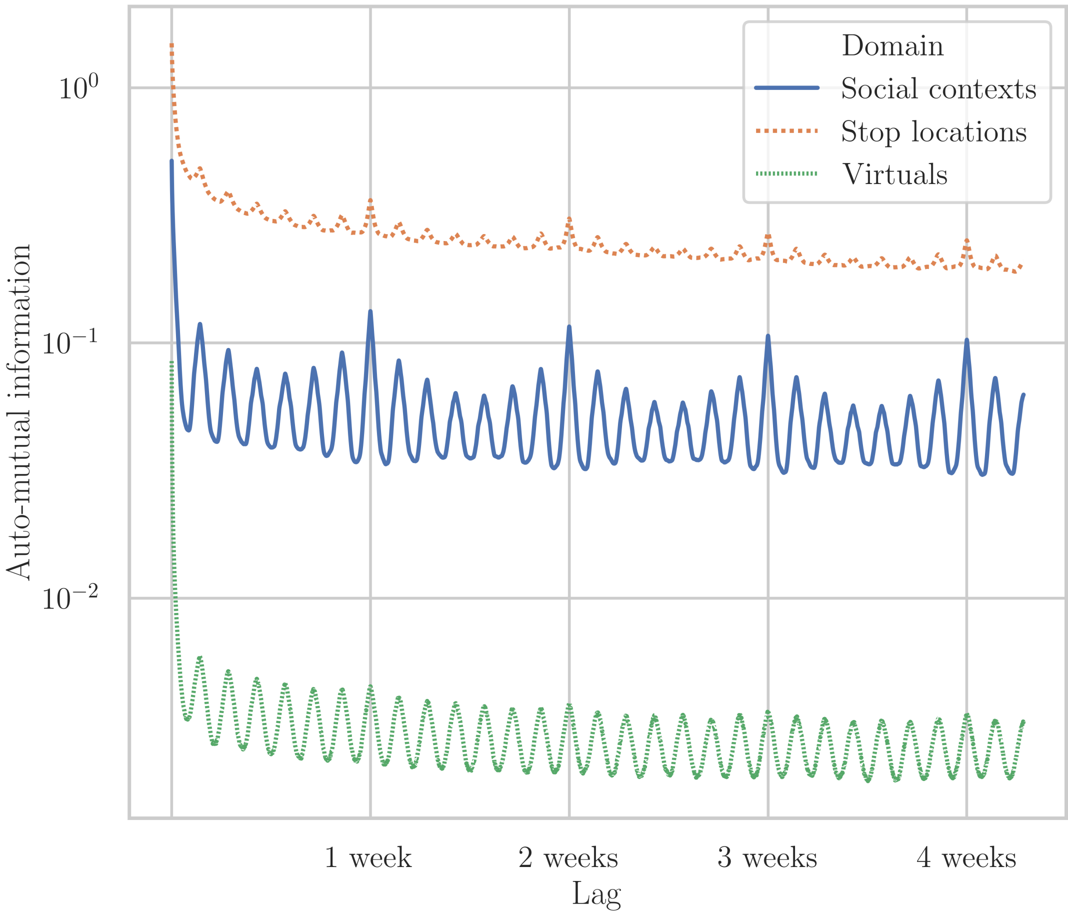

Figure 6.2: Average Auto-Mutual Information

The auto-mutual information averaged over all students measures how much mutual information past values held about the context a student was situated in. It is noteworthy that while how much the individual variables depended on past values differed by a couple orders of magnitude, they all showed a similar periodic pattern. In particular, daily and weekly periods were pronounced.

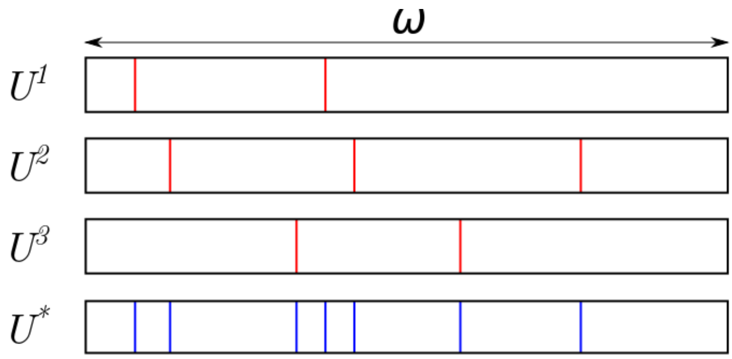

Figure 6.3: Example Visit Trains

Three example visit trains (U^1, U^2, U^3) and their corresponding master train (U^*) of length ω, where each tick represents one event observed at time t within period ω.

ratio of the standard deviation σ to the mean μ (Abdi, 2010):

$$c_v = \sigma/\mu \tag{6.1}$$

As the c_v is a unit less measure I could use the average $\overline{c_v}$ of all possible distinct sets of intervals of events to judge the variation of an individual's overall variation in inter-event times. $\overline{c_v}$ describes how regularly individuals re-visit locations, meet their peers, and call and text their social ties.

While $\overline{c_v}$ was a straightforward measure to assess the variation of inter-event data, multi-modal distributions might have a higher $\overline{c_v}$ without being intuitively less periodic. For example, if one meets someone very regularly every two days as well as every seven days, this would lead to a higher $\overline{c_v}$ than if one just meets every two days without one being necessarily more or less regular. Therefore, I also opted to quantify irregularity by calculating the inter-event irregularity (IEI, Williams, 2013) as this metric was more apt to deal with multi-periodicity. Williams (2013) used this metric to successfully study event data quite similar to the data I used for this chapter.

The IEI is calculated by first constructing a separate ensemble of visit trains (Figure 6.3) with window size ω, where ω denotes the length of period one is interested in (e.g. daily, weekly, monthly, etc. bins of event data), for each distinct set of events (in my case visits to a particular stop location, co-presence of a unique social group, or virtual interaction with the same person).

Let now $c_v(u) = \sigma(u)/\mu(u)$ be the c_v at a particular offset u from the beginning of the visit trains. IEI is then defined as:

$$IEI(U*) = \frac{1}{\omega} \int_0^{\omega} c_v(u)\, du \tag{6.2}$$

The intuition behind IEI is that it estimates regularity by assessing whether events happen at the same time in all time periods ω irrespective of a particular offset u. This way IEI is also able to deal with multi-periodicity better than just using the $\overline{c_v}$. By collapsing the visit trains $U^1, U^2, ..., U^n$ of an individual into a master train U^* (Figure 6.3), I could calculate IEI in one single pass over U^* by noting that $c_v(u)$ stays constant between subsequent events in U^* as the value of the integral only changes when an event occurs in U^* (Williams, 2013). While the original metric had been proposed for instantaneous events, it was straightforward to expand the metric to events that had an observed duration by adjusting how inter-event intervals were calculated.

In order to compute IEI, I needed to determine the appropriate length of the time window ω. As a lot of behaviour follows a weekly rhythm and Williams (2013) opted to focus on weekly periodicity in their analysis. However, longer periods might play an important role for understanding periodicity. I judged the appropriate length of the time by measuring how much the time series on average depends on past values. I used the average auto-mutual information of student's time series (Figure 6.2) to assess how much current states of the system depend on past states. Figure 6.2 clearly shows a strong daily as and a weak weekly component of spatial and social behaviour between a student's behaviour and their lagged time series. A Fourier transform of the auto-mutual information series, reveals a fairly strong daily period, but no periods longer than two days (Figure 6.4). As I could not spot longer periods than one week in either plot, I decided to set ω to one week.

6.4.2 Assessing Predictability

As I argued above, measuring periodicity alone might not be sufficient to determine whether an individual's behaviour is regular. I also wanted to assess how predictable a student's behaviour was or in other words how much uncertainty a student's time series contained. The idea is that even when a individual might not be periodic in their behaviour, they might still be regular in the series of locations they visit, or with whom they meet, or text and call.

If entropy is viewed as a measure of surprise, then one way to quantify the amount of uncertainty of a time series is to estimate the entropy of said time series. Ceteris paribus, a time series with a lower entropy is more predictable than a time series with a higher entropy. I thus in a first step calculated the Shannon entropy (Equation 3.1) of a user's time series.

However, calculating the Shannon entropy does not take into account any temporal dependencies. Other studies such as Alessandretti (2018) by using only the Shannon entropy failed to account for the temporal ordering of observations, which is crucial for understanding the temporal regularity of behaviour.

Several different metrics were proposed and used in the literature to estimate the entropy of a time series and take temporal ordering into account. Among current approaches were sample entropy (Tang, Lv, Yang, & Yu, 2015), permutation entropy (Tang et al., 2015), diffusion entropy (Scafetta & Grigolini, 2002), and an entropy estimator based on Lempel-Ziv complexity (Song, Qu, et al., 2010).[1]

[1] For a more detailed review see for example Tang et al. (2015)

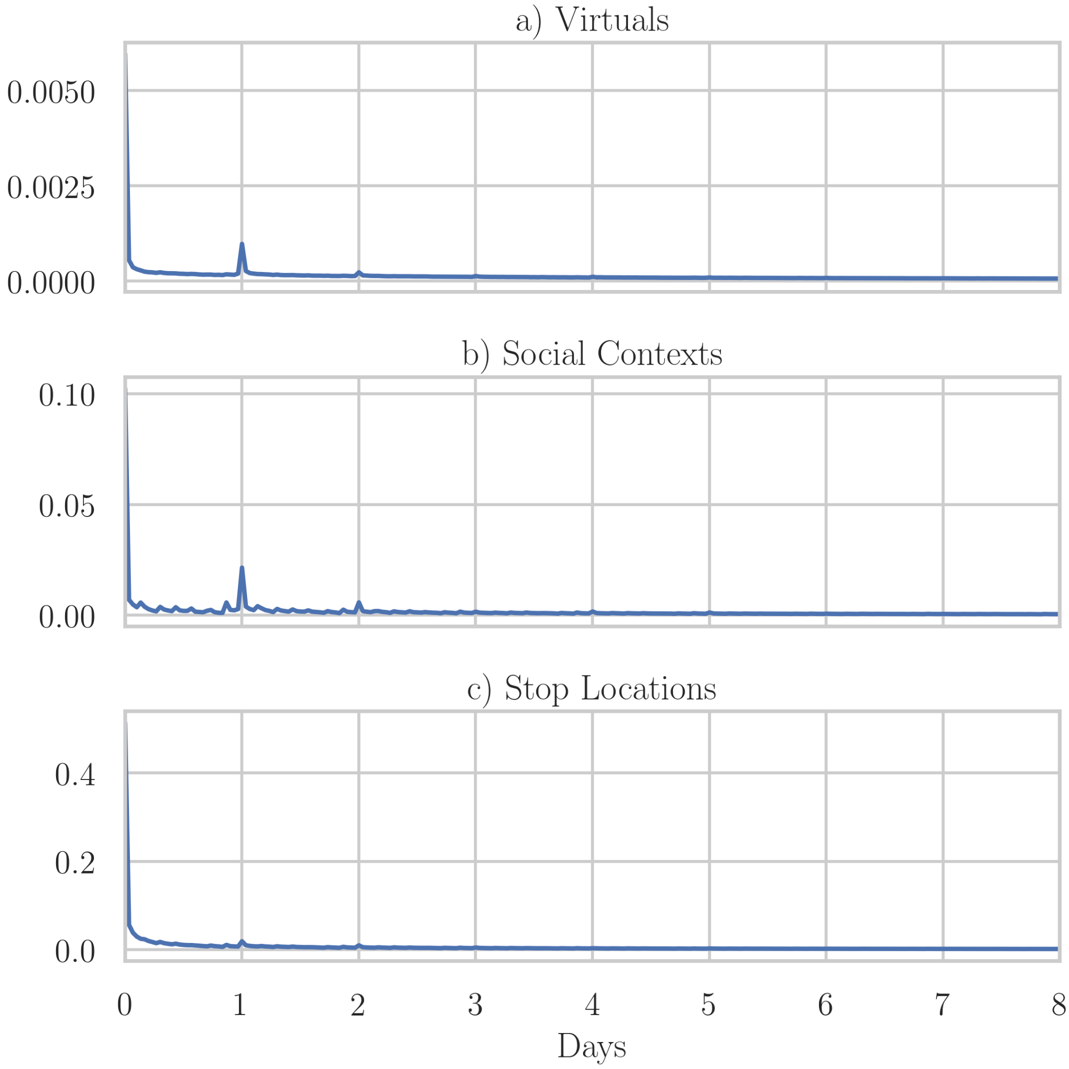

Figure 6.4: Frequency Plot of the Average Auto-Mutual Information Series

By applying the Fourier transform to the auto-mutual information series, I can visualise the frequency spectrum of the data. It is noteworthy that while Figure 6.2 indicates weekly periodicity, the longest period visible in this figure is 48 hours.

While a comparison for the effectiveness of various estimators existed for other fields (Hansen, Wei, Shieh, Fourcade, & Isableu, 2017), to my knowledge no such comparison existed for the field of human behavioural dynamics. Thus, I decided to use the most common one in the field of human behavioural dynamics, the Lempel-Ziv based entropy estimator (hereafter H_{LZ}). H_{LZ} is defined as follows:

$$H_{LZ} = - \sum_{T' \subset T} P(T') \, log_2 \, P(T') \tag{6.3}$$

where $P(T')$ is the probability of finding a particular time-ordered subsequence T' in the time series of the individual (Song, Qu, et al., 2010).

As discussed above the Shannon entropy does not account for temporal patterns and indeed I found that for my data that the Shannon entropy significantly overestimated the actual entropy of the time series (Table 6.1). Furthermore Song, Qu, et al. (2010) showed that H_{LZ} was robust to missing observations. Last, H_{LZ} is directly related to the predictability of the time series via the Fano inequality (Song, Qu, et al., 2010), where the inequality relates to the information lost in a noisy information channel (Lu et al., 2013).

	Avg. H	Avg. H_{LZ}	p
Virtual	4.2444	2.1610	<0.001
Social Contexts	6.3301	3.4300	<0.001
Stop Locations	4.3539	2.0546	<0.001

Table 6.1: Comparison of H and H_{LZ}

p-values of whether μ_H is different from $\mu_{H_{LZ}}$ are estimated using 10,000 bootstrap samples, where I follow the procedure for resampling for a hypothesis test as described by Boos (2003). The key point is that one needs to make sure that the resampling is performed under the appropriate null hypothesis. In my case, the null hypothesis for the pooled t statistic is $\mu_H - \mu_{H_{LZ}} = 0$.

6.4.3 Assessing Contextual Regularity

The last aspect of regularity, I was interested in was, whether a student's observed behaviour stayed the same at the same locations and social settings. In particular, I was looking at how regular the duration of visits as well as interaction with a student's social groups was. I could use $\overline{c_v}$ of the duration of the visits to a student's stop location and meetings with a distinct social group, as a metric to judge the variation of event durations.

Furthermore, I was able to assess how consistent students were with whom they meet at locations, and with whom they interacted virtually in different spatial and social settings. By calculating I (Equation 3.2) between a user's time series of stop locations, social contexts, and virtual interactions, I could quantify how consistent a user's behaviour in one sphere was with the other sphere. In other words, how much did the entropy of the behaviour in one sphere decrease by knowing the behaviour in the other sphere. While several authors found evidence for a strong link between the spatial and social behaviour (Section 2.4), little was known of how personality might mediate the link between spatial and social behaviour.

6.4.4 Scale and Statistical Power

When studying the effects of personality traits on behaviour, small effect sizes were relatively common due to the inability to perfectly map traits as well as heterogeneity of behaviour (Butcher, Graham, & Ben-Porath, 1995). Consequently, sample size played an important role for detecting the effects of personality on behaviour. A common approach in the literature (Ross et al., 2009; Chorley et al., 2015) was to split the data into quantiles (most commonly terciles) based on the personality traits and compare the differences in behaviour for the top and bottom groups. However, I decided against this approach as it is unclear how to best pick the number of quantiles to split the data in. Trivially, if there exists a linear relationship between the variables, the more quantiles one uses the bigger the differences between the groups. Even though the Copenhagen network study in total consists of 847 individuals, I only had personality traits for 508 students. With this sample size I was however still able to detect effect sizes of about 0.1 at the 95% significance level (Hulley, Cummings, Browner, Grady, & Newman, 2013).

6.5 Findings

6.5.1 Periodicity

As illustrated by Figure 6.2, on average, the periodicity of behaviour of the students followed a rhythmic pattern. Overall, there was strong periodicity with respect to daily and weekly patterns of behaviour; especially for the social context a student was in. Observing that I calculated the $\overline{c_v}$ for all distinct inter-event intervals for each student. Table 6.2 depicts the correlation between $\overline{c_v}$ and personality.

I found a significant positive relationships between agreeableness, conscientiousness, and extraversion leading to a higher $\overline{c_v}$ for virtual interaction. As theorised (**H2** and **H4**), both extraversion and agreeableness led to less virtual regularity. Unexpectedly, and contradictory to **H3** conscientiousness was also correlated with greater virtual irregularity. Although, this might be partly explained by the fact that virtual communications were not solely driven by the students themselves.

However, as discussed above, the $\overline{c_v}$ might be misleading if the distribution of inter-event intervals is multi-modal. I thus also calculated the IEI with a window length of one week and again analysed the correlation between the metric and personality traits. And indeed I could observe differences between the IEI and the $\overline{c_v}$. Unexpectedly, and in contradiction to **H1** and **H2**, individuals that score high on extraversion or openness were

	Virtuals	Social Contexts	Stop Locations
Extraversion	0.20*	0.00	0.07
Openness	0.04	0.01	0.04
Agreeableness	0.12*	0.03	0.08
Conscientiousness	0.11*	0.05	−0.03
Neuroticism	−0.02	0.01	−0.07

Table 6.2: Correlation Between $\overline{c_v}$ of the Inter-Event Intervals and Personality

Only $\overline{c_v}$ of virtual behaviour seemed to be significantly correlated with personality traits. *:$p < 0.05$

more regular with respect to how they visited stop locations. One possible explanation is that people with either of those traits had more varied social ties and thus had to balance more social appointments, which in turn led to relatively stable schedule with respect to spatial locations. Furthermore, students that scored high on conscientiousness had again a higher IEI score for virtual ties.

	Virtuals	Social Contexts	Stop Locations
Extraversion	−0.02	0.03	−0.11*
Openness	−0.05	−0.04	−0.13*
Agreeableness	0.00	0.03	−0.07
Conscientiousness	0.10*	0.04	0.00
Neuroticism	−0.01	0.02	0.01

Table 6.3: Correlation Between IEI and Personality

While stop locations seem to be significantly negatively correlated with both extraversion and openness, conscientiousness was positively correlated with virtual behaviour. *:$p < 0.05$

6.5.2 Predictability

In a first step, I calculated the Shannon entropy for each student and the respective sphere of behaviour. While the Shannon entropy was not the most apt metric to assess the regularity of a time series (as shown in Table 6.1), it allowed me to compare the results of my study with previous research that looked at H with respect to personality traits.

	Virtuals	Social Contexts	Stop Locations
Extraversion	0.20*	0.03	0.21*
Openness	0.01	0.02	0.01
Agreeableness	−0.01	0.03	0.02
Conscientiousness	−0.01	0.01	0.05
Neuroticism	−0.01	−0.09*	−0.05

Table 6.4: Correlation Between H and Personality

Only extraversion was significantly correlated with both less predictability for virtual behaviour as well as the set of visited stop locations. *:$p < 0.05$

The results as reported in Table 6.4 were consistent with previous findings. I found that more extroverted students have more varied social interactions via phone calls and texts as well as visit a greater variety of stop locations as predicted by **H2**. These findings suggested that indeed more extraverted students had more varied social ties and thus might have indeed more social commitments. I also found that neuroticism was negatively correlated with the variety of social contexts a student can be observed in. This was in line with my original hypothesis (**H5**) that neuroticism leads to more social regularity.

	Virtuals	Social Contexts	Stop Locations
Extraversion	0.15*	−0.03	0.15*
Openness	0.01	0.02	0.02
Agreeableness	−0.04	0.02	−0.02
Conscientiousness	−0.03	−0.01	0.00
Neuroticism	0.01	−0.07	−0.03

Table 6.5: Correlation Between H_{LZ} and Personality

As discussed above (Table 6.1 and is apparent in this figure as well, H overestimates the irregularity of a time series as it does not account for the temporal ordering of the observations. *:$p < 0.05$

Next I looked at how predictable a student's time series was by calculating H_{LZ} (Table 6.5). Recall that a higher H_{LZ} can be interpreted as a lower overall predictability of a time series. I was able to observe a significant positive correlation between extraversion

	Social Contexts	Stop Locations
Extraversion	0.04	0.16*
Openness	0.09*	0.06
Agreeableness	−0.04	0.09*
Conscientiousness	0.00	0.02
Neuroticism	−0.03	−0.10*

Table 6.6: Correlation Between $\overline{c_v}$ of the Event Durations and Personality

Mostly $\overline{c_v}$ of event durations of the stop locations was correlated with personality traits and not social contexts. Only openness was correlated with $\overline{c_v}$ of the durations of the social contexts. *:$p < 0.05$

and H_{LZ} for both virtual communications as well as the stop locations and a negative correlation between neuroticism and H_{LZ}. Overall and in support of **H2**, I discovered that more extroverted students had a significantly lower predictability for both the stop locations they visited and their virtual interactions.

6.5.3 Regularity of Contextual Behaviour

Last, I analysed the regularity of behaviour of students at different locations and settings. In particular, I looked at the duration of time spent at locations and with friends and whether they socialised with the same friends online and in real life in the same settings. To see how varied the distribution of durations of events was for students I could again calculate the $\overline{c_v}$. However, in this case, I calculated the $\overline{c_v}$ with respect to the distribution of durations of social encounters and time spent at stop locations (Table 6.6). I found that more extroverted and more agreeable students were significantly more varied in the amount of time they spend at the locations they visited; further confirming **H2** and **H4**. In agreement with **H1**, individuals that scored high on openness were also more likely to vary the amount they spent with different social groups. Furthermore, there was also evidence for **H5** as I could also observe a significant negative correlation between neuroticism and the $\overline{c_v}$ for event duration.

	Virt.-Soc. Con.	Virt.-Stop Loc.	Soc. Con.-Stop Loc.
Extraversion	0.26*	0.27*	0.18*
Openness	0.02	0.00	0.03
Agreeableness	0.10*	0.10*	0.05
Conscientiousness	0.13*	0.12*	0.13*
Neuroticism	-0.05	-0.02	-0.08

Table 6.7: Correlation Between I and Personality

Interestingly both extraversion and conscientiousness had a relatively large correlation with all possible variations of I. Whereas I hypothesised that conscientious students would be more regular, the result is surprising for extraversion. *:$p < 0.05$

To assess how strongly behaviour was linked between different domains, I calculated I between a student's time series of virtual interactions, stop locations, and social contexts. Interestingly individuals that scored higher on agreeableness, extraversion, or conscientiousness all had an increased coupling between at least two different spheres than individuals that scored lower on those personality traits. I also found a negative relationship between neuroticism and $I(\textit{social contexts}, \textit{stop locations})$.

While these findings suggested that the link between social, spatial, and virtual behaviour was indeed mediated by personality, the dynamics at play did not agree with my original hypotheses. While I would expect conscientiousness to be associated with a higher contextual regularity (**H3**), both extraversion and agreeableness were, in contrast to **H2** and **H4**, also associated with a higher contextual regularity. It is noteworthy that extraversion, which had been mostly been associated with less social and spatial regularity in my analysis, so far still had a significant bearing on the dependency between all types of behaviour.

6.6 Discussion

After a thorough search of the relevant literature I believe that my study is the first to attempt to systematically assess the effect of personality traits on the regularity of behaviour. With this study I tried to bridge the divide between more computationally oriented studies that have focused on regularity of behaviour and work that has focused on the effects of personality on behaviour. The two main finding of this chapter are: First, while regularity was indeed influenced by personality characteristics, its effects were relatively modest and only observable at a significant level for a subset of the personality traits. Overall extroverted students seemed to be slightly less regular than introverted students and more neurotic students slightly more regular. Second, there appeared to be a strong relationship for both extraversion and conscientiousness with contextual regularity, meaning that there was a lot of mutual information between virtual interactions, social contexts, and stop locations.

However, my study did not constitute an experiment with randomly sampled test and control groups, rather my study should be viewed of an analysis of digital traces of behaviour and personality in the "wild". As my data consisted of relatively sparse events of interactions with social groups, visits at stop locations as well as texts and calls, the findings of my study have thus to be viewed in this context. The regularity of the behaviour of students could be more or less affected by personality traits than the regularity of the general population is shaped by personality traits. Moreover, I my study did not take into account individual characteristics of the students such as gender as splitting the data would have further reduced the power of my study to find the small effect sizes common for studying the effects of personality traits (Butcher et al., 1995). Furthermore, while most variables were normally distributed, both the IEI and H of the social contexts as well as I deviated from a normal distribution, which needs to be taken into consideration when interpreting my results.

With respect to my hypothesis, I found the most support for **H2**, that means that extraversion was linked to less social and spatial regularity. Furthermore, there was a decent amount of evidence for **H5**, that neuroticism was correlated with more social regularity. Students with higher scores for extraversion had less predictable time series as well higher $\overline{c_v}$ for both inter-event intervals as well as overall event durations. I also discovered some evidence for **H1** and **H3** that openness and agreeableness were associated with a lower virtual regularity.

While not all my results supported all my hypotheses and the effect sizes were generally relatively small, I would like to point out that differences with respect to predictability and various possible confounding variables are rather rare. For example, a much bigger study (Song, Qu, et al., 2010) had not found any difference in regularity with respect to gender, age, home location, language group, population density, and rural versus urban environments.

Interestingly, the relationship between weekly periodicity and personality was, for extraversion, openness, agreeableness, and conscientiousness, opposite to what I initially theorised. Possibly, this was due to how a more varied social circle outside the study imposes a higher constraint on scheduling meetings and visits to certain stop locations.

Last, I found that behaviour in different contexts was significantly affected by personality. Although, it was known for a while that social and spatial behaviour of individuals was intrinsically linked (Alessandretti, 2018), to the best of my knowledge my study was the first that tried to understand how personality might play a role. In particular, I discovered that extraversion, agreeableness, and conscientiousness were associated with a higher degree of interdependence between spatial and social behaviour. Furthermore, as there were deviations from a normal distribution for I and this might have affected my findings partially.

To summarise, I found several significant relationships between regularity and personality. However, how exactly personality shaped regularity was clearly more complicated than I initially theorised given my findings with respect to weekly periodicity as well as mutual information between different spheres of behaviour. I believe that understanding the precise dynamics of how personality affects regularity could be an interesting route for future research. Especially since several studies also used digital traces of behaviour to predict personality. My results suggest that personality traits have implications not only for understanding human behaviour but also for models of human mobility and social networks.

I may not have gone where I intended to go, but I think I have ended up where I needed to be.

Douglas Adams

7

Conclusion

Everyday lives have increasingly become mediated by digital technology. Internet enabled mobile phones allow virtual communication with almost anyone instantly, online social networks shape how we socialise, and travel cards allow seamless trips on public transportation networks. All those digitally mediated interactions leave digital traces of behaviour behind. Studying those traces led to a deluge of new quantitative studies about human behaviour (Chapter 2).

And while the results of those new studies were not always revolutionary—some were re-discovering previous results on a grander scale—the fact that we could study the behaviour of thousands if not millions of individuals at a granular scale was. The new torrent of data was then not only leading to more detailed understanding of human behaviour, but also to a more detailed grasps of how behaviour changes with time, how processes spread, and how we can use regularities for prediction. The main dynamics of both social and mobility behaviour were fairly well understood in the literature (Chapter 2), including the influence of a variety of mediating factors such as age, gender, and socio-economic status (Section 2.5).

7.1 Contribution

However, several other possible contextual and mediating factors were comparatively purely understood. The contribution of this thesis consequently lay in studying three mediating factors that shape the dynamics of social networks and mobility behaviour.

In Chapter 4, I tried to understand what role different factors such as when, where, and with whom students met had for predicting future encounters between the students. I phrased the problem as a link-prediction problem in a time varying graph and found that who else was present at an encounter and the wider network topology were the most salient features for understanding future encounters between students. While whom else students met at the same time also improved my prediction, whereas when students meet had only a negligible impact on my prediction.

In Chapter 5, I analysed the longer term aggregated patterns of social and mobility behaviour of the students. In particular, I looked at the monthly pattern of a student's behaviour. I measured social behaviour with the set of a student's encounters and the entropy of said set and a student's mobility behaviour via a student's radius of gyration and the entropy of their visited locations. By using Granger-causality as a framework, I discovered that a student's past mobility had only a very limited impact on future mobility as well as social behaviour. However, both the total amount of encounters and the diversity of social connections with other students had a positive influence on future social and mobility behaviour. Someone that was very social in one month, tended to be very social the next month as well. In contrast, someone that met a lot of different people appeared to travel less the next month. Interestingly, the entropy of physical encounters was negatively associated with the diversity of locations a student visited the month before. One possible explanation is that increased mobility was not necessarily due to choice but due to higher fixed time commitments at other locations, which left less time to socialise within the peer group.

In Chapter 6, I answered the question whether personality traits shaped the spatial and social regularity of students. While evidence existed that personality shaped social and spatial behaviour and that both social and spatial behaviour were to a certain extent regular (Chapter 2), whether personality might mediate the regularity of behaviour was unclear. Overall, I found that extraversion was linked to less social and spatial regularity, whereas neuroticism was correlated with more social regularity. The effects were, however, relatively modest.

In summary, I showed that the social and spatial behaviour of the students, in the dataset I used, is to varying degrees dependent on mediating and contextual factors ranging from places to personality. When analysing both social networks and mobility behaviour it is thus prudent to account for contextual, potentially confounding variables.

7.2 Limitations & Future Directions

I would however like to highlight that my study was conducted using a dataset with a very specific subset of the population (Chapter 3). Moreover, the data I used for my thesis was neither a randomly selected subset of the general population nor selected in a way to be either representative of the overall population. My research is likely not directly transferable to a wider population and thus further research is needed before extrapolating freely to a wider population.

Furthermore, while I discovered several statistically significant associations in the data, I would like to stress that neither my analysis nor the data I used for my thesis provided irrefutable prove of causal relationships. Even though throughout my thesis I controlled for a variety of possible confounding factors such as the individual characteristics of students or the innate mobility pattern of individuals, the space of possible confounding variables is for all practical purposes infinite (Nagarajan et al., 2013). While any unobserved confounding variable might affect the results, I followed the reasoning of Druckman and Kam (2009) and believe that in the absence of any empirical or theoretical indication that variance within my population might affect the results my findings can still be insightful, when keeping the limitations of the sample and analysis in mind.

Given the limitations of my study the first avenue for future research is thus to replicate the findings of this thesis with different and/or wider samples of data. A second, straightforward extension of this thesis is to try and analyse other external factors that might shape the social and mobility behaviour of individuals.

In Chapter 5, I showed that longer-term dynamics play a role in shaping behaviour, however to the best of my knowledge a methodological assessment of how social ties and mobility co-evolve over various time scales has not been done yet. Friendship formation and dissolution is clearly not instantaneous. Some ties might exist for a much longer period, whereas others are more fleeting. Similarly mobility is shaped by different time scales: daily, weekly, or even longer patterns might vary considerably for the same individual. Thus, another possible route for future work could be to study the co-evolution of social ties and mobility at various timescales.

A

Appendices

A.1 The Interplay of Long-Term Social & Mobility Behaviour

A.1.1 VAR Estimators

While the OLS estimator is guaranteed to be the best linear unbiased estimator in the limit, the OLS estimator does not necessarily be the best estimator for my given VAR problem as defined in Chapter 5. Here I present various other possible estimators for my VAR system of equations.

OLS with Wald test

By using matrix notation I can succinctly write the ordinary least squares (OLS) estimate for all A_τ as

$$\hat{\mathbf{A}}^{OLS} = (\mathbf{X}^{past\top}\mathbf{X}^{past})^{-1}\mathbf{X}^{past\top}\mathbf{X}^{future} \tag{A.1.1}$$

where each column of $\mathbf{X}^{past}$ represents a particular lagged observations and $\mathbf{X}^{future}$ represents the observations at time t.

Once I have estimated $\hat{\mathbf{A}}^{OLS}$ I can use standard statistical tests to assess individual coefficients improve my prediction or not. In other words, I can test whether the inclusion of a coefficient $a \in \hat{\mathbf{A}}^{OLS}$ significantly reduces my mean squared error (MSE).

Commonly a Wald test is performed to test whether the "full" model that has access to all coefficients performs statistically better than a "restricted" model that is missing information about a set of coefficients (corresponding to all lagged observations of one variable, Luetkepohl, 2005). In a first step, I bootstrap the Wald statistics to determine the distribution of the statistic following the approach of Boubtane, Coulibaly, and Rault (2013). In a second step, I prune the $\hat{\mathbf{A}}^{OLS}$ to only include the set of coefficients whose Wald score is in the top 0.95 of my bootstrapped Wald statistic.

S3L

There exists a multitude of approaches for modelling longitudinal data with structural equation models (SEM) as well. For a comprehensive review see Rosel and Plewis (2008) or Newsom (2015), but common approaches include cross-lagged effects models (whose system of linear equations can be expressed as VAR(1) models), time-series based approaches such as autoregressive moving average (ARMA) models and growth curve models.

I decided against using SEMs directly for two reasons: First, SEMs are generally not used in the context of prediction, but are rather used to build a model that fits the data well based on theory. Recall that I am especially interested in testing my causal models using prediction of unseen data. Second, the above mentioned approaches for estimating SEM models have in common that they usually rely only a single run of learning with regard to the model parameters, which can be unstable in regards to the learned parameters. Small changes to the finite sample may lead to completely different causal estimates (Rahmadi et al., 2018). This is especially problematic in the case of small sample sizes or noisy data.

Thus, I have decided to use a novel approach for using SEMs to uncover causal structures, stable specification search for longitudinal data (S3L) instead of trying to apply SEMs directly to model my problem (Rahmadi et al., 2018; Rahmadi et al., 2017). The algorithm is based on stable specification search, that is given a population of models regard only those model parameters as stable that occur in the majority of all models (Meinshausen & Bühlmann, 2010) .

Not only does S3L avoid the problem of unstable causal estimates it also compares favourably to other state-of-the-art structure learning algorithms for longitudinal data such as FGES, PC-stable, CPC, CPC-stable, and PC-Max (Rahmadi et al., 2018). The basic idea of S3L is to sub-sample the data D into subsets of size $D/2$, and then use those subsets to find SEMs that are both parsimonious as well as have a high model fit.

As finding pareto-optimal models SEMs for each subset D that have a high fit and a low model complexity is a hard non-convex optimization problem, a genetic algorithm $NSGAII$ is used to optimise an initial population of models. The result of this optimization phase is the set M of pareto-optimal models. The causal models in set M are represented as completed partially directed acyclic graphs (CPDAG), where directed edges represent a causal relationship between A and B and undirected edges represent a correlation between A and B, whose causal relationship cannot be reliable inferred from the data alone.

In a next step all undirected edges and directed edges of all CPDAGs of M are pruned according to two criteria:

1. To avoid false positives only *stable* edges over the model population are considered. In other words the edge has to appear in at least π_{thr} of all models. I err on the side of caution and set π_{thr} conservatively to 0.8.

2. In order to avoid over fitting only edges that appear in a model whose model complexity is less than π_{BIC} are considered to be parsimonious. π_{BIC} is computed by grouping all models according to their model complexity (i.e., how many edges/coefficients the model has) and calculating the average BIC for each group of models with the same complexity resulting in the set Ψ of average BIC model scores. π_{BIC} is then set to $min(\Psi)$.

LASSO regression

I can also change the standard cost function of the OLS regression to find a parsimonious set of coefficients. Recall that the standard cost function OLS regression tries to minimise is the mean squared error (MSE) on the training data. Effectively LASSO regression performs L1 regularization (i.e. it adds the sum of the absolute value of the coefficients to its cost function). Hence, LASSO regression tries to balance minimising the MSE with finding a parsimonious set of independent variables. In a sense LASSO only keeps the "important" coefficients and sets the less relevant coefficients to zero. It is the inherent feature selection procedure of a LASSO regression that allows it to be used to infer "causal" networks of variables (Nagarajan et al., 2013). One straightforward way to use LASSO is to use cross validation to first find the value of λ that minimises the MSE for the training set (Nagarajan et al., 2013), where λ is the weight of the L1 regularization parameter. In a second step I can then use the CV-optimal value for λ to estimate my VAR system of variables.

Another way to use LASSO regression for estimating my VAR model is to use the stability selection procedure by Meinshausen and Bühlmann (2010). In a traditional setting, I chose one element of the set of possible models given the regularization/hyper-parameter λ

$$\{\hat{S}^{\lambda}; \lambda \in \Lambda\} \tag{A.1.2}$$

where Λ is the set of regularization parameters.

Instead of optimising λ directly stability selection picks a region of Λ based on pre-defined error bounds on the number of false positives. Furthermore, the original data D are perturbed many times (for example by sub-sampling) and one chooses all variables that occur in a large fraction of the resulting selection sets based on the bounded region of Λ. For details of how the error bounds are defined see Meinshausen and Bühlmann (2010). Accordingly, the set of stable parameters is defined as:

$$\hat{S}^{stable} = \{k : max_{\lambda \in \Lambda} \hat{\prod}_k^{\lambda} \geq \pi_{thr}\} \tag{A.1.3}$$

The idea is again to sub-sample the data D into subsets that are usually of size $D/2$, fit the LASSO model for the pre-selected region of Λ, and only include coefficients in the final model if they appear in at least π_{thr} of all models. Meinshausen and Bühlmann (2010) show that using stability selection in conjunction with LASSO considerably improves the accuracy and precision in uncovering "causal" relationships than using LASSO with cross-validation (especially for high-dimensional data).

James-Stein shrinkage estimator

Opgen-Rhein and Strimmer (2007) propose to apply James-Stein-type shrinkage to efficiently estimate the coefficients of a VAR system for which only data from a limited amount of time points is available. The OLS estimates of the coefficients of a VAR system can be obtained by Equation A.1.1. I can then rewrite Equation A.1.1 in terms of the $(n-1)$ multiple of the empirical covariance matrix $\mathbf{S} = \mathbf{\Phi}^\top \mathbf{\Phi}$, where $\mathbf{\Phi}$ is the joint matrix $\mathbf{\Phi} = [\mathbf{X}^{train} \mathbf{X}^{future}]$. I note that $\mathbf{S}$ has two sub-matrices $\mathbf{S}_1 = \mathbf{X}^{train\top} \mathbf{X}^{train}$ and $\mathbf{S}_2 = \mathbf{X}^{train\top} \mathbf{X}^{future}$, yielding

$$\hat{\mathbf{A}}^{OLS} = (\mathbf{S}_1)^{-1} \mathbf{S}_2 \tag{A.1.4}$$

$\mathbf{S}$ can now be replaced by a James-Stein shrinkage estimate $\mathbf{S}^*$, which allows us to determine the sub-matrices $\mathbf{S}_1^*$ and $\mathbf{S}_2^*$. Finally yielding the shrinkage estimate of the coefficients of the VAR system,

$$\hat{\mathbf{A}}^{Shrink} = (\mathbf{S}_1^*)^{-1} \mathbf{S}_2^* \tag{A.1.5}$$

It is unlikely that any of the components of $\hat{\mathbf{A}}^{Shrink}$ are exactly zero. Thus, I need to statistically test whether the entries of $\mathbf{A}^{Shrink}$ are vanishing. Instead of testing the regression coefficients directly, Opgen-Rhein and Strimmer (2007) propose to test the corresponding partial correlation coefficients as this facilitates smaller sample sizes as well as allows to account for dependencies between the coefficients. Once I have computed the partial correlation coefficients I can use the local false discovery rate approach as proposed by Efron (2005) to determine salient coefficients.

	$\epsilon_{Loc.ent.}$	$\epsilon_{R.of\,gyr.}$	$\epsilon_{Phys.enc.}$	$\epsilon_{Peer\,ent.}$	$\epsilon_{Virt.int.}$
Loc. ent.$_t$	-0.04*	-0.06*	0.11*	0.06*	0.01
R. of gyr.$_t$	-0.06*	-0.03	0.00	-0.07*	0.00
Phys. enc.$_t$	0.03	0.00	0.18*	0.13*	0.04*
Peer ent.$_t$	0.0	-0.02	0.13*	0.19*	0.02
Virt. int.$_t$	0.04*	-0.02	0.06*	0.04*	0.10*

Table A.1.1: Correlation Between X_t and e_t

The table depicts the correlation between the dependent variables X_t and the residuals ϵ. While several correlations were statistically significant, they were however mostly relatively small with the exception of *physical encounters*$_t$ and $\epsilon_{Phys.enc.}$ and *Peer entropy*$_t$ and $\epsilon_{Peer\,ent.}$. This might indicate the existence of a possible *social* confounding variable driving both dynamics. *: $p < 0.05$

A.1.2 Validation *OLS* model

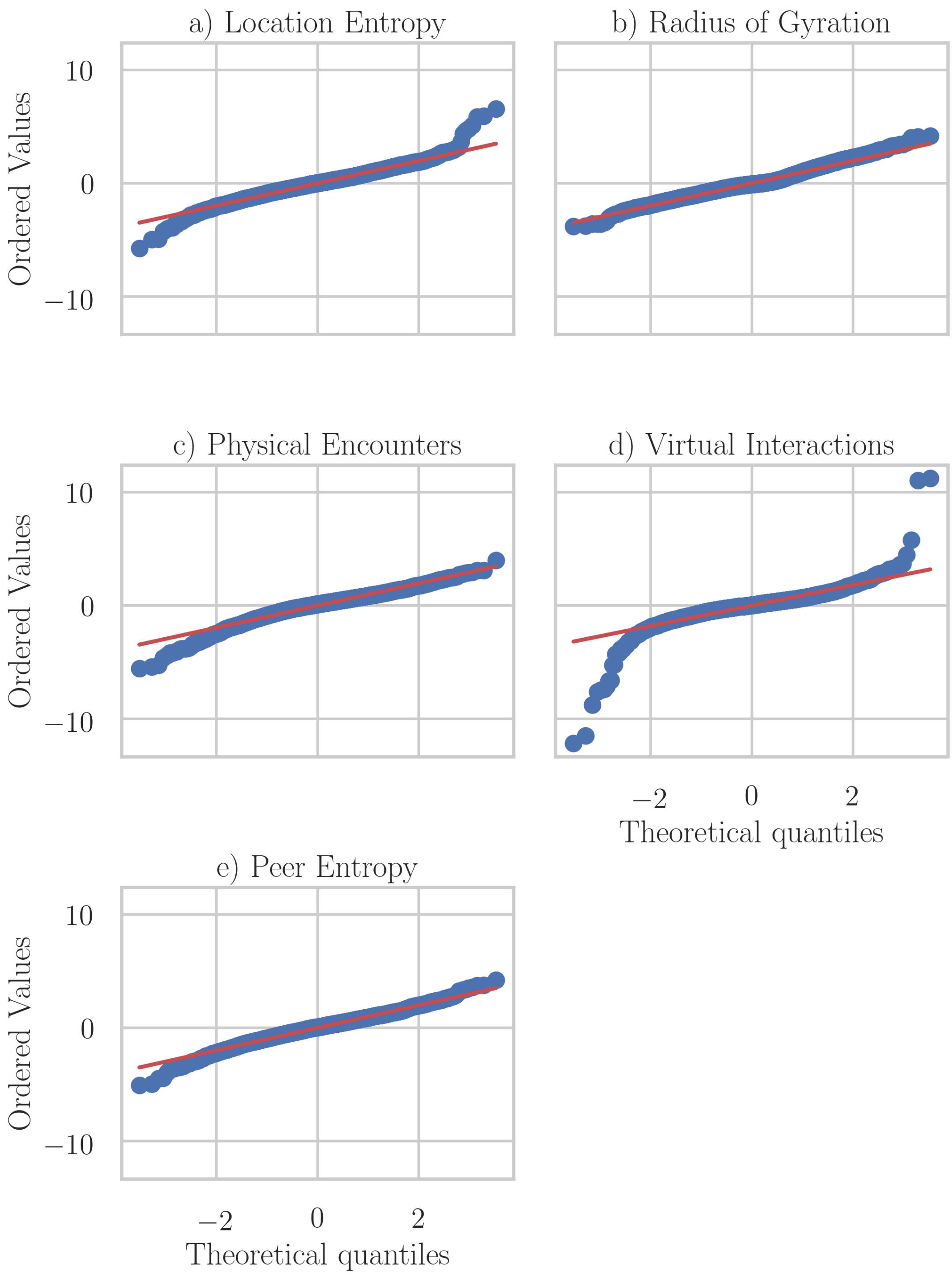

Figure A.1.1: Normal Q-Q Plot Residuals *OLS* Model

The plot depicts the residuals of the final *OLS* model. While the residuals were for all variables mostly normally distributed, for *virtual interactions* the smaller and larger residuals did not follow a normal distribution.

A.2 Personality & Regularity of Behaviour

This section shows the normal Q-Q plots of the all dependent ariables I used for assessing regularity in Chapter 6. While most variables were normally distributed, both the IEI and H of the social contexts as well as I were deviating substantially from a normal distribution.

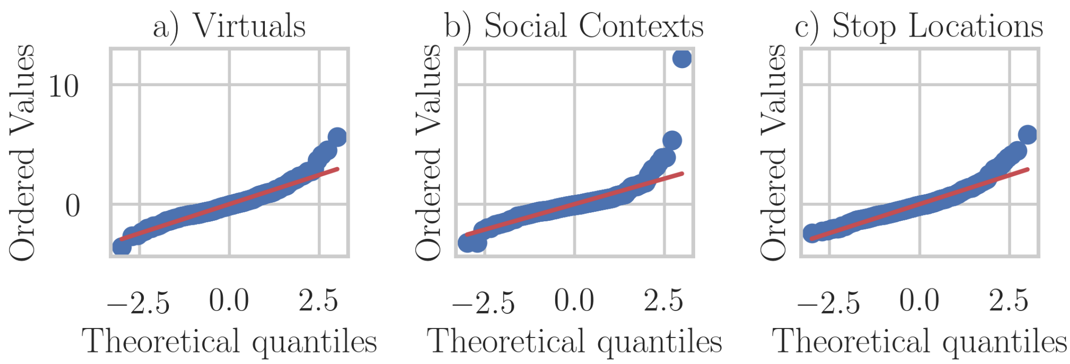

Figure A.2.1: Normal Q-Q plot of $\overline{c_v}$ of the Inter-Event Intervals

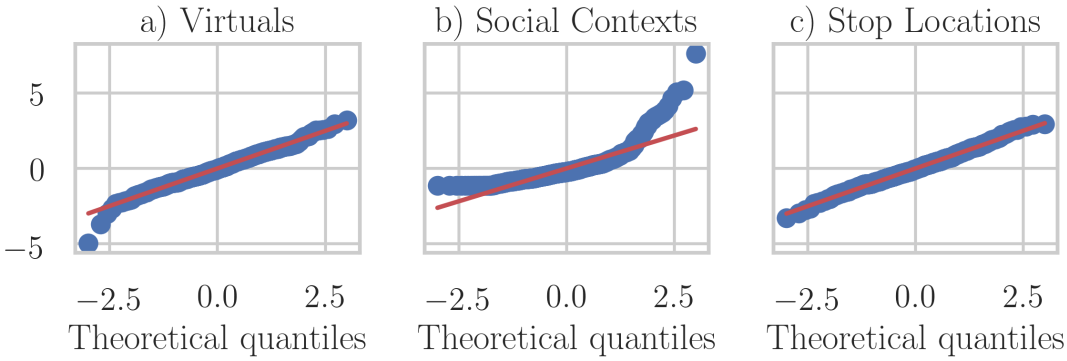

Figure A.2.2: Normal Q-Q plot of IEI

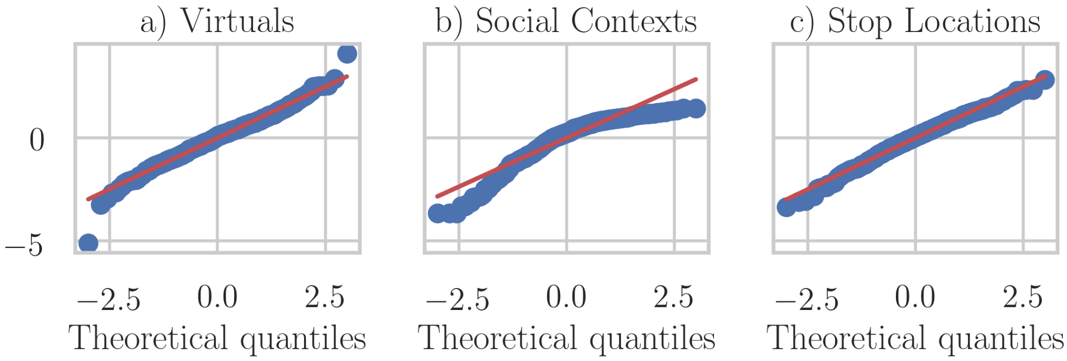

Figure A.2.3: Normal Q-Q plot of H

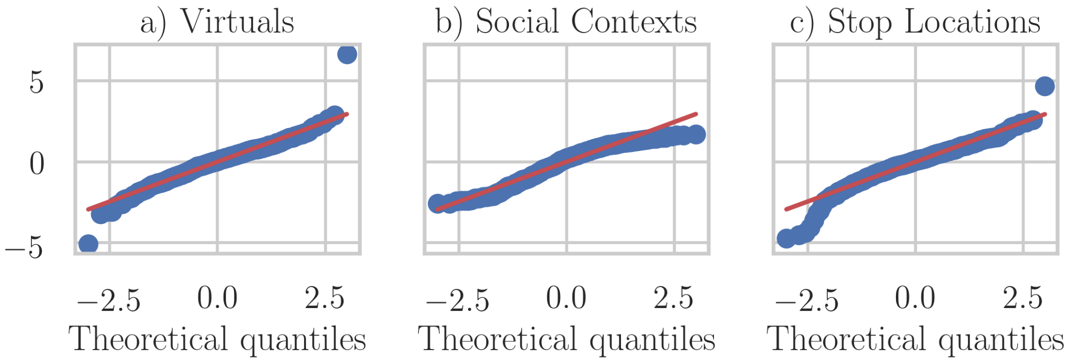

Figure A.2.4: Normal Q-Q plot of H_{LZ}

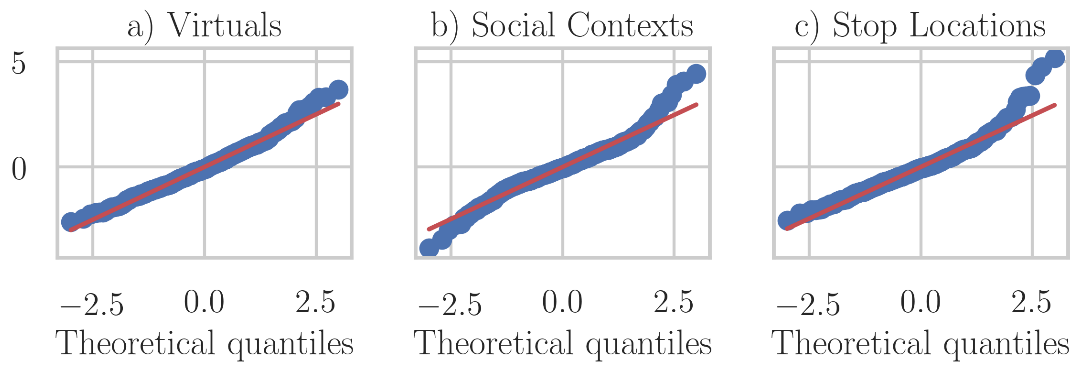

Figure A.2.5: Normal Q-Q plot of $\overline{c_v}$ of the Event Durations

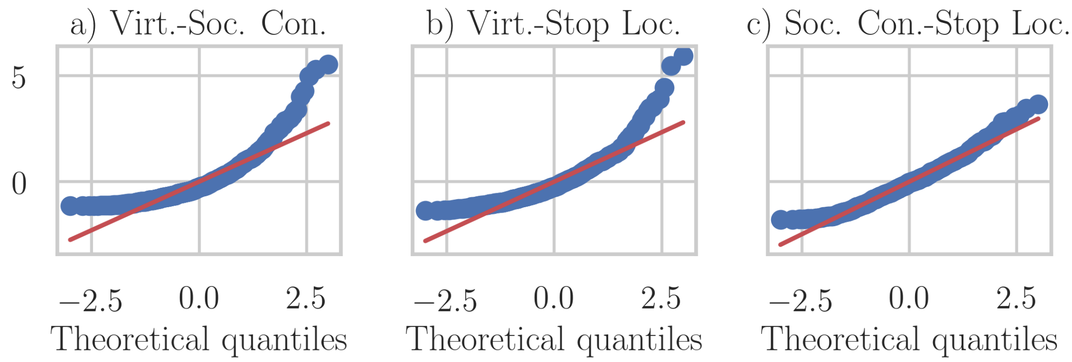

Figure A.2.6: Normal Q-Q plot of I

List of References

Abdi, H. (2010). Coefficient of variation. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 169–171). Thousand Oaks: Sage. (Cit. on p. 135).

Adamic, L. & Glance, N. (2005). The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. *LinkKDD '05 Proceedings of the 3rd international workshop on Link discovery*, 36–43. (Cit. on p. 10).

Adams, J., Faust, K., & Lovasi, G. S. (2012). Capturing context: Integrating spatial and social network analyses. *Social Networks*, *34*(1), 1–5. (Cit. on pp. 49, 66).

Agneessens, F., Waege, H., & Lievens, J. (2006). Diversity in social support by role relations: A typology. *Social Networks*, *28*(4), 427–441. (Cit. on p. 100).

Agosto, D. E. & Hughes-Hassell, S. (2005). People, places, and questions: An investigation of the everyday life information-seeking behaviors of urban young adults. *Library and Information Science Research*, *27*(2), 141–163. (Cit. on p. 100).

Agresti, A. & Barbara, F. (2009). *Statistical Methods for the Social Sciences* (4th). Upper Saddle River: Pearson. (Cit. on p. 64).

Ahas, R., Aasa, A., Silm, S., & Tiru, M. (2010). Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C*, *18*(1), 45–54. (Cit. on p. 40).

Ajrouch, K. J., Blandon, A. Y., & Antonucci, T. C. (2005). Social networks among men and women: The Effects of Age and Socioeconomic Status. *Journals of Gerontology Series: Social Sciences*, *60B*(6), S311–S317. (Cit. on pp. 18, 47).

Alessandretti, L. (2018). *Individual mobility in context: from high resolution trajectories to social behaviour* (Doctoral dissertation, City, University of London). (Cit. on pp. 3, 5, 6, 11, 18, 34–36, 39, 41, 42, 50, 53, 95, 96, 110, 124, 137, 147).

Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S., & Baronchelli, A. (2018). Evidence for a Conserved Quantity in Human Mobility. *Nature Human Behaviour*, *2*, 485–491. (Cit. on pp. 14, 40–42, 128, 129).

Ali, M. M., Amialchuk, A., & Rizzo, J. A. (2012). The influence of body weight on social network ties among adolescents. *Economics and Human Biology*, *10*(1), 20–34. (Cit. on pp. 18, 29, 48).

Allport, G. W. (1966). Traits revisited. *American Psychologist*, *21*(1), 1–10. (Cit. on p. 124).

Altmann, A., Tol Si, L., Sander, O., & Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics Original Paper*, *26*(10), 1340–1347. (Cit. on p. 71).

Amichai-Hamburger, Y. & Vinitzky, G. (2010). Social network use and personality. *Computers in Human Behavior*, *26*(6), 1289–1295. (Cit. on p. 125).

Amini, A., Kung, K., Kang, C., Sobolevsky, S., & Ratti, C. (2014). The Impact of Social Segregation on Human Mobility in Developing and Industrialized Regions. *EPJ Data Science*, *3*(6), 1–20. (Cit. on pp. 39, 51).

Ara, S. & Van Alstyne, M. (2011). The Diversity-Bandwidth Trade-off. *American Journal of Sociology*, *117*(1). (Cit. on p. 20).

Arnaboldi, V., Conti, M., Passarella, A., & Pezzoni, F. (2013). Ego networks in Twitter: An experimental analysis. *Proceedings - IEEE INFOCOM*, 3459–3464. (Cit. on pp. 18, 21, 22, 50).

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, *49*, 45–53. (Cit. on pp. 10, 13, 56).

Ashton, M. C. & Lee, K. (2008). The prediction of Honesty-Humility-related criteria by the HEXACO and Five-Factor Models of personality. *Research in Personality*, *42*, 1216–1228. (Cit. on p. 125).

Backlund, V. P., Saramaeki, J., & Pan, R. K. (2014). Effects of temporal correlations on cascades: Threshold models on temporal networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *89*(062815), 1–8. (Cit. on pp. 30, 31).

Backstrom, L., Sun, E., & Marlow, C. (2010). Find me if you can: improving geographical prediction with social and spatial proximity. *Proceedings of ICWWW'10*, 61–70. (Cit. on pp. 43, 66).

Bagrow, J. P. & Lin, Y.-R. (2012). Mesoscopic Structure and Social Aspects of Human Mobility. *PLoS ONE*, *7*(5), e37676. (Cit. on pp. 37, 40).

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *PNAS*, *106*(51). (Cit. on p. 10).

Balmaceda, J. M., Schiaffino, S., & Godoy, D. (2014). How do personality traits affect communication among users in online social networks? *Online Information Review*, *38*(1), 136–153. (Cit. on p. 126).

Bapierre, H., Jesdabodi, C., & Groh, G. (2015). Mobile Homophily and Social Location Prediction Measures of Social Cohesion. *arXiv*, 1–17. (Cit. on pp. 34, 42, 76).

Barabasi, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, *435*, 207–211. (Cit. on pp. 11, 30).

Barabasi, A.-L. (2012). Luck or reason. *Nature*, *489*, 8–9. (Cit. on pp. 23, 29).

Barabasi, A.-L. & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(5439), 509–513. (Cit. on pp. 3, 23).

Barbosa, H., Lima-Neto, F. B. D., Evsukoff, A., & Menezes, R. (2015). The effect of recency to human mobility. *EPJ Data Science*, *4*(21), 1–14. (Cit. on pp. 33, 37).

Baronchelli, A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in Cognitive Science. *Trends in Cognitive Sciences*, *17*(7), 348–360. (Cit. on pp. 34, 35).

Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *PNAS*, *101*(11). (Cit. on pp. 10, 16).

Bazzani, A., Giorgini, B., Rambaldi, S., Gallotti, R., & Giovannini, L. (2010). Statistical laws in urban mobility from microscopic GPS data in the area of Florence. *Journal of Statistical Mechanics: Theory and Experiment*, *2010*(5), P05001. (Cit. on pp. 34, 36).

Becker, R., Cáceres, R., Hanson, K., Isaacman, S., Loh, J., Martonosi, M., ... Volinsky, C. (2013). Human mobility characterization from cellular network data. *Communications of the ACM*, *56*(1), 74–82. (Cit. on pp. 33, 38).

Beiro, M. G., Panisson, A., Tizzoni, M., & Cattuto, C. (2016). Predicting human mobility through the assimilation of social media traces into mobility models. *EPJ Data Science*, *5*(30), 1–15. (Cit. on pp. 35, 45).

Belot, M. & Ermisch, J. (2009). Friendship Ties and Geographical Mobility: Evidence from the Great Britain. *Journal of the Royal Statistical Society. Series A*, *172*(2), 427–442. (Cit. on pp. 44, 45).

Berg, P. V. D., Arentze, T., & Timmermans, H. (2012). A multilevel path analysis of contact frequency between social network members. *Journal of Geographical Systems*, *14*. (Cit. on pp. 42, 96).

Berry, D. M. (2011). The computatonal turn: Thinking about the digital humanities. *Culture Machine*, *12*, 1–22. (Cit. on pp. 12, 13).

Bertrand, M. & Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*, *94*(4), 991–1013. (Cit. on p. 28).

Bianconi, G., Darst, R. K., Iacovacci, J., & Fortunato, S. (2014). Triadic closure as a basic generating mechanism of communities in complex networks. *Physical Review E*, *90*(042806), 1–10. (Cit. on pp. 24, 78, 88).

Bird, C., Gourley, A., Devanbu, P., Gertz, M., & Swaminathan, A. (2006). Mining Email Social Networks. *Proceedings of the 2006 International Workshop on Mining Software Repositories - MSR '06*, 137–143. (Cit. on p. 15).

Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, *4*(10), 1–55. (Cit. on p. 10).

Blonder, B., Wey, T. W., Dornhaus, A., James, R., & Sih, A. (2012). Temporal dynamics and network analysis. *Methods in Ecology and Evolution*, *3*(6), 958–972. (Cit. on p. 14).

Boccaletti, S., Bianconi, G., Criado, R., del Genio, C., Jomez-Gardenes, J., Romance, M., ... Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, *544*(1). (Cit. on p. 15).

Boessen, A., Hipp, J. R., Butts, C. T., Nagle, N. N., & Smith, E. J. (2017). The built environment, spatial scale, and social networks: Do land uses matter for personal network structure? *Environment and Planning B: Urban Analytics and City Science*, *45*(3), 1–17. (Cit. on pp. 18, 49, 66).

Bollen, J. (2011). Happiness Is Assortative in Online Social Networks. *Artificial Life*, *17*(3), 237–251. (Cit. on pp. 27, 28).

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*(7415), 295–298. (Cit. on pp. 27, 28).

Boos, D. D. (2003). Introduction to the Bootstrap World. *Statistical Science*, *18*(2), 168–174. (Cit. on p. 139).

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, *323*(April). (Cit. on pp. 2–4).

Boubtane, E., Coulibaly, D., & Rault, C. (2013). Immigration, unemployment and GDP in the host country: Bootstrap panel Granger causality analysis on OECD countries. *Economic Modelling*, *33*, 261–269. (Cit. on p. 154).

Bourdage, J. S., Lee, K., Ashton, M. C., & Perry, A. (2007). Big Five and HEXACO model personality correlates of sexuality. *Personality and Individual Differences*, *43*, 1506–1516. (Cit. on p. 123).

Bragazzi, N. L. & Del Puente, G. (2014). A proposal for including nomophobia in the new DSM-V. *Psychology Research and Behavior Management*, *7*, 155–160. (Cit. on p. 12).

Bramoullé, Y., Currarini, S., Jackson, M. O., Pin, P., & Rogers, B. W. (2012). Homophily and long-run integration in social networks. *Journal of Economic Theory*, *147*(5), 1754–1786. (Cit. on pp. 24, 30).

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. (Cit. on p. 72).

Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature*, *439*, 462–465. (Cit. on pp. 10, 35).

Brown, C., Noulas, A., Mascolo, C., & Blondel, V. (2013, September). A Place-Focused Model for Social Networks in Cities. *Proceedings of ICSC'13*, 75–80. (Cit. on pp. 49, 66, 95).

Bruederl, J. & Ludwig, V. (2015). Fixed-effects panel regression. In H. Best & C. Wolf (Eds.), *Regression analysis and causal inference* (pp. 327–357). Los Angeles. London. New Delhi, Singapore. Washington, DC: Sage. (Cit. on pp. 106, 113).

Burge, J., Lane, T., Link, H., Qiu, S., & Clark, V. P. (2009). Discrete dynamic bayesian network analysis of fMRI data. *Human Brain Mapping*, *30*(1), 122–137. (Cit. on p. 106).

Burke, M. & Kraut, R. E. (2014). Growing closer on facebook. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 4187–4196. (Cit. on pp. 10, 22, 43).

Burrows, R. & Savage, M. (2014). After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society*, *1*(1), 1–6. (Cit. on p. 13).

Busemeyer, J. R. & Townsend, J. T. (1993). Decision Field Theory: A Dynamic-Cognitive Approach to Decision Making in an Uncertain Environment. *Psychological Review*, *100*(3), 432–459. (Cit. on p. 31).

Butcher, J. N., Graham, J. R., & Ben-Porath, Y. S. (1995). Methodological Problems and Issues in MMPI, MMPI-2, and MMPI-A Research. *Psychological Assessment*, *7*(3), 320–329. (Cit. on pp. 140, 146).

Butts, C. T., Acton, R. M., Hipp, J. R., & Nagle, N. N. (2012). Geographical variability and network structure. *Social Networks*, *34*(1), 82–100. (Cit. on pp. 49, 66).

Cadwalladr, C. (2018, November). Parliament seizes cache of Facebook internal papers. (Cit. on p. 13).

Calabrese, F., Smoreda, Z., Blondel, V. D., & Ratti, C. (2011). Interplay between Telecommunications and Face-to-Face Interactions: A Study Using Mobile Phone Data. *PLoS ONE*, *6*(7), e20814. (Cit. on p. 42).

Calvo-Armegol, A. & Jackson, M. O. (2004). The Effects of Social Networks on Job Hunting. *American Economic Review*, *90*(3), 426–54. (Cit. on p. 23).

Campbell, K. E., Marsden, P. V., & Hurlbert, J. S. (1986). Social resources and socioeconomic status. *Social Networks*, *8*(1), 97–117. (Cit. on pp. 18, 47).

Campos, C. F., Heap, S. H., & de Leon, F. L. L. (2017). The political influence of peer groups: Experimental evidence in the classroom. *Oxford Economic Papers*, *69*(4), 963–985. (Cit. on p. 28).

Candia, J., González, M. C., Wang, P., Schoenharl, T., Madey, G., & Barabási, A. L. (2008). Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, *41*(224015), 1–11. (Cit. on pp. 16, 30).

Carpenter, D. P., Esterling, K. M., & Lazer, D. M. J. (2004). Friends, Brokers, and Transitivity: Who Informs Whom in Washington Politics? *The Journal of Politics*, *66*(1), 224–246. (Cit. on p. 10).

Carr, D., Jaffe, K. J., & Friedman, M. A. (2008). Perceived Interpersonal Mistreatment Among Obese Americans: Do Race, Class, and Gender Matter? *Obesity*, *16*(Suppl2), S60–S68. (Cit. on p. 48).

Carrasco, J. A. & Cid-Aguayo, B. (2012). Network capital, social networks, and travel: An empirical illustration from Concepción, Chile. *Environment and Planning A*, *44*(5), 1066–1084. (Cit. on pp. 95, 98).

Carrasco, J. A., Hogan, B., Wellman, B., & Miller, E. J. (2008). Agency in Social Acitivity Interactions: The Role of Social Networks in Time. *Tijdschrift voor Economische en Sociale Geografi*, *99*(5), 562–583. (Cit. on p. 42).

Carrasco, J., Miller, E., & Wellman, B. (2008). How Far and with Whom Do People Socialize?: Empirical Evidence About Distance Between Social Network Members. *Transportation Research Record: Journal of the Transportation Research Board*, *2076*, 114–122. (Cit. on pp. 44, 47, 51).

Carstensen, L. L. (1992). Social and emotional patterns in adulthood. *Psychology and Aging*, *7*(3), 331–338. (Cit. on p. 47).

Castillo-Manzano, J. I., Pozo-Barajas, R., & Trapero, J. R. (2015). Measuring the substitution effects between High Speed Rail and air transport in Spain. *Journal of Transport Geography*, *43*, 59–65. (Cit. on p. 96).

Cattell, V. (2001). Poor people, poor places, and poor health: The mediating role of social networks and social capital. *Social Science and Medicine*, *52*(10), 1501–1516. (Cit. on p. 23).

Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J.-F., & Vespignani, A. (2010, January). Dynamics of person-to-person interactions from distributed RFID sensor networks. *PloS one*, *5*(7), e11596. (Cit. on pp. 11, 15, 16, 19).

Centellegher, S., Lopez, E., Saramaeki, J., & Lepri, B. (2017). Personality traits and ego-network dynamics. *PLoS ONE*, *12*(3), e0173110. (Cit. on pp. 18, 19, 21, 22, 48, 50).

Centola, D. & van de Rijt, A. (2015). Choosing your network: Social preferences in an online health community. *Social Science and Medicine*, *125*, 19–31. (Cit. on pp. 27–29).

Ceruzzi, P. E. (2003). *A History of Modern Computing* (Second Edi). Cambridge, MA. London.: MIT Press. (Cit. on p. 5).

Chen, R., Resnick, S. M., Davatzikos, C., & Herskovits, E. H. (2012). Dynamic Bayesian network modeling for longitudinal brain morphometry. *NeuroImage*, *59*(3), 2330–2338. (Cit. on p. 106).

Cheng, Z., Caverlee, J., Lee, K., & Sui, D. Z. (2011). Exploring Millions of Footprints in Location Sharing Services. *The International Conference on Weblogs and Social Media*, 81–88. (Cit. on pp. 34, 35, 40).

Chipman, J. S. (2014). Gauss-Markov Theorem. (Cit. on p. 106).

Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility: user movement in location-based social networks. *KDD '11 Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1082–1090. (Cit. on pp. 6, 34, 40–42, 44, 45, 66, 74, 95, 96, 122).

Chorley, M. J., Whitaker, R. M., & Allen, S. M. (2015). Personality and location-based social networks. *Computers in Human Behavior*, *46*, 45–56. (Cit. on pp. 53, 124–127, 129, 140).

Christakis, N. A. & Fowler, J. H. (2007). The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine*, *357*(4), 370–379. (Cit. on pp. 27, 28).

Clark, A. F., Scott, D. M., & Yiannakoulias, N. (2013, May). Examining the relationship between active travel, weather, and the built environment: a multilevel approach using a GPS-enhanced dataset. *Transportation*, *41*(2), 325–338. (Cit. on pp. 39, 52, 53).

Clauset, A. & Eagle, N. (2007). Persistence and periodicity in a dynamic proximity network. *DIMACS Workshop on Com- putational Methods for Dynamic Interaction Networks*. (Cit. on pp. 7, 11, 33).

Clauset, A., Moore, C., & Newman, M. E. J. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, *453*(7191), 98–101. (Cit. on p. 25).

Cochran, M. M. & Brassard, J. A. (1979). Child Development and Personal Social Networks. *Society for Research in Child Development*, *50*(3), 601–616. (Cit. on p. 2).

Connor, M. C. O. & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, *43*, 971–990. (Cit. on p. 123).

Crandall, D. J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., & Kleinberg, J. (2010). Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(52), 22436–22441. (Cit. on pp. 44, 67).

Croissant, Y. & Millo, G. (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software*, *27*(2), 1–43. (Cit. on p. 113).

Csaji, B. C., Browet, A., Traag, V. A., Delvenne, J. C., Huens, E., Van Dooren, P., ... Blondel, V. D. (2013). Exploring the mobility of mobile phone users. *Physica A: Statistical Mechanics and its Applications*, *392*(6), 1459–1473. (Cit. on pp. 33, 37, 40, 41).

Currarini, S., Jackson, M. O., & Pin, P. (2010). Identifying the roles of race-based choice and chance in high school friendship network formation. *Proceedings of the National Academy of Sciences*, *107*(11), 4857–4861. (Cit. on pp. 28, 29).

Cuttone, A., Lehmann, S., & Larsen, J. E. (2014). Inferring human mobility from sparse low accuracy mobile sensing data. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp '14 Adjunct*, 995–1004. (Cit. on p. 58).

D'Ancona, M. (2018, July). With more elections looming, fake news must be tackled now. (Cit. on p. 11).

Davis, J. & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *ICML '06 Proceedings of the 23rd international conference on Machine learning*, 233–240. (Cit. on p. 85).

Dawid, H. & Gemkow, S. (2013). How Do Social Networks Contribute to Wage Inequality? Insights from an Agent-Based Analysis. *Industrial and Corporate Change*, *23*(5), 1171–1200. (Cit. on p. 23).

De Domenico, M., Lima, A., & Musolesi, M. (2013, December). Interdependence and predictability of human mobility and social interactions. *Pervasive and Mobile Computing*, *9*(6), 798–807. (Cit. on pp. 6, 10, 45, 66, 95, 96).

De Montis, A., Barthelemy, M., Chessa, A., & Vespignani, A. (2007). The structure of interurban traffic: a weighted network analysis. *Environment and Planning B: Planning and Design*, *34*, 905–924. (Cit. on p. 10).

De Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Scientific reports*, *3*(1376), 1–5. (Cit. on p. 34).

Decuyper, A., Rutherford, A., Wadhwa, A., Bauer, J.-M., Krings, G., Gutierrez, T., ... Luengo-Oroz, M. A. (2014). Estimating Food Consumption and Poverty Indices with Mobile Phone Data. *arXiv*. (Cit. on p. 47).

Di Ciommo, F., Comendador, J., López-Lambas, M. E., Cherchi, E., & Ortúzar, J. d. D. (2014). Exploring the role of social capital influence variables on travel behaviour. *Transportation Research Part A: Policy and Practice*, *68*, 46–55. (Cit. on p. 100).

Dill, J., Broach, J., Deutsch-Burgner, K., Xu, Y., Guensler, R., Levinson, D. M., & Tang, W. (2015). *Multiday GPS Travel Behavior Data for Travel Analysis*. Federal Highway Administration. (Cit. on p. 50).

DiMaggio, P. & Garip, F. (2011). How Network Externalities can exacerbate intergroup inequality. *American journal of sociology*, *116*(6), 1887–1933. (Cit. on p. 27).

Dong, W., Lepri, B., & Pentland, A. ((2011). Modeling the co-evolution of behaviors and social relationships using mobile phone data. *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia - MUM '11*, 134–143. (Cit. on p. 42).

Doreian, P. & Conti, N. (2012). Social context, spatial structure and social network structure. *Social Networks*, *34*(1), 32–46. (Cit. on pp. 25, 66).

Druckman, J. N. & Kam, C. D. (2009). Students as Experimental Participants: A Defense of the "Narrow Data Base". *SSRN*, 1–33. (Cit. on p. 152).

Duck, S. (1999). *Relating to Others* (2nd) (O. U. Press, Ed.). Buckingham. (Cit. on p. 31).

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, *16*(4), 681–735. (Cit. on pp. 2, 18, 21).

Dunbar, R. I. M. (1998). Theory of mind evolution language. In *Approaches to the evolution of language* (pp. 92–110). (Cit. on pp. 20, 21, 50).

Dunbar, R. I. M. & Spoors, M. (1995). Social networks, support cliques, and kinship. *Human Nature*, *6*(3), 273–290. (Cit. on pp. 18, 46).

Dunbar, R. I. (2009). The social brain hypothesis and its implications for social evolution. *Annals of Human Biology*, *36*(5), 562–572. (Cit. on pp. 2, 20, 21).

Eagle, N., Macy, M., & Claxton, R. (2010). Network diversity and economic development. *Science*, *328*(5981), 1029–1031. (Cit. on p. 47).

Eagle, N. & Pentland, A. (2006). Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing*, *10*(4), 255–268. (Cit. on pp. 10, 11, 15, 34).

Eagle, N. & Pentland, A. S. (2009, April). Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, *63*(7), 1057–1066. (Cit. on pp. 11, 44, 45, 74).

Efron, B. (2005). *Local False Discovery Rate*. Stanford University. (Cit. on p. 158).

Eichler, M. (2012). Causal Inference in Time Series Analysis. In *Causality: statistical perspectives and applications* (pp. 327–352). Hoboken, NJ: Wiley. (Cit. on p. 105).

Ellegard, K. & Svedin, U. (2012). Torsten Haegerstrand's time-geography as the cradle of the activity approach in transport geography. *Journal of Transport Geography*, *23*, 17–25. (Cit. on p. 5).

Elman, C. & O'Rand, A. M. (2007). The effects of social origins, life events, and institutional sorting on adults' school transitions. *Social Science Research*, *36*(3), 1276–1299. (Cit. on p. 28).

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). Density-Based Clustering Methods. *Proceedings of KDD'96*, *2*, 226–231. (Cit. on p. 75).

Etminani-Ghasrodashti, R. & Ardeshiri, M. (2015). Modeling travel behavior by the structural relationships between lifestyle, built environment and non-working trips. *Transportation Research Part A: Policy and Practice*, *78*, 506–518. (Cit. on p. 51).

Ewing, R. & Cervero, R. (2001). Travel and the Built Enviornment: A Synthesis. *Transportation Research Record*, *1780*(1). (Cit. on p. 52).

Ewing, R. & Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, *76*(3), 265–294. (Cit. on p. 52).

Fehr, E., Fischbacher, U., & Gaechter, S. (2002). Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms. *Human Nature*, *13*(1), 1–25. (Cit. on p. 31).

Feld, S. L. (1981). The Focused Organization of Social Ties. *American Journal of Sociology*, *86*(5), 1015–1035. (Cit. on pp. 3, 24, 25, 28).

Feld, S. L. (1982). Social Structural Determinants Of Similarity Among Associates. American Sociological Review. *New York*, *47*(6), 797–801. (Cit. on p. 28).

Finneran, L. & Kelly, M. (2003). Social networks and inequality. *Journal of Urban Economics*, *53*(2), 282–299. (Cit. on p. 23).

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3-5), 75–174. (Cit. on p. 26).

Fortunato, S. & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, *659*, 1–44. (Cit. on p. 26).

Freeman, L. (2001). The Effects of Sprawl on Neighborhood Social Ties. *APA*, *67*(1), 69–77. (Cit. on p. 62).

Freeman, L. (2011). The Development of Social Network Analysis - with an Emphasis on Recent Events. In J. Scott & P. J. Carrington (Eds.), *The sage handbook of social network analysis* (pp. 26–39). Los Angeles. London. New Delhi, Singapore. Washington, DC: SAGE. (Cit. on pp. 2–4).

Frei, A. & Axhausen, K. W. (2007). *Size and structure of social network geographies*. IVT. ETH. (Cit. on pp. 95, 98).

Frias-Martinez, V., Soto, V., Virseda, J., & Frias-Martinez, E. (2013). Can Cell Phone Traces Measure Social Development? *Third Conference on the Analysis of Mobile Phone Datasets, NetMob*. (Cit. on p. 51).

Frias-Martinez, V. & Virseda, J. (2012). On the relationship between socio-economic factors and cell phone usage. *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development - ICTD '12*, 76–84. (Cit. on p. 51).

Friggeri, A., Inria, E. N. S. D. L., Lambiotte, R., Kosinski, M., Fleury, E., & Inria, E. N. S. D. L. (2012). Psychological Aspects of Social Communities. *Proceedings International Conference on Social Computing/International Conference on Privacy, Security, Risk and Trust*, 195–202. (Cit. on pp. 123, 132).

Fyhri, A., Hjorthol, R., Mackett, R. L., Fotel, T. N., & Kyttä, M. (2011). Children's active travel and independent mobility in four countries: Development, social contributing trends and measures. *Transport Policy*, *18*(5), 703–710. (Cit. on p. 51).

Gee, L. K., Jones, J., & Burke, M. (2017). Social Networks and Labor Markets: How Strong Ties Relate to Job Finding on Facebook's Social Network. *Journal of Labor Economics*, *35*(2), 485–518. (Cit. on p. 20).

Ghoshal, G. & Holme, P. (2006). Attractiveness and activity in Internet communities. *Physica A: Statistical Mechanics and its Applications*, *364*, 603–609. (Cit. on pp. 16, 19).

Girvan, M. & Newman, M. E. J. (2002). Community structure in social and biological networks. *99*(12), 7821–7826. (Cit. on p. 26).

Glover, T. D. & Parry, D. C. (2009). A third place in the everyday lives of people living with cancer: Functions of Gilda's Club of Greater Toronto. *Health and Place*, *15*(1), 97–106. (Cit. on p. 74).

Goh, K.-I. & Barabasi, A.-L. (2008). Burstiness and memory in complex systems Burstiness and memory in complex systems. *EPL*, *81*(48002), 1–5. (Cit. on pp. 30, 31).

Goldberg, L. R., John, O. P., Kaiser, H., Lanning, K., & Peabody, D. (1990). An Alternative "Description of Personality": The Big-Five Factor Structure. *59*(6). (Cit. on p. 56).

Goncalves, B., Perra, N., & Vespignani, A. (2011). Modeling users' activity on twitter networks: Validation of Dunbar's number. *PLoS ONE*, *6*(8), e22656. (Cit. on pp. 10, 20).

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008a). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782. (Cit. on pp. 10, 34, 122, 124, 128).

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008b). Understanding individual human mobility patterns Supplementary Material. *Nature*, *453*. (Cit. on p. 64).

González-Bailõn, S. (2013). Social science in the era of big data. *Policy and Internet*, *5*(2), 147–160. (Cit. on p. 12).

Grabowicz, P. A., Ramasco, J. J., Gonçalves, B., & Eguíluz, V. M. (2014). Entangling mobility and interactions in social media. *PLoS ONE*, *9*(3), e92196. (Cit. on pp. 10, 42, 45, 96).

Granger, C. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3). (Cit. on pp. 95, 104).

Granovetter, M. (1973). The Strength of Weak Ties. (Cit. on pp. 2, 3, 19, 20, 22, 101).

Green, J., Jones, A., & Roberts, H. (2014). More than A to B: the role of free bus travel for the mobility and wellbeing of older citizens in London. *Ageing and society*, *34*(3), 472–494. (Cit. on p. 98).

Greenwood, M. J. & Hunt, G. L. (2003). The Early History of Migration Research. *International Regional Science Review*, *26*(1), 3–37. (Cit. on p. 4).

Hägerstrand, T. (1970). What about People in Regional Science? *Papers in Regional Science*, *24*(1), 7–21. (Cit. on pp. 4, 5, 33, 65, 92).

Hakim, M. M. & Merkert, R. (2016). The causal relationship between air transport and economic growth: Empirical evidence from South Asia. *Journal of Transport Geography*, *56*, 120–127. (Cit. on p. 96).

Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D*, *10*, 427–444. (Cit. on p. 52).

Hansen, C., Wei, Q., Shieh, J.-s., Fourcade, P., & Isableu, B. (2017). Sample Entropy, Univariate, and Multivariate Multi-Scale Entropy in Comparison with Classical Postural Sway Parameters in Young Healthy Adults. *Frontiers in Human Neuroscience*, *11*(206), 1–10. (Cit. on p. 139).

Hasan, S., Schneider, C. M., Ukkusuri, S. V., & González, M. C. (2013). Spatiotemporal Patterns of Urban Human Mobility. *Journal of Statistical Physics*, *151*(1-2), 304–318. (Cit. on pp. 34, 41).

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014, February). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, *41*(3), 260–271. (Cit. on pp. 10, 35).

Hays, R. B. & Oxley, D. (1986). Social Network Development and Functioning During a Life Transition. *Journal of Personality and Social Psychology*, *50*(2), 305–313. (Cit. on p. 46).

Helliwell, J. F. & Putnam, R. D. (2004). The social context of well-being. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1449), 1435–1446. (Cit. on p. 20).

Hill, R. & Dunbar, R. I. M. (2003). Social network size in humans. *Human Nature*, *14*(1). (Cit. on pp. 18, 21).

Ho, T. K. (2002). A Data Complexity Analysis of Comparative Advantages of Decision Forest Constructors. *Pattern Analysis & Applications*, *5*(2), 102–112. (Cit. on p. 71).

Holme, P. (2003). Network dynamics of ongoing social relationships. *Europhysics Letters*, *64*(3), 427–433. (Cit. on pp. 27, 30).

Holme, P. (2015). Modern temporal network theory: a colloquium. *Eur. Phys. J. B*, *88*(234), 1–30. (Cit. on pp. 12, 14, 49).

Holme, P. & Liljeros, F. (2014). Birth and death of links control disease spreading in empirical contact networks. *Scientific Reports*, *4*(4999), 1–8. (Cit. on p. 15).

Holme, P. & Saramaki, J. (2012). Temporal networks. *Physics Reports*, *519*(3), 97–125. (Cit. on pp. 14, 49).

Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social relationships and mortality risk: A meta-analytic review. *PLoS Medicine*, *7*(7), e1000316. (Cit. on p. 20).

Hong, J., Shen, Q., & Zhang, L. (2014). How do built-environment factors affect travel behavior? A spatial analysis at different geographic scales. *Transportation*, *41*(3), 419–440. (Cit. on p. 52).

Hossmann, T., Spyropoulos, T., & Legendre, F. (2011). A complex network analysis of human mobility. *2011 IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS 2011*, 876–881. (Cit. on pp. 18, 25, 27).

House, T., Read, J. M., Danon, L., & Keeling, M. J. (2015). Testing the hypothesis of preferential attachment in social network formation. *EPJ Data Science*, *4*(1), 1–13. (Cit. on p. 23).

Huang, L., Zhu, X., Ye, X., Guo, W., & Wang, J. (2016). Characterizing street hierarchies through network analysis and large-scale taxi traffic flow: a case study of Wuhan , China. *Environment and Planning B: Planning and Design*, *43*(2). (Cit. on p. 38).

Huang, Y., Shen, C., & Contractor, N. S. (2013). Distance matters: Exploring proximity and homophily in virtual world networks. *Decision Support Systems*, *55*(4), 969–977. (Cit. on pp. 27, 42).

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., & Newman, T. B. (2013). *Designing Clinical Research* (Fourth Edi) (R. Gaertner, Ed.). Philadelphia: Lippincott Williams & Wilkins. (Cit. on p. 141).

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, *9*(3), 90–95. (Cit. on p. 5).

Hyndman, R. J. & Athanasopoulos, G. (2013). *Forecasting: principles and practice*. Melbourne: OTexts. (Cit. on p. 79).

Igarashi, T., Takai, J., & Yoshida, T. (2005). Gender differences in social network development via mobile phone text messages: A longitudinal study. *Journal of Social and Personal Relationships*, *22*(5), 691–713. (Cit. on pp. 18, 46).

Iribarren, J. L. & Moro, E. (2009). Impact of human activity patterns on the dynamics of information diffusion. *Physical Review Letters*, *103*(038702), 1–4. (Cit. on pp. 15, 19).

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., & Varshavsky, A. (2011). Identifying important places in people's lives from cellular network data. In K. Lyons, J. Hightower, & E. M. Huan (Eds.), *Pervasive computing* (pp. 133–151). Springer. (Cit. on pp. 37, 39, 53).

Isella, L., Stehle, J., Barrat, A., Cattuto, C., Pinton, J.-F., & Van den Broeck, W. (2011, March). What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, *271*(1), 166–80. (Cit. on pp. 11, 67).

Jeong, H., Néda, Z., & Barabási, A. L. (2003). Measuring preferential attachment in evolving networks. *Europhysics Letters*, *61*(4), 567–572. (Cit. on p. 23).

Jiang, B. (2007). A topological pattern of urban street networks : Universality and peculiarity. *Physica A*, *384*. (Cit. on p. 38).

Jiang, B., Yin, J., & Zhao, S. (2009, August). Characterizing the human mobility pattern in a large street network. *Physical Review E*, *80*(021136), 1–11. (Cit. on pp. 34, 38, 52).

Jiron, P. (2007). Unravelling invisible inequalities in the city through urban daily mobility. The case of Santiago de Chile. *Swiss Journal of Sociology*, *33*(1), 45–68. (Cit. on pp. 99, 120).

Jo, H.-h., Karsai, M., Kertesz, J., & Kaski, K. (2012). Circadian pattern and burstiness in mobile phone communication. *New Journal of Physics*, *14*, 1–17. (Cit. on p. 33).

Johansen, A. (2004). Probing human response times. *Physica A: Statistical Mechanics and its Applications*, *338*(1-2), 286–291. (Cit. on pp. 10, 30).

Jones, E., Oliphant, T., & Peterson, P. (2001). {SciPy}: Open source scientific tools for {Python}. (Cit. on p. 5).

Jurdak, R., Zhao, K., Liu, J., Jaoude, A. M., & Cameron, M. (2015). Understanding Human Mobility from Twitter. *PloS one*, *10*(7), e0131469. (Cit. on pp. 34, 35, 37).

Kalish, Y. & Robins, G. (2006). Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure. *Social Networks*, *28*(1), 56–84. (Cit. on pp. 18, 48).

Kanai, R., Bahrami, B., Roylance, R., & Rees, G. (2011). Online social network size is reflected in human brain structure. *Proceedings of the Royal Society B*, *279*(1732), 1–8. (Cit. on pp. 15, 129).

Kang, C., Gao, S., Lin, X., Xiao, Y., Yuan, Y., Liu, Y., & Ma, X. (2010). Analyzing and geo-visualizing individual human mobility patterns using mobile call records. *2010 18th International Conference on Geoinformatics, Geoinformatics 2010*, 1–7. (Cit. on pp. 50, 51).

Karsai, M., Iniguez, G., Kaski, K., & Kertész, J. (2014). Complex contagion process in spreading of online innovation. *Journal of the Royal Society Interface*, *11*(20140694), 1–8. (Cit. on pp. 15, 20, 28).

Karsai, M., Kaski, K., Barabasi, A. L., & Kertesz, J. (2012). Universal features of correlated bursty behaviour. *Scientific Reports*, *2*(397), 1–7. (Cit. on pp. 15, 30, 31).

Karsai, M., Kivelä, M., Pan, R. K., Kaski, K., Kertész, J., Barabási, A. L., & Saramäki, J. (2011). Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *83*(025102(R)), 1–4. (Cit. on pp. 27, 30, 32).

Karsai, M., Perra, N., & Vespignani, A. (2014). Time varying networks and the weakness of strong ties. *Scientific Reports*, *4*(4001), 1–7. (Cit. on pp. 16, 19, 20, 22, 30).

Kaufmann, V., Bergman, M. M., & Joye, D. (2004). Motility: Mobility as Capital. *International Journal of Urban and Regional Research*, *28*(4), 745–756. (Cit. on p. 102).

Kawachi, I. & Berkman, L. F. (2001). Social Ties and Mental Health. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, *78*(3), 458–467. (Cit. on p. 20).

Kelly, R. L. (1983). Hunter-Gatherer Mobility Strategies. *Journal of Anthropological Research*, *39*(3), 277–306. (Cit. on p. 4).

Kim, K. & Altmann, J. (2017). Effect of homophily on network formation. *Communications in Nonlinear Science and Numerical Simulation*, *44*, 1339–1351. (Cit. on pp. 16, 30).

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, *3*, 262–267. (Cit. on pp. 12, 13).

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data and Society*, *1*(1), 1–12. (Cit. on p. 12).

Klimek, P. & Thurner, S. (2013). Triadic closure dynamics drives scaling laws in social multiplex networks. *New Journal of Physics*, *15*(063008), 1–9. (Cit. on pp. 23, 24).

Klinger, T. & Lanzendorf, M. (2016). Moving between mobility cultures: what affects the travel behavior of new residents? *Transportation*, *43*(2), 243–271. (Cit. on pp. 51, 52).

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *PNAS*, *110*(15), 2–5. (Cit. on pp. 123, 129).

Kossinets, G. & Watts, D. J. (2006). Empirical Analysis of an Evolving Social Network. *Science*, *311*(5757), 88–90. (Cit. on pp. 10, 15, 21, 24, 25).

Kossinets, G. & Watts, D. J. (2009). Origins of Homophily in an Evolving Social Network. *American Journal of Sociology*, *115*(2), 405–450. (Cit. on pp. 15, 24, 25, 29).

Kovanen, L., Karsai, M., Kaski, K., Kertesz, J., & Saramaki, J. (2011). Temporal motifs in time-dependent networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2011*(11), P11005. (Cit. on pp. 15, 32).

Kovanen, L., Kaski, K., Kertesz, J., & Saramaki, J. (2013). Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. *PNAS*, *110*(45), 18070–18075. (Cit. on pp. 15, 32).

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790. (Cit. on pp. 15, 28).

Krause, C. M. & Zhang, L. (2018). Short-term travel behavior prediction with GPS, land use, and point of interest data. *Transportation Research Part B: Methodological*, *forthcomin*, 1–13. (Cit. on pp. 34, 53).

Krings, G., Calabrese, F., Ratti, C., & Blondel, V. D. (2009). Urban Gravity: a Model for Intercity Telecommunication Flows. *Journal of Statistical Mechanics: Theory and Experiment*, *7*, L07003. (Cit. on pp. 33, 43).

Krings, G., Karsai, M., Bernhardsson, S., Blondel, V. D., & Saramäki, J. (2012). Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science*, *1*(4), 1–16. (Cit. on pp. 14, 33, 110).

Kumpula, J. M., Onnela, J., Saramäki, J., Kertész, J., & Kaski, K. (2009). Model of community emergence in weighted social networks. *Computer Physics Communications*, *180*(4), 517–522. (Cit. on pp. 20, 22, 25).

Kumpula, J. M., Onnela, J. P., Saramäki, J., Kaski, K., & Kertész, J. (2007). Emergence of communities in weighted networks. *Physical Review Letters*, *99*(228701), 1–4. (Cit. on pp. 20, 24, 25).

Kwan, M.-P. (2000). Gender differences in space-time constraints. *Area*, *32*(2), 145–156. (Cit. on p. 50).

Lambiotte, B. R. & Kosinski, M. (2014). Tracking the Digital Footprints of Personality. *Proceedings of the IEEE*, *102*(12), 1934–1939. (Cit. on pp. 123, 125, 126).

Lambiotte, R., Blondel, V. D., de Kerchove, C., Huens, E., Prieur, C., Smoreda, Z., & Van Dooren, P. (2008). Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, *387*(21), 5317–5325. (Cit. on pp. 43, 44, 66).

Larsen, J., Urry, J., & Axhausen, K. (2006). *Mobiliites, Networks, Geographies*. Hampshire. Burlington: Ashgate. (Cit. on pp. 6, 43, 44, 66).

Latham, P. E. & Roudi, Y. (2009). Mutual information. *Scholarpedia*, *4*(1), 1658. (Cit. on p. 62).

Laurent, G., Saramäki, J., & Karsai, M. (2015). From calls to communities: a model for time-varying social networks. *European Physical Journal B*, *88*(11), 1–10. (Cit. on pp. 24, 25, 30).

Lazer, D., Brewer, D., Christakis, N., Fowler, J., & King, G. (2009). Life in the network: the coming age of computational social. *Science*, *323*(5915), 721–723. (Cit. on pp. 6, 11–13, 122).

Lee, K., Ogunfowora, B., & Ashton, M. C. (2005). Personality Traits Beyond the Big Five: Are They Within the HEXACO Space? *Journal of Personality*, *73*(5), 1437–1463. (Cit. on p. 125).

Lenormand, M., Louail, T., Cantu-Ros, O. G., Picornell, M., Herranz, R., Arias, J. M., . . . Ramasco, J. J. (2015). Influence of sociodemographic characteristics on human mobility. *Nature Publishing Group*, *5*(10075), 1–14. (Cit. on pp. 34, 50, 51).

Levin, D. Z. & Cross, R. (2004). The Strength of Weak Ties You Can Trust: The Mediating Role of Trust in Effective Knowledge Transfer. *Management Science*, *50*(11), 1477–1490. (Cit. on pp. 19, 20, 22).

Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., & Tomkins, A. (2005). Geographic routing in social networks. *PNAS*, *102*(33), 11623–11628. (Cit. on p. 43).

Lima, A. (2016). *Digital Traces of Human Mobility and Interaction: Models and Application* (Doctoral dissertation, University of Birmingham, UK). (Cit. on pp. 15, 16).

Lima, A., Stanojevic, R., Papagiannaki, D., Rodriguez, P., González, M. C., Antonio, L., . . . Gonzalez, M. C. (2016). Understanding individual routing behaviour. *Journal of The Royal Society Interface*, *13*(116), 20160021. (Cit. on pp. 4, 21, 34, 36, 40, 128).

Liu, Y., Sui, Z., Kang, C., & Gao, Y. (2014). Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS ONE*, *9*(1), e86026. (Cit. on pp. 10, 38).

Lorrain, F. & White, H. C. (1971). Structural equivalence of individuals in social networks. *The Journal of Mathematical Sociology*, *1*(1), 49–80. (Cit. on p. 2).

Lovejoy, K. & Handy, S. (2011). Social networks as a source of private-vehicle transportation: The practice of getting rides and borrowing vehicles among Mexican immigrants in California. *Transportation Research Part A: Policy and Practice*, *45*(4), 248–257. (Cit. on p. 100).

Lu, X., Wetter, E., Bharti, N., Tatem, A. J., & Bengtsson, L. (2013). Approaching the Limit of Predictability in Human Mobility. *Scientific reports*, *3*(2923), 1–9. (Cit. on pp. 35, 40, 139).

Lu, Y. & Zhang, L. (2015). Imputing trip purposes for long-distance travel. *Transportation*, *42*(4), 581–595. (Cit. on p. 53).

Lu, Y. E., Roberts, S., Lio, P., Dunbar, R. I. M., & Crowcroft, J. (2009). Size matters: Variation in personal network size, personality and effect on information transmission. *Proceedings - 12th IEEE International Conference on Computational Science and Engineering, CSE 2009*, *4*(217141), 188–193. (Cit. on pp. 15, 18).

Luetkepohl, H. (2005). *New introduction to multiple time series analysis* (3 ed.). Berlin: Springer. (Cit. on pp. 103, 105, 106, 154).

Luo, F., Cao, G., Mulligan, K., & Li, X. (2016). Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography*, *70*, 11–25. (Cit. on pp. 34, 52).

Luo, S., Morone, F., Sarraute, C., Travizano, M., & Makse, H. A. (2017). Inferring personal economic status from social network location. *Nature Communications*, *8*(15227), 1–7. (Cit. on p. 47).

Maddala, G. S. & Wu, S. (1999). A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test. *Oxford Bulletin of Economics and Statistics*, *61*(s1), 631–652. (Cit. on p. 113).

Mair, H. (2009). Club Life: Third Place and Shared Leisure in Rural Canada. *Leisure Sciences*, *31*(5), 450–465. (Cit. on p. 74).

Mao, H., Shuai, X., Ahn, Y. Y., & Bollen, J. (2015). Quantifying socio-economic indicators in developing countries from mobile phone communication data: applications to Côte d'Ivoire. *EPJ Data Science*, *4*(1), 1–16. (Cit. on p. 47).

Marchetti, C. (1994). Anthropological Invariants in Travel Behavior. *Technological Forecasting and Social Change*, *47*(1), 75–88. (Cit. on p. 4).

Marchetti, S., Giusti, C., Pratesi, M., Salvati, N., Giannotti, F., Pedreschi, D., ... Gabrielli, L. (2015). Small area model-based estimators using big data sources. *Journal of Official Statistics*, *31*(2), 263–281. (Cit. on p. 51).

Marsden, P. V. (1988). Homogeneity in Confiding Relations. *Social Networks*, *10*, 57–76. (Cit. on p. 27).

Marsden, P. V. (1990). Network Data And Measurement. *Annual Review of Sociology*, *16*, 435–463. (Cit. on p. 11).

Mattisson, K., Håkansson, C., & Jakobsson, K. (2014). Relationships Between Commuting and Social Capital Among Men and Women in Southern Sweden. *Environment and Behavior*, *47*(7), 734–753. (Cit. on pp. 99, 120).

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th python in science conference* (pp. 51–56). (Cit. on p. 5).

McPherson, J. M. & Smith-Lovin, L. (1987). Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups. *American Sociological Review*, *52*(3), 370. (Cit. on pp. 27, 28).

McPherson, M., Smith-lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*(2001), 415–444. (Cit. on pp. 27, 30).

Meinshausen, N. & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society*, *72*(4), 417–473. (Cit. on pp. 106, 155, 157, 158).

Meo, P. D., Ferrara, E., Fiumara, G., & Provetti, A. (2014). On Facebook, most ties are weak. *Communications of the ACM*, *57*(11), 78–84. (Cit. on pp. 15, 20, 22, 25).

Metcalf, S. S. (2013). Modeling Social Ties and Household Mobility Modeling Social Ties and Household Mobility. *Annals of the American Association of Geographers*, *104*(1), 40–59. (Cit. on p. 45).

Mezard, M. & Montanari, A. (2009). *Information, Physics, and Computation*. New York: Oxford University Press. (Cit. on p. 61).

Milgram, S. (1967). The Small-World Problem. *Psychology Toda*, *1*(1), 61–67. (Cit. on pp. 26, 27).

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chlovskii, D., & Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science*, *298*(5594), 824–827. (Cit. on p. 25).

Min, B., Goh, K.-I., & Vazquez, A. (2013). Spreading dynamics following bursty activity patterns. *Understanding Complex Systems*, *036102*, 161–174. (Cit. on pp. 16, 31).

Miritello, G., Lara, R., Cebrian, M., & Moro, E. (2013). Limited communication capacity unveils strategies for human interaction. *Scientific Reports*, *3*(1950), 1–7. (Cit. on pp. 15, 16, 20–22, 39, 47, 50).

Miritello, G., Moro, E., & Lara, R. (2011). Dynamical strength of social ties in information spreading. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *83*(045102(R)), 1–4. (Cit. on p. 32).

Moinet, A., Starnini, M., & Pastor-Satorras, R. (2015). Burstiness and Aging in Social Temporal Networks. *Physical Review Letters*, *114*(108701), 1–5. (Cit. on pp. 19, 30).

Mok, D., Wellman, B., & Carrasco, J. (2010). Does Distance Matter in the Age of the Internet? *Urban Studies*, *47*(13), 2747–2783. (Cit. on pp. 43, 44).

Montjoye, Y.-A., Quoidbach, J., Robic, F., & Pentland, A. (2013). Predicting Personality Using Novel Phone-Based Metrics. In *Social computing, behavioral-cultural modeling and prediction* (pp. 48–55). Dordrecht, London, New York: Springer. (Cit. on pp. 48, 53, 123, 125, 127, 129).

Moody, J. (2001). Race, School Integration, and Friendship Segregation in America. *American Journal of Sociology*, *107*(3), 679–716. (Cit. on p. 28).

Moore, G. (1990). Structural Determinants of Men's and Women's Personal Networks. *American Sociological Review*, *55*(5), 726–735. (Cit. on p. 46).

Morelli, S. A., Ong, D. C., Makati, R., Jackson, M. O., & Zaki, J. (2017). Empathy and well-being correlate with centrality in different social networks. *Proceedings of the National Academy of Sciences*, *114*(37), 9843–9847. (Cit. on pp. 123, 129).

Muchnik, L., Pei, S., Parra, L. C., Reis, S. D. S., Andrade, J. S., Havlin, S., & Makse, H. A. (2013). Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Scientific reports*, *3*(1783), 1–7. (Cit. on p. 19).

Murakami, E. & Jennifer, Y. (1997). Daily Travel by Persons with Low Income Daily Travel by Persons with Low Income. *African American Mobility Symposium*, 1–27. (Cit. on p. 51).

Nagarajan, R., Scutari, M., & Lebre, S. (2013). *Bayesian Networks in R.* (Cit. on pp. 106, 112, 152, 157).

Narayan, P. K. & Smyth, R. (2005). Electricity consumption, employment and real income in Australia evidence from multivariate Granger causality tests. *Energy Policy*, *33*(9), 1109–1116. (Cit. on p. 105).

Narayan, P. K. & Smyth, R. (2009). Multivariate granger causality between electricity consumption, exports and GDP: Evidence from a panel of Middle Eastern countries. *Energy Policy*, *37*(1), 229–236. (Cit. on p. 105).

Neutens, T., Schwanen, T., & Witlox, F. (2011). The Prism of Everyday Life: Towards a New Research Agenda for Time Geography. *Transport Reviews : A Transnational Transdisciplinary Journal*, *31*(1), 25–47. (Cit. on p. 5).

Neutens, T., Witlox, F., & Demaeyer, P. (2007). Individual accessibility and travel literature review on time geography possibilities : A. *European Journal of Transport and Infrastructure Research*, *7*(4), 335–352. (Cit. on p. 5).

Newman, M. E. J. (2008). Mathematics of Networks. In P. Macmillan (Ed.), *The new palgrave dictionary of economics* (2nd, pp. 1–8). London: Palgrave Macmillan. (Cit. on p. 23).

Newman, M. E. (2001). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *64*(016131), 1–8. (Cit. on pp. 20, 23, 25).

Newman, M. E. (2003). Properties of highly clustered networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *68*(026121), 1–6. (Cit. on pp. 24, 25).

Newman & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(026113), 1–16. (Cit. on p. 82).

Newsom, J. T. (2015). *Longitudinal Structural Equation Modeling*. New York, NY: Routledge. (Cit. on p. 155).

Nguyen, T. & Szymanski, B. K. (2012). Using Location-Based Social Networks to Validate Human Mobility and Relationships Models. *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 1215–1221. (Cit. on pp. 15, 43, 44, 95).

Nieves, J. & Osorio, J. (2013). The role of social networks in knowledge creation. *Knowledge Management Research & Practice*, *11*(1), 62–77. (Cit. on p. 99).

Noe, N., Whitaker, R. M., Chorley, M. J., & Pollet, T. V. (2016). Birds of a feather locate together? Foursquare checkins and personality homophily. *Computers in Human Behavior*, *58*, 343–353. (Cit. on pp. 53, 123, 125–127).

Norris, C. J., Larsen, J. T., & Cacioppo, J. T. (2007). Neuroticism is associated with larger and more prolonged electrodermal responses to emotionally evocative pictures. *Psychophysiology*, *44*(5), 823–826. (Cit. on p. 127).

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012, January). A tale of many cities: universal patterns in human urban mobility. *PloS one*, *7*(5), e37027. (Cit. on pp. 10, 34, 36, 38, 39, 43, 52).

Noulas, A., Shaw, B., Lambiotte, R., & Mascolo, C. (2015). Topological Properties and Temporal Dynamics of Place Networks in Urban Environments. *WWW '15 Companion Proceedings of the 24th International Conference on World Wide Web*, 431–441. (Cit. on pp. 66, 67).

Oldenburg, R. & Brissett, D. (1982). The Third Place. *Qualitative Sociology*, *5*(4), 265–284. (Cit. on p. 74).

Oliveira, R. D. (2011). Towards a Psychographic User Model From Mobile Phone Usage. *ACM CHI*, *WIP*, 2191–2196. (Cit. on pp. 48, 123).

Olson, M. A. & Wyner, A. J. (2018). Making Sense of Random Forest Probabilities: a Kernel Perspective. *arXiv*, 1–35. (Cit. on p. 72).

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., ... Barabási, A.-L. (2007). Structure and tie strengths in mobile communication networks. *PNAS*, *104*(18), 7332–7336. (Cit. on pp. 10, 16, 19, 20, 22).

Opgen-Rhein, R. & Strimmer, K. (2007). Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process. *BMC Bioinformatics*, *8*(Suppl 2), S3. (Cit. on pp. 106, 158).

Oviedo Hernandez, D. & Titheridge, H. (2015). Mobilities of the periphery: Informality, access and social exclusion in the urban fringe in Colombia. *Journal of Transport Geography*, *55*. (Cit. on p. 100).

Palla, G., Barabási, A. L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature*, *446*(7136), 664–667. (Cit. on p. 26).

Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, *435*(7043), 814–818. (Cit. on pp. 25, 26).

Pan, X., Hou, L., & Liu, K. (2017). Social influence on selection behaviour: Distinguishing local- and global-driven preferential attachment. *PloS one*, *12*(4), e0175761. (Cit. on p. 28).

Pappalardo, L. (2016). The origin of heterogeneity in human mobility ranges. *CEUR Workshop Proceedings*, *1558*, 3–6. (Cit. on p. 37).

Pappalardo, L., Pedreschi, D., Smoreda, Z., & Giannotti, F. (2015). Using big data to study the link between human mobility and socio-economic development. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, 871–878. (Cit. on pp. 39, 51).

Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature Communications*, *6*(8166), 1–8. (Cit. on pp. 37, 39).

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... Seligman, M. E. P. (2015). Automatic Personality Assessment Through Social Media Language. *Journal of Personality and Social Psychology*, *108*(6), 934–952. (Cit. on p. 123).

Pedregosa, F. & Varoquaux, G. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830. (Cit. on pp. 5, 59, 71, 75).

Peng, T. Q., Liu, M., Wu, Y., & Liu, S. (2016). Follower-Followee Network, Communication Networks, and Vote Agreement of the U.S. Members of Congress. *Communication Research*, *43*(7), 996–1024. (Cit. on p. 28).

Peng, W., Baowen, X. U., Yurong, W. U., & Xiaoyu, Z. (2015). Link Prediction in Social Networks : the State-of-the-Art. *SCIENCE CHINA Information Sciences*, *58*(011101), 1–38. (Cit. on pp. 68, 71, 78).

Perez, F. & Granger, B. E. (2007). IPython: A System for Interactive Scientific Computing. *Computing in Science & Engineering*, *9*(3), 21–29. (Cit. on p. 5).

Perkins, J. M., Subramanian, S. V., & Christakis, N. A. (2015). Social networks and health: A systematic review of sociocentric network studies in low- and middle-income countries. *Social Science and Medicine*, *125*, 60–78. (Cit. on pp. 2, 23).

Perra, N., Gonçalves, B., Pastor-Satorras, R., & Vespignani, A. (2012). Activity driven modeling of time varying networks. *Scientific Reports*, *2*(469), 1–7. (Cit. on pp. 15, 16, 19).

Pham, T., Sheridan, P., & Shimodaira, H. (2015). PAFit: A statistical method for measuring preferential attachment in temporal complex networks. *PLoS ONE*, *10*(9), 1–18. (Cit. on p. 23).

Phithakkitnukoon, S., Smoreda, Z., & Olivier, P. (2012). Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data. *PLoS ONE*, *7*(6), e39253. (Cit. on pp. 38, 42, 43).

Picornell, M., Ruiz, T., Lenormand, M., Ramasco, J. J., Dubernet, T., & Frías-Martínez, E. (2015). Exploring the potential of phone call data to characterize the relationship between social network and travel behavior. *Transportation*, *42*(4), 647–668. (Cit. on pp. 44, 53).

Pollet, T. V., Roberts, S. G. B., & Dunbar, R. I. M. (2011). Extraverts Have Larger Social Network Layers. *Journal of Individual Differences*, *32*(3), 161–169. (Cit. on pp. 18, 48, 126).

Powell, J., Lewis, P. A., Roberts, N., García-Fiñana, M., & Dunbar, R. I. M. (2012). Orbital prefrontal cortex volume predicts social network size: An imaging study of individual differences in humans. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1736), 2157–2162. (Cit. on p. 21).

Preciado, P., Snijders, T. a. B., Burk, W. J., Stattin, H., & Kerr, M. (2012). Does proximity matter? Distance dependence of adolescent friendships. *Social Networks*, *34*(1), 18–31. (Cit. on pp. 43, 44, 97).

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific reports*, *3*(1684), 1–6. (Cit. on p. 10).

Prendergast, L. S. & Williams, R. D. (1981). Individual travel time budgets. *Transportation Research Part A: General*, *15*(1), 39–46. (Cit. on p. 51).

Prinzie, A. & Van Den Poel, D. (2011). Modeling complex longitudinal consumer behavior with Dynamic Bayesian networks: An Acquisition Pattern Analysis application. *Journal of Intelligent Information Systems*, *36*(3), 283–304. (Cit. on p. 106).

Putnam, R. D. (2000). *Bowling Alone: The collapse and revival of American community*. New York. London. Toronto. Sydney. Singapore: Simon and Schuster. (Cit. on pp. 19, 22, 99, 120).

Putzke, J., Gloor, P. a., Fischbach, K., & Schoder, D. (2010). The Evolution of Interaction Networks in Massively Multiplayer Online Games. *Jais*, *11*(2), 69–94. (Cit. on p. 27).

Puura, A., Silm, S., & Ahas, R. (2017). The Relationship between Social Networks and Spatial Mobility: A Mobile-Phone-Based Study in Estonia. *Journal of Urban Technology*, *25*(2), 7–25. (Cit. on pp. 42, 43, 51, 52).

Quercia, D., Bodaghi, M., & Crowcroft, J. (2012). Loosing "Friends" on Facebook. *Proceedings of Web Science*, 251–254. (Cit. on pp. 48, 123, 126, 129).

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna. (Cit. on p. 5).

Rabari, C. & Storper, M. (2015). The digital skin of cities: Urban theory and research in the age of the sensored and metered city, ubiquitous computing and big data. *Cambridge Journal of Regions, Economy and Society*, *8*(1), 27–42. (Cit. on pp. 12, 13).

Radicchi, F., Baronchelli, A., & Amaral, L. A. (2012). Rationality, irrationality and escalating behavior in lowest unique bid auctions. *PLoS ONE*, *7*(1), 5–12. (Cit. on p. 10).

Rahmadi, R., Groot, P., Hc Van Rijn, M., Ajg Van Den Brand, J., Heins, M., Knoop, H., & Heskes, T. (2018). Causality on Longitudinal Data: Stable Specification Search in Constrained Structural Equation Modeling. *Statistical Methods in Medical Research*, *27*(12), 3814–3834. (Cit. on pp. 106, 155).

Rahmadi, R., Groot, P., Heins, M., Knoop, H., & Heskes, T. (2017). Causality on cross-sectional data: Stable specification search in constrained structural equation modeling. *Applied Soft Computing Journal*, *52*, 687–698. (Cit. on pp. 106, 155).

Ranjan, G., Zang, H., Zhang, Z.-L., & Bolot, J. (2012, December). Are call detail records biased for sampling human mobility? *ACM SIGMOBILE Mobile Computing and Communications Review*, *16*(3), 33. (Cit. on pp. 41, 76).

Ravasz, E. & Barabási, A. L. (2003). Hierarchical organization in complex networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *67*(026112), 1–7. (Cit. on p. 25).

Ravenstein, E. G. (1885). The Laws of Migration. *Journal of the Statistical Society of London*, *48*(2), 167–235. (Cit. on p. 4).

Reades, J., Calabrese, F., & Ratti, C. (2009). Eigenplaces: analysing cities using the space-time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, *36*, 824–836. (Cit. on p. 10).

Reis, H. T., Nezlek, J., & Wheeler, L. (1980). Physical attractiveness in social interaction. *Journal of Personality and Social Psychology*, *38*(4), 604–617. (Cit. on pp. 18, 48).

Reis, H. T. & Wheeler, L. (1991). Studying Social interaction. *Advances in Experimental Social Psychology*, *24*, 269–318. (Cit. on p. 69).

Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S. J., & Chong, S. (2011). On the levy-walk nature of human mobility. *IEEE/ACM Transactions on Networking*, *19*(3), 630–643. (Cit. on pp. 34–36).

Roberts, S. G. B., Dunbar, R. I. M., Pollet, T. V., & Kuppens, T. (2009). Exploring variation in active network size: Constraints and ego characteristics. *Social Networks*, *31*. (Cit. on pp. 16, 21).

Rocha, L. E. C. & Blondel, V. D. (2013). Bursts of Vertex Activation and Epidemics in Evolving Networks. *PLOS Computational Biology*, *9*(3), e1002974. (Cit. on pp. 15, 16, 31).

Rocha, L. E. C., Liljeros, F., & Holme, P. (2011). Simulated Epidemics in an Empirical Spatiotemporal Network of 50,185 Sexual Contacts. *PLOS Computational Biology*, *7*(3), e1001109. (Cit. on p. 31).

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the Mechanics of Information Diffusion Across Topics: Idioms , Political Hashtags , and Complex

Contagion on Twitter. *Proceedings of the 20th International Conference on World Wide Web WWW'11 (ACM,* 695–704. (Cit. on pp. 15, 28).
Rosel, J. & Plewis, I. (2008). Longitudinal data analysis with structural equations. *Methodology*, *4*(1), 37–50. (Cit. on p. 155).
Rosenbaum, M. S., Ward, J., Walker, B. A., & Ostrom, A. L. (2007). A Cup of Coffee With a Dash of Love. *Journal of Service Research*, *10*(1), 43–59. (Cit. on p. 74).
Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G., & Orr, R. R. (2009). Personality and motivations associated with Facebook use. *Computers in Human Behavior*, *25*(2), 578–586. (Cit. on p. 140).
Roth, C., Kang, S. M., Batty, M., & Barthélemy, M. (2011, January). Structure of urban movements: polycentric activity and entangled hierarchical flows. *PloS one*, *6*(1), e15923. (Cit. on pp. 34, 37).
Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The Structure of Founding Teams: Homophily, Strong Ties, and Isolation among U.S. Entrepreneurs. *American Sociological Review*, *68*(2), 195. (Cit. on p. 28).
Sadilek, A. & Krumm, J. (2012). Far Out: Predicting Long-Term Human Mobility. *Proceeedings of the 26th AAAI Conference on Artificial Intelligence*, 814–820. (Cit. on pp. 10, 40).
Salon, D. & Gulyani, S. (2010). Mobility, poverty, and gender: Travel 'choices' of slum residents in Nairobi, Kenya. *Transport Reviews*, *30*(5), 641–657. (Cit. on p. 50).
Sander, J., Schupp, J., & Richter, D. (2017). Getting together: Social contact frequency across the life span. *Developmental Psychology*, *53*(8), 1571–1588. (Cit. on pp. 18, 47).
Sapiezynski, P., Stopczynski, A., Gatej, R., & Lehmann, S. (2015). Tracking human mobility using WiFi signals. *PLoS ONE*, *10*(7), e0130824. (Cit. on pp. 4, 59).
Saramaki, J., Leicht, E. A., Lopez, E., Roberts, S. G. B., Reed-Tsochas, F., & Dunbar, R. I. M. (2014). Persistence of social signatures in human communication. *Proceedings of the National Academy of Sciences*, *111*(3), 942–947. (Cit. on pp. 21, 22).
Saramäki, J. & Moro, E. (2015). From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. *European Physical Journal B*, *88*(164), 1–10. (Cit. on pp. 19, 20, 31, 33).
Savage, M. & Burrows, R. (2007). The Coming Crisis of Empirical Sociology. *Sociology*, *41*(5), 885–899. (Cit. on p. 13).
Scafetta, N. & Grigolini, P. (2002). Scaling detection in time series: Diffusion entropy analysis. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *66*(036130), 1–10. (Cit. on p. 137).
Scellato, S., Musolesi, M., Mascolo, C., & Latora, V. (2010). On nonstationarity of human contact networks. *Proceedings - International Conference on Distributed Computing Systems*, 105–111. (Cit. on pp. 33, 119, 128).
Scellato, S., Noulas, A., Lambiotte, R., & Mascolo, C. (2011). Socio-spatial Properties of Online Location-based Social Networks. *Proceedings ofthe 5th international AAAI conference on weblogs and social media - ICWSM'11*, 329–336. (Cit. on pp. 42, 66, 80).

Scellato, S., Noulas, A., & Mascolo, C. (2011). Exploiting Place Features in Link Prediction on Location-based Social Networks Categories and Subject Descriptors. *Proceedings of KDD'11*, 1046–1054. (Cit. on pp. 6, 7, 49, 66, 67, 74, 92, 95).

Schelling, T. C. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology*, *1*. (Cit. on p. 5).

Schlich, R., Schoenfelder, S., Axhausen, K. W., & Hanson, S. (2004). Structures of leisure travel: Temporal and spatial variability. *Transport Reviews*, *24*(2), 219–237. (Cit. on p. 41).

Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., González, M. C., Interface, J. R. S., & Couronne, T. (2013). Unravelling daily human mobility motifs. *Journal of the Royal Society Interface*, *10*(20130246), 1–8. (Cit. on pp. 10, 36, 40, 41).

Schönfelder, S. & Axhausen, K. W. (2003). Activity spaces: Measures of social exclusion? *Transport Policy*, *10*(4), 273–286. (Cit. on p. 40).

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Transactions on Database Systems*, *42*(3), 19. (Cit. on p. 60).

Schwanen, T., Lucas, K., Akyelken, N., Cisternas Solsona, D., Carrasco, J. A., & Neutens, T. (2015). Rethinking the links between social exclusion and transport disadvantage through the lens of social capital. *Transportation Research Part A: Policy and Practice*, *74*, 123–135. (Cit. on pp. 97–101).

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., . . . Ungar, L. H. (2013). Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLoS ONE*, *8*(9), e73791. (Cit. on p. 123).

Scott, J. & Carrington, P. J. (Eds.). (2011). *The SAGE handbook of social network analysis*. (Cit. on p. 26).

Seabold, S. & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *9th python in science conference* (pp. 57–61). (Cit. on p. 5).

Seaborn, C., Attanucci, J., & Wilson, N. H. M. (2009). Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data. *Transportation Research Record: Journal of the Transportation Research Board*, *2121*(1), 55–62. (Cit. on pp. 34, 35).

Seibert, S. E., Kraimer, M. L., & Liden, R. C. (2001). a Social Capital Theory of Career Success. *The Academy of Management Journal*, *44*(2), 219–237. (Cit. on p. 102).

Seidel, M.-D. L., Polzer, J. T., & Stewart, K. J. (2000). Friends in High Places: The Effects of Social Networks on Discrimination in Salary Negotiations. *Administrative Science Quarterly*, *45*(1), 1–24. (Cit. on p. 47).

Sekara, V. & Lehmann, S. (2014). The strength of friendship ties in proximity sensor data. *PLoS ONE*, *9*(7), e100915. (Cit. on pp. 58, 68).

Sekara, V., Stopczynski, A., & Lehmann, S. (2016). The fundamental structures of dynamic social networks. *PNAS*, *113*(36), 9977–9982. (Cit. on pp. 4, 3, 7, 11, 14, 26, 33, 50, 59, 67, 69, 75, 92, 110, 129, 131).

Selfhout, M., Burk, W., Branje, S., Denissen, J., van Aken, M., & Meeus, W. (2010). Emerging Late Adolescent Friendship Networks and Big Five Personality Traits: A

Social Network Approach. *Journal of Personality*, *78*(2), 509–538. (Cit. on pp. 48, 126).

Seth, A. (2007). Granger Causality. *Scholarpedia*, *2*(7), 1667. (Cit. on pp. 104, 113).

Sevtsuk, A. & Ratti, C. (2010). Does Urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *Journal of Urban Technology*, *17*(1), 41–60. (Cit. on pp. 33, 40, 41).

Shaw, B. A., Krause, N., Liang, J., & Bennett, J. (2007). Tracking changes in social relations throughout late life. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *62*(2), S90–S99. (Cit. on p. 47).

Shi, L., Wu, L., Chi, G., & Liu, Y. (2016). Geographical impacts on social networks from perspectives of space and place: an empirical study using mobile phone data. *Journal of Geographical Systems*, *18*(4), 359–376. (Cit. on pp. 42–44, 96).

Silm, S. & Ahas, R. (2014). Ethnic Differences in Activity Spaces: A Study of Out-of-Home Nonemployment Activities with Mobile Phone Data. *Annals of the American Association of Geographers*, *104*(3). (Cit. on pp. 39, 52).

Silvis, J. & Niemeier, D. (2009). Social Network and Dwelling Characteristics That Influence Ridesharing Behavior of Seniors. *Transportation Research Record: Journal of the Transportation Research Board*, *2118*, 47–54. (Cit. on p. 100).

Simmel, G. (1969). The Metropolis and Mental Life. In *Classic essays on the culture of cities* (pp. 47–61). Englewood Cliffs: Prentice-Hall. (Cit. on p. 4).

Skopek, J., Schulz, F., & Blossfeld, H. P. (2011). Who contacts whom? Educational homophily in online mate selection. *European Sociological Review*, *27*(2), 180–195. (Cit. on pp. 27, 29).

Small, M. L. (2013). Weak ties and the core discussion network: Why people regularly discuss important matters with unimportant alters. *Social Networks*, *35*(3), 470–483. (Cit. on p. 22).

Sobolevsky, S., Sitko, I., Des Combes, R. T., Hawelka, B., Arias, J. M., & Ratti, C. (2014). Money on the move: Big data of bank card transactions as the new proxy for human mobility patterns and regional delineation. The case of residents and foreign visitors in Spain. *Proceedings - 2014 IEEE International Congress on Big Data, BigData Congress 2014*, 136–143. (Cit. on pp. 10, 34).

Song, C., Koren, T., Wang, P., & Barabasi, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature Physics*, *6*(10), 818–823. (Cit. on pp. 34, 36, 37, 40).

Song, C., Qu, Z., Blumm, N., & Barabasi, A.-L. (2010, February). Limits of predictability in human mobility. *Science*, *327*(5968), 1018–1021. (Cit. on pp. 7, 10, 14, 40, 61, 122, 124, 128, 129, 137, 139, 147).

Song, C., Wang, D., & Barabasi, A.-L. (2012). Connections between Human Dynamics and Network Science. *arXiv*. (Cit. on pp. 18, 20).

Soto, V., Frias-Martinez, V., Virseda, J., & Frias-Martinez, E. (2011). Prediction of socioeconomic levels using cell phone records. *Proceedings of UAMP 2011: User Modelling, Adaption and Personlization*, 377–388. (Cit. on p. 51).

Staiano, J., Pianesi, F., Lepri, B., Sebe, N., Aharony, N., & Pentland, A. (2012). Friends Don't Lie - Inferring Personality Traits from Social Network Structure. *Proceedings* *of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, 321–330. (Cit. on pp. 18, 123).

Stanley, J., Hensher, D. A., Stanley, J., Currie, G., William, H., Vella-brodrick, D., ... Vella-brodrick, D. (2011). Social Exclusion and the Value of Mobility Social Exclusion and the Value of Mobility. *Journal of Transport Economics and Policy*, *45*(2), 197–222. (Cit. on pp. 7, 98).

Starnini, M., Baronchelli, A., & Pastor-Satorras, R. (2013). Modeling human dynamics of face-to-face interaction networks. *Physical Review Letters*, *110*(168701), 1–5. (Cit. on pp. 11, 16, 19).

Stehle, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.-F., ... Vanhems, P. (2011, August). High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. *PLoS ONE*, *6*(8), e23176. (Cit. on pp. 16, 67).

Stiller, J. & Dunbar, R. I. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, *29*(1), 93–104. (Cit. on p. 21).

Stokes, P. A. & Purdon, P. L. (2017). A study of problems encountered in Granger causality analysis from a neuroscience perspective. *PNAS*, *114*(34), E7063–E7072. (Cit. on p. 105).

Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M. M., Larsen, J. E., & Lehmann, S. (2014, January). Measuring large-scale social networks with high resolution. *PloS ONE*, *9*(4), e95978. (Cit. on pp. 11, 15, 34, 56).

Stouffer, S. A. (1940). Intervening Opportunities: A Theory Relating Mobility and Distance. *American Sociological Review*, *5*(6), 845–867. (Cit. on pp. 36, 38).

Stumpf, M. P. H. & Porter, M. A. (2012). Critical truth about power laws. *Science*, *335*(6069), 665–667. (Cit. on p. 18).

Sun, L., Axhausen, K. W., Lee, D.-h., & Huang, X. (2013). Understanding metropolitan patterns of daily encounters. *PNAS*, *34*(110), 13774–13779. (Cit. on pp. 34, 41, 44).

Swap, W. C. (1977). Interpersonal Attraction and Repeated Exposure to Rewarders and Punishers. *Personality and Social Psychology Bulletin, 3*, 248–251. (Cit. on p. 28).

Tacken, M. (1998). Mobility of the elderly in time and space in the Netherlands: An Analysis of the Dutch National Travel Survey. *Transportation*, *25*, 379–393. (Cit. on p. 51).

Takano, M. & Fukuda, I. (2017). Limitations of Time Resources in Human Relationships Determine Social Structures. *Palgrave Communications*, *3*(17014), 1–9. (Cit. on pp. 20–22).

Takhteyev, Y., Gruzd, A., & Wellman, B. (2012). Geography of Twitter Networks. *Social Networks*, *34*(1), 73–81. (Cit. on pp. 10, 15, 42, 43).

Tang, L., Lv, H., Yang, F., & Yu, L. (2015). Complexity testing techniques for time series data: A comprehensive literature review. *Chaos, Solitons & Fractals*, *81*, 117–135. (Cit. on p. 137).

Tizzoni, M., Sun, K., Benusiglio, D., Karsai, M., & Perra, N. (2015). The scaling of human contacts and epidemic processes in metapopulation networks. *Scientific Reports*, *5*(15111), 1–11. (Cit. on pp. 15, 18, 19).

Toole, J. L., Herrera-yaqu, C., Schneider, C. M., Gonza, M. C., Herrera-Yague, C., Schneider, C. M., & Gonz, M. C. (2015). Coupling human mobility and social ties. *Journal of the Royal Society, Interface*, *12*(20141128), 1–9. (Cit. on pp. 42–45, 47, 66, 76).

Tselios, V. (2014). The Granger-causality between income and educational inequality: a spatial cross-regressive VAR framework. *Annals of Regional Science*, *53*(1), 221–243. (Cit. on p. 105).

Tuhin, P., Kevin, S., Nathaniel, O., Scott, B., & Nazeem, M. (2016). Scaling Behavior of Human Mobility Distributions. In J. A. Miller, D. O. Sullivan, & N. W. Eds (Eds.), *Geographic information science* (pp. 145–162). (Cit. on pp. 35, 36).

Ubaldi, E., Perra, N., Karsai, M., Vezzani, A., Burioni, R., & Vespignani, A. (2016). Asymptotic theory for the dynamic of networks with heterogenous activity and tie allocation. *Scientific reports*, *6*(35724), 1–10. (Cit. on pp. 16, 19, 30).

Urry, J. (2002). Mobility and Proximity. *Sociology*, *36*(2), 255–274. (Cit. on pp. 94, 97).

Utsunomiya, K. (2016). Social capital and local public transportation in Japan. *Research in Transportation Economics*, *59*, 434–440. (Cit. on p. 99).

van den Berg, P., Arentze, T., & Timmermans, H. (2009). Size and Composition of Ego-Centered Social Networks and Their Effect on Geographic Distance and Contact Frequency. *Transportation Research Record: Journal of the Transportation Research Board*, *2135*, 1–9. (Cit. on pp. 41, 43, 46).

van den Berg, P., Kemperman, A., & Timmermans, H. (2014). Social Interaction Location Choice : A Latent Class Modeling Approach Social Interaction Location Choice : A Latent Class Modeling Approach. *Annals of the American Association of Geographers*, *104*(5), 959–972. (Cit. on pp. 41, 43).

van der Walt, S. & Aivazis, M. (2011). The NumPy Array: A Structure for Efficient Numerical Computation, Computing in Science & Engineering. *Computing in Science and Engineering*, *13*(2), 22–30. (Cit. on p. 5).

van Rossum, G. (1995). *Python tutorial, Technical Report CS-R9526*. Centrum voor Wiskunde en Informatica (CWI). Amsterdam. (Cit. on p. 5).

Vemuri, V. (1978). *Modeling of Complex Systems*. New York. San Francisco. London: Academic Press. (Cit. on p. 5).

Verbrugge, L. M. (1977). Structure of Adult Friendship Choices. *Social Forces*, *56*(2), 576–597. (Cit. on p. 44).

Vigier, A. (2014). Meeting friends of friends and homophily: a complementarity. *Econ Theory Bull*, *2*, 45–52. (Cit. on pp. 27, 30).

Viry, G., Kaufmann, V., & Widmer, E. D. (2009). Social integration faced with commuting: More widespread and less dense networks. In T. Ohnmacht, H. Makskim, & M. Bergmann (Eds.), *Mobilities and inequality* (pp. 121–144). Ashgate. (Cit. on pp. 7, 95, 98, 99, 120).

Vlahovic, T. A., Roberts, S., & Dunbar, R. I. M. (2012). Effects of duration and laughter on subjective happiness within different modes of communication. *Journal of Computer-Mediated Communication*, *17*(4), 436–450. (Cit. on pp. 22, 43).

Volkovich, Y., Scellato, S., Laniado, D., Mascolo, C., & Kaltenbrunner, A. (2012). The Length of Bridge Ties: Structural and Geographic Properties of Online Social Interactions. *Proceedings of the 6th international AAAI conference on weblogs and social media - ICWSM 912*, 346–353. (Cit. on pp. 10, 43, 44).

Wan, N., Qu, W., Whittington, J., Witbrodt, B. C., Henderson, M. P., Goulding, E. H., … Lin, G. (2013). Assessing smart phones for generating life-space indicators. *Environment and Planning B: Planning and Design*, *40*. (Cit. on p. 33).

Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabási, A.-l., & Science, C. (2011). Human Mobility, Social Ties, and Link Prediction Categories and Subject Descriptors. *Proceedings of KDD'11*, 1100–1108. (Cit. on pp. 35, 42, 44, 67, 94, 96).

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., … Campbell, A. T. (2017). StudentLife: Using Smartphones to Assess Mental Health and Academic Performance of College Students. In J. M. Rehg, S. Murphy, A., & K. Santosh (Eds.), *Mobile health: sensors, analytic methods, and applications*. Springer. (Cit. on pp. 15, 34).

Wang, X. W., Han, X. P., & Wang, B. H. (2014). Correlations and scaling laws in human mobility. *PLoS ONE*, *9*(1), 1–10. (Cit. on pp. 35, 36).

Wang, Y., Kang, C., Bettencourt, L. M., Liu, Y., & Andris, C. (2015). Linked Activity Spaces: Embedding Social Networks in Urban Space. In M. Helbich, J. J. Arsanjani, & M. Leitner (Eds.), *Computational approaches for urban environments* (pp. 1–395). (Cit. on pp. 43, 49).

Wang, Z.-Z. & Zhu, J. J. H. (2014). Homophily versus preferential attachment: Evolutionary mechanisms of scientific collaboration networks. *International Journal of Modern Physics C*, *25*(1440014), 1–8. (Cit. on pp. 28, 29).

Waskom, M., Botvinnik, O., O'Kane, D., Hobson, P., Ostblom, J., Lukauskas, S., … Qalieh, A. (2018, July). mwaskom/seaborn: v0.9.0 (July 2018). (Cit. on p. 5).

Watts, D. J. (2007). A twenty-first century science. *Nature*, *445*(7127), 489–489. (Cit. on pp. 6, 11, 12, 14).

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440–393. (Cit. on pp. 3, 24–26).

Wehrli, S. (2008). Personality on Social Network Sites: An Application of the Five Factor Model. *ETH Zuerich Sociology Working Paper*, *7*, 1–16. (Cit. on pp. 18, 48).

Whitaker, R. M., Noe, N., Whitaker, R. M., & Allen, S. M. (2016). Personality Homophily and the Local Network Characteristics of Facebook Personality Homophily and the Local Network Characteristics of Facebook. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 386–393. (Cit. on p. 123).

Williams, M. J. & Musolesi, M. (2016). Spatio-temporal networks: reachability, centrality and robustness. *Royal Society 2Open Science*, *3*(160196), 1–20. (Cit. on pp. 15, 38).

Williams, M. J. (2013). *Periodic patterns in human mobility* (Doctoral dissertation). (Cit. on pp. 11, 107, 123, 124, 128, 129, 135, 136).

Williams, N. E., Thomas, T. A., Dunbar, M., Eagle, N., & Dobra, A. (2014, August). Measures of Human Mobility Using Mobile Phone Records Enhanced with GIS Data. *PLoS ONE*, *10*(7), e0133630. (Cit. on p. 34).

Wimmer, A. & Lewis, K. (2010). Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook. *American Journal of Sociology*, *116*(2), 583–642. (Cit. on p. 29).

Wrzus, C., Hänel, M., Wagner, J., & Neyer, F. J. (2013). Social network changes and life events across the life span: A meta-analysis. *Psychological Bulletin*, *139*(1), 53–80. (Cit. on pp. 18, 47).

Wu, L., Zhi, Y., Sui, Z., & Liu, Y. (2014). Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS ONE*, *9*(5), e97010. (Cit. on pp. 10, 34).

Wu, W., Wang, J., & Dai, T. (2016). The Geography of Cultural Ties and Human Mobility: Big Data in Urban Contexts. *Annals of the American Association of Geographers*, *106*(3), 612–630. (Cit. on pp. 39, 52, 53).

Xu, Y., Belyi, A., Bojic, I., & Ratti, C. (2017). How friends share urban space: An exploratory spatiotemporal analysis using mobile phone data. *Transactions in GIS*, *21*(3), 468–487. (Cit. on pp. 18, 49).

Xu, Y., Shaw, S. L., Zhao, Z., Yin, L., Fang, Z., & Li, Q. (2015). Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. *Transportation*, *42*(4), 625–646. (Cit. on pp. 12, 33, 37, 51, 53).

Yan, X. Y., Han, X. P., Wang, B. H., & Zhou, T. (2013). Diversity of individual mobility patterns and emergence of aggregated scaling laws. *Scientific Reports*, *3*(2678), 1–5. (Cit. on p. 35).

Yang, Y., Chawla, N. V., Basu, P., Prabhala, B., & La Porta, T. (2013). Link prediction in human mobility networks. *Proceedings of ASONAM 2013*, 380–387. (Cit. on pp. 7, 67, 73, 76, 80, 92).

Yannick, L., Eric, F., Alvarez-Hamelin, J. I., Carlos, S., & Karsai, M. (2016). Socioeconomic correlations and stratification in social-communication networks. *Journal of the Royal Society Interface*, *13*(20160598). (Cit. on p. 47).

Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering Regions of Different Functions in a City Using Human Mobility and POIs Categories and Subject Descriptors. *Proceedings of the 18th ACM SIGKDD international conference on knowledge discover and data mining*, 186–194. (Cit. on p. 39).

Yuan, Y., Raubal, M., & Liu, Y. (2012). Correlating mobile phone usage and travel behavior - A case study of Harbin, China. *Computers, Environment and Urban Systems*, *36*(2), 118–130. (Cit. on pp. 42, 43, 50, 51).

Zenou, Y. (2015). A Dynamic Model of Weak and Strong Ties in the Labor Market. *Journal of Labor Economics*, *33*(4), 891–932. (Cit. on p. 20).

Zhang, D., Zhu, P., & Ye, Y. (2016). The effects of E-commerce on the demand for commercial real estate. *Cities*, *51*, 106–120. (Cit. on p. 11).

Zhang, H. T., Zhu, T., Fu, D., Xu, B., Han, X. P., & Chen, D. (2018). Spatiotemporal property and predictability of large-scale human mobility. *Physica A: Statistical Mechanics and its Applications*, *495*, 40–48. (Cit. on p. 35).

Zhang, Y. Q., Li, X., Xu, J., & Vasilakos, A. (2015). Human interactive patterns in temporal networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *45*(2), 214–222. (Cit. on pp. 11, 32).

Zhao, K., Musolesi, M., Hui, P., Rao, W., & Tarkoma, S. (2015). Explaining the power-law distribution of human mobility through transportation modality decomposition. *Scientific Reports*, *5*, 1–7. (Cit. on pp. 35, 37, 38).

Zhao, K., Stehlé, J., Bianconi, G., & Barrat, A. (2011). Social network dynamics of face-to-face interactions. *Physical Review E*, *83*(56109), 1–18. (Cit. on pp. 12, 30, 67).

Zhaoyang, R., Sliwinski, M. J., Martire, L. M., & Smyth, J. M. (2018). Age differences in adults' daily social interactions: An ecological momentary assessment study. *Psychology and Aging*, *33*(4), 607–618. (Cit. on pp. 18, 47).

Zignani, M., Papandrea, M., Gaito, S., Giordano, S., & Rossi, G. P. (2014, March). On the key features in human mobility: Relevance, time and distance. *2014 IEEE International Confefrenc4e on Pervasive Computing and Communication Workshops*, 260–265. (Cit. on p. 38).