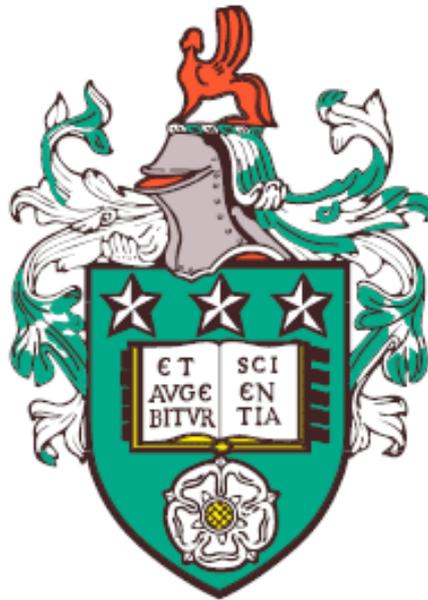


Monitoring and Modelling of Social Networks

Andrew Stuart Mellor



Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

University of Leeds

School of Mathematics

April 2017

The candidate confirms that the work submitted is their own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work of [1] and [2] forms the basis of Chapter 7. Chapter 8 details the analysis carried out in [3].

- [1] Heterogeneous out-of-equilibrium non-linear q -voter model with zealotry. Mellor, A.; Mobilia, M. & Zia, R. K. P.. *Physical Review E*, APS, 2017, 95, 012104

MM, RKPZ conceived the research/project. AM, MM, RKPZ conducted the analysis. AM ran all simulations and produced the figures. AM, MM, RKPZ contributed to the writing of the paper. AM was the corresponding author.

- [2] Characterization of the nonequilibrium steady state of a heterogeneous non-linear q -voter model with zealotry. Mellor, A.; Mobilia, M. & Zia, R. K. P.. *EPL (Europhysics Letters)*, 2016, 113, 48001

MM, RKPZ conceived the research/project. AM, MM, RKPZ conducted the analysis. AM ran all simulations and produced the figures. AM, MM, RKPZ contributed to the writing of the paper. AM was the corresponding author.

- [3] Influence of Luddism on innovation diffusion. Mellor, A.; Mobilia, M.; Redner, S.; Rucklidge, A. M. & Ward, J. A. *Physical Review E*, APS, 2015, 92, 012806

MM, SR conceived the research/project. AM, MM, SR, AR, JW conducted the analysis. AM ran all simulations and produced the figures. AM, MM, SR, AR, JW contributed to the writing of the paper. AM was the corresponding author.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Abstract

In this thesis we contribute to the understanding of online social networks, temporal networks, and non-equilibrium dynamics. As the title of this work suggests, this thesis is split into two parts, *monitoring* and *modelling* social networks. In the first half we look at current methods for understanding the behaviour and influence of individual users within a social network, and assess their robustness and effectiveness. In particular, we look at the role that the temporal dimension plays on these methods and the various representations that temporal networks can take. We introduce a new temporal network representation which describes a temporal network in terms of node behaviour which we use to characterise individuals and collectives. The new representation is illustrated with examples from the online social network Twitter. We model two particular aspects of social networks in the second half of this thesis. The first model, a generalisation of the popular Voter model, considers the dynamics of two opposite opinions in a heterogeneous society which differ by the resolve of their opinion. The second model investigates how the presence of ‘anti-bandwagon’ agents can prevent the spread of ideas and innovations on a social network, particularly on networks with restrictive topologies.

This contribution offers new ways to analyse temporal networks and online social media, and also provokes new and interesting questions for future research in the field.

Acknowledgements

First and foremost, thanks go to my supervisory team - Jon, Mauro, and Alastair. Your patience and willingness to offer your time has been second to none, and your comments have always been helpful and insightful. Thanks also for letting me do my own thing during my PhD whilst also knowing when to rein me in when I go off on a tangent. You have all helped shape me into the researcher I am today.

The funding and sponsorship from the EPSRC and Bloom Agency is greatly appreciated. Special thanks go to Peter Laflin for making this whole project work, offering interesting discussion and always asking questions while being supportive of my studies. I am also grateful to Amanda and Mary for taking interest in my work and offering many useful comments along the way.

Thanks to Sid and Royce for being stimulating and understanding collaborators. In particular Royce, for taking the time to explain everything in great detail and displaying an infectious child-like wonder at the new ideas and concepts we uncovered.

I would like to thank my friends for all their support during my time at Leeds. Particular mentions go to Steve and Colleen, who have helped keep me sane (or at least made me appear more sane, relatively speaking). Tom, Josh, Luke, Anja, Raphael, and Adam (and those who found freedom earlier, Chwas and Ricardo) - you have made my time at Leeds infinitely more fun, thank you!

I am hugely grateful to my parents, Mandy and David, and my brother, Richard, for their invaluable support throughout my whole life and education especially during my PhD when I have supposedly been “writing my book” or “publishing in a magazine”.

Finally, thanks to Mason Porter, who without organising a 7.30am exam revision supervision during my final weeks at Oxford, would not have set me off on this project and path.

Contents

Abstract	iii
Acknowledgements	v
Contents	vii
List of figures	xii
List of tables	xvii
Nomenclature	xix
1 Introduction	1
1.1 Social Networks in Modern Society	4
1.1.1 Challenges	8
1.1.2 Methodology	10
1.2 Thesis Outline	12
1.3 Contributions	13
I Monitoring	15
2 Temporal Networks	17
2.1 Temporal Networks	19

2.1.1	Temporal Network Notions	20
2.1.2	Representations	23
2.2	Temporal Network Analysis	27
2.2.1	Centrality Measures	27
2.2.2	Reachability and Connectivity	30
2.2.3	Inter-event Times	31
2.3	Discussion	33
3	Communicability Centrality	35
3.1	Communicability Centrality	37
3.1.1	Dynamic Communicability	39
3.1.2	Running Dynamic Communicability	41
3.1.3	Communicability in the Continuum Limit	43
3.2	Aggregation Errors in Communicability Calculation	43
3.2.1	Error Types	44
3.2.2	Temporal Network Partitions	47
3.2.3	Quantifying Errors	50
3.2.4	Summary	53
3.3	Efficient Calculation	53
3.3.1	Parallelisation	58
3.4	Discussion	58
4	The Temporal Event Graph	63
4.1	The Temporal Event Graph	66
4.1.1	Duality	69

4.2	Theoretical Properties of the TEG	75
4.3	Statistical Properties of the TEG	77
4.3.1	Component Sizes, Distribution, and Growth	78
4.3.2	Motif and Inter-event Time Distributions	80
4.3.3	Induced Aggregate Networks	84
4.4	Discussion	85
5	Temporal Events on Twitter	87
5.1	Twitter	88
5.1.1	What is Twitter?	89
5.1.2	Data Structures and APIs	93
5.2	Case Studies	95
5.2.1	The X Factor	96
5.2.2	Triathlon	100
5.2.3	Oxford	106
5.3	Discussion	107
II	Modelling	109
6	Model Methods	111
6.1	From the Master Equation to Mean-field	112
6.1.1	Master Equation	113
6.1.2	Fokker-Planck Equation	116
6.1.3	First Passage Processes	119
6.1.4	Mean-field Equations	122

6.2	Models on Static Networks	124
6.2.1	Intractability of the Master Equation	125
6.2.2	Standard Formulations	125
6.3	Discussion	127
7	The 2q-Voter Model with Zealotry	129
7.1	The 2q-Voter Model with Zealotry	133
7.1.1	Model Outline	133
7.1.2	Master Equation Formulation	134
7.1.3	Characterisation of the NESS	138
7.2	Continuum Descriptions and Exact Results	139
7.2.1	Fokker-Planck Equation	140
7.2.2	Mean-Field Analysis	141
7.2.3	Linear Stability and Phases	142
7.2.4	Numerically Exact Results for Small Systems	145
7.3	The Linear Gaussian Approximation (LGA)	147
7.3.1	Correlations and Probability Flows	149
7.3.2	Simulation Results and Assessment of Fit	151
7.4	Fluctuation-driven Switching Dynamics	153
7.5	Asymmetric Zealotry	157
7.6	Discussion	158
8	The LISA Model of Innovation Diffusion	161
8.1	The LISA Model	164
8.1.1	Model Outline	165

8.2	Mean-field Analysis	167
8.3	Random Graphs and Lattices	172
8.3.1	Erdős-Rényi Random Graphs	173
8.3.2	One-dimensional Lattices	179
8.4	Discussion	185
9	Summary and Outlook	189
9.1	Summary	189
9.2	Outlook	191
9.2.1	Using the Temporal Event Graph to Study Dynamical Processes on Networks	192
9.2.2	Network Filtering with the Temporal Event Graph	193
9.3	Final Remarks	195
	Bibliography	197

List of figures

1.1	The <i>Zachary Karate Club</i> network [4].	2
1.2	The <i>Florentine Families</i> network [5, 6].	3
1.3	An example network generated from Twitter.	11
2.1	A simple temporal network expressed as a sequence of three static networks.	18
2.2	A temporal network.	20
2.3	An example of bursty temporal behaviour.	23
2.4	Temporal network representations.	25
3.1	A simple temporal network.	40
3.2	The scenarios in which the temporal aggregation of events introduces error into the communicability calculation.	44
3.3	Static network choices which are used to generate temporal networks by repeated sampling of edges (with replacement).	51
3.4	Average correlation coefficients ρ_p (blue), ρ_s (green) between the broadcast vectors of the fixed-width partition of width Δt and true partition over an ensemble of 1000 synthetic networks.	52
3.5	Pictorial representation of the main loop of the parallel algorithm.	60

4.1	Contact networks and two possible dual network representations.	64
4.2	Illustration of the duality of temporal networks and the temporal event graph.	68
4.3	Inconsistent edge-labelled temporal event graphs.	71
4.4	The inverse algorithm for the TEG.	73
4.5	Temporal component dependence on Δt	79
4.6	Illustration of the temporal barcode associated with a Δt -TEG.	81
4.7	Examples of temporal networks and their temporal event graphs.	82
4.8	The IET distributions for the random temporal network.	83
5.1	The four possible colourings of the ABBC motif.	88
5.2	Examples of tweets.	89
5.3	Growth of monthly active users on Twitter from 2010 to 2016.	91
5.4	Example tweets and the corresponding temporal network.	95
5.5	Temporal barcodes of the 5s-TEG for the X Factor dataset.	97
5.6	The growth of the largest component of the TEG with Δt for the X Factor dataset, as a fraction of all events.	98
5.7	The IET and motif distributions of the TEG for three XFactor shows.	99
5.8	Temporal barcodes of the Δt -TEG for the Triathlon dataset.	101
5.9	The motif and IET distributions across the largest (by number of events) components of the 5min-TEG of the Triathlon dataset.	102
5.10	Component motif and IET entropy, calculated from the corresponding distributions with components scaled by the number of events they contain.	103
5.11	Aggregate graphs of temporal components from the 5min-TEG of the Triathlon dataset.	105

7.1	An illustration of the $2qVZ$ dynamics.	134
7.2	An illustration of the violation of detailed balance.	137
7.3	Criticality in the symmetric $2qVZ$ ($z_{\pm} = z$).	144
7.4	Typical time series of the $2qVZ$	145
7.5	Numerically exact stationary distributions and probability currents in both the low zealotry regime, $z < z_c$, and high zealotry regime, $z > z_c$	147
7.6	Numerically exact properties of the $2qVZ$ NESS with $(N, S, Z, q_1, q_2) = (280, 100, 40, 1, 2)$ (low zealotry regime).	148
7.7	$C_{ij} = \langle \xi_i \xi_j \rangle$ as a function of system size N	152
7.8	Stationary probability currents.	153
7.9	The LGA prediction of the two-point correlation function $\tilde{C}(\tau)$ compared to simulations of varying susceptible population size in the high and low zealotry regimes with $s_{1,2} = s$, $z_{\pm} = z$ and $(q_1, q_2) = (1, 2)$	154
7.10	Switching behaviour of the $2qVZ$ ($z < z_c$).	155
7.11	The mean switching time τ_s between the two stable fixed points for systems of size $N = 100$ with $(q_1, q_2) = (1, 2)$	156
7.12	A summary of the low zealotry regime with asymmetric zealotry with parameters $(N, S, Z_+, Z_-, q_1, q_2) = (258, 100, 30, 28, 1, 2)$ and consequently $\delta = 0.01$	158
8.1	Schematic depiction of the LISA model.	165
8.2	Evolution of a realisation of the LISA model on a complete graph of 10^6 nodes with $I_0 = 0.8$ and Luddism parameter $r = 0.9$	169
8.3	Dependences of the final-state densities L_{∞} , I_{∞} and A_{∞} for a complete graph of 10^4 nodes and $I_0 = 0.9$	171
8.4	The evolution, averaged over 100 realisations, of the LISA model on an ER graph with $N = 10^3$ nodes, $k = 10$, and $I_0 = 0.8$	174

8.5	Dependence of the final densities L_∞, I_∞ and A_∞ on the average degree for ER graphs with $N = 10^3$ nodes.	176
8.6	Variance in the stationary state of the density of adopters, $N_A(\infty)/N$, obtained over an ensemble of 40 realisations of the model on 75 randomly generated networks for $N = 100 - 1000$	177
8.7	The mean-field steady state predictions (8.10) over the parameter space (γ, r) for $k/N = 0.025$ and $I_0 = 0.9$	178
8.8	Time dependence of the densities in each state for a one-dimensional lattice of size $N = 10^5$ averaged over 100 realisations.	179
8.9	Final simulated average proportions of adopters, ignorants and Luddites for varying values of γ , averaged over 100 simulations.	181
9.1	The CCDFs of the IET distributions for the SIS temporal network with constant probability of infection from each neighbour and constant recovery rate (once infected).	194
9.2	Using the TEG to filter the Twitter firehose.	195

List of tables

3.1	A summary of the temporal network partitions, the restrictions on α , the average number of time intervals, and the errors associated with each.	54
4.1	The set of all possible two-event motifs \mathbb{M} , given by their contact sequence, description, label, and label properties ξ_{in} , ξ_{out} , and ξ_{switch}	69
5.1	Information contained within a tweet.	94
5.2	Datasets of Twitter conversations considered in this chapter.	96
5.3	Descriptions of the largest components of the 1h-TEG for the Oxford dataset.	106

Commonly Used Acronyms and Terminology

Acronyms

IET	: Inter-event time
DAG	: Directed acyclic graph
TEG	: Temporal event graph
(C)CDF	: (Complementary) cumulative density function
SI(S)	: Susceptible-Infected(-Susceptible) [model]
PCA	: Principal component analysis
ER	: Erdős-Rényi [graph]
VM	: Voter model
NESS	: Non-equilibrium steady state
ME	: Master equation
FP	: Fixed point
FPE	: Fokker-Planck Equation
LGA	: Linear Gaussian approximation

Terminology

Throughout this thesis we will consider static and temporal networks, as well as their duals.

A static network is described mathematically as $G = G(V, E)$, where $V \subset \mathbb{N}$ is the set of *nodes*, and $E \subseteq V \times V$ is the set of *edges*.

A temporal network is described by $G = G(V, T, D, E)$ where $V \subset \mathbb{N}$ is a set of *nodes*, $T \subset \mathbb{R}_0^+$ the set of times, $D \subset \mathbb{R}_0^+$ is the set of durations, and $E \subseteq V^2 \times T \times D$ is the set of *events*.

A temporal network dual is described by $\mathcal{G} = \mathcal{G}(\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = E$ is the set of events (known as *event vertices* or simply *vertices*) and $\mathcal{E} \subseteq \mathcal{V}^2$ is the set of *edges*.

1

Introduction

Networks are ubiquitous in nature and society; from the complex interactions of proteins that make up the human body to the roads and rails that form our transport network across the globe, and from the many neurons making up our nervous system to the people we interact with on a daily basis or call our friends. Regardless of the context, networks can be modelled abstractly as a collection of *nodes* (proteins, junctions, people) which are connected by a collection of *edges* (interactions, roads, relationships). The study of these objects is commonly known as *complex network analysis* [7, 8] which has its foundations in the more abstract *graph theory* [9]. As a result, the word network is often used interchangeably with graph.

The interaction of human beings and the resulting collective behaviour is particularly fascinating. This kind of interaction drives many systems that we see today, such as the stock market, war, politics, and the behaviour of large crowds. Understanding the

way humans behave collectively and the relationships they form is commonly known as *social network analysis* [10]. Although the first use of graph theory is thought to be that of Leonhard Euler’s study on the seven bridges of Königsberg [11, 12], the graphical study of social networks only appeared relatively recently in the early 20th century [13].

As the bridges inspired Euler to develop the notion of paths on networks, advancements in the field of network analysis came through the study of empirically observed data. Perhaps the most famous example of a social network is that of the Zachary Karate Club [4], seen in Figure 1.1. The network features 34 members of a karate club. Two members are connected by an edge if they were friends outside of the club. Due to an ongoing disagreement between the club president (node 33) and the karate instructor (node 0), the club ultimately split into two separate clubs, denoted by the red and green nodes. The unique nature of this network, being a social system which fragmented into two, has

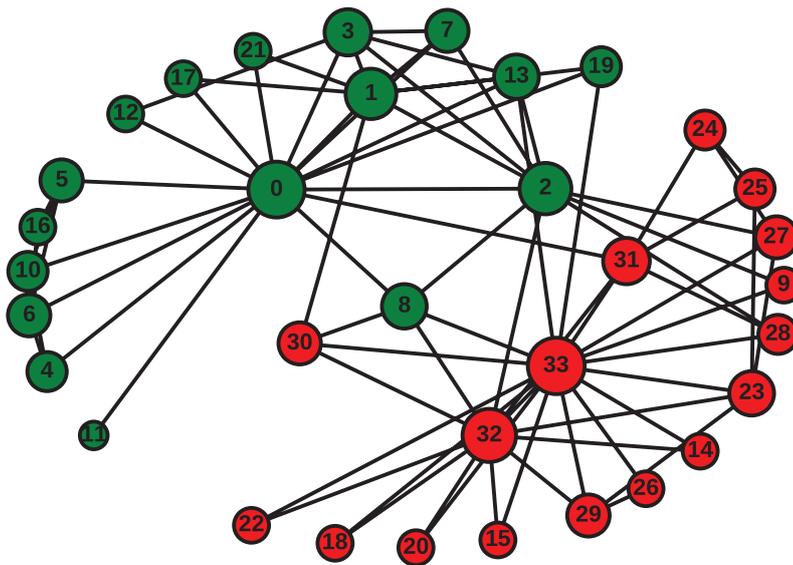


Figure 1.1: The *Zachary Karate Club* network [4].

made it an ideal testing ground for community detection algorithms¹. In this network the question is ‘can we predict how the club would split from the observed friendships alone?’

Another early example of a social network from the 15th century (although the analysis was conducted over 500 years later), is the marriage relationships between oligarch

¹ The network has been so widely used has that there now exists the ‘Karate Club Club’, a club for those first to include the Karate club network in their talk at conferences in network science.

Florentine families in Italy [5, 6]. Here the nodes represent the families, and an edge connects two nodes if the families are connected through marriage. The network was

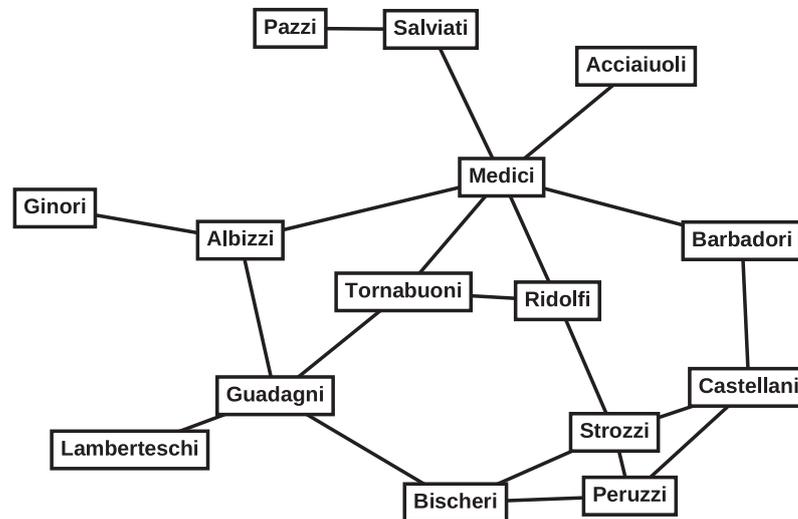


Figure 1.2: The *Florentine Families* network [5, 6].

created to understand the relative importance, influence (or centrality) of these families over the time period - which family is the most powerful? This type of question can be probed by considering the particular properties of the nodes in the network, often referred to as *centrality measures*, of which there are many [7, 14–20]. Also collected was a network of business and financial transactions between the families. These two networks can be studied in combination as a multilayer network [21].

Another notable example of social network analysis is Milgram’s (subsequently named) ‘six degrees of separation’ experiment [22]. Milgram’s experiment was simple; ask participants to deliver a letter to another named person in the United States, without knowledge of their address. If they did not know the named person, they were to pass the letter on to another person who they thought may be better suited to deliver the letter successfully. The average path length of letters which made it to their destination was roughly six. This strikingly small number (given that the population of the US was 200 million at the time) was one of the first observations of what is now known as the *small world effect* which more formally states that the average path length grows logarithmically with the network size.

In these examples, the collection of empirical data sparked pertinent questions, which

led to new breakthroughs and understanding of complex networks. The process of collecting this data was laborious however; information on the Florentine families was only available due to strict record keeping in Florence at the time, the friendships of the karate club were gathered over three years, and in Milgram's experiment of the 296 letters sent out, 232 never reached their destination.

As the strict record keeping in Florence helped generate social networks of Florentine families, modern social network analysis now benefits from the proliferation of the internet, digital record keeping, and excesses of data which allow us to study social networks in brand new ways.

1.1 Social Networks in Modern Society

Since the turn of the millennium, the meaning of a 'social network' has changed. Social networks now more commonly refer to online social networks such as Twitter² and Facebook³.

What exactly is an online social network? Although platforms vary, on a basic level users create an account and can declare other users as their 'friend' (often bidirectional) or choose to 'follow' the activity of another user (unidirectional). Beyond this declaration, users can interact with each other through messaging or by sharing content such as images, videos, and URLs. As these social networks are online, every interaction a user has with the network is recorded which provides invaluable opportunities for research (if the data is publicly available).

Social networks are now an influential part of modern society. Facebook has 1.87 billion monthly active users (users who have logged in at least once in a monthly period) as of 31 December 2016 [23] which is nearly a quarter of the current world population and a larger fraction of the population who are able to access the network. Twitter has 313 million monthly active users as of June 30 2016 [24] but boasts over a billion monthly unique views of pages which contain embedded messages from the Twitter network.

²www.twitter.com

³www.facebook.com

Online social networks have taken on a myriad of roles within society including reconnecting and chatting with friends, organising social groups and activities, and sharing pictures or ideologies [25]. Beyond their original uses, they are now places to find breaking news stories, buy and sell items, and discuss politics [26, 27] (as showcased by the recent UK EU referendum [28] and the 2017 US presidential election [29]).

The wealth of available data on online social networks has instigated many studies from a mathematical and social perspective. With metadata available on the users themselves two studies were able to investigate the prevalence of homophily (the preference of individuals to associate with other individuals of similar traits) in a sample of the Facebook social network [30, 31], finding that age was the main factor for connections to be formed between users. Also, in an online experiment paralleling that of Milgram's in the 1970s, Facebook's own research team estimated that the average shortest path length on their social network was approximately 3.5 [32], suggesting that the small-world effect is even more profound online than offline. One of the first studies into the Twitter social network [33] obtained 41.7 million user profiles, and 1.47 billion social relations in order to investigate the spread of information on the network, and the relative influence of the users on the network (much like the Florentine families, only with a network over a million times larger).

There have since been many studies using data from Facebook and Twitter, as well as other popular online social networks such as LinkedIn, Sina Weibo, and Reddit [34–50]. Some studies of note aim to understand the social network as an entity, characterising the distribution of number of connections [33], the degree of homophily [51], and the features of nodes which make them influential on the network [37]. Others use social networks as a proxy to make forecasts on things outside the network such as detecting earthquakes [47], or attempting to predict the stock market [35] or political elections [42]. Care needs to be exercised with these latter studies as the users of social media are not necessarily a representative sample of the whole of society, and their online behaviour may not be reflective of their behaviour off the social network [52]. In this thesis we will primarily be interested in social networks as a social and mathematical entity, aiming to characterise their properties and features.

Advertising

An important aspect which has helped drive the popularity of online social networks is that they are predominantly free to use. Platform providers generate their revenue through data collection and licensing, and the inclusion of advertisements. The creation of online social networks has driven a change in the way advertisements are presented, as highlighted in a report by the Academy of Marketing [53]:

“Today’s consumers are technologically smart, visually orientated, and information overloaded; and they lead an ‘on-the go’ life style, which can make them hard to reach (Leek & Christodoulides, 2009) ...

...It [mobile and internet advertising] allows individual personalization (Yanis, 2008; Jayawardhena et al., 2009), individual targeting and at the same time, the collection of customer data.”

Whereas advertising in the past has relied on a scatter gun approach of reaching far and wide, online social networks offer advertisers the opportunity to understand their target audience through data collection and, furthermore, allow companies to engage with consumers on a social level [54].

By 2020, internet advertising is forecast to make up to 40% of all UK advertising revenue [53] which is a significant part of a sector which currently contributes approximately £13 billion to the UK economy [55]. To succeed in this sector companies and advertisers need to be able to fully utilise social networks by understanding how individuals and collectives behave whilst using these networks and be able to use social network data to efficiently spend advertising resources to achieve the greatest impact.

Bloom Agency

This body of work is funded by an EPSRC CASE studentship, part sponsored by the data led, full service integrated marketing agency, *Bloom Agency* (subsequently referred to as *Bloom*). Bloom are at the forefront of studying company and brand engagement with social media and are actively engaged in ongoing research in this area [56–60].

Using their in-house analytics suite *Whisper* (based upon temporal centrality measures [61]), Bloom are able to monitor online conversations in real-time, identify the key influencers in conversations, and uncover opportunities to drive and shape the conversation around brands. Bloom has worked with many high profile clients to help engage their current and potential customers, details of which can be found in [62]. In particular their work with Sky Sports and ITV has shown that engaging users during television programmes with a 'second screen' social experience can help characterise audiences, assess the impact of televised marketing campaigns, and provide a richer experiences for these audiences.

Working closely with Bloom over the course of this project has highlighted some of the many challenges they face on a daily basis. The main observations drawn from time working alongside Bloom are:

- **Real-time analysis is crucial.** The attention of social media users is quickly lost and so action needs to be taken quickly in order to be effective. Furthermore new content is adopted rapidly by users and often acted upon within minutes. This means that information can be spread quickly, or controversial content can soon be met with outrage.
- **Client needs and demands can vary wildly.** The clients of Bloom are all different and each has their own motivations to interact with social media. Some however want to interact with social media but have no fixed objectives. It is therefore important that any analysis of social networks should be able to describe them in generality, but also be adaptable to answer specific, focused questions.
- **Identifying key conversation drivers (or influencers) is important.** While potentially many thousands of users will participate in an online discussion around a particular topic, the conversation is often centred around a small number of users who attract a significant share of other users messages.
- **Tracking the spread of opinion and sentiment is difficult.** With content on social media being limited to short messages it can be difficult to decipher the opinions of users. Some attempts have been made to measure the sentiment of

individual messages, however due to the short message length and the difficulty to detect sarcasm this is often inconclusive. One should therefore look to study the mechanisms of idea transmission to help inform what data is collected.

- **Social media is noisy.** Despite the best efforts of social media providers, a significant percentage of content produced on these networks is autonomous, or *spam*⁴. It is therefore important to be able to identify this content so as to avoid any incorrect conclusions.
- **Conversation networks can potentially be classified.** Bloom's network visualisation tools show that, at least visually, similar network structures and components occur across different case studies. This suggests that these patterns can potentially be classified and used to understand new networks when they appear.

1.1.1 Challenges

There are a number of challenges when studying online social networks.

Mathematical

Social networks have often been described using the mathematical language of graph theory. Static graphs (or networks) are well studied, but their analysis is often limited to approximations of the network such as assuming homogeneity or ignoring degree-degree correlations (whether hub nodes are likely to be connected to other hubs, or to nodes of low degree) [63, 7, 64, 65]. The static network model is useful for studies of social networks where the connections are assumed to be fixed or to change slowly over time. By contrast, the networks studied by Bloom are generated over the space of hours and, as we will see in Chapter 5, change rapidly over time. We therefore look to use temporal networks to model the networks generated by social media. Temporal networks however are less understood and less tractable than their static network

⁴ Estimated between 8 – 15%, with disagreement between official and academic sources.

counterparts. The challenge is therefore to devise new mathematical methods to aid our understanding of temporal networks, and specifically to find ways to quantify the behaviour of individuals within the network.

Complexity

Much like the financial markets, social networks are constantly evolving. They both evolve over time as the habits of the users change and technological features improve. For example, social networks are now well equipped to facilitate the sharing of images and videos where they were once limited to text only [66]. They also react strongly to external stimuli such as world news or live events, driving conversations and provoking content creation. Social networks are perhaps even more complex than the financial markets despite both being driven by the interaction of many users. For example, there is no equivalent for the infamous Black-Scholes-Merton model [67] for social networks. Any approach to model social networks must be able to capture the inherent complexity of these systems, but must also be adaptable to changes in the way they behave.

Data

Online social networks produce large amounts of data at a high rate. This is due to the large number of users and interactions between them, but also due to the recording of every detail of these interactions; the time at which the interaction occurred, any media attached to the interaction, and full details of each interacting user. As a consequence there are a number of logistical issues with the handling of such data. Data needs to be collected, filtered, and stored. This raises questions about what information is needed to answer the research question at hand and how easily it can be stored.

In addition to being able to process the large amount of data produced by online social networks, there is the issue of being able to process the data in real-time or near-real-time. This was recognised by Innovate UK in their 2016 ‘Creative Industries Strategy Report’ [68] which states:

“Across the economy, data has become recognised as crucial to business success.

With increased connectivity has come an increase in available data, and more opportunities to manage and capture value from it. When this data is embedded in instant or near real-time decision making, it becomes even more powerful, with commentators forecasting a shift towards embedding data in decision-making and strategic processes.”

Therefore the challenge is not simply being able to process the large volume of data, but having the ability to process it quickly and efficiently so that it can be utilised in real-time decision making.

A flavour of the complexity of online social networks is captured in Figure 1.3 which shows the interactions between users of the Twitter social network who mention the word ‘Leeds’ over the period of an hour on a typical Saturday afternoon. This relatively small sample of the Twitter network already has a large number of nodes and has non-trivial structure. Other keywords, especially those surrounding popular and current topics (politics, sports, etc.), can generate much larger networks in a similar time frame. At the centre of this network is what is often referred to as the ‘hairball’. The term has its origins in our inability to visualise these networks (which is a research area in itself) however it has become a metaphor for the complexity of these networks and the difficulty of studying them.

The subject of this thesis is therefore to develop tools (both mathematical and computational) to *untangle the hairball* of online social networks, and use them to study the behaviour of individuals and collectives.

1.1.2 Methodology

Our approach to studying social networks takes two contrasting but complementary perspectives.

The first addresses the challenges faced by Bloom and the wider community by developing methods to understand *influence* and *behaviour* on social networks, and tools to be able to conduct this analysis in real-time and at scale. This approach is primarily data-driven and aims to utilise the plethora of data available from online social networks.

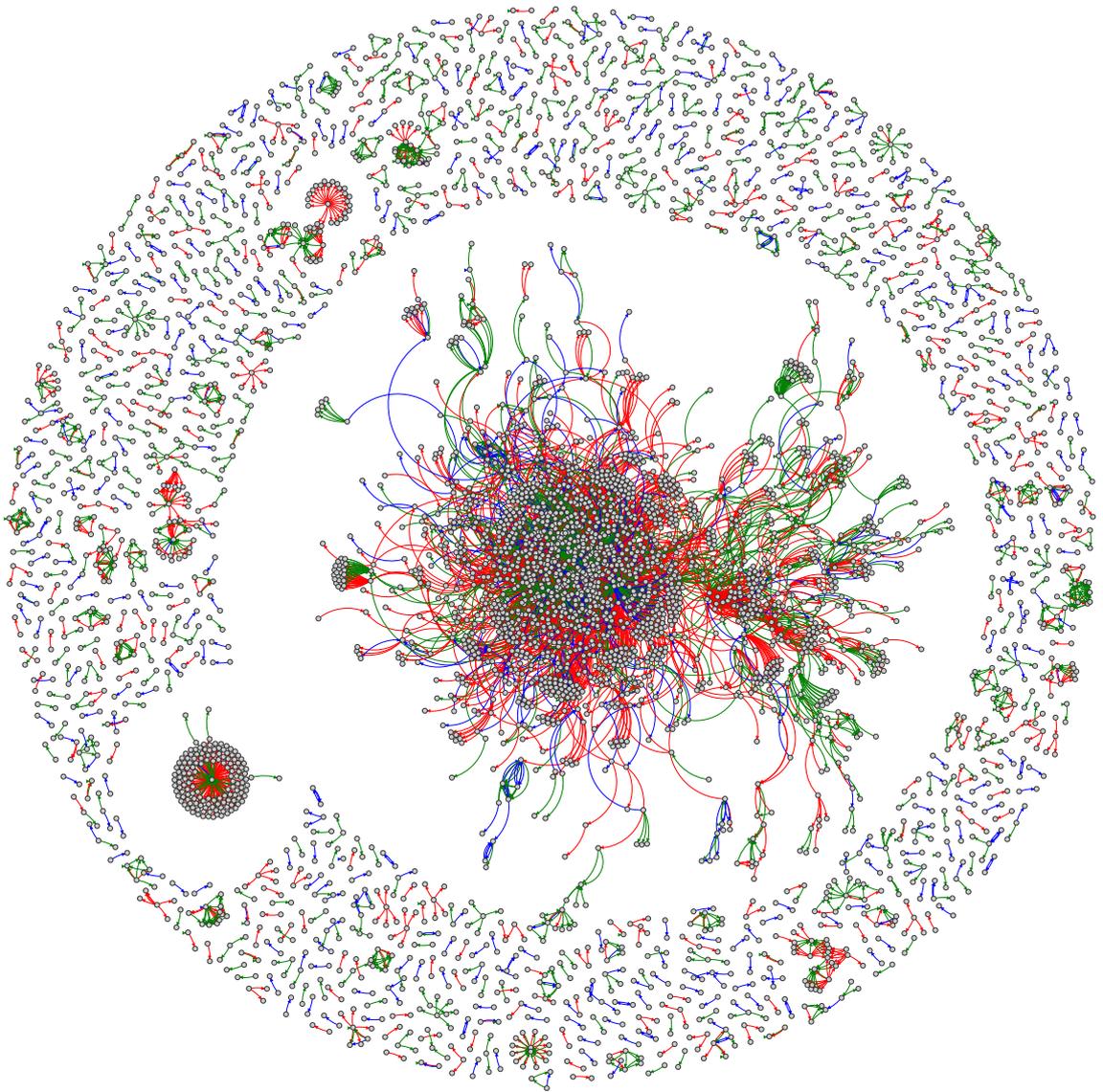


Figure 1.3: An example network generated from Twitter. The nodes are the users of the network and an edge is present between two nodes if one node has sent a message to the other. Green edges are messages, blue edges are replies to previous messages, and red edges are retweets (copies) of previous tweets.

We frame this approach in the context of temporal networks and study the evolution of the interaction patterns of the nodes. While static networks are well understood (although difficult to model), the study of temporal networks is still in its infancy and we lack the tools to analyse them without resorting to some form of approximation or aggregation. We therefore assess the current methods used to understand influence on temporal networks and, in particular, how the partitioning of the network in time can have consequences on the conclusions of the analysis. We also explore ways to capture the behaviour of individuals and collectives in a temporal network and subsequently develop a new behaviour-centric representation of a temporal network.

The second approach considers simple models of spreading behaviour on social networks. Considering simplistic mechanisms of interaction between individuals we study the macroscopic behaviour of the system as a whole which, if representative of real world systems, may clarify the mechanisms that drive interactions on social networks. By controlling the model parameters we look to find regimes of distinct behaviour, and by considering different network structures we aim to understand how the network topology affects the dynamics. In order to analyse these models formally we need to apply theory from statistical mechanics and dynamical systems.

1.2 Thesis Outline

This thesis is organised in two parts. Due to the diversity of topics in this thesis formal literature review is deferred to the relevant sections.

Part One

The first part of this thesis concerns the study of temporal networks, what tools are used to analyse them, and how we can apply them to online social networks. In Chapter 2 we summarise the current literature on temporal networks. In Chapter 3 we review the *communicability* centrality and analyse its performance from both an analytical and computational viewpoint before providing a novel and efficient algorithm for its calculation. In Chapter 4 we present a new representation of a temporal network, the

temporal event graph, and study its properties. Finally, in Chapter 5 we use the temporal event graph to study data collected from the Twitter social network.

Part Two

The second part of the thesis concerns models of human behaviour. In Chapter 6 we outline some of the methods used to analyse stochastic processes and dynamics on networks that are used in the subsequent chapters. In Chapter 7 we introduce the *2q-voter model with zealotry*, an extension to the classical voter model which exhibits non-trivial non-equilibrium behaviour. In Chapter 8 we introduce the *LISA model* of innovation diffusion. This model aims to understand how ‘anti-bandwagon’ behaviour can affect the spread of an innovation on a network.

Finally in Chapter 9 we present a summary of the thesis, discuss the impact of this research, and highlight future research directions.

1.3 Contributions

This thesis contributes to both the understanding of temporal networks and to the study of non-equilibrium social systems. The key contributions are summarised as follows:

- **Analysis of the communicability centrality measure.** We explore the behaviour of the dynamic communicability measure [61] when different temporal network aggregations are used, showing that the interpretation of the metric changes depending on the aggregation level. We provide an algorithm for the exact and efficient calculation of the measure which operates on an unaggregated temporal network, and can easily be parallelised.
- **Introduction of a new temporal network representation.** Building upon previous work on edge-based network representations [69, 70] and temporal motifs [71], we introduce a new representation for temporal networks, the *temporal event graph*. Through examples we show how this representation is

able to decompose the temporal network and provide a means of understanding the generating process of the network through temporal motif and inter-event time distributions. Furthermore we outline a method to sample from temporal networks by sampling entire components (of the temporal event graph) instead of individual events. This representation is not restricted solely to social networks and can in fact be generalised to any series of connected temporal events (purchases/trades, academic collaborations etc.).

- **Introduction of a non-equilibrium model of opinion dynamics.** We present a heterogeneous out-of-equilibrium non-linear voter model of opinion dynamics. Through numerics, simulations and analysis we characterise the behaviour of the model under various parameter regimes. The model is one of the first properly studied models of this class which exhibits non-vanishing stationary probability currents which are investigated by considering a linear Gaussian approximation (LGA) of the Fokker-Planck equation. We find that the model can be accurately described using a LGA for large systems and that the non-equilibrium currents give hints to the existence of ‘leaders’ and ‘followers’ within a society.
- **Study of the spread of innovation in a network context.** We study a simple model of the spread of innovation on a network, extending the widely used Bass model [72]. In this model we introduce a new class of agents called ‘Luddites’ who oppose the rapid spread of innovation. The novel mechanism for the creation of Luddites incorporates the rate of change of the states of neighbouring agents, rather than just their states. We find that the rapid spread of innovation can polarise a population and, on constrained topologies, the presence of innovation opposing agents can block pathways to innovation spread.

Part I

Monitoring

2

Temporal Networks

The study of static networks has been fruitful in helping understand the complicated relationships that exist in nature. Historical studies of networks have been in the static framework due to both the availability of data and the relative ease of analysis. Previous studies of social networks have looked at the networks of Facebook friendships [30, 31], the Twitter follower network [37], and self-reported networks of friendship [73, 74]. These networks capture relationships at a given time or over some time period. The networks studied are however constantly evolving [75–77], forming new edges as well as destroying others in response to node activity as well as external stimulæ. For example, unkept friendships on social media may result in ‘defriending’, or ‘unfollowing’ of nodes, or the outbreak of a contagious virus may temporally cause a node to disconnect itself from the network entirely as a means of quarantine.

Does this temporal information really make a difference? In Figure 2.1 we consider the

possibility of information transfer between two nodes, that is, the existence of a temporal walk between them (formally defined in Section 2.1.1). Considering the network as an

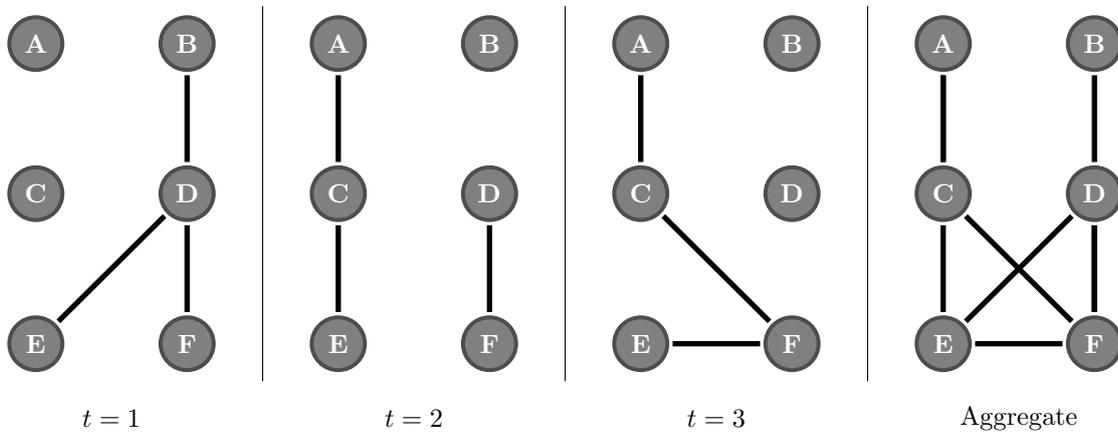


Figure 2.1: A simple temporal network expressed as a sequence of three static networks. Ignoring the temporal nature of the network and considering only an aggregated network suggests a temporal path exists between A and B. This path does not exist when the temporal nature of the network is taken into account.

aggregated static network, we see that there exists a path from A to B. However, with the addition of temporal information we know that that path $D \rightarrow B$ occurs prior to that of $A \rightarrow C$. Therefore a temporal path does not exist from A to B over the time period. As this example shows, the temporal structure of the network plays an important role in the dynamics of systems acting on the work, in the same way that the network topology is influential.

Chapter Outline

This chapter is intended to be a review of temporal network literature: terminology, concepts and recent studies. It contains no original material and serves to introduce the notation and terminology needed for subsequent chapters. In Section 2.1 we define a temporal network in its most granular terms and define a number of notions useful to the study of temporal networks. We also show a number of popular representations of temporal networks which capture the temporal dynamics of the network to varying extent. In Section 2.2 we summarise the types of analysis conducted on temporal

networks, illustrated with examples, before concluding in Section 2.3.

2.1 Temporal Networks

We consider temporal networks described by an *event sequence* (or contact sequence).

Definition 2.1.1 (Temporal network). Let $V \subset \mathbb{N}$ be a set of interacting nodes, $T \subset \mathbb{R}_0^+$, a non-empty ordered set of interaction times, and $D \subset \mathbb{R}_0^+$ a set of event durations. The temporal network is then defined as the quadruple

$$G = (V, T, D, E)$$

where $E \subset V^2 \times T \times D$ is the set of temporal events.

A temporal event $e \in E$ takes the form $e = (u, v, t, \delta)$, corresponding to a contact from u to v (and from v to u in the undirected case), initialised at time t , and lasting for a duration δ . Unless explicitly stated, we will consider temporal events to be directed.

This representation can fully describe the temporal nature of most networks. Some authors [78, 79] prefer to make the distinction between event sequences of zero duration and interval graphs where activity between two vertices is not instantaneous but persists over a set of time intervals. The difference between these modes of thinking is that often the restriction is placed on vertices such that they can be involved in only one event at any given time, whereas in the interval graph this is allowed. A clear example of this would be to model telephone calls by an event sequence as a traditional phone call only supports two people. By contrast a temporal network of physical proximity may allow multiple people to be connected to each other over some time interval. This dichotomy can confuse what is in fact the same object. For this reason we will consider all temporal networks as event sequences and relax any conditions on the number of events a vertex can be involved with, and state explicitly when this is not the case.

For the purposes of visualisation, a temporal network is often displayed as a static network where edges are labelled with the times and durations of events occurring along those edges, as in Figure 2.2.

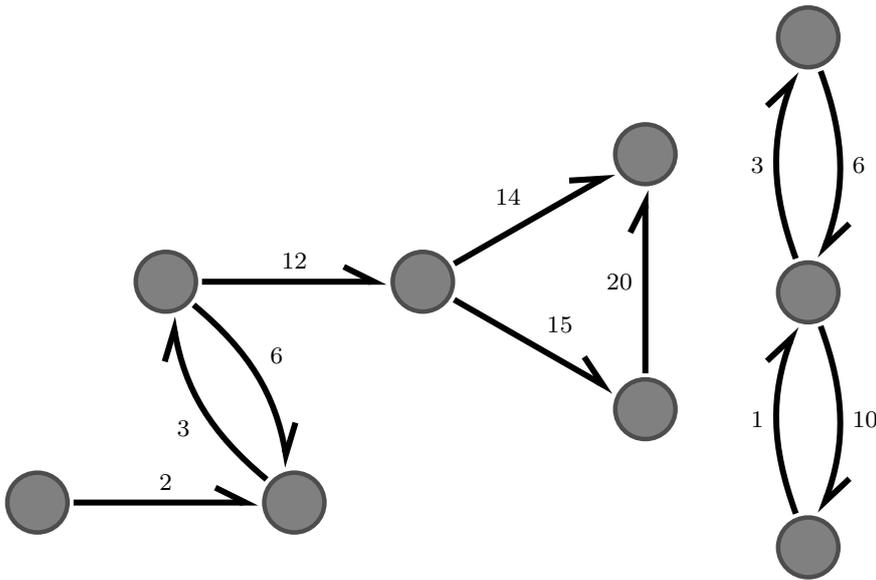


Figure 2.2: A temporal network. Events (edges) are labelled with times when that event occurs.

2.1.1 Temporal Network Notions

Temporal Walks

The edges of a static network constrain the dynamics of processes acting on the network. A sequence of edges where the end node of one edge is the start node of the subsequent edge is called a path, provided each node is visited at most once. Extending this notion to temporal networks one naturally requires a constraint on the traversal times between nodes.

Definition 2.1.2 (Temporal walk). A temporal walk between nodes i and j is a sequence of temporal events $((u_0, u_1, t_1), (u_1, u_2, t_2), \dots, (u_{n-1}, u_n, t_n))$ such that $u_0 = i$, $u_n = j$, and the sequence of event times is non-decreasing, i.e. $t_1 < t_2 < \dots < t_n$.

Alternatively, it is often useful to define a temporal walk as the sequence of nodes along the walk and times visited rather than the sequence of events. It is easy to see however that these two definitions are equivalent and will be used interchangeably.

Furthermore we define a temporal path to be a temporal walk such that each node is visited at most once.

Remark 2.1.3. There exists a temporal walk from i to j if and only if there exists a temporal path from i to j .

This is evident as the set of temporal paths is strictly a subset of the set of temporal walks, and a temporal walk can become a temporal path by the removal of events which cause a loop back to revisit a node. If we impose restrictions on the time spent at each node however, there may be instances where the restricted temporal walk exists, but the restricted path does not.

In the literature a temporal path has also been referred to as a *time-respecting* path [80], a *journey* [81], or a *non-decreasing* path [82]. It is also worth noting that the existence of a temporal walk is not transitive, that is, the existence of a temporal walk from $i \rightarrow j$ and from $j \rightarrow k$ does not imply the existence of a temporal walk from $i \rightarrow k$.

Inter-event Times (IETs)

In typical temporal networks nodes can participate in a number of different temporal events, and an event between two nodes can occur multiple times over the duration of the network. In telecommunication networks, two users may make regular phone calls on a daily or weekly basis, for example. The correlations in the timings of such events can have a significant impact on the temporal structure of the network and hence play a role in determining the number of temporal paths through the network. The notion of *inter-event time* characterises this behaviour.

Definition 2.1.4 (Inter-event time). The inter-event time between any two temporal events e_i, e_j is given by

$$\tau_{ij} = t_j - t_i$$

assuming that event e_j occurs after e_i ¹.

From this, there are a number of inter-event time distributions one can consider. First is the inter-event time distribution across the entire temporal network

$$\phi = \{\tau_{i,i+1} | e_i \in E\}$$

¹ For events with duration, the inter-event time is given by $\tau_{ij} = t_j - (t_i + \delta_i)$ provided $t_j > (t_i + \delta_i)$.

This is a global measure of activity patterns within the network. Two other useful distributions are the inter-event times of nodes being the source of an event, or the node being the target of an event. Let

$$E_i^{\text{out}} = \{e \in E \mid e = (i, k, t) \text{ for any } k \in V \setminus \{i\}, t \in T\},$$

$$E_i^{\text{in}} = \{e \in E \mid e = (k, i, t) \text{ for any } k \in V \setminus \{i\}, t \in T\}$$

be the sets of events involving node i as a source and target respectively. Then the corresponding inter-event times are given by

$$\phi_i^{\text{out}} = \{\tau_{jk} \mid e_j, e_k \in E_i^{\text{out}} \text{ and } \nexists e_l \in E_i^{\text{out}} \text{ s.t. } t_j < t_l < t_k\},$$

$$\phi_i^{\text{in}} = \{\tau_{jk} \mid e_j, e_k \in E_i^{\text{in}} \text{ and } \nexists e_l \in E_i^{\text{in}} \text{ s.t. } t_j < t_l < t_k\}.$$

This gives the behaviour of an individual in the network and also the behaviour of other nodes behaviour towards it. These two distributions can be combined to give a measure of the general participation of a node in the network

$$\phi_i = \{\tau_{jk} \mid e_j, e_k \in E_i \text{ and } \nexists e_l \in E_i \text{ s.t. } t_j < t_l < t_k\}$$

where $E_i = E_i^{\text{out}} \cup E_i^{\text{in}}$ is the set of all events for which i is either a source or target. In this case it is not possible to distinguish between the behaviour of the node itself and of others. Finally one can consider the inter-event times of contact between two nodes. The distribution

$$\phi_{ij} = \{\tau_{kl} \mid e_k = (i, j, t_k), e_l = (i, j, t_l)$$

$$\text{and } \nexists e_m \in E \text{ s.t. } e_m = (i, j, t_m) \text{ where } t_k < t_m < t_l\}$$

captures times between particular interactions occurring in the network.

In networks of human communication such inter-event times have been found to be *bursty* in nature - large gaps in activity are followed by periods of high activity [83, 84] (see Figure 2.3). Recognising human temporal behaviour plays an important role in identifying and detecting ‘bots’ in online conversations on Twitter (see Chapter 5). As a word of caution, it is not possible to fully determine the inter-event time distribution empirically from a finite time window as longer inter-event times are less likely to be sampled during the observation window. One study [85] investigates the bias introduced by finite time windows and corrects for this problem by modelling the event sequence by using stationary renewal processes.



Figure 2.3: An example of bursty temporal behaviour. Below, events are represented by a vertical line at the time they occur. The closer two lines are, the closer in time they occur, and the shorter the inter-event time. Above, a moving-window count of the number of events occurring.

2.1.2 Representations

Temporal networks do not have a unique representation. There is in fact a spectrum of representations of temporal networks from the fully aggregated static network, to the atomic event sequence. Each representation of a temporal network aggregates the temporal data to a different extent and each has its own advantages and disadvantages. The temporal representation is usually chosen to match the detail of the data under study and such that the analysis is tractable.

Network Snapshots

By far the most common representation of a temporal network is in the form of a sequence of static networks, representing the temporal network over a prescribed time interval. Let $P = \{t_{\min} = t^{(0)}, t^{(1)}, \dots, t^{(n)} = t_{\max}\}$ with $t^{(0)} < t^{(1)} < \dots < t^{(n)}$ be a set of points which partition the interval $[t_{\min}, t_{\max}]$. Note that these intervals do not need to be of fixed width, however a uniform partition over the time frame is usually adopted. The adjacency matrix for the network in the interval $I_k = (t^{(k-1)}, t^{(k)}]$ is defined such

that

$$(\mathbf{A}_k)_{ij} = \begin{cases} 1 & \text{if } \exists(i, j, t) \in E \text{ s.t. } t \in I_k \\ 0 & \text{otherwise} \end{cases}.$$

For events with duration the condition for a non-zero entry becomes $\exists(i, j, t, \delta) \in E$ s.t. $[t, t+\delta] \cap I_k \neq \emptyset$, that is, the event occurs at some point during the interval but may start before and/or finish after. The adjacency matrix can also be weighted according to the number of events between nodes in each interval (activity). The weighted adjacency matrix for the interval I_k is given by

$$(\mathbf{W}_k)_{ij} = |\{(i, j, t) \in E | t \in I_k\}|.$$

Adding weights to each interaction is especially useful if there are expected to be large numbers of interactions between pairs of nodes within each interval.

Representing the temporal network in this fashion has a number of advantages. As each snapshot of the network is static, then static analysis of each snapshot can be conducted and compared across the time frames. For instance, this could be used to track the clustering coefficient or average degree of the network over time. More importantly, temporal walks across the network can be realised by the multiplication of subsequent adjacency matrices. For example, $(A_k A_{k+1})_{ij} = 1$ if there exists a single temporal path from i to j via a third node in the interval $(t^{(k-1)}, t^{(k+1)}]$. The notion of temporal walks is important for calculating a number of centrality measures as well as in the study of dynamical processes acting on the network (in particular spreading dynamics). Aggregation of the temporal network causes some information to be lost but provides a reasonably proxy to the temporal network for most purposes.

Adjacency tensor (multiplex network)

The adjacency tensor (also know as the *time-varying graph* [86]) encodes the temporal network into a 4D tensor. The adjacency tensor \mathbf{A} is defined such that

$$\mathbf{A}_{i,k_i,j,k_j} = \begin{cases} 1 & \text{if } \exists(i, j, t, \delta) \in E \text{ s.t. } t \in I_{k_i} \text{ and } t + \delta \in I_{k_j} \\ 0 & \text{otherwise.} \end{cases}$$

In this representation the temporal network can be considered as a *multilayer* network [21] where each layer represents an interval of time. Nodes in one time interval can communicate with other nodes (and itself) in other time intervals. A typical example of this representation involves phone calls. A telephone call may be initialised in interval I_{k_i} but any possible information transfer (or at least action based on information transfer) may not occur until the call ends in the interval I_{k_j} . This is an improvement on considering information transfer to be instantaneous which is often unphysical. It is worth noting that this representation fully captures all information in an event sequence if the time frame is partitioned such that each event start and end time falls into its own interval.

Examples of these representations are given in Figure 2.4.

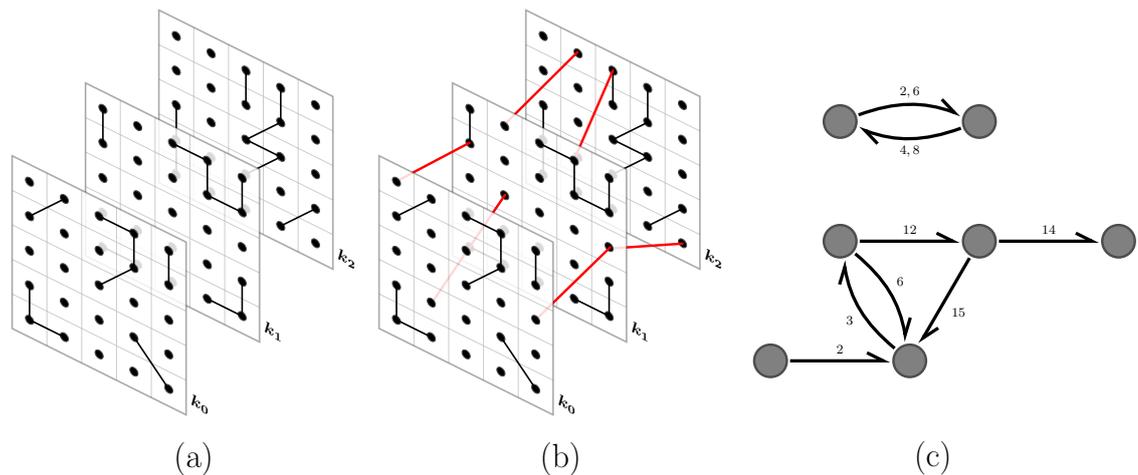


Figure 2.4: Temporal network representations. (a) Network snapshots, or a series of adjacency matrices. (b) Adjacency tensor. Similar to the network snapshots however nodes can interact with other nodes (and themselves) between intervals. (c) Contact sequence. Events are labelled by the time they occur and are not binned or aggregated.

Static networks

At the other end of the spectrum, there are a number of static representations which aim to capture a particular aspect of the temporal network. The *reachability graph* [87, 88]

is defined as

$$\mathbf{A}_{ij}^{\text{reach}} = \begin{cases} 1 & \text{if } \exists \text{ a temporal path from } i \rightarrow j \\ 0 & \text{otherwise.} \end{cases}$$

By explicitly incorporating temporal paths through the network the reachability graph does not falsely imply the existence of paths through the network where they do not exist, as is the case with the aggregate network (Figure 2.1). Information is lost in the length and route of temporal paths which can be vital for the analysis of spreading dynamics; a virus is more likely to spread making as few ‘hops’ between agents as possible, assuming a fixed probability of transmission on contact.

More simple approximations come in the form of *activity-weighted* and *time-weighted* graphs. Let $E_{ij} \subset E$ be the set of temporal events which take the form (i, j, t, δ) for any t, δ . The activity-weighted graph is defined by the weighted adjacency matrix

$$\mathbf{A}_{ij}^{\text{aw}} = |E_{ij}|$$

which gives a measure of how active each edge of the network is over the time frame. Similarly, the time-weighted graph captures the total duration two nodes are in contact and is defined by

$$\mathbf{A}_{ij}^{\text{tw}} = \sum_{(u,v,t,\delta) \in E_{ij}} \delta.$$

The use of static representations is not to be discounted. While the analysis of static graphs simplifies the true network by aggregating (or even ignoring) the temporal dimension, it is often useful in cases where the network evolves on timescales significantly slower than a process acting on the network. For example, the road network can be safely assumed to be static over a daily or weekly time frame. However, if the traffic flows of transport were modelled then the incorporation of temporal data would potentially be more enlightening due to it being able to capture the large fluctuation in traffic volumes along edges (streets) across the day.

2.2 Temporal Network Analysis

Static networks have been studied since Euler first considered the problem of the Königsberg bridges in 1736 and since then many methods have been devised to characterise the topological structure of these networks. The addition of a temporal degree of freedom to these networks requires these measures to be altered to utilise the extra information presented.

In this section, we highlight a number of different measures used to describe temporal networks and their inter-event time distributions.

2.2.1 Centrality Measures

Centrality measures capture the importance of nodes or edges with respect to a particular property of the network. For static networks, centrality measures have been employed to great effect in assessing node vulnerability to epidemic spread [19], ranking stations in train transport infrastructures [89] or identifying influential users of a social network [90], as well as in many other areas. Perhaps one of the most influential centrality measures is that of Google's *PageRank* [91], used as part of their search engine technology to rank websites by their importance and relevance to search terms.

For temporal networks similar notions of centrality can be constructed [92–94, 80]. Some temporal centrality measures are defined on a series of static networks, whereas others fully utilise the unaggregated temporal information available. The simplest generalisation is to replace the notion of paths in static networks with time-respecting paths. Here we outline some of the more common centralities used in the literature.

Degree

The degree of a node is the most fundamental centrality, capturing the number of edges from that node to any other node. For a sequence of temporal events, the degree of a node u at a time t is given by the number of events $(u, \cdot, t_i, \delta_i)$ such that $t_i < t < t_i + \delta_i$.

When events become instantaneous ($\delta_i \rightarrow 0$) it is useful to consider the degree of a node within a time interval,

$$C_d(u, k) = \sum_j (\mathbf{A}_k)_{uj},$$

i.e. the degree of the node in each temporal snapshot.

Temporal betweenness

Betweenness centrality on static networks is based on shortest paths, in particular, measuring the number of shortest paths which pass through a focal node. Generalising to temporal networks, the temporal betweenness centrality [92] is based on either the shortest or fastest temporal paths through the temporal network. The temporal betweenness is explicitly given by

$$C_b(u) = \frac{\sum_{u \neq v \neq w} \eta_u(v, w)}{\sum_{u \neq v \neq w} \eta(v, w)}$$

where $\eta_u(v, w)$ is the number of shortest (fastest) paths between v and w passing through u and $\eta(v, w)$ is the total number of shortest (fastest) paths between v and w . Here η is calculated over the complete window of observation of the temporal network. One can also allow C_b to evolve over time by only counting temporal paths that occur before a time t , or in an interval around t .

The temporal betweenness centrality suffers from the same issues as its static predecessor, namely that the focus on strictly shortest (fastest) paths is unrealistic for many real world systems; the shortest paths may only be fractionally shorter than alternatives and often full information of the network required to take the shortest path is not present.

Temporal closeness

On static networks the closeness centrality measures the inverse total distance to all other vertices in the network. Swapping shortest paths for shortest temporal paths the

temporal closeness centrality [93] is given by

$$C_c(u) = \frac{N - 1}{\sum_{u \neq v} d_t(u, v)}$$

where $d_t(u, v)$ is the path or temporal distance between u and v . The temporal distance, more commonly known as the *latency*, captures the time taken for a path from u to reach v , provided it reaches v before time t . As with static networks, the closeness of a disconnected network is undefined. With temporal networks it is even more unlikely that a temporal path between two nodes exists for the entire time period, and latencies may be infinite. An alternate temporal closeness based on reciprocal latencies is given by

$$C_e(u) = \frac{1}{N - 1} \sum_{u \neq v} \frac{1}{d_t(u, v)} \quad (2.1)$$

where $\frac{1}{d_t(u, v)}$ is defined to be zero if there does not exist a time respecting path from u to v arriving before time t [95]. This can be seen as a measure of *temporal efficiency* - smaller latencies lead to a larger value.

Communicability

Temporal betweenness and closeness are based upon the shortest paths through the network. A contrasting assumption is that traversal across the network is random, and any quantity that resides on the network diffuses across it. In this setting it makes sense to consider all time-respecting paths of all lengths. However, a quantity moving along longer paths may be more susceptible to loss, or is less likely to use a longer path to begin with. Therefore longer walks should carry less weight than shorter walks. This idea is the basis of the communicability centrality. First defined for static networks to understand a physical system of coupled springs [96], the communicability centrality has found many uses in applications such as understanding brain function [97, 98] and assessing infection vulnerability [99, 100].

A full discussion of communicability is deferred to Chapter 3 where it is examined in more detail.

2.2.2 Reachability and Connectivity

We are often concerned with how a process evolves on a temporal network, how quickly and widely can a disease or information spread across the population. These properties are dependent not only on the activity of particular nodes or contact events but also the time ordering of these events [80, 70].

Revisiting Figure 2.1 we see that while a path exists between A and B in the aggregated network, a temporal path does not. In this instance B is not reachable from A during the observed time window, however A is reachable from B. So, if A carried an infection at the start of the time period then the infection would not be able to reach B.

The reachability of a pair of nodes can be characterised by the latency between them [87]. One can subsequently calculate the average latency τ across all pairs of nodes which are connected by a temporal path. Averaging in this fashion can be misleading as a node that can reach the entire network may have the same average latency as one which can reach only a small fraction. This can be captured by the reachability ratio f ; the fraction of nodes in the network reachable from a node at a point in time. The pair (τ, f) then characterises both aspects of reachability. A single measure of reachability is given by the harmonic mean of the latencies between all node pairs [101]. This is simply the average temporal efficiency of the network (2.1) seen previously. The temporal efficiency of the network suffers the same issue as when considering τ alone in that it is impossible to distinguish between nodes with low f and those with high τ .

Beyond reachability, the static notions of connectedness provide another view on the connectivity of networks. These too have been generalised to the temporal network setting [102].

Definition 2.2.1 (Strong connectedness). Two nodes u, v of a temporal network are strongly connected if there exists a temporal path from u to v and also a temporal path from v to u .

Definition 2.2.2 (Weak connectedness). Two nodes u, v of a temporal network are weakly connected if there exists a temporal path from u to v and from v to u on the underlying undirected temporal network where event directions are ignored.

Connectedness can also be thought of as reciprocated reachability. Note that connectedness is a reflexive and symmetric relation but it is not transitive as information on the temporal paths from u to v and v to w is not informative of paths from u to w . As strong/weak connectedness is not an equivalence relation the strongly/weakly connected components can not be defined in terms of the connectedness of node pairs but is instead given by:

Definition 2.2.3 (Strongly/Weakly connected component). A set of nodes of a temporal network is a temporal strongly/weakly connected component of the network if each node of the set is strongly/weakly connected to all other nodes in the set.

Studying the temporal components of 100k Facebook users over the course of six months revealed that while the static aggregate network has a largest strongly connected component of 32% of all nodes, the temporal strongly connected component was less than 0.15% of all nodes [102]. This represents a drastic disparity between what is often assumed to be a network of mutual reachability (the static aggregation) and the true level of mutual connectivity. This same study found that nodes which belonged to multiple strongly connected components played a pivotal role within the temporal network.

Despite the promise of temporal components, finding them is proven to be an NP-complete problem. This unfortunately limits this type of study to smaller networks.

2.2.3 Inter-event Times

The timings between events have been of interest to researchers across many fields. In particular the IETs of human behaviour have garnered a number of studies in recent years [103, 104, 83, 105, 106]. Previously it was difficult to analyse the IETs of interaction due to a lack of recording. The careful documentation of the correspondence of Charles Darwin and Albert Einstein allowed researchers to study the response times to letters [107]. Surprisingly the response times of these authors were found to share the same scaling law as with modern email, albeit with a different scaling exponent. Now, all activity on the internet or over cellular phone networks can potentially be tracked and

monitored. Examples include, but are not limited to, phone calls [104], emails [105], and website clicks [106].

The IET distributions are of interest because they give clues to the behaviour of the agents interactions between themselves and with the environment. The IETs of telephone calls exhibit a circadian pattern [104, 83, 105] along with weekly activity patterns. The IETs are also bursty in nature. The burstiness of human interaction is seen as an important characteristic and can be quantified by a burstiness parameter [84] given by

$$B = \frac{\sigma_\tau - m_\tau}{\sigma_\tau + m_\tau}$$

where m_τ and σ_τ are the mean and standard deviation of the IET distribution respectively. For finite temporal networks the variance is always finite $B \in (-1, 1)$. A completely bursty distribution gives $B = 1$, exponentially distributed IETs lead to $B = 0$ and for completely periodic IETs $B = -1$. This parameter is not robust to finite time window sampling, namely the value $B = 1$ can only be achieved when the time window to sample IETs is infinite. In [108] the parameter has subsequently been improved upon to be independent of finite sample effects. This bursty behaviour can be replicated by a simple model of task scheduling [83] although this may not be the only mechanism at play.

The IETs are not only interesting from a behavioural viewpoint. The timings of events can have major effects on any process which occurs on the temporal network [109, 110]. For susceptible-infected (SI) models of epidemic spread, the burstiness of human contact sequences have been shown to hinder the spread of infection [110, 63]. This behaviour is confirmed with numerical [111, 112] and analytical results [110, 113]. The slow spreading on temporal networks with heavy tails can be explained by the *waiting time* of the process on each node, that is, the time between receiving an incoming event to sending an event outwards afterwards. The expected waiting time for power-law distributed IETs is greater than for less heavy tailed distributions. This means that any spreading process potentially has long periods of inactivity before it can move around the network. However, long waiting times are not the only factor in the speed of spreading on the network. The temporal ordering of events has also been shown to have a significant

effect on the proliferation of a process [114]. This effect occurs in the bulk of the distribution where there are more events in contrast to the tail.

The effect of bursty IETs is not so clear cut. There are different data sets and process spreading mechanisms where the bursty nature of the IETs have been shown to instead accelerate the spread of the processes [115, 116]. It is therefore not possible to generalise the effect that temporal networks have on network dynamics and instead each process must be considered on its own.

Finally, care needs to be exercised when inferring IET distributions from data. Sampling IETs over a finite time window leads to a linear cut-off to the observed IET distribution at the end of the time window [85], assuming that the IET distribution is generated from a stationary renewal process. This can lead to qualitatively different inferences of the IET distribution (a power-law being confused for an exponential, for example) and major quantitative differences in the distribution moments, particularly those sensitive to the tail of the distribution. One possible way to account for this sampling bias is to use the non-parametric Kaplan-Meier estimator [117].

2.3 Discussion

In this chapter we have described a temporal network and the various representations it can take. Incorporating temporal information into network problems highlights the shortcomings of static aggregated networks; there can be static paths between nodes that exist where temporal paths do not. This has major consequences for the study of dynamical processes on these networks.

Along with the notions of walks and paths, centrality measures have been generalised from static to temporal networks. Centrality measures continue to play an important part in assessing node roles within the network structure. The properties of these centralities are not fully understood (owing to a lack of suitable null model for temporal networks) and some measures are dependent on the level of temporal aggregation of the network. In the next chapter we further explore one particular measure, the communicability centrality.

As with static networks, the reachability and connectivity of nodes plays an important role in how processes evolve on the network. Furthermore, the temporal ordering of events and the inter-event times are equally crucial aspects to consider for spreading dynamics. This motivates the work of Chapter 4 where we use inter-event times as the basis of a novel event-centric temporal network representation.

The work outlined in this chapter is a small survey of recent research into temporally evolving networks. There are many different aspects that are not covered such as temporal communities [118], models of network evolution [119–121], and adaptive networks [122], all of which form part of a rich and rapidly growing field.

3

Communicability Centrality

Centrality measures are widely used to identify the key nodes and edges in social networks, internet structure, biological networks, and transport infrastructure [18, 123, 124, 91]. Historically this type of analysis has been restricted to static networks, however as seen in Chapter 2 efforts have been made towards extending these measures to temporal networks. In this chapter we focus on the *communicability* centrality, a generalisation of the *Katz* centrality [17]. While many network properties concern shortest paths (or *geodesics*) through the network, information spreading in a network does not necessarily follow these paths [125], neither does it need to explore the full network before arriving at its destination. Communicability centrality incorporates information on all possible walks through the network through a weighted count of all walks from one node to another.

Communicability has seen a number of uses across different fields. By modelling the

activity of the brain as a temporal network, using fMRI scans the communicability centrality highlighted a primary network of nodes which, under a targeted iterative attack (node removal) reduced brain function quicker than random attack [98]. Modelling the brain instead as anatomically distinct regions, another study [97] was able to use a reduced communicability matrix (by singular value decomposition) to distinguish between patients who had previously suffered a stroke and those who had not. As a walk counting measure, communicability centrality is well suited to studying spreading processes which are path-dependent, such as the susceptible-infected (SI) model. Communicability centrality has been measured in digital networks such as email [99], as well as proximity networks of patients and staff at an emergency department of a hospital [100]. In both cases it identified a set of top spreaders of infection, but it did not always outperform other, less complex, measures. Other novel uses of communicability centrality have been to identify critical edges in opinion spreading dynamics [126, 127], and to devise a ranking system for competitive sport [128].

Communicability Centrality in Social Media

Modern day social networks evolve rapidly and typically involve millions of nodes. With an estimated seven thousand new pieces of content (tweets) posted every second on the Twitter social network [129] it can be difficult to fully quantify the network structure and node behaviour. In this sense it is useful to focus on a select number of individuals who are *influential* or important. This view is also taken by businesses, who with limited resources, look to find agents who can proliferate a brand message or who may be susceptible to a viral campaign. Communicability provides a means to measure the influence of users in a social network, taking into account both activity levels and the network structure. Furthermore the communicability measure is in fact two distinct measures which, as shown in Section 3.1, describe the ability of a user to broadcast and receive information. Due to its relative infancy it is important to fully understand the measure and how it should be used for further study of social networks.

The utility of communicability has been proved on applying the measure to temporal networks extracted from the micro-blogging service Twitter. The communicability

measure has been able to identify key influencers in these networks [57]. The extraction of these influencers in these cases are insensitive to small parameter changes, and the set and ranking of influences generated from the communicability metric were highly correlated to the rankings of a panel of professional social media analysts [58]. The success of the measure is such that it has been implemented commercially at scale, allowing the analysis of thousands of messages in near real-time [56].

Chapter Outline

In Section 3.1 we review the literature surrounding communicability and give definitions for the different flavours of the metric for both static and temporal graphs. In Section 3.2 we analyse how the aggregation of the temporal network changes the interpretation of the communicability metric and introduces errors. We propose various partitions of the temporal network to remove these errors while maintaining computational efficiency. We present a series and parallel algorithm to calculate the communicability metric efficiently and free from aggregation error in Section 3.3, before finally concluding in Section 3.4.

3.1 Communicability Centrality

The key idea of communicability centrality is to consider all walks through the network and downweigh longer walks such that their contribution is less than shorter walks. For now, consider only a static network described by the adjacency matrix \mathbf{A} . Communicability makes use of the fact that walks can be counted via powers of the adjacency matrix. For example, the entries of the adjacency matrix \mathbf{A}^1 gives the possible walks of length one between any two nodes. The square of the adjacency matrix, \mathbf{A}^2 , counts the number of walks of length two, which is seen clearly as $(\mathbf{A}^2)_{ij} = \sum_k \mathbf{A}_{ik} \mathbf{A}_{kj}$. This counts the number of walks from i to j via an intermediate node k . Following the same logic we can generalise to see that \mathbf{A}^k captures all walks of length k .

One possible way to downweigh longer walks is to scale by the factorial of the walk

length. This leads us to define the *exponential communicability matrix*

$$\mathbf{Q}_{\text{exp}} = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!} = e^{\mathbf{A}}, \quad (3.1)$$

where we have used the fact that the infinite sum is the Taylor expansion of the matrix exponential and that $\mathbf{A}^0 = \mathbf{I}$.

Another possible weighting is by a constant factor $\alpha < 1$. With this weighting the *resolvent communicability matrix* can be written neatly as the matrix resolvent

$$\mathbf{Q}_{\text{res}} = \sum_{k=0}^{\infty} \alpha^k \mathbf{A}^k = (\mathbf{I} - \alpha \mathbf{A})^{-1}. \quad (3.2)$$

With this scheme careful consideration needs to be given to the parameter α to ensure the convergence of the infinite sum. To ensure convergence we require that $\alpha < 1/\rho(\mathbf{A})$ where $\rho(\mathbf{A})$ is the spectral radius of \mathbf{A} , i.e., the largest eigenvalue by absolute size. Both weighting schemes are useful for different applications, however for reasons that will become clear in the next section we will focus on the resolvent communicability for the remainder of the chapter.

Choosing α

How should one choose the walk weighting parameter α ? In the limit of $\alpha \rightarrow 0$, terms of $\mathcal{O}(\alpha^2)$ become negligible meaning that the communicability centrality reduces to counting node degree over time. Due to this consideration α is chosen to be suitably large so that the communicability centrality has minimal correlation with node degree.

For the closely related Katz centrality [17] it is suggested that α should lie within the interval $\left[\frac{1}{2\rho(\mathbf{A})}, \frac{1}{\rho(\mathbf{A})} \right)$. Choices in the literature range from $\alpha = 1/(2\rho(\mathbf{A}))$, or inspired by the damping factor of Google's PageRank algorithm [91], $\alpha = 0.85/\rho(\mathbf{A})$. One approach, aims to chose α to match as closely as possible the resolvent and exponential schemes of communicability [130]. By minimising the norm of difference between the vectors $e^{\mathbf{A}}\mathbf{1}$ and $(\mathbf{I} - \alpha\mathbf{A})^{-1}\mathbf{1}$ where $\mathbf{1} = (1, 1, \dots, 1)^T$ one arrives at the expression

$$\alpha = \frac{1 - e^{-\rho(\mathbf{A})}}{\rho(\mathbf{A})}.$$

Using this value for α means that there is minimal difference between the exponential and resolvent based communicability centrality however there is no reason for these to be closely aligned other than for consistency.

3.1.1 Dynamic Communicability

Many systems such as communication networks or the spread of disease in a population are not well captured by static networks but are instead well represented by a temporal (or dynamic) network, as discussed in Chapter 2. The notion of communicability is easily extended to the temporal setting and takes the form of the *dynamic communicability* matrix [61].

We describe the temporal network as a series of adjacency matrices \mathbf{A}_k (see Chapter 2) capturing the aggregated connections in ordered intervals I_k (which are not necessarily of equal length). As with static networks, the products of adjacency matrices describe the number of dynamic or temporal walks through the network, provided that adjacency matrices are multiplied in time order such that $\mathbf{A}_{k_1}\mathbf{A}_{k_2} \implies k_2 > k_1$. This condition preserves and encodes the arrow of time in the system as matrix multiplication is non-commutative in general. As an example, the matrix

$$\mathbf{A}_{k_1}\mathbf{A}_{k_2}\dots\mathbf{A}_{k_n}, \quad k_1 \leq k_2 \leq \dots \leq k_n, \quad (3.3)$$

has an i, j element which counts the number of dynamic walks of length n from i to j where the m th step takes place in the interval I_{k_m} . This product also allows the walk to make multiple steps within one interval. As an example, take the temporal network given in Figure 3.1. There are only four possible temporal walks of length greater than one. These are given by the matrix products

$$\begin{aligned} (\mathbf{A}_1^2)_{ij} &= \delta_{i1}\delta_{j3}, \\ (\mathbf{A}_1\mathbf{A}_2)_{ij} &= \delta_{i2}\delta_{j4}, \\ (\mathbf{A}_1\mathbf{A}_3)_{ij} &= \delta_{i1}\delta_{j3}, \\ (\mathbf{A}_1^2\mathbf{A}_3)_{ij} &= \delta_{i1}\delta_{j4}, \end{aligned}$$

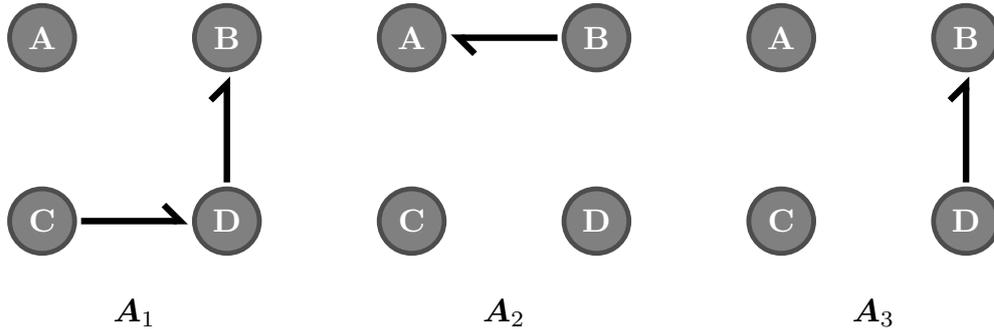


Figure 3.1: A simple temporal network.

where δ_{ij} is the Kronecker delta. All other (order preserving) combinations of these matrices are identically zero.

By a simple expansion, it is clear to see that all possible permutations of (3.3) are included in the product of matrix resolvents $(\mathbf{I} - \alpha \mathbf{A}_k)^{-1}$ with the correct weighting of α corresponding to the walk length. This motivates the definition of the dynamic communicability matrix

$$\mathbf{Q}_n = \prod_{k=1}^n [\mathbf{I} - \alpha \mathbf{A}_k]^{-1}, \quad (3.4)$$

(where multiplication is on the right) or defined recursively,

$$\mathbf{Q}_{k+1} = \mathbf{Q}_k [\mathbf{I} - \alpha \mathbf{A}_{k+1}]^{-1} \quad (3.5)$$

with $\mathbf{Q}_0 = \mathbf{I}$ so all nodes are weighted equally initially. With this definition walks can make infinitely many steps in each interval or may “wait” at a node across multiple intervals before making the next step. Again, care needs to be taken to ensure that the communicability matrix converges. This amounts to requiring that each matrix resolvent converges, i.e., $\alpha < 1/\max_k \rho(\mathbf{A}_k)$. The multiplication of the weighting factors show why resolvent communicability is preferred to the factorial weighting. The function $f(x) = a^x$ uniquely satisfies $f(x)f(y) = f(x+y)$ so that $(\alpha^m \mathbf{A}^m)(\alpha^n \mathbf{A}^n) = \alpha^{m+n} \mathbf{A}^{m+n}$. This means that walks of length m combined with walks of length n carry the correct weighting of a length $m+n$ walk. However this is not true for the factorial function as $\left(\frac{\mathbf{A}^m}{m!}\right) \left(\frac{\mathbf{A}^n}{n!}\right) \neq \left(\frac{\mathbf{A}^{m+n}}{(m+n)!}\right)$.

The dynamic communicability matrix captures all products of the form (3.3) and so the entries \mathbf{Q}_{ij} capture how well information can be passed from node i to node j . In many

contexts we are not concerned with two individuals in particular, and instead want to understand how well a node can transmit and receive information to the entire network. This leads to the definition of the *broadcast* and *receive* vectors

$$\mathbf{b}(k) := \mathbf{Q}_k \mathbf{1} \quad \text{and} \quad \mathbf{r}(k) := \mathbf{Q}_k^T \mathbf{1} \quad (3.6)$$

respectively, where $\mathbf{1} = (1, 1, \dots, 1)^T$. These correspond to the row and column sums of the matrix \mathbf{Q}_k . Under the assumption that information is not passed from node to node with absolute certainty then the number of ways for information to be passed from one node to another becomes important. The broadcast vector counts the total number of temporal walks starting (or broadcast) from each node. Similarly the receive vector counts the total number of temporal walks ending (or received) at each node. We can therefore use $\mathbf{b}(k)$ and $\mathbf{r}(k)$ as a proxy to measure how well information can travel from or to a node respectively.

3.1.2 Running Dynamic Communicability

A further generalisation of communicability [131] argues that messages that were sent a long time ago may be less important than those sent more recently. Indeed, as temporal networks grow, the entries of the communicability matrix increase dramatically as the number of possible walks increases with each iteration, i.e., $\|\mathbf{Q}_{k+1}\| > \|\mathbf{Q}_k\|$ where $\|\mathbf{Q}\| = \mathbf{1}^T \mathbf{Q} \mathbf{1}$. This can become problematic when the number of intervals in the temporal network is large. Examples of where this can be an issue include hourly email or telephone over a period of months or years, or second-by-second social network messages over the course of a week. Firstly, this causes potential overflow errors when calculating the communicability matrix numerically if it is not normalised at each iteration. Secondly, nodes that have a high broadcast component early on in the time period will continue to be ranked highly. This is due to them having more potential future events to extend the current walks starting at that node than a node that has a similar broadcast score at a later time.

Combining the downweighing in both time and walk length, the *running dynamic*

communicability matrix \mathbf{S}_k [131] is defined iteratively by

$$\mathbf{S}_k = (\mathbf{I} + e^{-\beta\Delta t_k}\mathbf{S}_{k-1})(\mathbf{I} - \alpha\mathbf{A}_k)^{-1} - \mathbf{I}, \quad (3.7)$$

where $\mathbf{S}_{-1} = \mathbf{0}$. This is indeed a generalisation of \mathbf{Q} ; if $\beta = 0$, $\mathbf{S}_k + \mathbf{I} \equiv \mathbf{Q}_k$. Moreover, in the limit $\beta \rightarrow \infty$ the process is memoryless and the communicability matrix at any point in time is given by the matrix resolvent for that interval. Furthermore it fixes the issues faced in longer temporal networks as information transmission is allowed to decay as the age of the walk and nodes can have large broadcast and receive components at particular times and smaller components afterwards.

Finding a Null Model For Communicability

For static communicability we saw that there have been numerous suggestions for the parameter α . Similarly concerns around parameter choices are present for the running dynamic communicability, however there are now two parameters, α and β to consider and tune appropriately.

One study [132] investigates how the running communicability measure on a null model of a temporal network depends on the parameters α and β . By considering a null model where contacts are random between any two nodes and the contact times obey a Poisson process with uniform rate, they concluded that for the measure to remain finite, that “old” walks need to decay as quickly as new ones are added. By normalising the expected receive score of any node in the system, the resulting expression

$$\beta \approx \frac{2\alpha\mu}{N}$$

is obtained where N is the number of nodes and the average time between events is given by μ^{-1} . This gives a useful relation between the two parameters, however it relies on knowing the system size and event rate in advance which is not always the case. A further issue lies in the null model of the temporal network; observed real world networks are very far from random which means that this relation cannot be put into practice.

3.1.3 Communicability in the Continuum Limit

Taking a limit of smaller time steps in (3.7) results in a continuous time version of the metric. The continuous time *dynamic communicability* matrix $\mathbf{S}(t)$ is introduced in [94] and evolves in time according to

$$\mathbf{S}'(t) = -\beta\mathbf{S}(t) - [\mathbf{I} + \mathbf{S}(t)] \log(\mathbf{I} - \alpha\mathbf{A}(t))$$

where $\mathbf{S}(0) = \mathbf{0}$. The log denotes the matrix logarithm which is defined as the inverse of the matrix exponential.

The motivation for a continuous time metric is to remove any issues that arise due to the selection of a time frame duration and to capture more realistically any non-instantaneous types of communication. A continuously evolving network can be defined by an adjacency matrix where each entry \mathbf{A}_{ij} is itself a function of time. As an illustrative example imagine a phone call between two nodes i and j . If the phone call starts at t_0 and ends at t_k then the corresponding adjacency matrix entry will be

$$\mathbf{A}_{ij}(t) = \mathbf{A}_{ji}(t) = \begin{cases} 1 & \text{for } t \in [t_0, t_k] \\ 0 & \text{otherwise} \end{cases}$$

provided there is no further activity between the two nodes. This is not a smooth function, although one could equally define an adjacency matrix whose entries are smooth functions of time.

3.2 Aggregation Errors in Communicability Calculation

Despite its widespread use there have been few studies to understand the effect of partitioning time on communicability scores and how this influences the choice of parameters. In this section we describe and quantify the errors involved in the aggregation of temporal events into discrete time intervals and introduce the associated hierarchical partition of time which removes these errors.

3.2.1 Error Types

There are three distinct types of error introduced by the aggregation of temporal events into discrete time windows. The errors introduced vary in magnitude depending on the network topology. More fundamentally, the interpretability of communicability as a weighted count of walks changes, depending on the presence of cycles in each adjacency matrix.

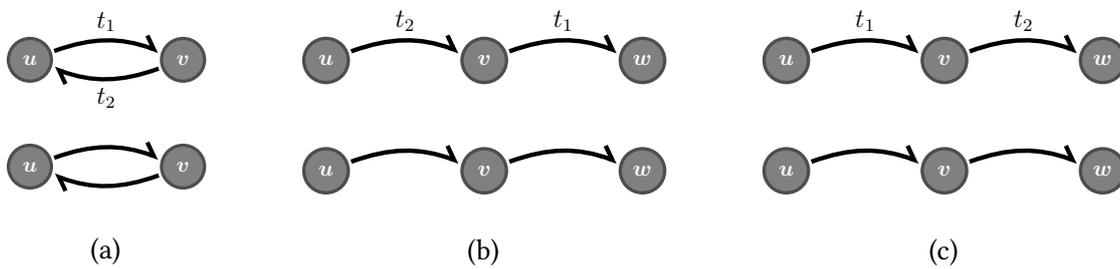


Figure 3.2: The scenarios in which the temporal aggregation of events introduces error into the communicability calculation. Here we assume $t_2 > t_1$. In each case, the network above represents the true, most granular behaviour while the network below gives all information available if the two events are aggregated. (a) Cycle error: when events are aggregated, an infinite walk cycling between the two nodes becomes possible. (b) Causality error: when events are aggregated, a walk from u to w becomes possible where it does not exist. (c) Time decay error: When events are aggregated, the age of the walk is unknown.

Cycle Error

The presence of cycles in a network results in the existence of walks of infinite length through the network. The simplest example of cycle formation is a reciprocated edge between two nodes (Fig. 3.2(a)). Suppose we have two subsequent events $e_1 = (u, v, t_1)$ and $e_2 = (v, u, t_2)$, where $t_2 > t_1$ which may form part of a larger temporal network. Suppose a time partition is chosen such that $t_1, t_2 \in I_k$ for some k , that is they contained

in the same interval. The two events are represented in a single adjacency matrix

$$\mathbf{A}_k = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Assuming $\alpha < 1$ and that neither node has participated in an event previously, the broadcast score of node u is

$$\mathbf{b}_u(k) = \sum_{s=1}^{\infty} \alpha^s = \frac{\alpha}{1 - \alpha}. \quad (3.8)$$

This can be calculated readily as the row sum of $\mathbf{Q} = \sum_s \alpha^s \mathbf{A}^s$.

Now instead, suppose that a time partition is chosen such that these two events occur in separate but subsequent time intervals I_k, I_{k+1} , forming two subsequent adjacency matrices

$$\mathbf{A}_k = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_{k+1} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

The broadcast score of node u is then given by

$$\mathbf{b}'_u(k+1) = \alpha + \alpha^2 = \alpha(1 + \alpha). \quad (3.9)$$

This is a factor $1 - \alpha^2$ of the aggregated broadcast score. For modest $\alpha = 0.5$, equating to a 50% transfer rate of information, the aggregated broadcast score overestimates the true broadcast score by 33%. For complicated network structures this overestimation can in fact be made arbitrarily large by including more cycles in the network.

While the inclusion of infinite cycles introduces numerical error it also raises the issue of the interpretation of the metric. When no cycles are present within any time interval the metric counts only walks along edges whose events have been physically observed in that interval. By contrast if a cycle is present, the metric counts an infinite number of walks along cycles in the same time interval. For example, the calculation of broadcast in (3.8) assumes that a message has been passed infinitely many times between the two nodes in the interval I_k , regardless of the interval length. The calculation (3.9) however only counts the two observed message passing events from the temporal network. This can lead to inconsistency in calculating the communicability metric as cycles may appear in some intervals but not others.

Causal Error

The second type of error, illustrated in Figure 3.2(b) occurs when temporal ordering is not preserved, leading to walks that are not feasible being counted. We refer to this as a causal error. Consider the case where there are two events $e_1 = (v, w, t_1)$ and $e_2 = (u, v, t_2)$. When both events are aggregated into one interval I_k the adjacency matrix becomes

$$\mathbf{A}_k = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

which has a non-zero entry in \mathbf{A}_k^2 , i.e., there is a path of length two. The broadcast score of node u is then

$$\mathbf{b}_u(k) = \alpha + \alpha^2.$$

Separating the two events into two intervals I_k, I_{k+1} preserves the relative temporal ordering of the two events, resulting in only paths of length one and the broadcast score of node u being

$$\mathbf{b}'_u(k+1) = \alpha.$$

Time Decay Error

The final and most minor error introduced by aggregation concerns only the dynamic communicability metric and the suppression of temporal walks by their age. In Figure 3.2(c) there are two events $e_1 = (u, v, t_1)$ and $e_2 = (v, w, t_2)$. If the events are aggregated into one interval, the broadcast score of node u is

$$\mathbf{b}_u(k) = \alpha + \alpha^2.$$

compared to

$$\mathbf{b}'_u(k+1) = (\alpha + \alpha^2)e^{-\beta\Delta t_{k+1}}$$

if the events appear in their own interval. This error will only become significant if β and the interval length is large. Consider a temporal network which is aggregated by day. An event occurring at 11.59PM could be considered a day older than an event occurring at 00.01AM the following day although there is relatively little time between the two.

3.2.2 Temporal Network Partitions

The examples above highlight the potential errors which can be introduced when applying the communicability metric to aggregated temporal networks. We now outline four particular partitions of time, their effect on the errors in communicability metric, and simple algorithms to calculate these partitions. For simplicity we assume the full temporal network is known in advance, however in most cases these partitions are easily extended to instances where it is not, such as in a real-time implementation.

Fixed-Length Interval Partition

The fixed-length interval partition divides the temporal network into intervals of equal length.

This partition is the most commonly used in the literature [58, 132, 99]. Partitioning the network in this manner has a number of advantages that make it attractive for study. One advantage is that the number of intervals can be calculated in advance and so the run-time of the algorithm can be estimated easily. The fixed-length interval partition is also the most simple as the partition requires no knowledge of the temporal network beyond the times at which events occur. However, to ensure convergence of the communicability matrix one needs to calculate the spectral radius of each adjacency matrix which is an $\mathcal{O}(N^2)$ operation. For convergence we require $\alpha < 1/\max_k(\rho(\mathbf{A}_k))$ where $\rho(\cdot)$ is the spectral radius. The maximum spectral radius possible for a network involving N nodes is $N - 1$, meaning for large systems the parameter α is potentially extremely small and the communicability metric will correlate strongly with the degree of each node.

The restrictions on α by the network topology make it difficult to assign a physical

meaning to the parameter and impossible to compare values across different temporal networks. For Google's PageRank centrality [91], the corresponding model parameter is the perceived (and to some extent measured) probability of following a hyperlink on a web page, rather than typing in a new internet address. No such meaning can be given here.

For real-time implementation there is an added caveat; if the latest interval has an adjacency matrix with a larger spectral radius than those previously calculated then α needs to be reduced to conform to the new restriction. As a result, all previous communicability scores need to be recalculated with the new parameter which can become a computationally intensive task.

The fixed-length interval partition has the potential to carry all three types of calculation error.

Acyclic Partition

An acyclic partition of the temporal network consists of a set of intervals covering the time frame such that each adjacency matrix which encodes the interactions over each interval is acyclic. This is equivalent to requiring that each adjacency matrix is *nilpotent*, i.e., there exists an m such that $\mathbf{A}^m = \mathbf{0}$. Following previous notation the partition is formally given as the set $\{I_0, I_1, \dots, I_n\}$ such that for all $k = 0, \dots, n$ there exists an m such that $\mathbf{A}_k^m = \mathbf{0}$.

The eigenvalues of a nilpotent matrix are special in that they are all zero [133]. Consequently the spectral radius of each matrix is trivially zero. The restriction placed on α in the calculation of the communicability matrix reduces to α being a positive finite constant. With free reign to choose α this allows a physical meaning to be attached to the parameter as well as allowing the use of a constant parameter across data sets. In the information transfer setting α can be seen as a probability of successfully passing information across an edge provided $0 \leq \alpha \leq 1$.

The average length interval in an acyclic partition depends wholly on the underlying system being studied. For electronic instantaneous communication (Twitter, email, etc.)

the most likely cause of cycle creation is through reciprocated messages. This can happen in the order of tens of seconds for short messages, to hours for longer messages [134, 135].

The acyclic partition removes the error of counting infinitely many paths within a time frame, however it does not guarantee that causality is preserved. As the partition produces variable length intervals the decay of paths is also uneven.

There are many ways to detect cycle formation in growing networks. Bender et al. [136] and Haeupler et al. [137] use a two way search to maintain a topological ordering of events as well as to detect cycles. The best of these algorithms work at $O(M_k^{3/2})$ where $M_k \ll M$ is the number of events in a given time step. This calculation is insignificant compared to the matrix inversion and multiplication involved in the calculation of communicability at each time frame.

Causality-Preserving Partition

A causality-preserving partition ensures that within each interval the causal relationship of any two adjacent events, i.e. events that share at least one node, is preserved. This equates to enforcing that all events for which a node is a source occur after events where the same node is a target. This condition also prevents the formation of cycles and hence is a stricter partition than the acyclic partition. As we have explicitly removed the causality error described earlier and there are no cycles in each adjacency matrix the only error realised with this partition is the incorrect time decay of temporal walks.

True Partition

The true partition of the temporal network is where each event occurs exclusively in its own interval. This assumes that no two events can occur at the same time. However, in real life data temporal events are recorded at discrete times, such as every second. This makes it impossible to always guarantee a true partition is possible, however we will make the assumption that events occur at unique times or that the chance of two events

occurring at the same time is negligible¹.

We will consider this partition of the network as the ‘ground-truth’ when calculating the communicability centrality.

3.2.3 Quantifying Errors

To quantify the errors introduced with the time aggregation of temporal networks (summarised in Table 3.1), we calculate the dynamic communicability matrix \mathbf{Q} for the fixed-time and true partitions across three types of synthetically generated temporal networks.

To build a temporal network from a static network $G = (V, E)$ (where V is the set of vertices and E the set of edges), at each iteration we randomly pick an edge $(i, j) \in E$ and assign it a time corresponding to the iteration number k . This results in a temporal event (i, j, k) which forms part of the temporal network. The iteration can be repeated until the desired number of temporal events have been created. Allowing time to evolve discretely has the advantage that we know how many events occur within a time window and also that no two events can co-occur. In this assessment we will consider three network structures: a complete graph (all-to-all), an ensemble of Erdős-Rényi (ER) graphs, and an ensemble of ER-like acyclic graphs, examples of which are shown in Figure 3.3. The inclusion of the acyclic graph leads to an acyclic temporal network, allowing us to understand the effects of loops in the fixed partition scheme without creating an acyclic partition.

For each temporal network generated in this fashion we calculate the dynamic communicability matrix \mathbf{Q} , and the broadcast vector \mathbf{b} at the end of the temporal network², using the true partition and fixed intervals of varying width. We could similarly repeat the analysis for the receive vector, however the results would be identical.

¹ If two events occur at the same time but do not share any nodes then they can be ordered arbitrarily, as far as the calculation of communicability centrality is concerned.

² Since we do not calculate the running communicability, i.e., $\beta = 0$, the final broadcast score gives us the most information.

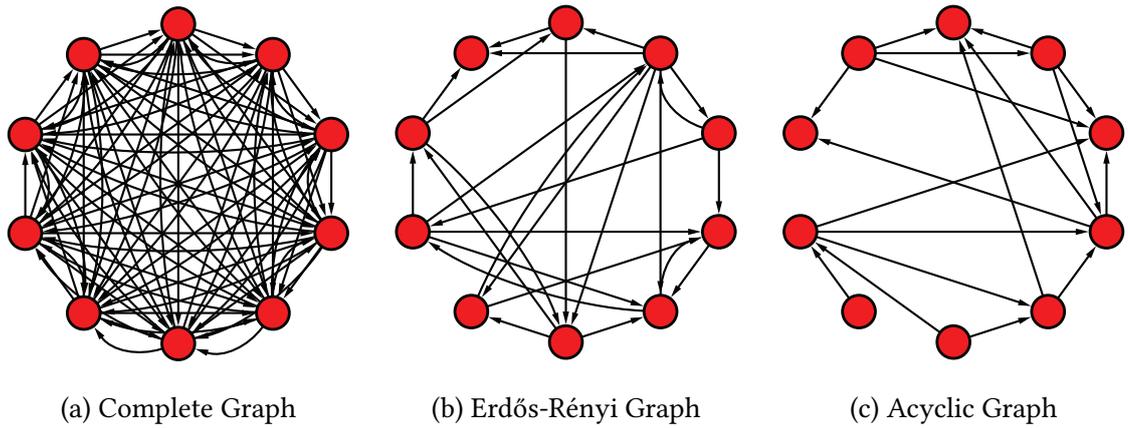


Figure 3.3: Static network choices which are used to generate temporal networks by repeated sampling of edges (with replacement).

The restrictions on the parameter α for the fixed-width partition require it to be chosen carefully so that the fixed-width and true partitions can be compared. The restriction on α is dependent on the window size Δt , and so to use a constant value we need to choose the smallest α across all values of Δt . This leads to a small value of α such that the broadcast vector is strongly correlated with the aggregated node degree. In order to address this we pick a fixed value for α and truncate the matrix resolvent to ensure convergence. For a fixed-width partition of width $\Delta t \in \mathbb{Z}$ we use the truncated resolvent

$$(\mathbf{I} - \alpha \mathbf{A})^{-1} \simeq \sum_{k=0}^{\Delta t} \alpha^k \mathbf{A}^k.$$

This ensures that if a dynamic walk of length Δt occurs over the Δt events in the partition then it will be captured. This means that the set of walks counted using the true partition is a subset of walks counted in the fixed-width partition communicability and so we can assess how many extra walks are being counted using the fixed-width partition.

To measure the differences between the different partitions we compare the broadcast vectors using the Pearson correlation coefficient, and Spearman's rank correlation coefficient. The Pearson correlation between two variables X and Y is given by

$$\rho_p(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

where $\text{Cov}(X, Y)$ is the covariance $\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$, and μ_i and σ_i are the mean and standard deviation of i . This gives the linear dependence

between the variables X and Y . Here, X and Y are the broadcast vectors calculated using the different methods. The correlation $\rho_p \in [-1, 1]$ takes values ± 1 for total positive/negative correlation and zero if there is no linear correlation³. Quite often we are interested only in the relative rankings of nodes. For this purpose the Spearman's rank correlation, given by

$$\rho_s(X, Y) = \rho_p(\text{rg}(X), \text{rg}(Y))$$

where $\text{rg}(X)_i$ is the rank of the raw value X_i in the vector, is the most suitable.

For this study we consider static graphs with 200 vertices, and the number of edges is variable between graph types. We generate temporal networks by drawing 1000 samples from the static network, with replacement. The results, averaged over 1000 graphs per graph type, are given in Figure 3.4. Naturally we see good agreement across

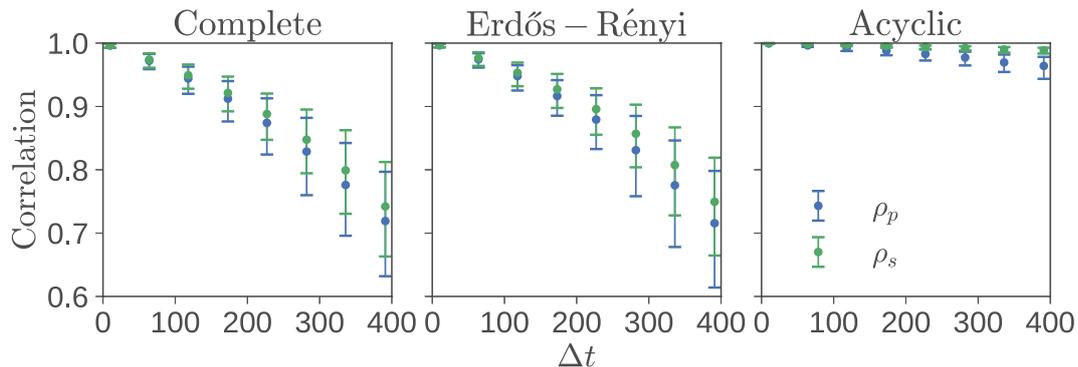


Figure 3.4: Average correlation coefficients ρ_p (blue), ρ_s (green) between the broadcast vectors of the fixed-width partition of width Δt and true partition over an ensemble of 1000 synthetic networks. a) The complete graph, b) ER graphs with parameter $p = 0.3$, c) acyclic graphs. Here the downweighing parameter $\alpha = 0.5$. Data points are plotted with circles, and the 5th and 95th percentiles are given by the error bars.

all graphs for Δt sufficiently small as the two methods converge. For the complete graph and the ER graphs we see a similar drop off in correlation as Δt increases, both in the Pearson and Spearman correlations. For the acyclic graph however we see very

³ Note that the two variables can have very strong non-linear dependence (e.g. $Y^2 + X^2 = 1$) and have a Pearson coefficient of zero.

little drop in the correlation between the partitions, and in particular the Spearman's rank correlation remains above 0.95 even when the temporal network is partitioned into three ($\Delta t = 300$). This makes it apparent that the inclusion of infinite walks in the communicability calculation for fixed-width partitions is the primary difference between the two partitions. This echoes the earlier notion that the two are measuring fundamentally different quantities, observed and imaginary walks through the network.

For a more robust analysis we could also consider other correlation measures such as Kendall's tau coefficient, or studying the correlation between only the top 10% highest ranked nodes. In applying the communicability metric to real examples of thousands of nodes often only the highest rank nodes are analysed. Therefore, repeating this analysis on only a subset of the broadcast vector should inform whether the errors in communicability appear uniformly across the nodes, or that higher or lower rank nodes are most affected.

3.2.4 Summary

In summary, the choice of time partition has a major effect on both the nature of the metric (Table 3.1) and the relative rankings that it assigns to nodes (Fig. 3.4).

The partition should be considered and chosen before the other parameters α and β are fixed. In practice one also needs to take into consideration the computational intensity of using each partition. However, as we show in Section 3.3, calculating the communicability from an unaggregated temporal network can be just as computationally simple as using an aggregated network.

3.3 Efficient Calculation

As we have seen in the previous section, the aggregation of temporal networks into discrete time intervals distorts the interpretation of the communicability measure and introduces quantitative errors. Ideally, the communicability matrix should be calculated on the full, unaggregated temporal network, however this raises a number of

Partition	α	# Time intervals	Errors
Fixed Time	$\alpha < 1/\max_k(\rho(\mathbf{A}_k))$	$T/\Delta t$	Cycle Time decay Causal
Acyclic	Unrestricted	T/λ_a	Time decay Causal
Causality-Preserving	Unrestricted	T/λ_c	Time decay
True	Unrestricted	T (or M)	No error

Table 3.1: A summary of the temporal network partitions, the restrictions on α , the average number of time intervals, and the errors associated with each. T is the total duration of the study, Δt is the specified time interval, λ_a is the average time for a cycle to form, λ_c is the average time for a node to be the target of an event after being the source of an event, and M is the total number of events.

computational challenges which, without a suitable algorithm, could drastically increase the computational time required. For example, the iterative step in the communicability calculation requires a matrix inverse ($\mathcal{O}(N^3)$) and matrix multiplication (also $\mathcal{O}(N^3)$). In a temporal network with $|E|$ events (and therefore $|E|$ iterations required), the total complexity is $\mathcal{O}(|E|N^3)$, where the number of the events is usually at least of the same order as N . By contrast choosing a fixed interval the total complexity is $\mathcal{O}(|P|N^3)$, where P is the partition of the temporal space, which can be chosen at will and $|P| \ll |E|$.

Online social networks are extremely fast-paced and reactive, and often the aim is to analyse and respond to events in near-real-time. Due to this constraint the speed of calculation is paramount, and is a key determinant in the effectiveness of any algorithm. Thankfully when each event belongs solely to one interval there are efficient ways to calculate the resolvent and the subsequent update of the communicability matrix. We introduce a novel implementation below.

For an event $(u_k, v_k, t_k) \in I_k$, the adjacency matrix \mathbf{A}_k is given by

$$(\mathbf{A}_k)_{ij} = \begin{cases} 1 & \text{if } (i, j) = (u_k, v_k) \\ 0 & \text{otherwise.} \end{cases}$$

or more compactly, $(\mathbf{A}_k)_{ij} = \delta_{i,u_k} \delta_{j,v_k}$. As $\mathbf{A}_k^m = \mathbf{0}$ for $m \geq 2$, on expansion the matrix resolvent $(\mathbf{I} - \alpha \mathbf{A}_k)^{-1}$ becomes $\mathbf{I} + \alpha \mathbf{A}_k$. Considering the columns of the dynamic communicability matrix

$$\mathbf{Q}_k = [(\mathbf{q}_0)_k \ (\mathbf{q}_1)_k \ \cdots \ (\mathbf{q}_N)_k],$$

the effect of right multiplication by $\mathbf{I} + \alpha \mathbf{A}_k$ (from the iteration (3.5)) surmounts to a single column operation

$$(\mathbf{q}_i)_{k+1} = \begin{cases} (\mathbf{q}_i)_k + \alpha (\mathbf{q}_{u_k})_k & \text{if } i = v_k \\ (\mathbf{q}_i)_k & \text{otherwise.} \end{cases}$$

This can be seen from the individual entries, $(\mathbf{Q}_{k+1})_{ij} = \sum_s (\mathbf{Q}_k)_{is} (\delta_{s,j} + \alpha \delta_{s,u_k} \delta_{j,v_k})$, where the right most term is only non-zero when $j = v_k$.

For the dynamic communicability matrix, this row operation is all that is required. However for the running dynamic communicability matrix (where time decay is introduced) more work is needed. This consideration, found in [132], is derived below. The running dynamic communicability iteration is given by

$$\mathbf{S}_k = [\mathbf{I} + e^{-\beta \Delta t_k} \mathbf{S}_{k-1}] [\mathbf{I} - \alpha \mathbf{A}_k]^{-1} - \mathbf{I}, \quad k = 0, 1, 2, \dots$$

where $\mathbf{S}_{-1} = \mathbf{0}$ and $\Delta t_k = t_k - t_{k-1}$. Again, we consider the update of the matrix by a single event (u_k, v_k, t_k) . The adjacency matrix, as before is $(\mathbf{A}_k)_{ij} = \delta_{i,u_k} \delta_{j,v_k}$.

By reducing the matrix resolvent the iteration becomes

$$\mathbf{S}_k = [\mathbf{I} + e^{-\beta \Delta t_k} \mathbf{S}_{k-1}] [\mathbf{I} + \alpha \mathbf{A}_k] - \mathbf{I}, \quad k = 0, 1, 2, \dots,$$

or equivalently the entries are described by

$$(\mathbf{S}_k)_{ij} = \sum_k \left[(\delta_{i,k} + e^{-\beta \Delta t_k} (\mathbf{S}_{k-1})_{ij}) (\delta_{k,j} + \alpha \delta_{k,u_k} \delta_{j,v_k}) \right] - \delta_{i,j}.$$

We then consider each case separately:

$$\begin{aligned}
u_k \neq i \neq j \neq v_k : (\mathbf{S}_k)_{ij} &= e^{-\beta\Delta t_k} (\mathbf{S}_{k-1})_{ij} \\
u_k \neq i = j \neq v_k : (\mathbf{S}_k)_{ii} &= e^{-\beta\Delta t_k} (\mathbf{S}_{k-1})_{ii} \\
i = u_k \text{ and } j \neq v_k : (\mathbf{S}_k)_{u_k j} &= e^{-\beta\Delta t_k} (\mathbf{S}_{k-1})_{u_k j} \\
i \neq u_k \text{ and } j = v_k : (\mathbf{S}_k)_{iv_k} &= e^{-\beta\Delta t_k} (\mathbf{S}_{k-1})_{iv_k} \\
&\quad + \alpha e^{-\beta\Delta t_k} (\mathbf{S}_{k-1})_{iu_k} \\
i = u_k \text{ and } j = v_k : (\mathbf{S}_k)_{u_k v_k} &= e^{-\beta\Delta t_k} (\mathbf{S}_{k-1})_{u_k v_k} \\
&\quad + \alpha e^{-\beta\Delta t_k} (\mathbf{S}_{k-1})_{u_k u_k} + \alpha
\end{aligned}$$

which can be combined to give,

$$(\mathbf{S}_k)_{ij} = e^{-\beta\Delta t} (\mathbf{S}_{k-1})_{ij} + \alpha e^{-\beta\Delta t} (\mathbf{S}_{k-1})_{iu_k} \delta_{j,v_k} + \alpha \delta_{i,u_k} \delta_{j,v_k}.$$

Letting $\mathbf{S}_k = [(\mathbf{s}_0)_k \ (\mathbf{s}_1)_k \ \dots \ (\mathbf{s}_N)_k]$, this iteration corresponds to:

1. Downweighing each entry of \mathbf{S} by $e^{-\beta\Delta t}$.
2. Adding $\alpha(\mathbf{s}_{u_k})_k$ to column $(\mathbf{s}_{v_k})_k$.
3. Adding α to $(\mathbf{S}_{k-1})_{u_k v_k}$.

Steps 2 and 3 are $\mathcal{O}(N)$ and $\mathcal{O}(1)$ respectively, however step 1 is $\mathcal{O}(N^2)$. From a computational perspective this is not ideal. For $M \propto N$ events (each node must ‘introduce themselves’ to the network by creating an event), the overall complexity of the algorithm is $\mathcal{O}(N^3)$. To remedy this we note that at upon the addition of an event, steps 2 and 3 only require that the two columns involved are correctly downweighed. This means that the exponential downweighing can be applied as and when needed, rather than updating the entire matrix on the addition of a new event. This reduces the overall complexity of each iteration to $\mathcal{O}(N)$, however, we require that the times each column was last updated, t_i^* for $i = 1, \dots, N$, be recorded.

Using this method we are able to recover the receive vector at each iteration as it is a column sum and all entries in a column are updated at the same time. However, as the

broadcast vector is a row sum of the communicability matrix we need to ensure each column is correctly downweighed. This requires a full update of the communicability matrix. To record the broadcast vector at time t then each column i needs to be multiplied by $e^{t-t_i^*}$ which is an unavoidable $\mathcal{O}(N^2)$ operation. The severity of this operation can be controlled by increasing or reducing the recording rate of the broadcast and receive vectors.

Algorithm 1 provides a method to calculate the running communicability matrix, based on the above. The algorithm takes as input a set of temporal events, a set of times to record the broadcast and receive vectors, and the parameters α and β . The complexity of the algorithm is $\mathcal{O}(MN + KN^2)$ where M is the number of events to be processed, and K is the number of measurements of the broadcast and receive vectors taken. This algorithm assumes that the total number of nodes is known in advance, however it can be easily extended to deal with an increasing N (simply by adding extra rows and columns onto \mathbf{S}).

We can compare this to the standard fixed-time partition algorithm. The full multiplication of matrices, and calculation of the resolvent are both $\mathcal{O}(N^3)$ operations. The calculation of the spectral radius and matrix downweighing are both $\mathcal{O}(N^2)$. Assuming that the number of partitions are chosen to correspond to the number of measurements required, the full algorithm is $\mathcal{O}(KN^3)$. Provided M is $o(N^2)$, a reasonable assumption given that social networks tend to be sparse, the event-based algorithm should be quicker for sufficiently large N . In truth the complexity of matrix operations here are unoptimised and matrix multiplication and inversion can be reduced to $\mathcal{O}(N^{2.373})$ using specialised algorithms, bringing the two algorithms closer together. Further consideration needs to be taken when performing real-time analysis. If the parameter α changes upon an update for the fixed-time partition then all previous iterations need to be recalculated, a problem avoided using the event-based algorithm.

This new algorithm requires that the full communicability matrix is stored at all times which can be memory intensive for large temporal networks, however there is not currently an algorithm to calculate the running dynamic communicability without this requirement. This algorithm however is easily parallelised, allowing the storage in

memory to be distributed over multiple machines.

3.3.1 Parallelisation

The event-based algorithm lends itself easily to parallelisation. There are three basic approaches, one involving shared memory and a further two using distributed memory. In brief, the shared memory approach queues the edges and assigns the addition of events (the `ADDEVENT` function from Algorithm 1) to each processor, maintaining that the temporal ordering with respect to each node is preserved. The broadcast and receive vectors can be periodically read from the shared matrix.

By contrast the distributed memory methods parallelises the communicability matrix rather than the event set by distributing the rows or columns over each processor. As Algorithm 1 uses column operations the distributed rows approach (Algorithm 2) is preferable as it minimises data transfer between processors. For example, if the columns were distributed then the addition of one column to another requires an entire column to be passed from one processor to another. With distributed rows column operations can be performed independently between the processors, see Figure 3.5.

Reading the broadcast and receive vectors requires a small amount of effort. The broadcast scores for each node can be calculated on the corresponding processor and then combined to give the full broadcast vector. For the complete receive vector, each processor produces a receive vector which are then summed.

3.4 Discussion

The communicability metric offers an intuitive and simple way to assess the influence of nodes in a temporal network. This metric has been useful in identifying nodes of importance both in terms of their ability to spread information as well as receive it. In particular, it seems a suitable metric to study online social networks formed from high frequency events taking place between a large number of nodes, and where the underlying structure of node connections is poorly understood.

Algorithm 1 Series calculation of communicability.

Require: E = Event sequence

Require: I = Recording times

Require: N = Number of nodes

1: **Initialize:**

$\mathbf{S} \leftarrow \mathbf{0}$

▷ an $N \times N$ matrix

$\mathbf{t}^* = (t_0^* \ t_1^* \ \dots \ t_N^*) \leftarrow \mathbf{0} \in \mathbf{M}_{1 \times N}(\mathbb{R})$

▷ Column update times

$t \leftarrow 0$

▷ System time

2: **function** COMMUNICABILITY(E, I, α, β)

3: **for** $(u_k, v_k, t_k) \in E$ **do**

4: **if** $\exists i_0 \in I$ s.t $t < i_0 < t_k$ **then**

▷ Event occurs after a recording point

5: $t \leftarrow i_0$

6: UPDATEMATRIXANDRECORDVECTORS($\mathbf{S}, \mathbf{t}^*, t$)

7: ADDEVENT($\mathbf{S}, (u_k, v_k, t_k), t$)

return \mathbf{S}

8: **function** UPDATEMATRIXANDRECORDVECTORS(\mathbf{S}, \mathbf{t}^*)

9: $\mathbf{S} \leftarrow \mathbf{S} - \mathbf{I}$

10: $[\mathbf{S}]_{is} \leftarrow [\mathbf{S}]_{is} \times e^{-\beta(t-t_s^*)}, \forall s \in \{1, \dots, N\}$

11: $t_s \leftarrow t, \forall s \in \{1, \dots, N\}$

12: $\mathbf{b}(t) \leftarrow \mathbf{S}\mathbf{1}$

13: $\mathbf{r}(t) \leftarrow \mathbf{S}^T \mathbf{1}$

14: $\mathbf{S} \leftarrow \mathbf{S} + \mathbf{I}$

15: **function** ADDEVENT($\mathbf{S}, (u_k, v_k, t_k), t$)

16: $t \leftarrow t_k$

17: $[\mathbf{S}]_{u_k u_k} \leftarrow [\mathbf{S}]_{u_k u_k} - 1$

18: $[\mathbf{S}]_{v_k v_k} \leftarrow [\mathbf{S}]_{v_k v_k} - 1$

19: $[\mathbf{S}]_{i u_k} \leftarrow [\mathbf{S}]_{i u_k} \times e^{-\beta(t-t_{u_k}^*)}$

20: $[\mathbf{S}]_{i v_k} \leftarrow [\mathbf{S}]_{i v_k} \times e^{-\beta(t-t_{v_k}^*)}$

21: $t_{u_k}^* \leftarrow t$

22: $t_{v_k}^* \leftarrow t$

23: $[\mathbf{S}]_{i v_k} \leftarrow [\mathbf{S}]_{i v_k} + \alpha [\mathbf{S}]_{i u_k}, \forall i$

24: $[\mathbf{S}]_{u_k v_k} \leftarrow [\mathbf{S}]_{u_k v_k} + \alpha$

25: $[\mathbf{S}]_{u_k u_k} \leftarrow [\mathbf{S}]_{u_k u_k} + 1$

26: $[\mathbf{S}]_{v_k v_k} \leftarrow [\mathbf{S}]_{v_k v_k} + 1$

$$\begin{array}{l}
\text{SLAVE 1} \\
\text{SLAVE 2} \\
\vdots \\
\text{SLAVE } N/2
\end{array}
\begin{array}{c}
\left[\begin{array}{ccccccc}
S_{1,1} & S_{1,2} & \cdots & \boxed{S_{1,u}} & \cdots & \boxed{S_{1,v}} & \cdots & S_{1,N} \\
S_{2,1} & S_{2,2} & \cdots & \boxed{S_{2,u}} & \cdots & \boxed{S_{2,v}} & \cdots & S_{2,N}
\end{array} \right] \\
\left[\begin{array}{ccccccc}
S_{3,1} & S_{3,2} & \cdots & \boxed{S_{3,u}} & \cdots & \boxed{S_{3,v}} & \cdots & S_{3,N} \\
S_{4,1} & S_{4,2} & \cdots & \boxed{S_{4,u}} & \cdots & \boxed{S_{4,v}} & \cdots & S_{4,N}
\end{array} \right] \\
\vdots \\
\left[\begin{array}{ccccccc}
S_{N-1,1} & S_{N-1,2} & \cdots & \boxed{S_{N-1,u}} & \cdots & \boxed{S_{N-1,v}} & \cdots & S_{N-1,N} \\
S_{N,1} & S_{N,2} & \cdots & \boxed{S_{N,u}} & \cdots & \boxed{S_{N,v}} & \cdots & S_{N,N}
\end{array} \right]
\end{array}$$

Figure 3.5: Pictorial representation of the main loop of the parallel algorithm (Algorithm 2). The rows of the communicability matrix \mathbf{S} are split between the number of processors (here there are $N/2$ processors with 2 rows each). For each event (u, v, t) to be processed, each slave processor independently performs the column operations on columns u and v as in Algorithm 1.

We assessed the interpretation of the metric when applying it to different levels of temporal aggregation of the network and found that errors can be introduced as aggregation is increased. We showed that these errors can lead to significant differences in both the values and relative rankings of the broadcast vectors. In order to remedy this issue we proposed an efficient, parallelisable algorithm which calculates the ground truth communicability matrix with no temporal aggregation and exploits the properties of the matrix resolvent when the adjacency matrix has only one entry. This also had the added effect that the once restricted parameter α was no longer restricted and could be assigned a physical meaning and be consistent between studies.

An important point to realise is that the communicability centrality tells us that a node may be important in the temporal network, however it reveals little information about the behaviour of that node which resulted in a high broadcast or receive score. This thought leads us onto the next chapter where we classify the behaviour of nodes and communities of nodes in a temporal network.

Algorithm 2 Pseudocode for the parallel calculation of the communicability iteration with. Note that execution is not necessarily by line order.

Require: E = Event sequence

Require: T = Set of times to record broadcast and receive vectors

MASTER NODE

Assign each slave node a fraction of the total rows.

For each event, $(u, v, t) \in E$:

Broadcast the edge, (u, v, t) to all slave nodes.

Broadcast the times at which those columns were last updated, (t_u^*, t_v^*) , to all slave nodes.

For each recording time, $t \in T$:

Inform all slave nodes that an update is required.

Broadcast the times at which all columns were last updated.

Gather the broadcast scores of each node and record them.

Gather the receive score fragments, piece them together, then record them.

SLAVE NODES

For each event and update times, $(u, v, t), (t_u^, t_v^*)$ received:*

Perform the required column operations (Algorithm 1).

For each set of recording times received:

Update all columns by the relevant factor, $e^{t-t_i^*}$.

Sum each row and column and offer results to the master node.

4

The Temporal Event Graph

In Chapter 2 we saw many different descriptions of temporal networks, the most common being a series of static adjacency matrices representing the connections of the network within a specific time interval. These representations are node-centric, in that they capture the connections between the nodes. In this chapter we provide an alternative representation which describes the connections between events, also known as the *dual*.

Perhaps the most basic kind of this representation is the line graph [138]. The vertices of the line graph are the edges of the underlying network, see Figure 4.1(a). An edge is present between vertices of the line graph when they both share a node. This type of graph has been used in static networks of sexual contact, and the study of disease spread on these networks [69]. In Figure 4.1(a) we see two networks of relationships (top) and the corresponding line graph representations (bottom). Despite both contact networks

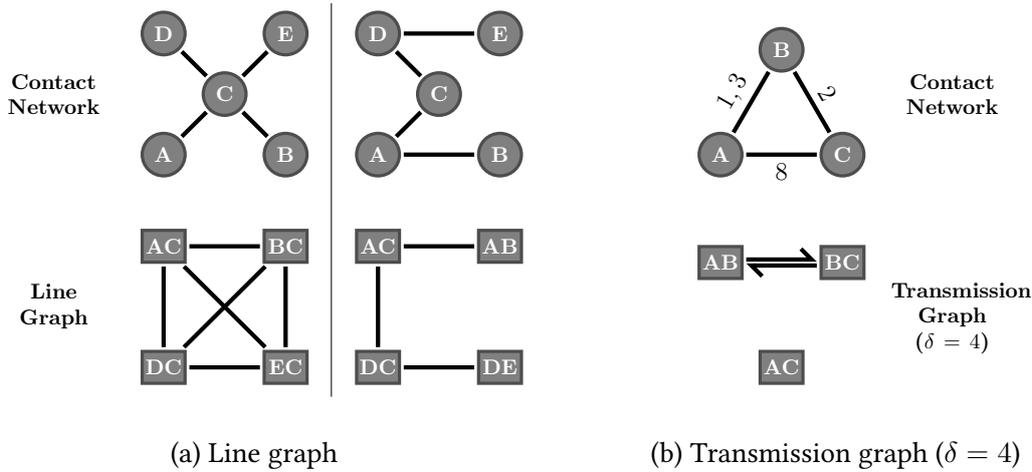


Figure 4.1: Contact networks and two possible dual network representations.

having the same edge density of $d = 0.4$ ¹, from the line graph it is evident that it is much easier for a disease to spread on the left network than the right. What the line graph captures is the *concurrency* of the contact network, which can be defined as the edge density of the line graph [138].

The equivalent edge-centric graph for temporal networks is the transmission graph [70], constructed to “represent epidemiologically meaningful paths through a network of contacts through time.” As with the line graph, edges in the temporal contact network are vertices in the transmission graph (see Figure 4.1(b)). A vertex i is connected to a vertex j if the contacts i and j share a node, and j occurs after i starts but no later than a time δ after i occurs². For example, in Figure 4.1(b) there is a connection between vertices AB and BC as they share a node and those edges are active less than a time δ apart. Similarly, there is a connection from BC to AB but not from either to AC as that edge appears over a time δ later. This construction creates a growing network of $N(N-1)$ vertices, corresponding to all possible node pairs in the contact network. One may also define an interval transmission graph where links are removed after a period of inactivity. These graphs draw inspiration from the line graph as links capture possible paths of transmission. The transmission graphs however are able to incorporate the

¹ The edge density is defined as $d = \frac{2E}{N(N-1)}$ where N and E are the number of nodes and edges respectively.

² This is the definition of Δt -adjacency, given in the next section.

dynamic nature of the contact sequence in contrast to the line graph which considers only static aggregations of contact. However, for both the transmission graph and interval transmission graph the temporal aspect of the network has been aggregated, so the original temporal network cannot be recovered. In other words, a transmission graph representation does not uniquely define a temporal network. For example, in Figure 4.1(b) the event occurring between A and C can occur at any time greater than 7 and the transmission graph would be identical.

In Chapter 3 we used the communicability centrality to capture the importance of nodes with respect to their ability to broadcast and receive information. By switching from a node-centric perspective to an event-centric perspective we can assess the behaviour of nodes with respect to the events they participate in. In particular we want to quantify the types of interactions they have with other nodes and the frequency in which they occur. To this end we introduce the *temporal event graph* (TEG), a static graph representation of a temporal network which encodes these behaviours and, unlike the line and transmission graphs, uniquely describes the temporal network.

The TEG can be seen as a descendent of the transmission graph, however, a derivative of the TEG is also used in the calculation of temporal motifs [71] from which we draw the basic concepts and notation.

Chapter Outline

In Section 4.1 we define the TEG and its preliminaries, as well as showing that it uniquely defines a temporal network. In Section 4.2 we give some further theoretical properties of the TEG and in Section 4.3 describe the statistical properties of the TEG which we can use to assess the structure and connectivity of the temporal network. Finally we discuss the merits of the TEG and possible avenues for future research in Section 4.4.

4.1 The Temporal Event Graph

As in Chapter 2 we consider temporal networks as a sequence of temporal events E where an individual event $e_i = (u_i, v_i, t_i) \in E$ corresponds to an interaction of node u_i with node v_i at time t_i (here assuming interaction is instantaneous).

The systems most suited to this representation are communication networks (letter and email correspondence, phone calls, social media etc.) and proximity networks (human contact networks) [139]. Examples of temporal networks from online social networks will be given in the next chapter. To define the TEG we first need to be able to relate two events in a meaningful way, capturing the relationship of the nodes and the temporal proximity of the events. One such relation is that of Δt -adjacency [71].

Definition 4.1.1. Two time-ordered events e_i, e_j are said to be Δt -adjacent if they share at least one node ($\{u_i, v_i\} \cap \{u_j, v_j\} \neq \emptyset$) and the time between the two events (inter-event time) is no greater than Δt , i.e. $t_j - t_i < \Delta t$.

With this definition we can formally define the TEG.

Definition 4.1.2. For a temporal network $G = G(V, E, T)$, the Δt -Temporal Event Graph, hereby known as the Δt -TEG, is a directed graph $\mathcal{G} = \mathcal{G}(\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = E$ and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. The graph is defined such that there is a vertex for each event in E and each vertex is connected to the *subsequent* Δt -adjacent event of each node in that event. Let

$$S(u_i) = \{k \mid \underbrace{(\{u_i\} \cap \{u_k, v_k\} \neq \emptyset)}_{\text{Share a node}} \text{ and } \underbrace{(0 < t_k - t_i < \Delta t)}_{\text{Occur within } \Delta t \text{ of each other}}\},$$

be the set of subsequent Δt -adjacent events for the node u_i with the equivalent set defined for v_i . Then the set of edges in the TEG is then given by

$$\mathcal{E} = \{(e_i, e_j) \mid (j = \min\{S(u_i)\}) \text{ or } (j = \min\{S(v_i)\})\}.$$

This construction means that each vertex has an out-degree and in-degree of at most two (see Lemma 4.2.1).

The Δt -TEG consists of one or more connected temporal components (or maximal temporal subgraphs [71]), that is, for each pair of vertices in a component there exists a sequence of events between them such that all pairs of consecutive events are Δt -adjacent. Of particular interest is the Δt -TEG in the limit $\Delta t \rightarrow \infty$, hereby referred to as the TEG. The examples in Figure 4.2 show how the TEG is constructed from an event sequence. To avoid ambiguity we use the terms nodes and events for the temporal network, and event vertices and edges for the TEG.

There are two important functions of the edge set to consider. Firstly the function $\tau : \mathcal{E} \rightarrow \mathbb{R}_0^+$, given by $\tau((e_i, e_j)) = t_j - t_i$ describes the inter-event time (IET) between the two events. Since there is an edge for each node in an event the distribution of IETs describes the times between individual node activity, either interacting or being the target of an interaction. The function $\mu : \mathcal{E} \rightarrow \mathbb{M}$ where $\mathbb{M} = \{\text{ABAB}, \text{ABBA}, \text{ABAC}, \text{ABCA}, \text{ABBC}, \text{ABCB}\}$ is the set of two-event motifs (Table 4.1) which describes the relative positions of the nodes between events. These motifs are given a descriptive name which is indicative of the behaviour associated with each pattern. This behaviour is associated with the motif in its entirety and not a particular node within the motif. For example, the ABAC motif is described as the broadcast motif as node A is sending messages to multiple other nodes. The ABBA motif is the reciprocal motif, as messages from A to B are then reciprocated. Let f_{ij} be an enumeration of the ordered sequence of nodes (u_i, v_i, u_j, v_j) (not necessarily distinct) mapped to the corresponding alphabetic character³, then $\mu((e_i, e_j)) = f_{ij}(u_i)f_{ij}(v_i)f_{ij}(u_j)f_{ij}(v_j)$. For example, the edge $((5, 10, t_0), (10, 12, t_1))$ becomes ABBC under the action of μ . It is also possible for the motif function μ to incorporate other event data, as we will see in Chapter 5.

There are three properties of the motif set, $(\xi_{\text{out}}, \xi_{\text{in}}, \xi_{\text{switch}})$, which are required in Section 4.1.1. For event pairs involving three distinct nodes we define ξ_{out} to be the label and position of the node which appears in both events, ξ_{in} to be the label and position of the shared node in the later event, and $\xi_{\text{switch}} = 1$ if $\xi_{\text{out}} = \xi_{\text{in}}$ and -1 otherwise. For example, in the motif ABBC the node labelled B is carried forward from

³ e.g. $f_{ij}(u_i) = \text{A}, f_{ij}(v_i) = \text{B}, \dots$

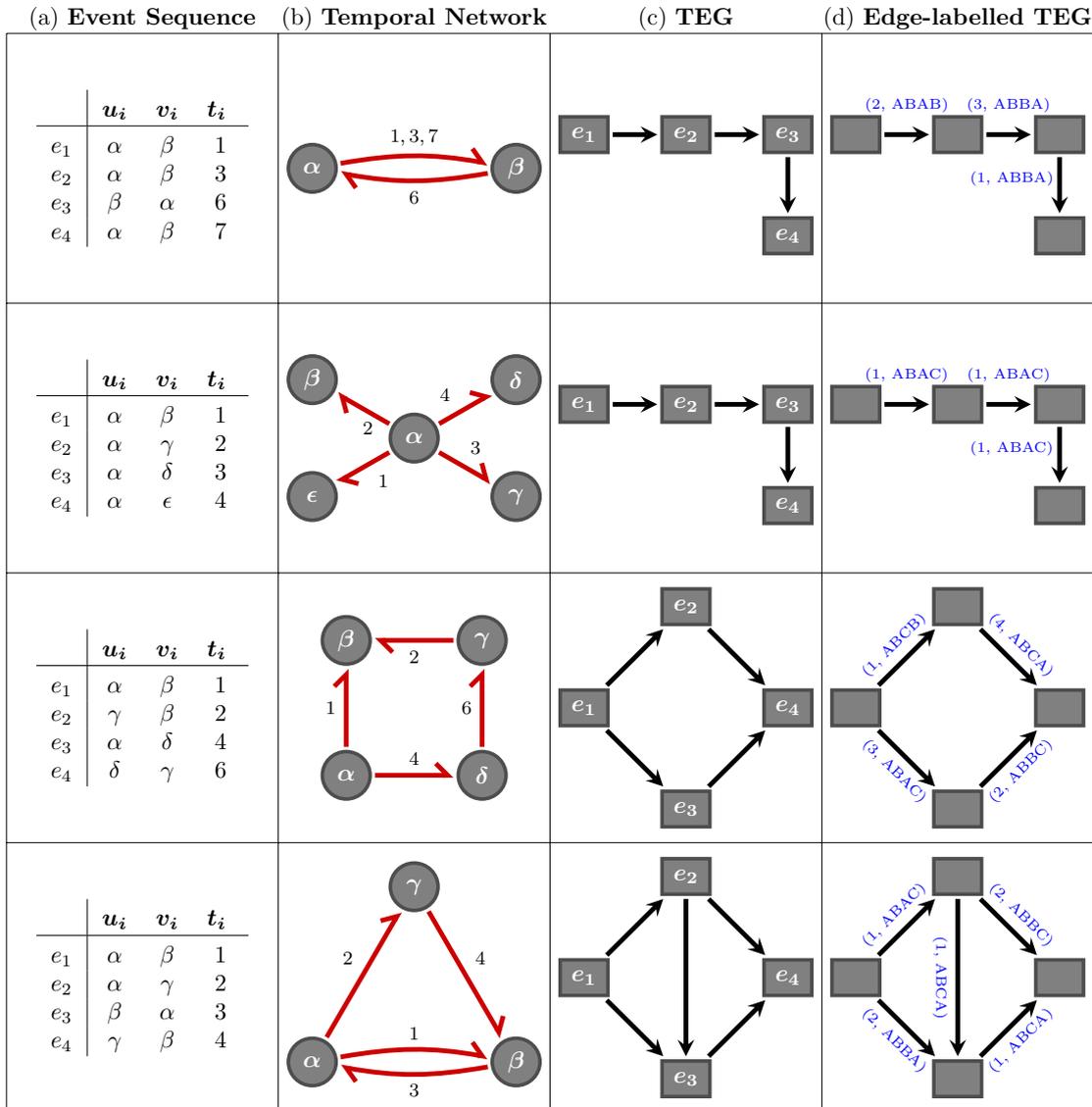


Figure 4.2: Illustration of the duality of temporal networks and the temporal event graph. (a) Four simple temporal networks (event sequences) involving four events. (b) Pictorial representations of the temporal networks. Event labels represent the instantaneous time when that event occurred between two nodes. (c) The TEG for each temporal network. (d) The corresponding *edge-labelled* TEGs (Def. 4.1.3). Edges are labelled with the tuple (τ, μ) , the inter-event time and motif respectively. Note in the bottom example the next two events for node A are connected to the first event. This is consistent as the ABBA edge occurred *after* that of the ABAC, i.e., node A's subsequent event was $A \rightarrow C$ and node B's subsequent event was $B \rightarrow A$ (coincidentally A's next event).

Motif	Name	Shorthand	ξ_{out}	ξ_{in}	ξ_{switch}
$A \rightarrow B, B \rightarrow A$	Reciprocal	ABBA	AB	BA	-1
$A \rightarrow B, A \rightarrow B$	Repeated	ABAB	AB	AB	1
$A \rightarrow B, A \rightarrow C$	Broadcast	ABAC	A•	A•	1
$A \rightarrow B, C \rightarrow A$	Non-sequential	ABCA	A•	•A	-1
$A \rightarrow B, B \rightarrow C$	Message Passing	ABBC	•B	B•	-1
$A \rightarrow B, C \rightarrow B$	Receiving	ABCB	•B	•B	1

Table 4.1: The set of all possible two-event motifs \mathbb{M} , given by their contact sequence, description, label, and label properties ξ_{in} , ξ_{out} , and ξ_{switch} .

the first event so $\xi_{\text{out}}(\text{ABBC}) = \bullet\text{B}$ and takes the first position in the second event so that $\xi_{\text{in}}(\text{ABBC}) = \text{B}\bullet$. Subsequently as $\xi_{\text{out}} \neq \xi_{\text{in}}$, then $\xi_{\text{switch}}(\text{ABBC}) = -1$, the node labelled B has switched between being the target of an event to a source. For consistency we define $\xi_{\text{out}}(\text{ABAB}) = \text{AB} = \xi_{\text{out}}(\text{ABBA})$ and $\xi_{\text{in}}(\text{ABAB}) = \text{AB}$ and $\xi_{\text{in}}(\text{ABBA}) = \text{BA}$.

4.1.1 Duality

The TEG contains both event information and the connectivity of events in terms of adjacency. We can consider a TEG without the event information, defined purely by the connectivity information and edge functions.

Definition 4.1.3. The *edge-labelled TEG* is the static graph defined by the adjacency pair (A^τ, A^μ) where

$$A_{ij}^\tau = \begin{cases} \tau(e_i, e_j) & \text{if } (e_i, e_j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases},$$

is the weighted adjacency matrix consisting of IETs and

$$A_{ij}^\mu = \begin{cases} \mu(e_i, e_j) & \text{if } (e_i, e_j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}.$$

is the matrix containing edge motif labels.

Not all permutations of the vertices and edges of an edge-labelled TEG describe a temporal network. There are four conditions required for an edge-labelled TEG to represent a temporal network. We call graphs which satisfy the following conditions *consistent* graphs.

C1 Event times must be consistent across all paths: Let P_{ij} be the set of all directed paths between vertices i and j . A path $p_\alpha \in P_{ij}$ is the sequence of edges in the path. The sum of inter-event times along all paths must be equal, that is

$$\sum_{(k,l) \in p_\alpha} A_{kl}^\tau = \sum_{(k,l) \in p_\beta} A_{kl}^\tau \text{ for all } p_\alpha, p_\beta \in P_{ij}.$$

C2 Nodes in each event have only one subsequent event: For each pair of out-edges $(i, k), (i, l)$ of a vertex we require $\xi_{\text{out}}(A_{ik}^\mu) \neq \xi_{\text{out}}(A_{il}^\mu)$.

C3 Nodes in each event cannot be overprescribed: For each pair of in-edges $(k, i), (l, i)$ of a vertex we require $\xi_{\text{in}}(A_{ki}^\mu) \neq \xi_{\text{in}}(A_{li}^\mu)$.

C4 Edge types and nodes must be consistent across multiple paths: If there exists an edge (i, j) such that there exists a secondary path $p \in P_{ij}$ via at least one other vertex then

$$A_{ij}^\mu = \begin{cases} \text{ABAB} & \text{if } \prod_{(k,l) \in p} \xi_{\text{switch}}(A_{kl}^\mu) = 1 \\ \text{ABBA} & \text{if } \prod_{(k,l) \in p} \xi_{\text{switch}}(A_{kl}^\mu) = -1 \end{cases}.$$

Conversely if there is a vertex with two in edges, $(i, j), (k, j)$, with $A_{ij}^\mu \in \{\text{ABAB}, \text{ABBA}\}$ then there exists a path $p \in P_{ij}$ with $(k, j) \in p$ and $\prod_{(m,n) \in p} \xi_{\text{switch}}(A_{mn}^\mu) = \xi_{\text{switch}}(A_{ij}^\mu)$. Similarly for a vertex with two out edges $(i, j), (i, k)$ with $A_{ij}^\mu \in \{\text{ABAB}, \text{ABBA}\}$ then there exists a path $p \in P_{ij}$ with $(i, k) \in p$ and $\prod_{(m,n) \in p} \xi_{\text{switch}}(A_{mn}^\mu) = \xi_{\text{switch}}(A_{ij}^\mu)$.

Those graphs which do not satisfy these conditions are inconsistent in that they do not uniquely describe a temporal network, and attempting to recover the temporal network will lead to contradiction. Examples of inconsistent TEGs are given in Figure 4.3.

Inconsistent Edge-labelled TEGs

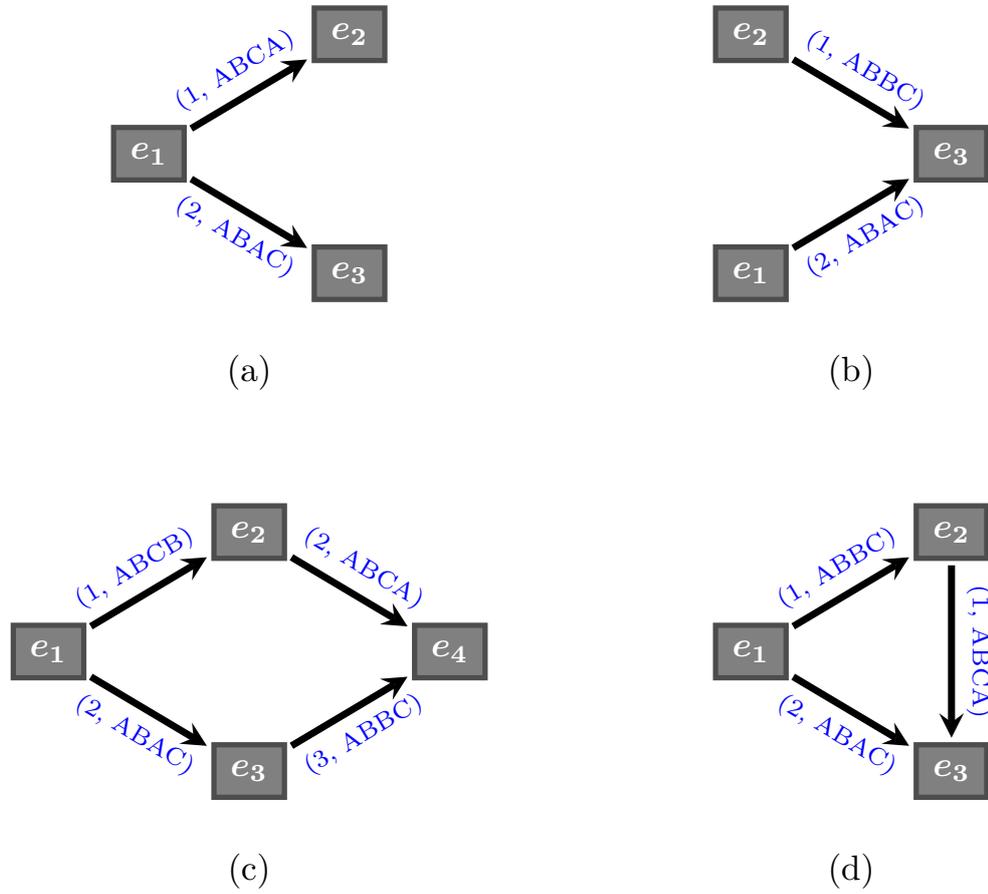


Figure 4.3: Inconsistent edge-labelled temporal event graphs. Edges are labelled with the tuple (τ, μ) . (a) The subsequent two events for node A are included as edges, breaking condition [C2]. (b) Both incoming edge types dictate the first node of the event which is contradictory (condition [C3]). (c) The inter-event times across multiple paths are not equal (condition [C1]). (d) The edge between events e_1 and e_3 is incorrectly labelled. By reconstructing the temporal network or using condition [C4] we see that $A_{13}^\mu = ABAB$.

For each connected component of an edge-labelled TEG we are able to reconstruct the temporal network using the following algorithm:

- (a) Find the maximal path from a root vertex (no incoming edges) to a leaf vertex (no outgoing edges) in the edge-labelled TEG using network of IETs, A^τ , allowing for backwards traversal along edges with opposite weight. (Fig. 4.4(a)). This can be achieved by finding the shortest path in the network $(A^\tau)^T - A^\tau$. One possible

algorithm is the Floyd-Warshall algorithm [140].

- (b) Label the first vertex in the path with $t = 0$ and subsequently propagate the event times through the edge-labelled TEG along the edges. For a vertex i , the time at which that event occurs is given by

$$t_i = \sum_{(m,n) \in P_{0i}} \left((A^\tau)^T - A^\tau \right)_{mn}$$

To be able to do this we require the condition [C1] otherwise the existence of multiple paths between vertices leads to a contradiction in event times.

- (c) For events in time order, resolve the nodes in each event from the incoming edges (Fig. 4.4(b,c)). We require condition [C3] here otherwise there can be a conflict on resolving a node position. If a node in an event is unprescribed (the event has zero or one incoming edge) then the unprescribed nodes are given a new label.

Condition [C2] is required by definition of the edge-labelled TEG to enforce that the subsequent event of each node is connected by an edge. Without it, the subsequent two edges for one node could be given. Finally condition [C4] ensures that the edge-labelled TEG is uniquely labelled (Fig. 4.3(d)).

Lemma 4.1.4. *The maximal path (allowing for backwards traversal along edges with negative weight) through the edge-labelled TEG includes the earliest and latest event in the temporal network.*

Proof. Let $p_{max} = (e_0, \dots, e_k)$ be the sequence of vertices in the maximal path. Suppose there exists an event $e_* \notin p_{max}$ such that $t_* < t_i$ for $i = 0, \dots, k$. Then, as the TEG is connected, there exists a path p_{*i} (ignoring edge directions) from $e_* \rightarrow e_i, \forall e_i \in p_{max}$. Then $l(p_{*i}) > l(p_{0i})$ where $l(\cdot)$ is the length of the path, and hence the path $e_* \rightarrow e_i \rightarrow e_k$ is longer than p_{max} . This is a contradiction and hence the maximal path through the TEG must contain the earliest event in the temporal network. A similar but opposite argument shows that the latest event is also contained in the maximal path. \square

The existence of an inverse algorithm highlights a duality between the edge-labelled TEG and the temporal network.

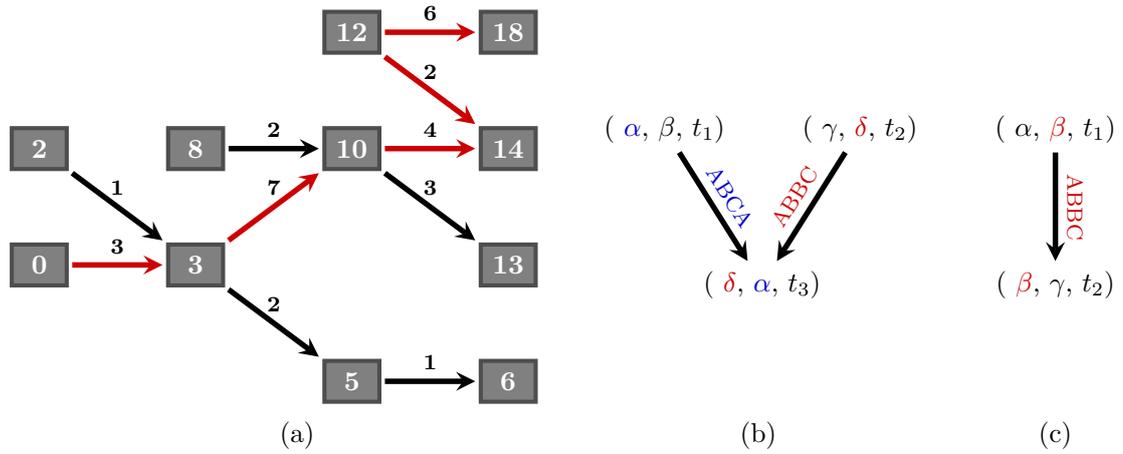


Figure 4.4: The inverse algorithm for the TEG. (a) The maximal path between root and leaf vertices (red) through the TEG with edges labelled with IETs. Once the maximal path has been found, the root vertex is assigned time $t = 0$ and the remainder of times are found by propagation along edges. (b) The resolution of an event from two incoming edges. Each incoming edge determines one of the nodes in the later event. (c) The resolution of an event with one incoming edge. In this case only one node is prescribed and so the other is given a new label.

Theorem 4.1.5. *Let X be the set of all temporal networks translated in time such that the first event occurs at $t = 0$, nodes are labelled in order of appearance, and such that the time-aggregated graph of connections is connected. Let Y be the set of all consistent connected temporal event graphs. Then there exists a bijection $f : X \rightarrow Y$, that is, an edge-labelled TEG uniquely describes a temporal network in X .*

Proof. Trivially for each temporal network there exists only one edge-labelled TEG as the nodes in each event have at most one subsequent event⁴ and the functions τ and μ are deterministic. The proof rests on the existence of the inverse algorithm f^{-1} , outlined above. We consider a general event $e_i = (u_i, v_i, t_i)$ in the temporal network, and its representative vertex x in the edge-labelled TEG. By the translation of the temporal network, this event occurs t_i time units after the first event. By finding the maximal path through the edge-labelled TEG we find the first event in the temporal network (Lemma 4.1.4), and can hence find the time which x occurs relative to this first event,

⁴ Here we assume that no two events occur at the time same.

that is, t_i . The event is now is correctly placed in time. To recover the nodes of the event u_i and v_i , assume the nodes in all previous events have been correctly determined in order of appearance. There are three possible cases:

1. Event e_i has no incoming edges. In this case neither of these nodes have previously interacted and can be enumerated.
2. Event e_i has one incoming edge prescribing one node. In this case a new node is involved and is enumerated accordingly (Fig. 4.4(c)).
3. Event e_i has one or two incoming edges prescribing both nodes. In this case the nodes are completely determined by previous events (Fig. 4.4(b)).

For the base case, the earliest event vertices have no incoming edges and are labelled freely. Subsequent event vertices must then have all incoming edges prescribed as they occur strictly earlier in time. Hence the nodes in e_i are correctly labelled, relative to the labelling of the previous events. As both nodes are labelled relative to previous events, and the time of the event is positioned relative to the first event, the event is recovered from the TEG. Since this argument holds for an arbitrary event in the temporal network, it holds for all. Therefore $f^{-1}(f(X)) = X$, and f is a bijection. \square

Corollary 4.1.6. *A temporal event graph \mathcal{G} , consisting of multiple connected components defines a temporal network up to a translation of time between components. If the events of \mathcal{G} are time stamped then \mathcal{G} uniquely defines a temporal network.*

Proof. By Theorem 4.1.5 for each connected component there exists a unique temporal network such that the earliest event occurs at $t = 0$. Trivially there exists an ensemble of temporal networks with the same TEG, dependent on the choice of earliest event time for each component. If the time of this event is given then the choice is removed and hence the TEG uniquely defines the temporal network. \square

Time translation between components may seem disconcerting, however these components are truly disconnected and do not share any nodes. This means that, assuming the network is not visible to those within it, any dynamics on the network

are completely independent across components⁵. Most digital communication channels that we will consider are hidden from an observer, e.g. email, SMS, telephone calls. Other sources of communication such as Twitter are in the public domain and so all messages are observable (although require active searching). Furthermore, adding event timestamps to the vertices fixes the temporal components in time, and so the graph then uniquely defines a temporal network.

This means that the temporal network can be uniquely defined within the time translation of components by the network of subsequent adjacent events, their IETs, and the motifs formed between them. As a result, considering the network in this formalism is equivalent to studying the temporal network as the same information is contained in both.

4.2 Theoretical Properties of the TEG

Lemma 4.2.1. *Each vertex in the TEG has at most in-degree two and out-degree two.*

Proof. Consider an event vertex representing the event $e_i = (u_i, v_i, t_i)$. From our definition we let

$$A_u^+(i) = \{k | (\{u_i\} \cap \{u_k, v_k\} \neq \emptyset) \text{ and } (0 < t_k - t_i < \Delta t)\},$$

$$A_v^+(i) = \{k | (\{v_i\} \cap \{u_k, v_k\} \neq \emptyset) \text{ and } (0 < t_k - t_i < \Delta t)\}$$

be the subsequent Δt -adjacent events for the nodes u_i and v_i respectively. The set of edges in the TEG is given by

$$\mathcal{E} = \{(e_i, e_j) | j = \min(A_u^+(i)) \text{ or } j = \min(A_v^+(i))\}.$$

⁵ In the case where the network is visible, observing the network usually prompts a response that is directed towards the observed agents, subsequently connecting the two components. There may be cases where nodes in one component observe nodes in another and act upon that information without any interaction with the component. In these cases it is important to include the time stamp of each event in the TEG.

Therefore, for each e_i there exists up to two events whose indices are the minima of each set. These two minima do not need be unique, nor exist, and so there are at most two out edges.

For the edge in-degree, the previous Δt -adjacent events for the nodes u_i and v_i are

$$A_u^-(i) = \{k | (\{u_i\} \cap \{u_k, v_k\} \neq \emptyset) \text{ and } (0 < t_i - t_k < \Delta t)\},$$

$$A_v^-(i) = \{k | (\{v_i\} \cap \{u_k, v_k\} \neq \emptyset) \text{ and } (0 < t_i - t_k < \Delta t)\}$$

We can analogously define the edge set as

$$\mathcal{E} = \{(e_j, e_i) | j = \max(A_u^-(i)) \text{ or } j = \max(A_v^-(i))\}.$$

By the same reasoning as with forward definition, nodes can have a maximum in-degree of at most two. \square

Lemma 4.2.2. *The TEG is a directed acyclic graph (DAG).*

Proof. For a graph G to be a DAG, each node in G must not have a directed path from that node back to itself. The edge set is given by

$$\mathcal{E} = \{(e_i, e_j) | j = \min(A_u^+(i)) \text{ or } j = \min(A_v^+(i))\}$$

where the set $A_u^+(i)$ contains only events e_k such that $t_k > t_i$ by definition. Suppose there exists a direct path from event i back to itself via a sequence of ordered events $e_{k_1}, e_{k_2}, \dots, e_{k_n}$. Then by transitivity this implies $t_i < t_{k_1} < t_{k_2} < \dots < t_{k_n} < t_i$, which is a contradiction. Hence no such path exists and the TEG is a DAG. Simply put, as edges travel strictly forward in time there can be no cycles in the graph. \square

Lemma 4.2.3. *The set of nodes in each component of the TEG are distinct, i.e., if there exist two components of the TEG C_1, C_2 with node sets $V_1, V_2, \subset V$ then $V_1 \cap V_2 = \emptyset$.*

Proof. Suppose $V_1 \cap V_2 \neq \emptyset$ and there exists a node $u \in V_1 \cap V_2$. Then there exists a set of events in C_1 which contain u with times $t_1^{(1)}, t_2^{(1)}, \dots, t_{n_1}^{(1)}$. Similarly there exists a set of events in C_2 which contain u with times $t_1^{(2)}, t_2^{(2)}, \dots, t_{n_2}^{(2)}$. Assuming that event times are distinct then there exists an ordering of these times. Regardless of the relative

ordering of these times there must exist a time $t_i^{(1)}$ followed by a time $t_j^{(2)}$ (or vice versa). These events share a node and the timing of the events are consecutive meaning the two events are *adjacent*. This implies there exists an edge between the two events by definition of the TEG, and C_1 and C_2 are one component. This contradicts the original statement and hence C_1 and C_2 must contain distinct nodes. \square

Note that this is not true in the Δt -TEG, even if the components completely overlap in time.

4.3 Statistical Properties of the TEG

In this Section we outline some of the statistical properties of the TEG, illustrated with examples from synthetic networks. These properties will be examined on real social networks in the next chapter.

The Δt -TEG provides a means of assessing the temporal structure of the network. In this section we consider the Δt -TEG as the weighted static network where edge weights are the IETs between events. This allows us to prune the network, based on edge weights (IETs). We consider the *weakly connected components* of the TEG where two vertices are in the same component if they are connected on the undirected graph induced by ignoring edge direction. Note that these components are maximal Δt -connected subgraphs [71] and describe the connectivity of the events themselves but cannot describe the connectivity of the nodes in general⁶. In fact, finding strongly connected components of nodes in temporal networks has been shown to be an NP-complete problem [102].

The Δt -TEG contains edges (i, j) where $A_{ij}^\tau < \Delta t$, using the notation of the edge-labelled TEG from the previous section. Let $C_i^{\Delta t}$ be the i th component of the Δt -TEG, where components are partially ordered by the number of events they contain

⁶ One can make a number of statements on the connectivity of the nodes, following edges with certain motif types. For example, a chain of ABBC edges implies a path from the source node of the first event to the target node of the final event. However this is difficult to do in fully generality.

such that $|C_0| \geq |C_1| \geq \dots$. These components are a natural decomposition of the temporal network and will be the focus of this section and subsequent study.

4.3.1 Component Sizes, Distribution, and Growth

The number and size of components in the Δt -TEG is dependent on Δt . A natural question is to ask how many components there are in a temporal network and how the events are distributed between them.

In the limit $\Delta t \rightarrow 0$ the TEG will be completely disconnected (assuming no two events occur at once), however it is not guaranteed that as $\Delta t \rightarrow \infty$ that a single component will form. In fact in the limit $\Delta t \rightarrow \infty$ the components of the TEG contain distinct sets of nodes (Lemma 4.2.3) and correspond to the connected components of the time-aggregated temporal network. For intermediate Δt the structure of the TEG has a complex dependency on both the connectivity of the nodes (who connects to who) and the times between subsequent connections.

To characterise the network structure we look at the component size distribution of the Δt -TEG. We are also interested in the size of the largest component $|C_0^{\Delta t}|$. In particular understanding the growth of $|C_0^{\Delta t}|$ as a function of Δt gives clues to the temporal structure of the network; what fraction of the whole network does it contain and for what value of Δt does it reach 95% of its total size?

As an example, we look at a randomly generated temporal network. To generate a temporal network of N nodes with M events with a prescribed IET distribution X we perform the iteration:

1. Increment t to $t + \tau$ where τ is drawn from X
2. Draw u, v from $\{1, \dots, N\}$ without replacement
3. Add event (u, v, t) to the temporal network

for each event, after initialising $t = 0$.

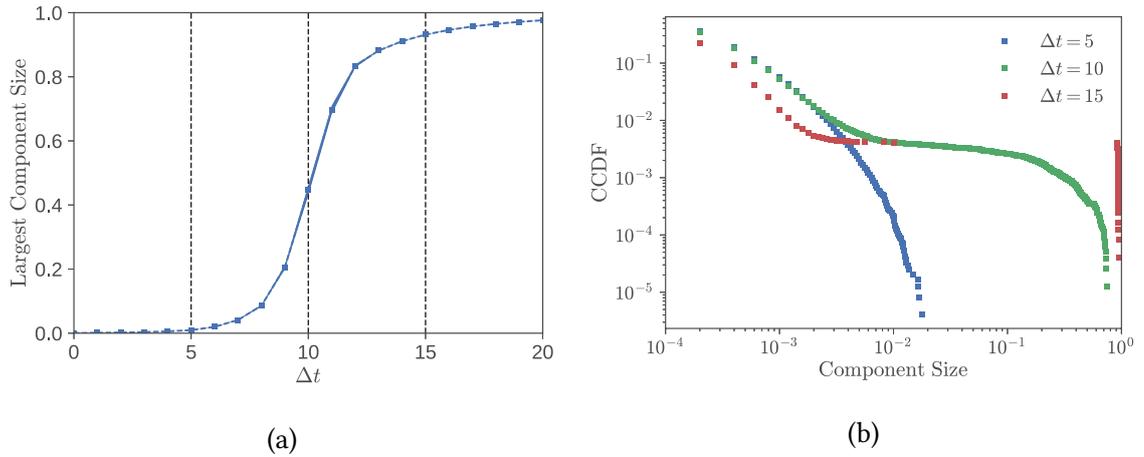


Figure 4.5: Temporal component dependence on Δt . (a) The size of the largest temporal component in the Δt -TEG as a fraction of the graph size for a random temporal network of 200 nodes and 5000 events. The largest component size has a sigmoidal dependence on Δt , with a sharp transitional period from being only a small fraction of all events ($< 10\%$), to containing almost all events ($> 90\%$). (b) The corresponding distribution of temporal component sizes for $\Delta t = 5, 10, 15$ constructed using an ensemble of random temporal networks. For $\Delta t = 5$ there are a range of component sizes however none which make up more than 10% of the network. For $\Delta t = 10$ components can take any size. For $\Delta t = 15$ components either make up the majority of the network, or are small isolated components.

In Figure 4.5 we see the results for a random graph where $N = 200$, $M = 5000$, and X is power-law distributed with probability density $P(x; a) = ax^{a-1}$, where $0 \leq x \leq 1$ and $a = 0.2$. Results are averaged over an ensemble of 100 temporal networks. The size of the largest component has a sigmoidal dependence on Δt , with only a small fraction of the TEG connected below a characteristic time, and the majority of events connected above (Fig. 4.5(a)). The average duration of the temporal network is 1000 meaning that when Δt is only 2% of the network duration, the majority of events are connected. Also, due to the random selection of nodes the largest component ultimately contains every event as $\Delta t \rightarrow \infty$. The distribution of temporal components (Fig. 4.5(b)) also display this transition. For $\Delta t = 5$ there is a continuous spectrum of component sizes although the maximum observed size is less than 10% of events. The probability of observing components any larger grows exponentially small. For $\Delta t = 10$ almost all possible

component sizes are observed. However, above the characteristic time at $\Delta t = 15$, the distribution is not continuous. Components either are a small fraction of the TEG, or are the majority fraction. There are no components of intermediate size.

Another aspect of a component size is its growth over time. As events are added to the temporal network, they may be added to one of the existing temporal components if they are Δt -connected to an event in those components. There may however be events introduced which are Δt -connected to two existing temporal components. These events cause the coalescence of the two components. By studying the growth of components over time we can observe the events which bring different parts of the network together understand how the network grows over time.

One way to visualise the temporal components is through a *temporal barcode*, as seen in Figure 4.6. This displays the components of the Δt -TEG, ordered by their size with the largest components at the bottom. Within each component, the individual events are plotted by a single vertical line. This visualisation allows us to see the duration of each component, its temporal position relative to other components, and the distribution of IETs within the component. While not enlightening for a random temporal network we will make use of the temporal barcode in the next chapter.

4.3.2 Motif and Inter-event Time Distributions

Beyond the structure of the temporal network, the TEG also prescribes the two-event motif types and IETs between any two adjacent events⁷. This allows us to aggregate both quantities and assess the distribution of the motifs and IETs not only as a whole, but within each component.

The simple temporal networks in Figure 4.7 have trivial motif distributions. In Figure 4.7(a) the only motif present is that of ABAC, reflective of the broadcasting type behaviour of node ϵ in this instant. If we were to consider the distribution of motifs in Figure 4.7(b) we would see an equal split between the ABAB and ABBA motifs. However,

⁷ For consistency with the work of [71] we will consider only *valid* motifs and their corresponding IETs.

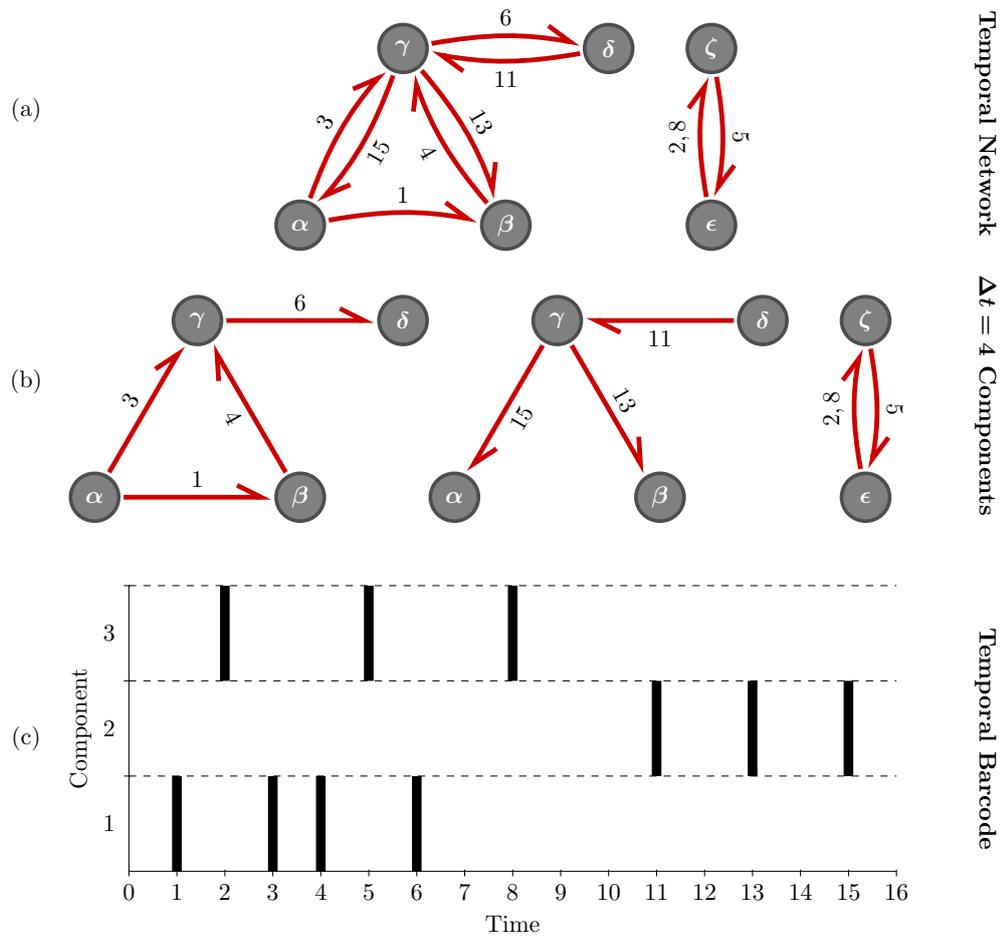
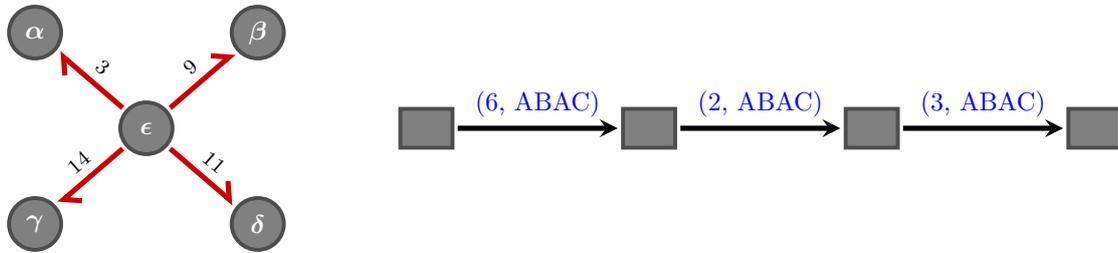


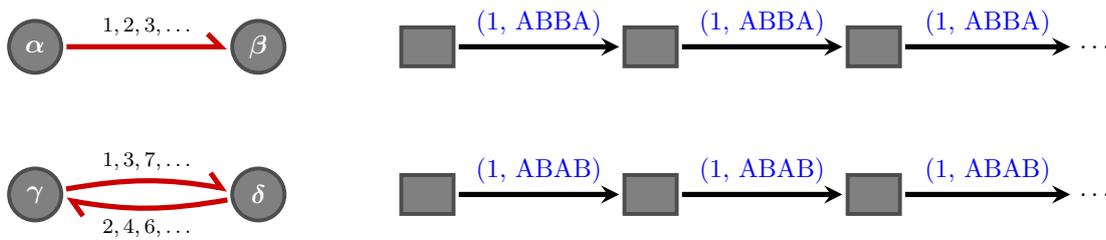
Figure 4.6: Illustration of the temporal barcode associated with a Δt -TEG. (a) A temporal network involving six nodes and nine events. Event labels represent the instantaneous time when that event occurred. (b) The temporal components of (a) when $\Delta t = 4$. (c) The temporal barcode of (b). There are three different components. Events in each component appear as black lines. Components 1 and 2 are distinct from 3 as they involve a distinct set of nodes. Components 1 and 2 are distinct as there is a gap greater than Δt between activity on the nodes.

considering the motif distribution of each component we see that there are in fact two distinct components containing either the ABAB or ABBA motif only. Without a suitable null model for the temporal network, analysing the motif distributions alone cannot give the significance of any observations [141, 142], and choosing a null model is non-trivial beyond time-shuffling and time-reversal [143, 144]. Comparing the temporal network with itself however allows us to gain information about the relative motif counts. Motif

counts can be compared across different node or event types, or even different intervals in the network, however, given the use of temporal components in the calculation of motif counts, comparing the motif distributions across temporal components is a natural way to proceed.



(a) (Left) a temporal network consisting of a central node messaging four other nodes in turn. (Right) the corresponding TEG.



(b) (Left) a temporal network consisting of two pairs of nodes. The bottom pair periodically reciprocate messages in turn, whereas in the top pair all messages are sent in one direction. (Right) The corresponding TEG.

Figure 4.7: Examples of temporal networks and their temporal event graphs.

Returning to the random temporal network example, and by considering the number of ways a particular motif can form, the motif distribution is given by

$$\Pr(x) = \begin{cases} \frac{1}{4N-6} & \text{for } x \in \{ABAB, ABBA\} \\ \frac{N-2}{4N-6} & \text{for } x \in \{ABBC, ABCB, ABAC, ABCA\}. \end{cases} \quad (4.1)$$

So, as $N \rightarrow \infty$, the ABAB and ABBA motifs are less likely to be observed and all other motifs are observed with equal probability. This illustrates why the random temporal network model is an unsuitable null model for social systems where one expects a degree of reciprocity.

Coupled to each motif, each edge in the TEG carries the IET between the two connected

events. This is the time between events which an individual node participates (described in Chapter 2). For the random temporal network, the time between consecutive events in the network is prescribed as part of the generating algorithm. In this case, the IETs are power-law distributed which is confirmed in Figure 4.8(a). In Figure 4.8(b) we plot the CCDF of the IETs of the TEG which instead are the times between consecutive events for each node. For real data, this distribution is a complex function of node interactivity and activity patterns. For the random temporal network however the distribution is geometric. This is due to each node having a constant probability of being in an event at each iteration.

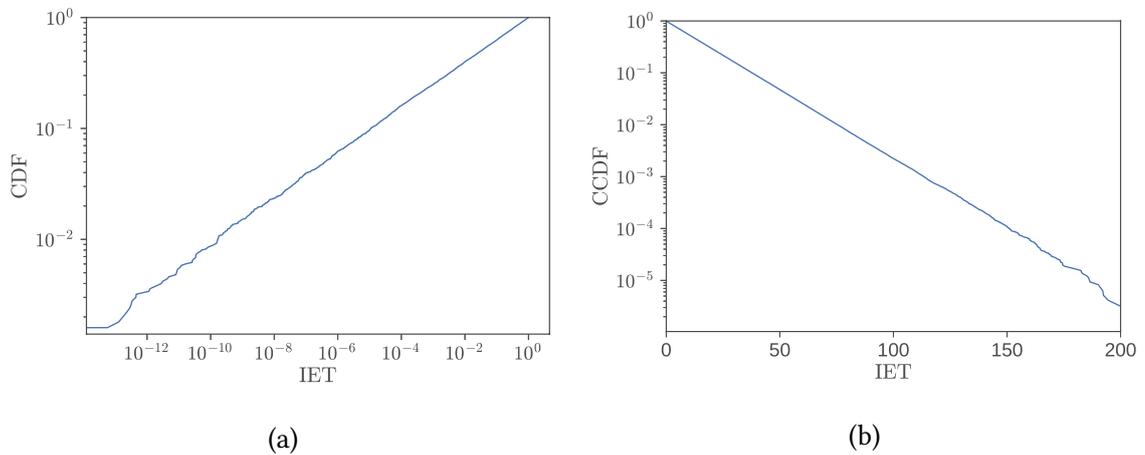


Figure 4.8: The IET distributions for the random temporal network. (Left) the CDF for the IET across the entire network. (Right) the CCDF for the IET distribution of the TEG, i.e., the time between consecutive events for each node.

Entropic Measures

For simple temporal events there are six possible motifs. However with generalisations of temporal networks such as allowing coloured edges or nodes there can be many more possible motifs. The full distribution of motifs can therefore be difficult to analyse and it is instead useful to use a single measure to capture the diversity (or predictability) of the motifs within a component. For this purpose we use Shannon's information entropy,

which for a distribution $\{p\}$ is given by

$$S(\{p\}) = - \sum_i p_i \log_2 p_i. \quad (4.2)$$

This takes a minimal value of zero when $p_k = 1$ and $p_i = 0$ for all $i \neq k$, corresponding to a fully predictable system, and a maximal value of $-\log_2 p$ when all $p_i = p$, corresponding to complete randomness. When there are six possible motifs the maximum entropy is $-\log_2 \left(\frac{1}{6}\right) \approx 2.58$. In both the examples in Figure 4.7 the entropy of each component is identically zero as each component consists of a single motif. In this sense, all these components are completely predictable. For the random temporal network there are four possible motifs occurring with equal probability in the large N limit. The entropy is therefore $-\log_2 \left(\frac{1}{4}\right) = 2$. Therefore, from a motif based standpoint, the random temporal network is not as random as possible.

Likewise, although less trivially, we can compute the entropy of the IET distribution. As the IET is a continuous variable we instead use the cumulative residual entropy (CRE) [145, 146] defined as

$$S_{\text{CRE}}(X) = - \int P(X > x) \log_2 P(X > x) dx, \quad (4.3)$$

where X is the IET distribution. The CRE shares many features with Shannon entropy (the CRE of a delta function is 0, for example), and for the purpose of this study provides a sufficient and consistent means to characterise the diversity in the IET distribution⁸.

4.3.3 Induced Aggregate Networks

In Chapter 2 we saw how a temporal network can be aggregated to form a single static network (or a series of static networks). The Δt -TEG provides a convenient way to decompose a larger temporal network, however being event-centric it can be difficult to assess the connectivity of the nodes within each component. This information can be extracted easily however by considering the static aggregation of the temporal component. The static network can then be analysed using standard methods to find

⁸ The variance can also be considered, however this performs poorly as a measure of diversity on distributions not well described by their mean, e.g. a bimodal distribution.

quantities of interest. In particular, we will be interested in the number of nodes, edge density, the fraction of reciprocated edges, and network diameter.

Studying the components of the decomposed network offers the advantage of understanding the role of nodes within a particular context, as opposed to consideration of the static graph of the full temporal network, which may be dense or noisy, or of fixed intervals which may dissect patterns of behaviour. Partitioning the random temporal network into intervals of fixed width results in a series of Erdős-Rényi (ER) static networks with edge forming parameter p dependent on the number of events in each partition. This gives the ‘Temporal ER Network’ as described in [147]. The aggregated networks of the TEG components by contrast are not in the class of ER graphs and their properties are yet to be determined.

4.4 Discussion

In this chapter we introduced the temporal event graph, a natural extension to the transmission graph [70], and an important precursor to the calculation of temporal motifs [71]. Furthermore, we showed that up to a translation in temporal components the TEG fully describes the temporal network. In this sense we can describe a temporal network by the behaviours of the nodes in the network; the motifs they participate in, and the times between their interaction. This combined approach offers a richer analysis of the temporal network than the study of the motif and IET distributions in isolation. We also showed in Section 4.3 various statistical properties of the TEG that we can use to classify temporal networks and the constituent components.

There are other, yet to be studied, uses of TEG. When calculating the motif distribution, Δt is often chosen heuristically and then results are checked for stability with respect to Δt . A fuller understanding of the TEG structure as a function of Δt should help inform the choice of Δt . Information from the component-wise motif and IET distributions may also better inform the choice of Δt as distinctive patterns of behaviour may appear as Δt is varied. We also neglected the motif conditional IET distributions, such as $\Pr(t|ABBA)$, the probability the IET was t given that the event formed an ABBA motif. We’ll explore

these further in Chapter 9 where we study them in the context of dynamical processes on networks.

In the next chapter we will use the TEG to model real world networks from the online social network, Twitter.

5

Temporal Events on Twitter

In the previous chapter we introduced the temporal event graph (TEG) which allowed us to dissect a temporal network, study the IET distributions, and characterise the behaviour through the motif distribution. Creating a meaningful null model for temporal networks is a difficult task (illustrated by the number of different approaches [132, 148, 149, 79]), and so we turn our attention to applying the TEG to real world networks and suitable randomised reference networks.

As we saw in Chapter 1 one of the major difficulties is ‘untangling the hairball’ of temporal events to be able to understand how the network formed and to identify nodes and components of interest. In this chapter we show that one way to do this is using the TEG and draw examples from the online social network, Twitter.

In order to capture the full amount of information provided by the social network we need to distinguish between different types of interaction between users. For example

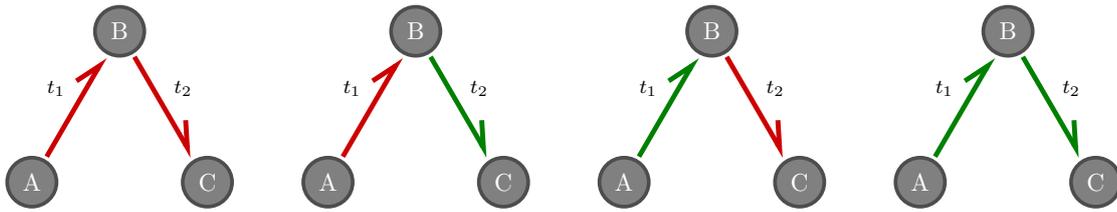


Figure 5.1: The four possible colourings of the ABBC motif.

in telecommunication networks one could differentiate phone calls or SMS messages, or as we see later, between messages and *retweets*. We therefore generalise motifs to incorporate *coloured* events which expands the space of possible motifs. In the absence of coloured edges there were six possible two-event motifs. When there are two possible event colourings there are 24 possible two-event motifs (Fig. 5.1), and in general with c distinct colourings there are a possible $6 \times c^2$ motifs.

Chapter Outline

In Section 5.1 we give an overview of the Twitter¹ social network and its relevance in modern society to everything from politics, news, advertising, and even early warning systems for earthquakes. We also show the various temporal networks we can extract from Twitter using their freely available API (application programming interface). Section 5.2 is devoted to case studies of samples from Twitter. In this section we show how we can use the temporal event graph (Chapter 4) to decompose the temporal network and find conversations (or lack of conversation) in the network. Finally, in Section 5.3 we discuss the possible impact of these studies and avenues for future research.

5.1 Twitter

The examples in this chapter are taken from the online social network and microblogging service, Twitter. All information and statistics presented in this section are valid as of March 2017 and will be subject to change.

¹ www.twitter.com

5.1.1 What is Twitter?

Twitter is a social network and microblogging service first introduced in 2006 [33]. Users can post short messages of up to 140 characters (known as a *tweet*) to the network (examples of which are given in Figure 5.2). These messages can contain URLs, images, and videos. All tweets are public to the entire social network and can be found by

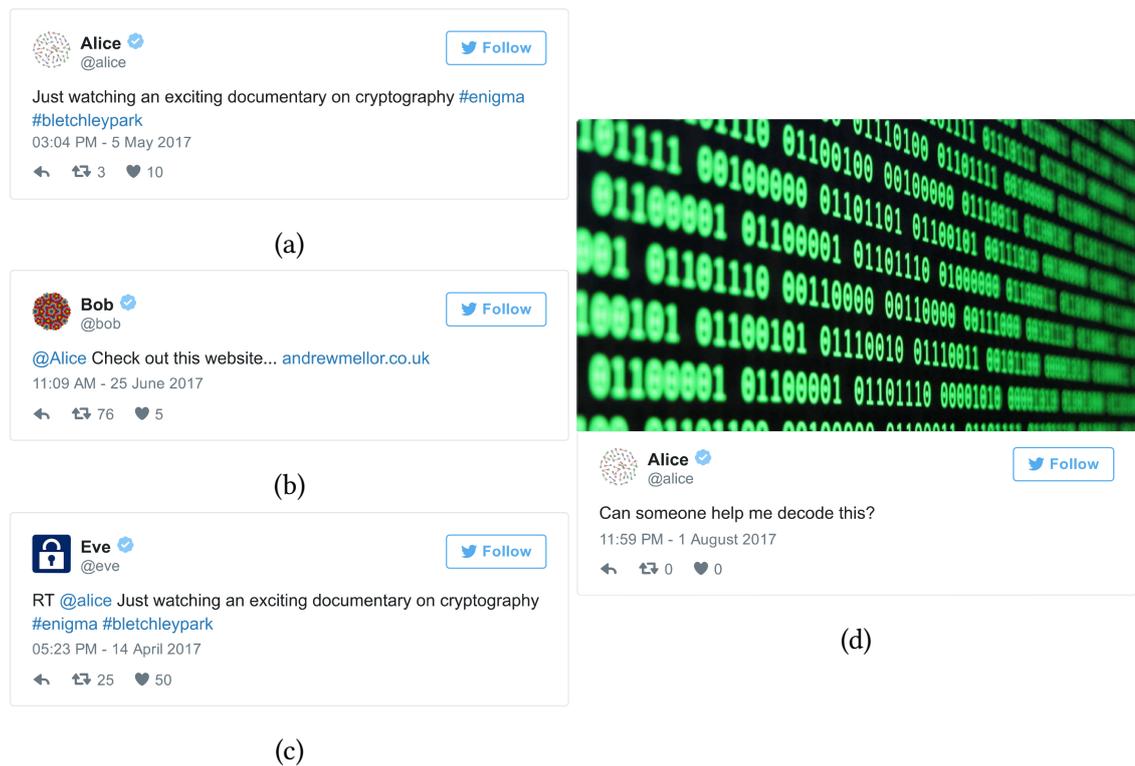


Figure 5.2: Examples of tweets. a) A singleton tweet which uses hashtags. b) A message from one user to another which shares a website link. c) A retweet of a previously seen tweet. d) A singleton tweet which shares a photo.

searching for keywords. In particular, words preceded by the hash symbol, # (also known as a *hashtag*), can be used to tag content as being relevant to a topic. Twitter allows users to search for all tweets containing a particular tag and provides a list of the most commonly used hashtags.

Users can choose to follow any other users, up to a maximum of 5,000 at any given time. This subscribes them to read any tweets that are produced by those that are followed. This relationship is not symmetric, i.e. you can follow without being followed by another

user. Because of this there can exist users who follow very few other users but are themselves followed by many. These are potentially influential users in the network as their messages can be reached by a relatively large fraction of other users. Furthermore, other users of the network can be *mentioned* by using the syntax @username in a tweet. Any number of users can be included (within the 140 character limit). This allows users to direct messages towards other users even if they are not followed by them.

Tweets can be categorised into four different types:

Singleton A tweet containing no mentions i.e., no use of @username.

Message A tweet containing mentions to one or more other users which is created independently of any other tweet.

Reply A tweet containing one or more mentions as a direct response to another tweet. This is not encoded in the tweet text but is available as metadata.

Retweet A direct copy of another tweet, preceded by the term 'RT @username' followed by the original tweet text. This is used as a means of extending a message to a new set of users while also being seen as an endorsement of the original tweet.

These distinctions are important for understanding the behaviour of users and the construction of temporal networks from Twitter data as will be discussed in Section 5.1.2.

Statistics

Over the course of its lifetime Twitter has become one of the leading online social networks with over 317 million monthly active users [129] (Fig. 5.3). By comparison the largest social network, Facebook, has 1.87 billion monthly active users [23]. There is a high level of activity on the network, with an estimated 500 million tweets per day [129]. This makes the platform both interesting to study but also difficult to manage computationally. The platform is globally relevant with 79% of all activity outside the US [24], although a notable omission is from China where access is restricted. This means that users are able to spread their messages widely across the globe.

82% of active users interact with the social network through a mobile device [24].

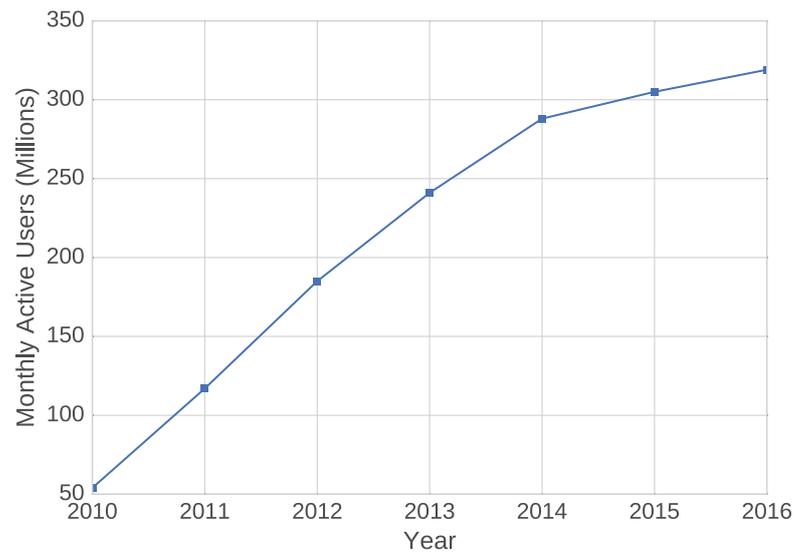


Figure 5.3: Growth of monthly active users from 2010 to 2016. Over this time period the social network has seen strong linear growth, however has grown slower since 2014. Monthly active users are defined as persons who have logged onto the social network at least once in a month (they do not need to post).

This is reflective of a shift in the way we interact with the internet in a move from desktop computers to mobile devices such as phones and tablets. This has occurred not only with social media, but can be seen across a number of sectors from banking to shopping. The mobility of internet access has introduced a number of phenomena including interaction with social media during live events occurring around the world. This ranges from tweeting about live events while on location (as has been the draw for Twitter as a news medium) to tweeting while watching a concert, sporting event, or the latest television show. In 2015, approximately 87% of people reported using a second digital device while watching television [150]. Because of this, the social media ‘buzz’ generated online before, during, and after a television show has become an important tool for assessing user engagement. This phenomenon is very apparent in the millennial generation (roughly those born between 1983 and 2001). 71% of them say tweeting during an event makes it more fun, 70% enjoy reading tweets while watching an event on TV, and 69% will use a hashtag to follow all the tweets related to an event [151]. Given the appetite for discussion of events on social media, there are many opportunities for advertisers to generate content with the ‘second screen’ in mind through event-specific

content and discussion.

Despite widespread use across the globe, Twitter has struggled to monetise its platform, making a loss of \$521 million in 2015 [152]. Approximately 90% of Twitter revenue is through advertising (\$1.99 billion of \$2.22 billion in 2015). For this advertising to be successful it requires a large and targetable user base. In their annual report [152], they list a number of risk factors to the platform that need to be addressed:

The platform must remain relevant: The platform needs to attract usage from celebrities, organisations, and subsequently users to increase the number of active users, all the while adapting to new trends and the movements of competitors.

An estimated 5% of active users are bots: As Twitter allows a degree of automation through its platform an industry of ‘fake’ accounts has appeared. These fake accounts are used to artificially boost follower numbers, generate higher activity levels for particular topics, and push an agenda or content piece. This is problematic for Twitter (and researchers) as it can be difficult to differentiate between human accounts and robotic (bot) accounts. If the latter come to dominate the social network then this will drive away users and advertisers.

Advertisers must be able to optimise campaigns: While Twitter has a large user base, not all content is suitable for everyone. Twitter needs to be able to better understand its users in order to give opportunities to companies to offer tailored advertisements. This includes targeting users, using particular advertising styles and content, and timing advertisements so that they have the greatest effect and largest potential audience.

The first risk is dependent on the executive choices and innovations that Twitter and other competitive platforms make, as well as the continued usage by high profile organisations and people. The latter two points pose two interesting questions that can be addressed using the framework of the TEG:

Can we systematically detect automated behaviour?

How can we understand user behaviour better and monitor it over time?

We address these questions in part in Section 5.2.

5.1.2 Data Structures and APIs

Twitter is an open platform and allows for access to samples of their data for public use.

APIs (Application Programming Interfaces)

There are two major APIs available to access Twitter data which are the *streaming* and *search* APIs [153].

The search API deals with individual requests for information about the platform or its users and to post data and updates. Functionality includes getting/posting a tweet, getting the followers of a user, or getting a list of the current trending topics.

The streaming API allows for the live streaming of all tweets which contain a set of keywords, involve a particular user, are geographically located within a bounding box, or a combination of all three. Use of this API is limited to collecting less than 1% of all tweets which occur. Sampling tweets at 1% is often referred to as the *garden hose* as opposed to the *firehose*, which is 100% of all tweets.

Regardless of which API is used, the data returned for an individual tweet has the same structure and is in JSON² format. In Table 5.1 we outline the main fields that we will consider that are contained within a tweet.

Each tweet also contains other metadata about the user, such as their follower counts and profile settings. In this chapter we will primarily be concerned with the interactions between users rather than the users themselves and so we ignore this information.

Networks

How is this data structured and how can we construct networks from it?

The easiest conceptual network is that of the *follower network*. This is a directed graph where a node is connected to another if they follow that other user. One can also define the ‘reverse’ network where the direction of the edge reads “is followed by”, which is

² Javascript Object Notation, a commonly used data format for data transfer over the internet.

Field	Description
tweet_id	A numeric identifier for the tweet.
user_name	The @username of the tweeter (also know as a <i>handle</i>).
user_id	A numeric identifier for the tweeter (persistent across username changes).
text	The main body of the text, up to 140 characters.
created_at	The time stamp of when the tweet was posted.
geolocation	The lat/lon coordinates of where the tweet originated (optional).
in_reply_to_tweet_id	The tweet_id of the tweet that this tweet was a reply to (optional).
retweeted_status_id	The tweet_id of the tweet that this tweet is a retweet of (optional).

Table 5.1: Information contained within a tweet.

more representative of information flow. The follower network does vary in time - users can follow other users or unfollow users at their own discretion. In many studies it is assumed to be static [37, 154, 155], or modelled by an adaptive network which evolves slowly over time [156, 34]. Due to restrictions of the Twitter API it is difficult to obtain the follower network through a method such as snowball sampling. The most followed user on the network has 85 million followers, which under current restrictions of one per minute would take approximately three years to collect, and due to having many followers, is more likely to appear in a sample.

The primary network that we will focus on is what is described as the *mentions* network. In this network the nodes are users and a directed temporal connection between nodes occurs when one node includes one or more @username mentions in their tweet. More specifically, when user A mentions user B (either as a message or reply) then an event $A \rightarrow B$ is created. If user A retweets a previous tweet of user B then an event $B \rightarrow A$ is created. The direction of edges here represents the flow of information in the network, as user A has actively pulled information to itself from B . Examples of how the mentions

network is created are given in Figure 5.4 where we have used the temporal network diagrams introduced in Chapter 2.

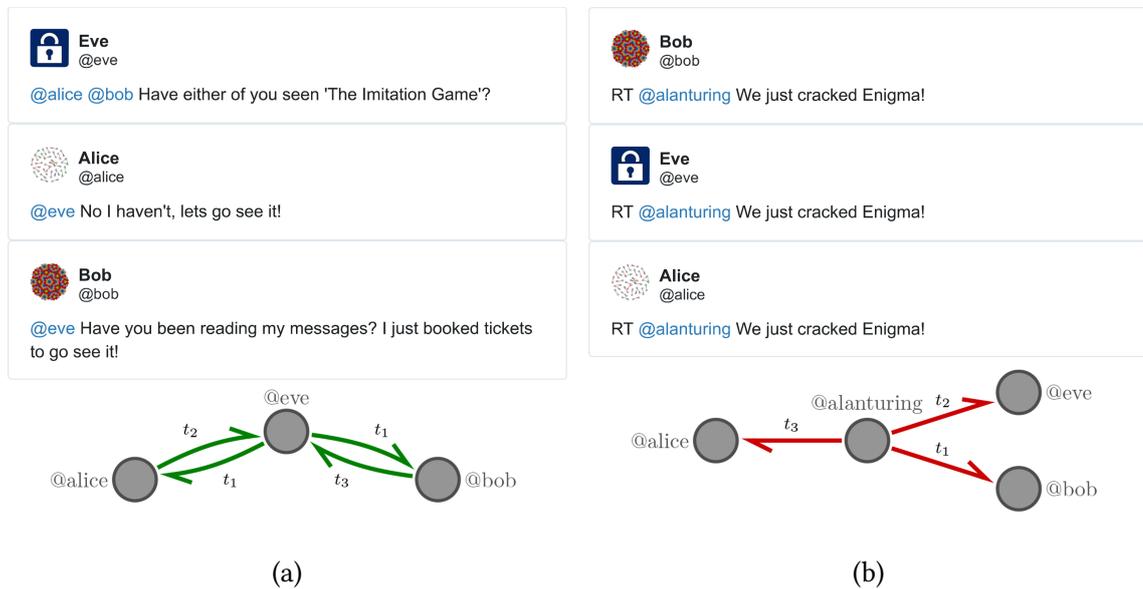


Figure 5.4: Example tweets and the corresponding temporal network. Tweets are ordered from top to bottom with times t_1, t_2, t_3 . a) A conversation between three users. b) Multiple retweets of a single tweet. Here the arrow direction represents the flow of information from original tweeter to retweeter.

This network captures the activity present in the network and, being activity-driven, captures the relationships between nodes within a particular time window. Consequentially the mentions network is a more ‘current’ network than the follower network, where connections may have been inactive for a long period of time.

5.2 Case Studies

The remainder of this chapter is devoted to analysing data collected from Twitter using the methods described in Chapter 4. We study three different datasets (described in Table 5.2) which span different timescales and topics. In each study we highlight a particular aspect of the TEG.

The data was collected by filtering tweets which contained the corresponding keywords listed in Table 5.2.

Name	N	M	Duration	Keywords	Comments
X Factor [†]	78k*	150k*	2h	xfactor	Reality TV show in the UK. 10 weeks of shows.
Triathlon [†]	53k	126k	1 week	leeds, triathlon	Build up to the ITU World Triathlon Leeds 2016
Oxford	246k	528k	1 month	oxford	Tweets mentioning Oxford in March 2017

* Average values per week.
[†] Data provided by Bloom Agency.

Table 5.2: Datasets considered in this section. Here N is the number of users in the data, and M is the number of tweets (events).

5.2.1 The X Factor

The X Factor is a reality television show which originated in the United Kingdom. The premise of the show is a singing contest where acts perform and are subsequently judged by a panel and the public. The show airs weekly (over 10 weeks) with one act being voted off each week. The show is spread across both Saturday and Sunday, however for the purposes of this study we will consider only the Saturday show which contains the live performances.

The show is of interest to advertisers: it drew in an average of 8.61 million viewers every week in 2014 [157]. Furthermore, as the show encourages audience participation through voting, the audience are engaged rather than being passive. The temporal networks generated by viewers mentioning the X Factor are driven by the events which occur during the show and as a result show sudden spikes in activity. Conversations are primarily about the performances however other topics include the judging panel, the adverts, and the viewers' rituals when sitting down to watch the show.

In Figure 5.5 we give the 5s-TEG barcodes (i.e. $\Delta t = 5$ seconds) for three of the shows (weeks 1, 6 and 9) which show the largest temporal components by number of events. From these we can see that the temporal networks are highly connected; in weeks 1 (a) and 9 (c) the duration of the temporal network is spanned by three or fewer components. By contrast the 5s-TEG for week 6 (b) consists of many overlapping components. This can be explained by considering the behaviour of the audience and the structure of the show. In week 1 the acts are unfamiliar leading to conversations centred around the X

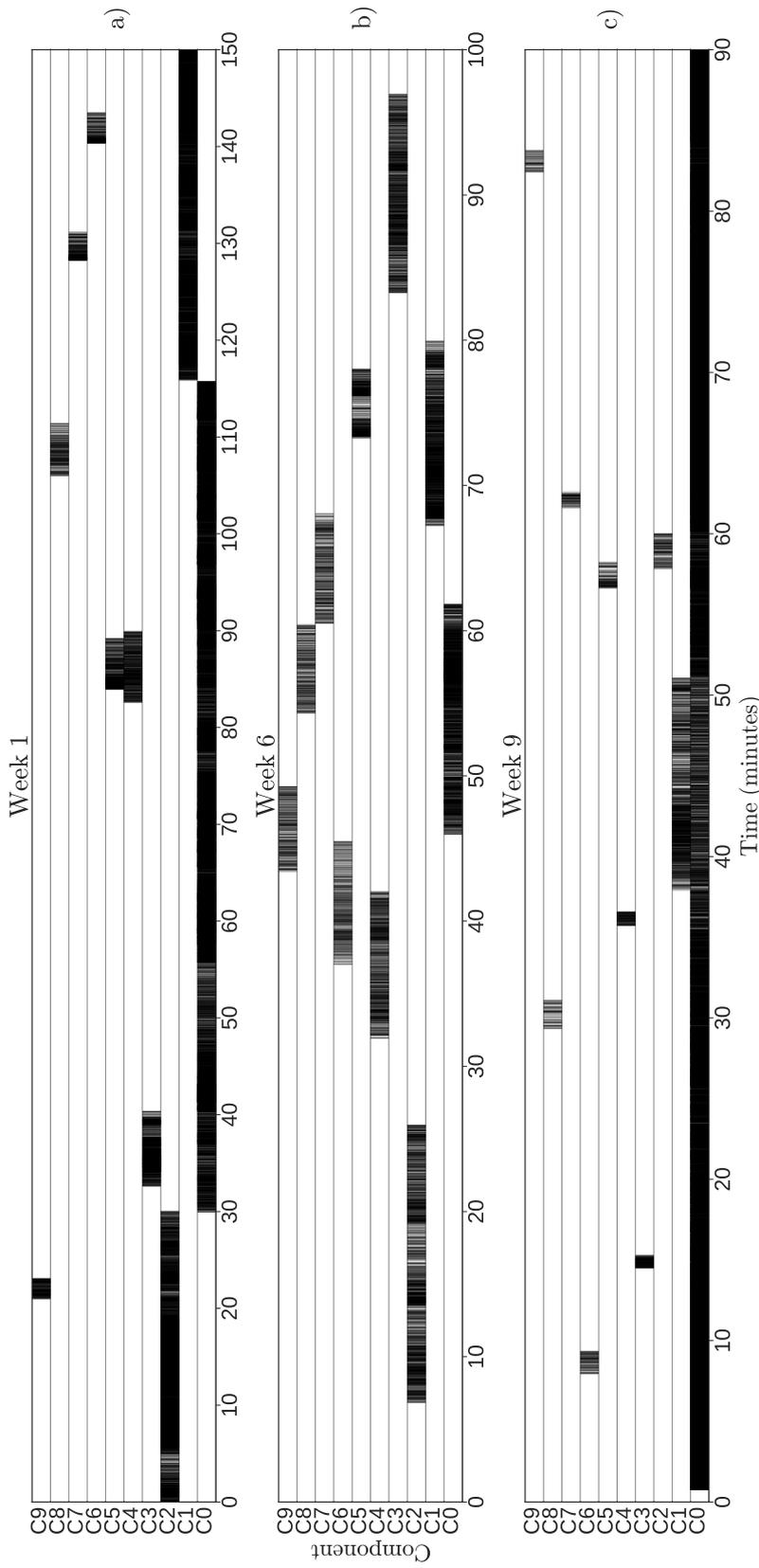


Figure 5.5: Temporal barcodes (see Chapter 4) of the 5s-TEG for the X Factor dataset. a) (Week 1) The first show sees the majority of conversation centred around the central account for X Factor with the TEG consisting predominantly of a single large component. b) (Week 6) Overlapping components which span the duration of the show, indicative of multiple separate conversations occurring, and breaks in activity. c) (Week 9) As the only four acts are left in the competition the conversations centralise once again, resulting in the majority of events belonging to a single component.

Factor account itself, or about all acts. In week 6 there the audience are likely to have developed an affinity to one or more act and will be therefore more likely to discuss them individually. Finally in week 9 there are sufficiently few acts left in the show that the conversations return to being centralised.

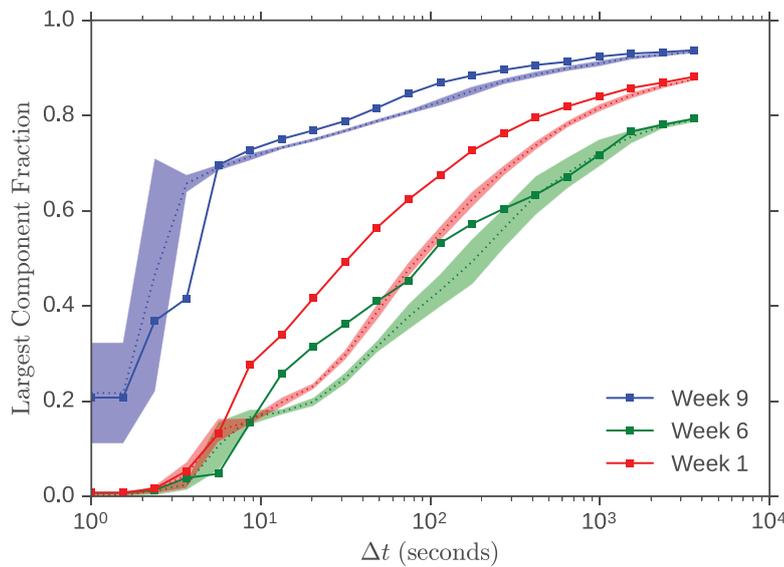


Figure 5.6: Solid (marked) The growth of the largest component of the TEG with Δt for the X Factor dataset, as a fraction of all events. (Dashed) largest component growth averaged over an ensemble of 30 time-shuffled networks with two standard deviation intervals. Notably, week 9 shows a rapid growth of the largest component in comparison to the other two weeks, and furthermore has the largest size for any value of Δt .

To investigate the structure of the Δt -TEG in more detail we assess its dependence on Δt (shown in Figure 5.6). This reaffirms our observations from the temporal barcodes; the temporal networks are highly connected, and as $\Delta t \rightarrow \infty$, the largest component makes up the majority of the network³. All three weeks show a sigmoidal dependence on Δt , which is also observed across the remaining weeks. This suggests a characteristic timescale on which the pivotal edges in the TEG (in terms of connectivity) form. Perhaps surprising is that for week 9 even when Δt is as small as 5 seconds the largest component contains over half the temporal events⁴. This was due to tweets during this show

³ Although the shows vary in the number of events, the number of events and largest component fraction are only weakly correlated.

⁴ The fraction of users in the component is roughly the same also.

predominantly being directed towards one or two users who were the target of a tweet every 5 seconds or less. We also observe that the growth of the largest component on time-shuffled networks occurs slower as a function of Δt for weeks 1 and 6, but not for week 9. This can be easily explained as the time shuffling destroys all temporal correlations between events and hence adjacent events are less likely to be Δt -adjacent. This explanation breaks down in week 9 however as there are enough events in the component (94k) which spans 5400 seconds, such that many events occur within a 5 second window.

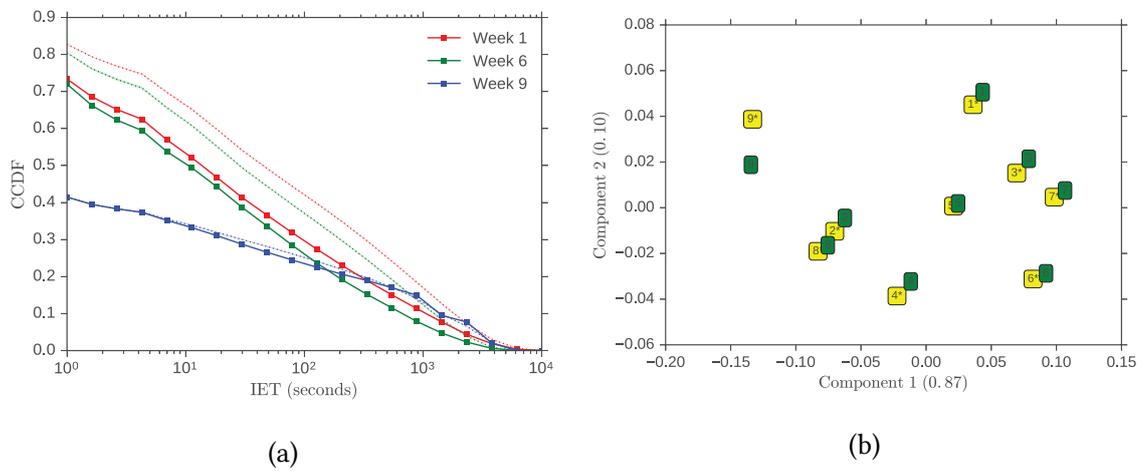


Figure 5.7: a) (Solid) The IET distribution of the TEG for three shows. (Dashed) the IETs averaged over an ensemble of 30 time-shuffled versions of the network. b) The two-event motif distribution for each show, reduced to two dimensions using principal component analysis. In green are the shows themselves (labelled with week number) and in yellow are the motif distributions averaged over an ensemble of 30 time-shuffled versions of the network (labelled with an asterisk).

In Figure 5.7 we characterise each show by the IET and motif distributions individually. Figure 5.7(a) shows that week 9 has a significant number of IETs which are of only 1 second, and weeks 1 and 6 have a similar distribution. Despite the average week 6 IET (230s) being less than week 1 (350), the largest component of the Δt -TEG for week 1 is larger than that of week 6 for any value of Δt (Fig. 5.6). This suggests that nodes in week 1 are more interconnected and centred around a small number of users and topics than in week 6. We can also look at the motif distribution of each show

(shown in Figure 5.7(b)). Here we calculate the two-event motif distribution for the TEG ($\Delta t \rightarrow \infty$) and use principal component analysis to reduce the distribution to two dimensions. The shows have notable differences in their motif distribution which again reaffirms that the TEGs have different structures. The motif distributions for the time-shuffled data (yellow) show little variation from the true data. This is due to the prevalence of symmetric motifs (such as ABAC) and the fact that time-shuffling changes only the temporal ordering.

5.2.2 Triathlon

This dataset focuses on the ITU World Triathlon hosted Leeds in 2016. In addition to the event itself, which occurred on Sunday 12th June, the dataset covers the entire week building up to the main event and consists of any tweet which contains at least one of the words ‘triathlon’ or ‘leeds’ during this period.

In Figure 5.8 we examine the structure of the Δt -TEG across different timescales. Despite filtering using a relatively broad keyword covering a city, the largest component of the TEG makes up a significant portion of the TEG (64710 of 79680 events) across the entire week (Fig. 5.8). The largest component sees near constant activity with periods of lesser activity naturally occurring between midnight and 6AM.

In order to investigate the largest component of the TEG we consider its structure when we reduce Δt to 30 minutes. This breaks the giant component into multiple components, the largest of which represent conversations which span the days of the week (Fig. 5.8(b)). In addition to the daily conversations there are other components which span the night time hours. These components represent distinct conversations which discuss topics unrelated to the triathlon. The largest component of the 30min-TEG is that of Sunday, the day of the triathlon.

Reducing Δt further to 5 minutes (Fig. 5.8(c)), we can decompose the ‘Sunday’ component even further. The largest component spans the daylight hours, although the activity level of the component varies throughout the day, peaking during the late afternoon when the elite race occurred. Further reduction of Δt sees this component

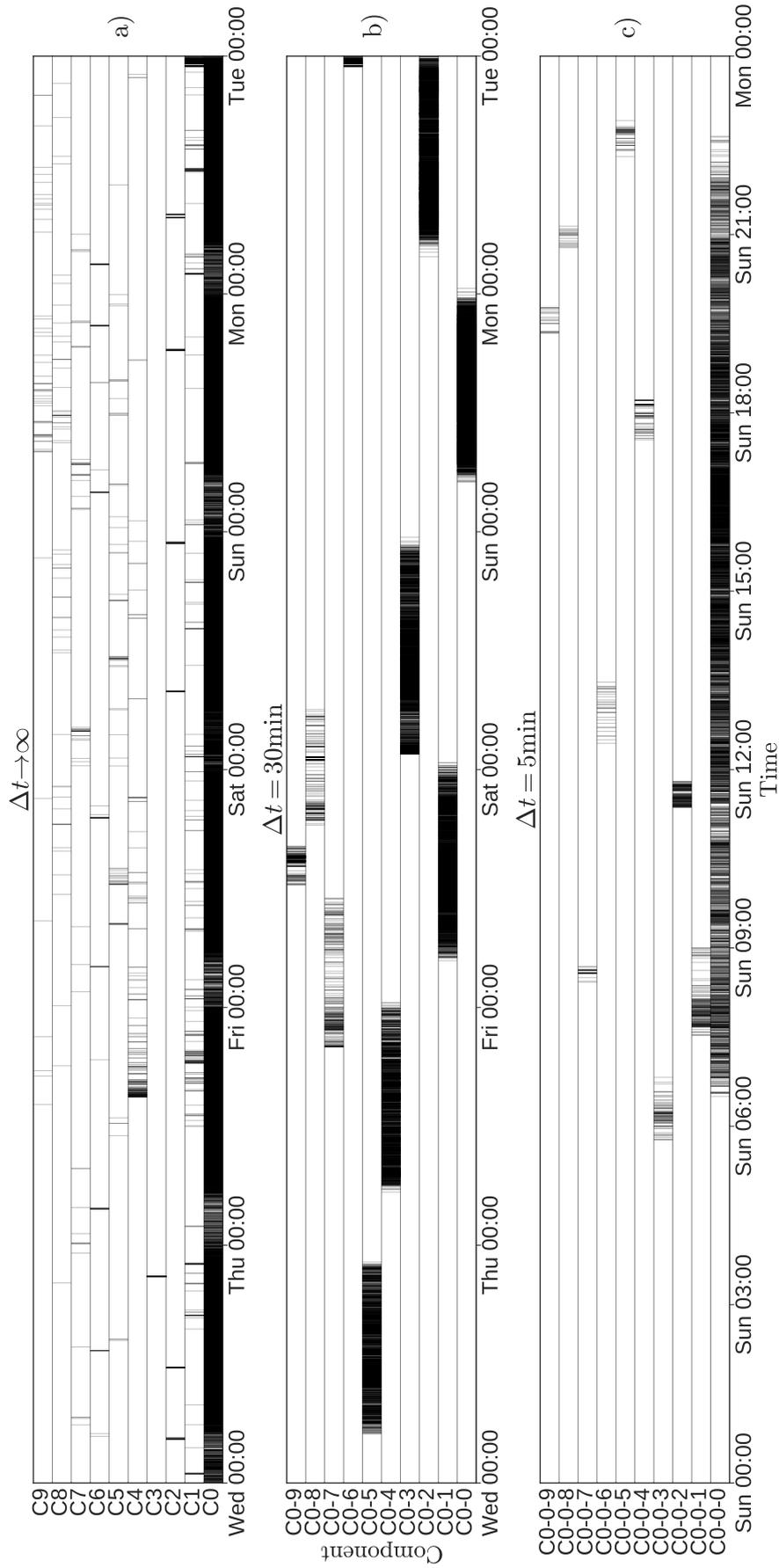


Figure 5.8: Temporal barcodes (see Chapter 4) of the Δt -TEG for the Triathlon dataset. As we reduce Δt , we focus on only the largest component of the previous Δt -TEG. a) $\Delta t \rightarrow \infty$. b) $\Delta t = 30$ minutes. c) $\Delta t = 5$ minutes.

split into a morning component focused on the amateur triathlon which preceded the main elite event, and the elite event itself. Alongside the largest component there are other components which co-occur. At this resolution the IET patterns of these smaller components become clear. From Figure 5.8 we can see that component C0-0-6 appears to show some periodic behaviour, which may be generated by an automated account. We want to be able to systematically detect this kind of behaviour, which we do next.

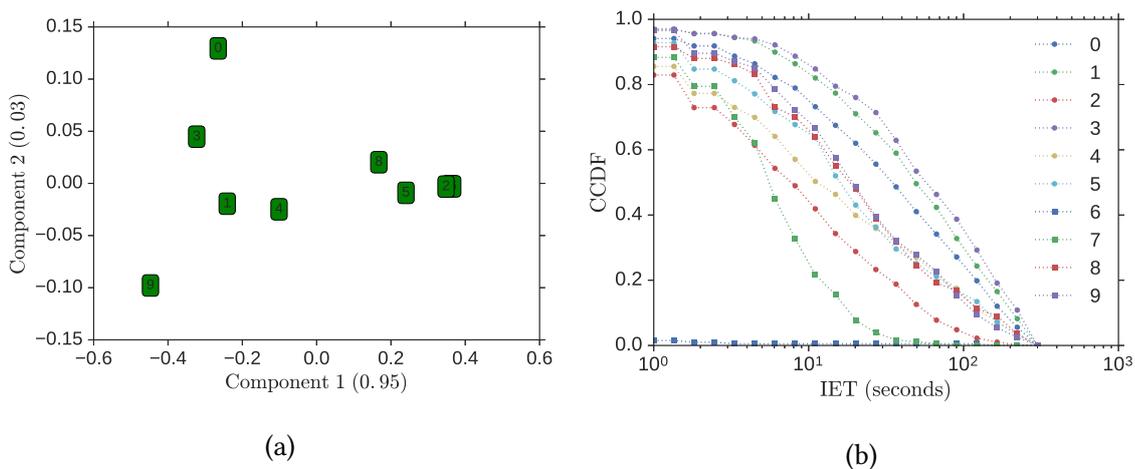


Figure 5.9: The motif and IET distributions across the largest (by number of events) components of the 5min-TEG of the Triathlon dataset. a) A reduction of the motif distribution to two dimensions using principal component analysis (PCA) with the explained variance percentage on the axes. The components differ predominantly by the presence of the ABAC motif where both events are retweets. Components aligned to this behaviour are towards the right. Variation in the second component is caused by a mixture of motifs. b) The IET distribution for each component. Components qualitatively share the same distribution, although with differing scaling. The notable exception is component 6 which has IETs of 1 or 2 only. All IETs are less than 300s (5 minutes) by construction of the 5min-TEG.

As we did with the X Factor shows, we can characterise the temporal components of the 5min-TEG using the motif and IET distributions, seen in Figure 5.9. The distinguishing feature of the motif distributions (Fig. 5.9(a)) is the presence of the ABAC retweeting motif. Conversations expressing only that motif are on the right, whereas the more diverse components are on the left. Components 0, 3, 1 and 9 are therefore less likely to

show retweeting behaviour and are hence more likely to contain original messages (and therefore original content). The IETs for the components (Fig. 5.9(b)) follow roughly the same distribution although with different scalings. The outlier in this instance is component 6 which has all IETs less than 5s. This means that this component would persist as Δt is decreased, whereas the others would decompose.

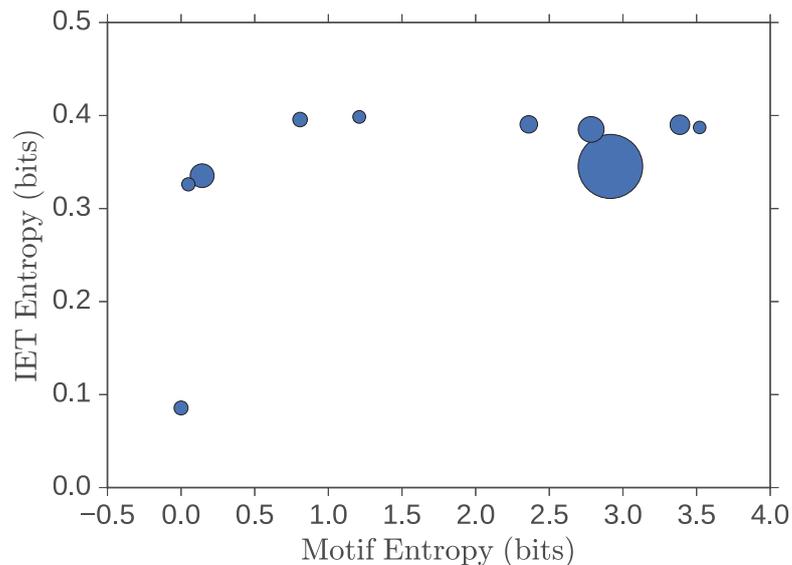


Figure 5.10: Component motif and IET entropy, calculated from the corresponding distributions with components scaled by the number of events they contain. Components with low motif and IET entropy are predictable and are most likely generated by automated accounts. Components with low motif entropy engage in predominately one behaviour, in this case message and retweet broadcasting (ABAC). Components with high motif entropy contain a diverse mix of behaviour.

The motif and IET distributions are useful to characterise the different temporal components, however as high dimensional objects they can be difficult to study, especially if they can not be significantly reduced in dimension. One way to measure the predictability of a component is through entropy (detailed in the previous chapter). This can be used to identify components which behave in a particular way, or are periodically active. This is extremely useful in identifying ‘spam’ accounts in the network. In particular this method can identify entire clusters of spam accounts which individually behave normally but operate on a global timer.

In Figure 5.10 we plot the motif and IET entropy for the top ten components of the 30min-TEG. The band of components with IET entropy between 0.3 and 0.4 represent conversations with a diverse array of IETs, and are therefore likely human generated. Components with low (or zero) motif entropy contain predominantly one behaviour. The component in the bottom left (having low motif and IET entropy) is formed of multiple spam accounts. The components in the top left consist of only retweeting behaviour of a single piece of content, i.e. there is no real conversation generated. On the right-hand-side are components which are diverse in behaviour and IETs, representing more natural conversations in the network⁵. These observations can inform a generalised method for the detection of spam accounts or clusters of spam accounts. If both the motif and IET entropies for a temporal component fall below a critical value (to be determined from larger scale studies) then these components can be flagged as possible automated accounts. They can be subsequently checked manually to confirm their authenticity, or checked algorithmically using auxiliary data such as text and user profiles.

Finally, we qualitatively assess the structure of the time-aggregated temporal components, seen in Figure 5.11. In Figure 5.11(a) there is a complex array of retweet stars with interconnected messages. This suggests that the component may be able to be decomposed further by reducing Δt . The behaviour of the nodes in component 2 (Fig. 5.11(b)) is evident from the aggregated graph; there are many retweets of a single central node and little other behaviour. This is reflected in the motif entropy which is approximately zero. The behaviour of component 5 (Fig. 5.11(c)) is similar to that of component 2 however there are multiple nodes being retweeted, and a selection of ‘bridging’ nodes which retweet two or more different users, effectively bringing these communities together. Looking at a component on the periphery, such as component 149 (Fig. 5.11(d)), we see almost a complete clique of four nodes with very little retweeting. This is more representative of a conversation in the network, and can be characterised by the high degree of reciprocity.

⁵ The theoretical maximum motif entropy for two-event motifs with three event colourings is $-\log_2\left(\frac{1}{6 \times 2^3}\right) \approx 5.58$, which occurs when all motifs appear equally.

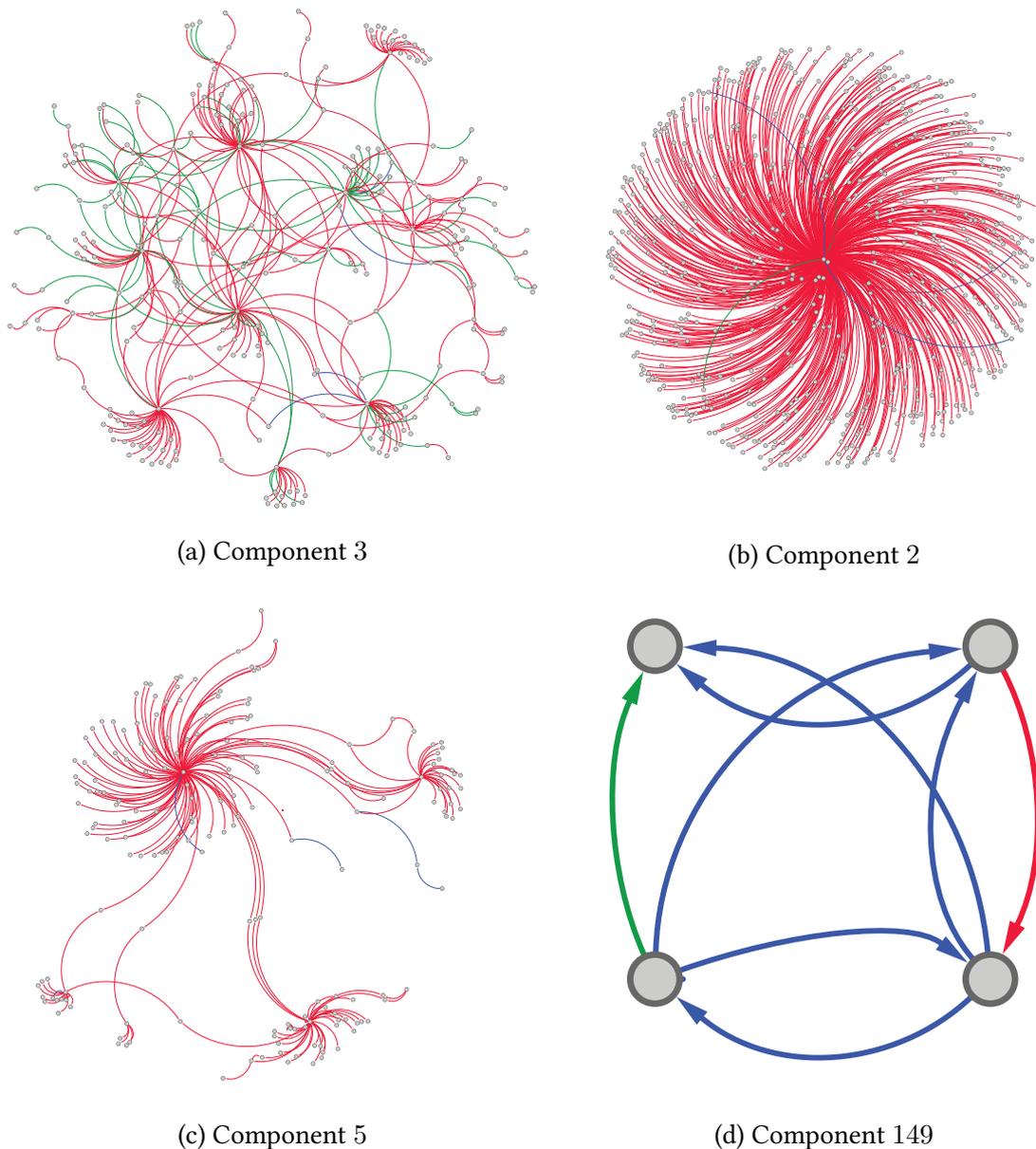


Figure 5.11: Aggregate graphs of temporal components from the 5min-TEG of the Triathlon dataset. Here red events are retweets, green are messages, and blue are replies. a) A complex mix of retweeting and messaging behaviour. Individual behaviours may become transparent with a reduction in Δt . b) Retweeting behaviour. A single user has been retweeted multiple times in the component duration which makes up the majority of the component. c) Retweeting behaviour with bridge nodes. Similar to the structure of (b) however there are multiple users retweeted and a select few ‘bridge’ users who have retweeted more than one user. d) Conversational behaviour between a small number of nodes, indicated by reciprocated links, a clique-like structure, and the abundance of reply tweets.

5.2.3 Oxford

The final dataset we study is a month long sample of all tweets containing the word ‘oxford’ during the month of March 2017. This data is different from the previous two studies which focused on a particular event (TV show or race), and our aim is instead to monitor the temporal network to pick out topics of conversation where and when they occur. As before, we can construct the Δt -TEG and vary Δt to assess the timescales of the temporal network. Here we focus on the 1h-TEG which decomposed the temporal network giving approximately 400 components with over 50 events in them⁶.

By studying the content of each event, the topic of each component can be discerned. This can be done systematically by analysing word and bigram frequencies (a widely studied problem itself under the name of topic modelling). In Table 5.3 we describe

#	N	M	% Retweets	Topic	Duration
0	8794	9667	88	Lawsuit where court decided based on use of Oxford comma	70h
1	4239	4517	98	The conditional offer for Malala Yousafzai at Oxford University	40h
2	3267	3341	98	US television host Rachel Maddow and her Oxford credentials	61h
3	3070	3290	100	Unknown (Thai)	44h
4	2999	3099	99	Young Malaysian politician debating Oxford offer	38h
5	2265	2381	77	The death of Inspector Morse creator, Colin Dexter	15h

Table 5.3: Descriptions of the largest components of the 1h-TEG for the Oxford dataset. Here N is the number of nodes (users) and M the number of events.

the top six components across the month by number of events. The topics include the offers for study for a prominent Malaysian politician and Nobel prize laureate Malala Yousafzai, the death of the creator of Inspector Morse (set in Oxford), and a lawsuit which was decided by the lack of an Oxford comma in legislation. Other components outside the top ten included news stories from across Oxford (fires, robberies, etc.) and the results of local sports teams.

While there are many methods of topic detection by means of natural language processing, what we highlight here is that the temporal components are often centred

⁶ A method for choosing Δt systematically has yet to be devised.

around one particular topic. This means that in calculating the components of the Δt -TEG we also can retrieve the topics discussed in the temporal network. Intuitively this makes sense, events which occur close together between the same nodes are likely to share some of the same information. One way to fully understand how temporal components and topics are related is to use the latest state of the art method in topic detection to decompose the temporal network into topical components and compare and contrast these with the temporal components of the Δt -TEG.

For this dataset the top components are primarily composed of retweets (all over 75%). Some topics however sparked more original conversation such as the emotive death of the Inspector Morse creator, or the absurdity of the use of the Oxford comma in a lawsuit. Are the top conversations on Twitter purely driven by retweets? Most likely not and in fact the high proportion of retweets in this case and the retweet motifs in the previous two cases highlights an issue with the method of sampling - retweets are more likely to be sampled. Sampling by keyword means that for any tweet containing the keyword, all retweets of that tweet will necessarily also contain the keyword. This is not true of replies or messages where the topic of conversation may only be mentioned once before being omitted or referred to using a pronoun. We look to address this issue by suggesting a new sampling method in Chapter 9.

5.3 Discussion

In this chapter we have looked at the social media platform and microblogging service, Twitter, and how it is used as a place to share content, discuss the latest events, and report news. In particular we have modelled the Twitter mentions network as a temporal event graph, and shown that it provides a natural way to decompose the temporal network.

In the case studies presented in this chapter we have used the TEG to quantify how centralised the temporal networks are in terms of the largest component size of the TEG. Here we saw how the conversation surrounding the X Factor was very centralised and very well connected even on a five second timescale. By contrast the Triathlon conversation exhibited multiple communities and timescales over which conversations

occurred. This new information offers the opportunity for advertisers and brands to decide where to push content, either centrally within the conversation or to attempt to bridge temporal components and communities to possibly drive further conversation.

We also showed that the IET and motif times can characterise different temporal networks and the temporal components within them. This means that behaviours can be quantified and isolated in a way not previously possible. In our case studies we also highlighted a number of interesting statistical properties of these networks, from motif and IET entropy to the properties of the aggregate graph. We also used event metadata such as text and tweet type to augment our understanding of the temporal networks and temporal components. This type of analysis can easily be extended to include more advanced descriptors and methods such as topic modelling, natural language processing, and image recognition for attached media. Using these new measures advertisers will be able to better classify individual user and collective behaviour, opening up new avenues for targeted marketing campaigns and strategies, all the while ensuring that resources are not wasted on autonomous accounts.

By collecting and understanding these new statistical features of networks we open up the possibility of applying statistical, predictive models to user behaviour and the generation of temporal networks. To this end, it may be possible to generate synthetic temporal networks with prescribed motif and IET distributions or predict the probability of an event occurring at some time in the future. One thing that has become clear by taking a behavioural view of the temporal network is that there is a bias towards sampling retweets from Twitter while potentially many other conversations are missed. Any analysis of data collected from Twitter in this fashion must therefore account for this bias, or better yet, avoid it to begin with.

These preliminary studies raise many questions about the structure of Twitter and the TEG in general: What is the temporal component size distribution for the entirety of Twitter? How long do we need to observe Twitter before the TEG consists of a single one giant component? What timescales do we see and how do we effectively pick Δt ? These questions can hopefully be addressed by studying the structure of the TEG on other well studied temporal networks and reference models.

Part II

Modelling

6

Model Methods

This chapter is intended to support the subsequent two chapters by introducing the methods required to study stochastic processes and probabilistic models. There are no new results in this chapter and so those familiar with the concepts of the master equation, Fokker-Planck equation, and dynamical processes on networks may wish to skip this chapter. Details on these methods can be found in [158–163] for stochastic processes, and in [64, 7, 65, 164, 165] for network dynamics.

Chapter Outline

In Section 6.1 we describe a general system using a state vector and see how the evolution of the system can be described exactly through a master equation. We then provide derivations of the most commonly used approximations of the system dynamics, namely the Fokker-Planck equation and the mean-field equations. This formulation will provide

the framework for Chapter 7.

In Section 6.2 we show how the methods of Section 6.1 are often unsuitable for systems evolving on network structures and provide a brief overview of the methods used to study these systems (some of which are used in Chapter 8). Finally, in Section 6.3 we discuss the approaches outlined in the chapter and provide references to additional reading material.

6.1 From the Master Equation to Mean-field

To begin modelling a system, we need to be able to fully describe the state of the system at any given time. The system can be described using a *state vector*

$$\mathbf{X}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{bmatrix},$$

which describes N different properties of the system. For chemical reactions involving multiple chemical species, the elements of \mathbf{X} may describe the number of molecules of each species (or their concentration), assuming that the chemicals are well-mixed. For more complicated systems, such as epidemic spread on a network, the state vector may describe the state of each individual in the system, whether they are infected or not.

Since our system evolves stochastically, we consider the probability that the system is in a state \mathbf{X} at time t , denoted $P(\mathbf{X}, t)$, and make the assumption that the set of joint probabilities

$$P(\mathbf{X}_1, t_1; \mathbf{X}_2, t_2; \dots)$$

exist. Furthermore we can also define the conditional probability densities

$$P(\mathbf{X}_1, t_1, \mathbf{X}_2, t_2; \dots | \mathbf{Y}_1, \tau_1; \mathbf{Y}_2, \tau_2; \dots),$$

where it is implicitly assumed that

$$t_1 > t_2 > \dots > \tau_1 > \tau_2 > \dots$$

An important class of stochastic processes is those with the Markov property, that is, the conditional probability is determined entirely by the knowledge of the most recent state, i.e.,

$$P(\mathbf{X}_1, t_1, \mathbf{X}_2, t_2; \dots | \mathbf{Y}_1, \tau_1; \mathbf{Y}_2, \tau_2; \dots) = P(\mathbf{X}_1, t_1, \mathbf{X}_2, t_2; \dots | \mathbf{Y}_1, \tau_1).$$

A process satisfying the Markov property is commonly called a *memoryless* process, as the behaviour of the system at any particular point in time is dependent only on the state vector at that time. For the remainder of this thesis we will consider only systems which satisfy the Markov property.

6.1.1 Master Equation

Once able to describe the system in terms of the state vector and the probability densities we look to understand how these densities evolve over time (here treating time as a continuum). Let $T_\tau(\mathbf{X}, \mathbf{X}')$ be the probability of a transition from state \mathbf{X}' to \mathbf{X} in a time interval of length τ . In the limit $\tau \rightarrow 0$, we enforce that only one state transition can occur and see that T_τ has the form

$$T_\tau(\mathbf{X}, \mathbf{X}') = (1 - a^{(0)}(\mathbf{X})\tau)\mathbb{1}_{\mathbf{X}=\mathbf{X}'} + W(\mathbf{X}, \mathbf{X}')\tau + \mathcal{O}(\tau^2) \quad (6.1)$$

where $W(\mathbf{X}, \mathbf{X}')$ is the transition probability per unit time from state \mathbf{X}' to \mathbf{X} . The first term is the probability that no other state transition takes place in the interval, with $a^{(0)}(\mathbf{X})$ ¹ given by

$$a^{(0)}(\mathbf{X}) = \int W(\mathbf{X}', \mathbf{X}) d\mathbf{X}'.$$

The *Chapman-Kolmogorov* equation (CKE) is an identity that must be obeyed by the conditional probabilities of any Markov process and is given by

$$P(\mathbf{X}_3, t_3 | \mathbf{X}_1, t_1) = \int P(\mathbf{X}_3, t_3 | \mathbf{X}_2, t_2) P(\mathbf{X}_2, t_2 | \mathbf{X}_1, t_1) d\mathbf{X}_2. \quad (6.2)$$

The CKE states simply that the probability of being in a state at a later time, given an initial state, is given by the integral of the probability over all intermediate jumps

¹ The use of the notation $a^{(0)}$ becomes apparent in Section 6.1.2

between the two states². In terms of the transition probabilities the CKE becomes

$$T_{\tau+\tau'}(\mathbf{X}_3, \mathbf{X}_1) = \int T_\tau(\mathbf{X}_3, \mathbf{X}_2) T_{\tau'}(\mathbf{X}_2, \mathbf{X}_1) d\mathbf{X}_2$$

assuming $\tau, \tau' > 0$ and that rates are time independent.

Now, using the small τ expansion for T_τ (6.1), the CKE becomes

$$\begin{aligned} T_{\tau+\tau'}(\mathbf{X}_3, \mathbf{X}_1) &= [1 - a^{(0)}(\mathbf{X}_3)\tau'] T_\tau(\mathbf{X}_3, \mathbf{X}_1) + \tau' \int W(\mathbf{X}_3, \mathbf{X}_2) T_\tau(\mathbf{X}_2, \mathbf{X}_1) d\mathbf{X}_2 \\ &+ \mathcal{O}((\tau')^2). \end{aligned}$$

Dividing through by τ' and taking the limit $\tau' \rightarrow 0$, this becomes the differential form of the CKE

$$\frac{\partial}{\partial \tau} T_\tau(\mathbf{X}_3, \mathbf{X}_1) = \int [W(\mathbf{X}_3, \mathbf{X}_2) T_\tau(\mathbf{X}_2, \mathbf{X}_1) - W(\mathbf{X}_2, \mathbf{X}_3) T_\tau(\mathbf{X}_3, \mathbf{X}_1)] d\mathbf{X}_2$$

which is often written as the *master equation* (ME)

$$\frac{\partial}{\partial t} P(\mathbf{X}, t | \mathbf{X}_0, t_0) = \int [W(\mathbf{X}', \mathbf{X}) P(\mathbf{X}', t | \mathbf{X}_0, t_0) - W(\mathbf{X}, \mathbf{X}') P(\mathbf{X}, t | \mathbf{X}_0, t_0)] d\mathbf{X}' \quad (6.3)$$

where we have changed notation $\mathbf{X}_3 \rightarrow \mathbf{X}$, and $\mathbf{X}_1 \rightarrow \mathbf{X}_0$. This highlights that the ME is defined for a transition probability $P(\mathbf{X}, t | \mathbf{X}_0, t_0)$ from an initial condition and not for $P(\mathbf{X}, t)$. However, in most cases reference to the initial condition of the system is omitted and the later notation is used (which we will use for the rest of the chapter).

If the state space is discrete and given by $\mathbf{n} = (n_1, \dots, n_N)$ then the ME can be written as a sum,

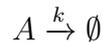
$$\frac{\partial}{\partial t} P(\mathbf{n}, t) = \sum_{\mathbf{n}'} [W(\mathbf{n}', \mathbf{n}) P(\mathbf{n}', t) - W(\mathbf{n}, \mathbf{n}') P(\mathbf{n}, t)]. \quad (6.4)$$

Here the meaning of the ME is very clear; the change in probability of being in any given state is the sum of all probability flows into and out of that state. This can be seen by the gain term (first term) and the loss term (second term) in Equation 6.4. In Chapter 7 we describe the evolution of two populations of ‘voters’ using a ME where the rates are dependent on the proportion of voters of a certain opinion.

² Here times are ordered by index in ascending order such that $t_1 < t_2 < t_3$.

Example - A Decay Process

Perhaps the most simple example of a master equation comes from chemistry. Consider the decay of a single chemical species A which occurs at a constant rate k , given by



where \emptyset are chemical species which we are not tracking and do not participate in the reaction. The rate constant is defined such that $k dt$ gives the probability that a randomly chosen molecule of A reacts in the time interval $[t, t + dt)$ where dt is an infinitesimally small time step. In any time interval, there are multiple things that may happen:

no reactions occur	with probability	$1 - A(t)k dt + \mathcal{O}(dt^2)$
exactly one reaction occurs	with probability	$A(t)k dt + \mathcal{O}(dt^2)$
two or more reactions occur	with probability	$\mathcal{O}(dt^2)$

Now, letting $P(n, t)$ be the probability that $A(t) = n$ and using the possibilities above (omitting probabilities of $\mathcal{O}(dt^2)$) we can write the master equation as

$$\frac{d}{dt}P(n, t) = k(n + 1)P(n + 1, t) - knP(n, t).$$

Here the right hand side consists of a gain and loss term. We can transition to state n if we were previously in state $n + 1$ and a molecule reacts. Similarly, we can transition out of state n if we were previously in that state and a molecule reacts. We cannot however transition from state $n - 1$ to state n .

This appears to be an infinite system of ordinary differential equations (ODEs), however with an initial condition $A(0) = n_0$, and due to molecules being removed only, there are only $n_0 + 1$ ODEs to solve. To incorporate the initial condition we set $P(n, t) = 0$ for all $n > n_0$.

For systems with a large state space or non-trivial transition probabilities the ME is generally too difficult to work with analytically³, however this case it is sufficiently simple that we can solve it with the initial condition $P(n_0, 0) = 1$, and $P(n, 0) = 0$ for

³ The ME is linear and the solution can easily be expressed in terms of a matrix exponential, however calculating moments or other macroscopic quantities requires a great deal of effort.

$n < n_0$ to give

$$P(n, t) = \exp(-knt) \binom{n_0}{n} [1 - \exp(-kt)]^{n_0-n}.$$

6.1.2 Fokker-Planck Equation

The ME is an integro-differential equation and is therefore difficult to work with. Fortunately we can simplify the ME to a second order differential equation, called the *Fokker-Planck equation* (FPE), with relative ease. To understand the derivation the Fokker-Planck equation we first consider a one-dimensional system and show that it can be generalised easily (in understanding, not notationally) to multiple dimensions. The one-dimensional ME is given by

$$\frac{\partial}{\partial t} P(x, t) = \int [W(x, x')P(x', t) - W(x', x)P(x, t)] dx'. \quad (6.5)$$

To simplify the ME we consider small changes in state and so it is useful to express the transition probability as a function of the jump size r from one configuration x' to another one x giving

$$W(x, x') = W(x'; r), \quad r = x - x'.$$

Rewriting the ME (6.5) in this notation leads to

$$\frac{\partial}{\partial t} P(x, t) = \int [W(x - r; r)P(x - r, t) - W(x; -r)P(x, t)] dr. \quad (6.6)$$

Naturally the change of variables introduces a change of sign, although this is absorbed into the boundary conditions if we assume a symmetric domain from $-\infty$ to ∞ .

We suppose that on the timescales considered the jumps in x are small, that is, $W(x', r)$ is a sharply peaked function of r but also varies slowly with respect to x' . Secondly we assume that $P(x, t)$ is continuous and a slowly varying function of x . Under these assumptions, we can Taylor expand the left term in (6.6) about x , giving

$$\begin{aligned} \frac{\partial}{\partial t} P(x, t) &= P(x, t) \int W(x; r) dr + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \int r^m \frac{\partial^m}{\partial x^m} [W(x; r)P(x, t)] dr \\ &\quad - P(x, t) \int W(x; -r) dr \\ &= \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \frac{\partial^m}{\partial x^m} \left\{ \left[\int r^m W(x; r) dr \right] P(x, t) \right\} \end{aligned}$$

where the first and last term cancel as the the sign of r is insignificant when integrating over the entire domain. It is useful to to define the *jump moments*

$$a^{(m)}(x, t) = \int r^m W(x; r) dr, \quad (6.7)$$

which lead to the *Kramers-Moyal* expansion (KME) of the ME

$$\frac{\partial}{\partial t} P(x, t) = \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \frac{\partial^m}{\partial x^m} [a^{(m)}(x, t) P(x, t)]$$

which, as a differential equation of infinite order, is identical to the ME. This means in principle that despite converting an integro-differential equation to a purely differential equation, the KME is equally as intractable as the ME. Typically, the KME is truncated, assuming that the terms $a^{(m)}(x, t)$ are zero or are negligible for $m > 2$. This leads to the Fokker-Planck equation

$$\frac{\partial}{\partial t} P(x, t) = - \frac{\partial}{\partial x} [a^{(1)}(x, t) P(x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [a^{(2)}(x, t) P(x, t)]. \quad (6.8)$$

The first and second jump moments are often referred to as the *drift* and *diffusion* terms respectively, more commonly denoted by,

$$f(x, t) := a^{(1)}(x, t) \quad \text{and} \quad d(x, t) := a^{(2)}(x, t). \quad (6.9)$$

Following the logic above, the derivation of the FPE can be extended to a multidimensional Markov process. The difficulty lies in utilising the multivariate Taylor expansion of the ME, which gives

$$\frac{\partial}{\partial t} P(\mathbf{X}, t) = \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \sum_{j_1 \dots j_m} \frac{\partial^m}{\partial x_{j_1} \dots \partial x_{j_m}} [a_{j_1 \dots j_m}^{(m)}(\mathbf{X}, t) P(\mathbf{X}, t)].$$

where the jump moments are now given by

$$a_{j_1 \dots j_m}^{(m)}(\mathbf{X}, t) = \int (x'_{j_1} - x_{j_1}) \dots (x'_{j_m} - x_{j_m}) W(\mathbf{X}, \mathbf{X}') d\mathbf{X}'.$$

Truncating at second order we arrive at the multidimensional FPE,

$$\frac{\partial}{\partial t} P(\mathbf{X}, t) = - \sum_i \frac{\partial}{\partial x_i} [a_i^{(1)}(\mathbf{X}, t) P(\mathbf{X}, t)] + \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} [a_{ij}^{(2)}(\mathbf{X}, t) P(\mathbf{X}, t)].$$

Provided the moments $a_i^{(1)}$ and $a_{ij}^{(2)}$ are linear, the FPE admits a Gaussian solution at stationarity (i.e., $\frac{\partial}{\partial t} P(\mathbf{X}, t) = 0$) [158, 160]. If the moments are non-linear (as we see in

Chapter 7) then we can linearise about a fixed point of the system. This gives a linear Gaussian approximation (LGA) of the stationary state probability density about the fixed point.

Chemical Fokker-Planck Equation

When working with a discrete state space further work is required to arrive at a Fokker-Planck approximation. We restrict ourselves to cases where the system can only transition between adjacent states, and again consider a one-dimensional system for clarity. This derivation was first considered for systems of chemicals where the number of molecules of a chemical species could either increase or decrease incrementally. In these systems, the number of molecules were typically large and so a continuum approximation is suitable.

The discrete ME for this system is given by

$$\frac{\partial}{\partial t}P(n, t) = H^-(n+1, t) - H^-(n, t) + H^+(n-1, t) - H^+(n, t) \quad (6.10)$$

where $H^\pm(n, t) = W(n \pm 1, n)P(n, t)$ give the probability flow from the above and below state respectively. To move to a continuum perspective by introducing a large number N and measure the state of the system in terms of a new variable $x = n/N$, which we treat as a continuous variable. Rewriting (6.10) in terms of the continuous variable x we get

$$\frac{\partial}{\partial t}P(n, t) = H^-(x + 1/N, t) - H^-(x, t) + H^+(x - 1/N, t) - H^+(x, t).$$

As we assume that N is large, we can approximate the first and third terms of the right hand side by a Taylor series about the point (x, t) ,

$$H^i\left(x \pm \frac{1}{N}\right) = H^i(x, t) \pm \frac{1}{N} \frac{\partial H^i}{\partial x}(x, t) + \frac{1}{2N^2} \frac{\partial^2 H^i}{\partial x^2}(x, t) + \mathcal{O}\left(\frac{1}{N^3}\right).$$

Substituting this into the ME, and truncating terms of $\mathcal{O}(N^{-3})$ gives

$$\frac{\partial}{\partial t}P(x, t) = -\frac{1}{N} \frac{\partial}{\partial x} [H^- - H^+] + \frac{1}{2N^2} \frac{\partial^2}{\partial x^2} [H^+ + H^-].$$

This is instantly recognisable as the FPE (6.8) where the drift and diffusion terms are given by the sum and difference of the transition rates respectively, i.e.,

$$\begin{aligned} f(x, t) &= (H^- - H^+)/[NP(x, t)] \\ d(x, t) &= (H^+ + H^-)/[N^2P(x, t)]. \end{aligned}$$

Quite often, when N is the size of the system being considered, a natural rescaling of time $t = \tau N$ leads to the FPE

$$\frac{\partial}{\partial \tau} P(x, \tau) = -\frac{\partial}{\partial x} [H^- - H^+] + \frac{1}{2N} \frac{\partial^2}{\partial x^2} [H^+ + H^-],$$

which explicitly highlights the scaling of the diffusive term with N .

The chemical Fokker-Planck equation can similarly be generalised to higher dimensional systems with relative ease (an example of which is used in Chapter 7).

6.1.3 First Passage Processes

In some instances we want to know the likelihood of ending up in a given state depending on the initial state. In this case the end position is known but the initial position is undetermined.

Backwards Fokker-Planck Equation

We can write the CKE (6.2) in one dimension as

$$P(x, t|z, \tau - d\tau) = \int P(x, t|y, \tau) P(y, \tau|z, \tau - d\tau) dy.$$

This equation is valid for any $d\tau$, however as before we will consider the small time limit $d\tau \rightarrow 0$. Assuming that the jumps are small we can Taylor expand $P(x, t|y, \tau)$ about the point $y = z$ to give

$$\begin{aligned} P(x, t|y, \tau) &= P(x, t|z, \tau) + (y - z) \frac{\partial}{\partial z} P(x, t|z, \tau) \\ &\quad + \frac{(y - z)^2}{2} \frac{\partial^2}{\partial z^2} P(x, t|z, \tau) + \mathcal{O}((y - z)^3). \end{aligned}$$

Substituting this into the right hand side of the CKE yields

$$\begin{aligned} P(x, t|z, \tau - d\tau) &= P(x, t|z, \tau) \times \int P(y, \tau|z, \tau - d\tau) dy \\ &+ \frac{\partial}{\partial z} P(x, t|z, \tau) \times \int (y - z) P(y, \tau|z, \tau - d\tau) dy \\ &+ \frac{\partial^2}{\partial z^2} P(x, t|z, \tau) \times \int \frac{(y - z)^2}{2} P(y, \tau|z, \tau - d\tau) dy + \mathcal{O}(d\tau^2). \end{aligned}$$

Using the fact that $P(z, \tau|y, \tau - d\tau) = d\tau W(z, y)$, we recognise the integrals as the jump moments (6.7) and (6.9). Dividing by $d\tau$ and rearranging the terms gives

$$\begin{aligned} \frac{P(x, t|z, \tau - d\tau) - P(x, t|z, \tau)}{d\tau} &= f(z, \tau) \frac{\partial}{\partial z} P(x, t|z, \tau) \\ &+ d(z, \tau) \frac{\partial^2}{\partial z^2} P(x, t|z, \tau) + \mathcal{O}(d\tau). \end{aligned}$$

In the limit $d\tau \rightarrow 0$, we obtain the *backward Kolmogorov equation*,

$$-\frac{\partial}{\partial \tau} P(x, t|z, \tau) = f(z, \tau) \frac{\partial}{\partial z} P(x, t|z, \tau) + d(z, \tau) \frac{\partial^2}{\partial z^2} P(x, t|z, \tau) \quad (6.11)$$

often referred to as the *backwards Fokker-Planck equation* (bFPE). The corresponding *forwards FPE* reads as

$$\frac{\partial}{\partial t} P(x, t) = -\frac{\partial}{\partial x} [f(x, t)P(x, t)] + \frac{\partial^2}{\partial x^2} [d(x, t)P(x, t)]$$

which for time independent coefficients, $f(x, t) \equiv f(x)$ and $d(x, t) \equiv d(x)$, has a stationary solution

$$P^*(x) \propto \frac{1}{d(x)} \exp \left[\int_0^x \frac{f(s)}{d(s)} ds \right] \quad (6.12)$$

which we use next to describe the first passage times.

First Passage Times

For a number of systems we are not concerned with whether it is in a particular state at a given time, but instead when the system first reaches that state. Examples span from reaching critical concentrations in chemistry, to activation of stock options or countermeasures when a stock price reaches a given target. In this section we introduce the tools to study the first passage problem in one dimension. This allows us to ask

questions such as “when will the system reach this state, given this initial condition?” and “how long does the system stay in this state on average?” We will see such a question in Chapter 7 where we calculate the time for the system to switch between two fixed points.

Let $h(y, t)$ be the probability that $x \in (-\infty, x_u)$ for all $0 < t' < t$, given that it started at y , i.e.,

$$h(y, t) = \int_{-\infty}^{x_u} P(x, t|y, 0) dx. \quad (6.13)$$

Assuming drift and diffusion are time-independent, i.e. $f(x, t) \equiv f(x)$, $d(x, t) \equiv d(x)$, we can shift time backwards giving

$$h(x, t) = \int_{-\infty}^{x_u} P(y, 0|x, -t) dy.$$

Using the transformation s to $-t$ in the backwards Kolmogorov equation (6.11) we obtain

$$\frac{\partial}{\partial t} P(x, 0|y, -t) = f(y) \frac{\partial}{\partial y} P(x, 0|y, -t) + d(y) \frac{\partial^2}{\partial y^2} P(x, 0|y, -t).$$

By integrating over x , and using (6.13) the equation becomes

$$\frac{\partial}{\partial t} h(y, t) = f(y) \frac{\partial}{\partial y} h(y, t) + d(y) \frac{\partial^2}{\partial y^2} h(y, t). \quad (6.14)$$

Let $\tau(y)$ be the average time to first leave the interval given that $P(y, 0) = 1$. The probability that x first leaves the interval in during $[t, t + dt)$ is given by

$$h(y, t) - h(y, t + dt) \approx -\frac{\partial}{\partial t} h(y, t) dt,$$

using the Taylor series expansion for $h(y, t + dt)$. Therefore, to compute the *average* time we integrate over all possible escape times giving

$$\tau(y) = -\int_0^{\infty} t \frac{\partial}{\partial t} h(y, t) dt = \int_0^{\infty} h(y, t) dt,$$

using integration by parts for the later equality⁴.

Integrating (6.14) with respect to t yields

$$-1 = f(y) \frac{d}{dy} \tau(y) + d(y) \frac{d^2}{dy^2} \tau(y) = \mathcal{G}_b(y) \tau(y)$$

⁴ $[th(y, t)]_0^{\infty} = 0$ as the process necessarily escapes $(-\infty, x_u)$ eventually, i.e., $h(y, \infty) = 0$.

where \mathcal{G}_b is the infinitesimal generator of the bFPE.

We now have an ODE for the average escape time which we can readily solve with suitable boundary conditions. The boundary conditions are domain specific, but usually incorporate a reflective boundary condition on the domain walls (e.g. particles contained in a box) with a second condition that the first passage time to an interval is exactly zero on the boundary (the process has already escaped the interval). These are prescribed by

$$\frac{d}{dy}\tau(-\infty) = 0 \quad (6.15)$$

for the reflective boundary condition, and

$$\tau(x_u) = 0 \quad (6.16)$$

on the interval boundary.

Using an integrating factor to integrate we get

$$\begin{aligned} \frac{d}{dy}\tau(y) &= -\exp\left[-\int_0^y \frac{f(z)}{d(z)} dz\right] \int_{-\infty}^y \frac{1}{d(x)} \exp\left[\int_0^x \frac{f(z)}{d(z)} dz\right] dx \\ &= -\frac{1}{d(y)P^*(y)} \int_{-\infty}^y P^*(x) dx, \end{aligned}$$

where we have used the stationary distribution (6.12). Finally, integrating over y and using (6.16) we arrive at the integral equation for the average first passage time,

$$\tau(y) = \int_y^{x_u} \frac{1}{d(z)P^*(z)} \int_{-\infty}^z P^*(x) dx dz.$$

Unfortunately this method cannot be easily extended to multiple dimensions, especially in non-equilibrium systems, however methods do exist for particular systems and domains [159].

6.1.4 Mean-field Equations

The ME describes the evolution of the state probability densities. With suitable manipulation of the ME one can derive equations for the moments of the distribution. In particular, the first moment describes the mean and from the second moment we can calculate the variance of the probability densities. For large discrete systems we have

seen that the second order terms scale as $1/N$. This means that for sufficiently large N we can provide a quantitatively accurate description of the system using the first moment alone⁵.

The average over a particular quantity $\phi(\mathbf{X})$ of the system at a time t is given by

$$\langle \phi(\mathbf{X}) \rangle(t) = \int \phi(\mathbf{X}) P(\mathbf{X}, t) d\mathbf{X}.$$

The standard way to find $\langle \phi(\mathbf{X}) \rangle$ is to multiply the ME (6.3) by $\phi(\mathbf{X})$ and integrate over all possible states \mathbf{X} . This results in an ODE for the evolution of $\phi(\mathbf{X})$ which allows us to easily identify any fixed points of the dynamics and their stability. However, in ignoring the fluctuations some dynamics of the model may be missed, such as fluctuation-driven switching between quasi-stable fixed points (as in the case in Chapter 7).

Example - A Decay Process

Revisiting the example of chemical decay in the previous section, we can derive an equation for the mean number of a chemical species we would expect to see over infinitely many realisations. The master equation was given by

$$\frac{d}{dt} P(n, t) = k(n+1)P(n+1, t) - knP(n, t). \quad (6.17)$$

In the discrete case, the mean number of the chemical species is given by

$$\langle n \rangle(t) = \sum_{n=0}^{\infty} nP(n, t)$$

which is the expectation of n . By multiplying (6.17) by n and summing over all n we arrive at

$$\begin{aligned} \frac{d\langle n \rangle}{dt}(t) &= \sum_{n=0}^{\infty} kn(n+1)P(n+1, t) - \sum_{n=0}^{\infty} kn^2P(n, t) \\ &= \sum_{n=0}^{\infty} k(n-1)nP(n, t) - kn^2P(n, t) \end{aligned}$$

where we have shifted the index $n \rightarrow n-1$ in the first summation. We can do this as $P(n, t) > 0$ for $0 < n < n_0$, and is zero elsewhere. Upon cancelling terms we are left

⁵ This description becomes exact in the limit $N \rightarrow \infty$.

with a differential equation for the mean

$$\frac{d\langle n \rangle}{dt}(t) = -k\langle n \rangle$$

which, with initial condition $\langle n \rangle(0) = n_0$, has solution

$$\langle n \rangle(t) = n_0 \exp(-kt). \quad (6.18)$$

Similarly, we can generate ODEs for the a^{th} moment by multiplying by n^a and summing over n . In this case calculating the a^{th} moment will require the calculation of all moments less than a ⁶. From the above equation it is clear that the chemical decays exponentially (which was not necessarily clear from the ME), however it is not possible to discern any information on the fluctuations using the macroscopic description (6.18) alone.

6.2 Models on Static Networks

In this section we consider systems which can be represented as a network. We consider a static network consisting of a set of nodes $V \subset \mathbb{N}$ and set of edges $E \subset V^2$, which can be represented by an adjacency matrix \mathbf{A} . We typically consider systems of variables x_i, y_i, \dots on each node in the network which are coupled to the variables of adjacent nodes. We assume that any coupling is restricted to pairs of adjacent nodes only, meaning no three or more body interactions can occur.

Perhaps the most widely recognised of this class of model is the susceptible-infected (SI) model where each node can be in either the susceptible (prone to infection) or infected state. Infection can spread from an infected node to an adjacent susceptible node with a fixed probability (or at a fixed rate). For this model, each node has a single variable x_i which is zero for a susceptible node, and one if infected.

⁶ In many cases calculating the a^{th} moment requires the $(a + 1)^{\text{th}}$ moment. This means the system is not closed and the $(a + 1)^{\text{th}}$ moment needs to be approximated through *moment closure* [166].

6.2.1 Intractability of the Master Equation

When we consider systems that are homogeneous and well mixed (as is often assumed for chemical reactions and large populations) we can fully describe the system by a simple count of nodes of each type. This means that the state space of the system is of the order of the number of nodes, or *system size*. By contrast, for a system on a network with m possible node states the state space is of size m^N where N is the system size. Therefore the number of possible states grows exponentially with the system size. This makes it incredibly difficult to enumerate all possible states, write down the corresponding ME, and subsequently solve it. Realistically this can only be done exactly for small systems and so it is therefore simpler to consider the evolution of the macroscopic variables (averages) instead of the full probability distribution.

6.2.2 Standard Formulations

One approach to modelling dynamics on networks is to start with the ME and make approximations and aggregations to reduce the problem. If such an approach is ‘bottom-up’, then the contrasting ‘top-down’ approach is to describe the macroscopic quantities of the system using an ODE and introduce network effects incrementally. As the network analysis of Chapter 8 uses only a naïve network approximation, we take this latter approach in this section.

A simple example of a dynamical system described by a single variable $x(t)$ evolves according to the first order ODE

$$\frac{dx}{dt} = f(x).$$

Here there are no network effects and the right hand side is independent of time (such a system is called *autonomous*).

Returning to the SI model example and letting x be the probability a node is infected, a well mixed system (i.e, complete graph) evolves according to

$$\frac{dx}{dt} = \beta x(1 - x),$$

where β is the rate of infection.

We now look to extend this description to incorporate the network topology. There are many ways to do this, each involving assumptions on the network topology and the correlations between node states.

Individual-based

The most general description is where each node is modelled individually. The state of each node is governed by a differential equation of the form

$$\frac{dx_i}{dt} = f_i(x_i) + \sum_j \mathbf{A}_{ij} g_{ij}(x_i, x_j). \quad (6.19)$$

Here f_i specifies the intrinsic dynamics of the node, i.e. how the node state would evolve independently of the network. The network interaction is governed through the adjacency matrix \mathbf{A} , and the coupling relation g_{ij} which gives the contribution from adjacent nodes.

As there are N coupled differential equations to solve which can be problematic, the system is often simplified by considering universal function of f and g , and if appropriate, reducing g to a function of the adjacent node only.

One particular example of this description is where x_i is the probability of node i being in a particular state [167]. In the SI model, the system evolves according to

$$\frac{dx_i}{dt} = \beta \sum_j \mathbf{A}_{ij} (1 - x_i) x_j.$$

This is an example where $g(x_i, x_j)$ is not a symmetric function as infection can only pass in one direction.

Degree Grouping

As with the ME formulation, the individual-based approach is often intractable for all but the simplest of models. One approximation that maintains the degree heterogeneity of the network is to aggregate node states by degree. Letting x_k be the average state of

a degree k node (or density of degree k nodes in a particular state) the system can be described by the coupled equations

$$\frac{dx_k}{dt} = f(x_k) + k \sum_{k'} P(k'|k)g(x_{k'}, x_k)$$

where $P(k'|k)$ is the probability that an edge originating at a degree- k node has a degree k' node at its other end. This approach has the benefit that degree-degree correlations are captured (also known as network assortativity). If no such correlations exist, or they are ignored, then $P(k'|k) = k'P(k')/\langle k \rangle$ where $\langle k \rangle$ is the average degree of the network.

Mean-field Approximation

Finally we introduce the most assumptive description of a process on a network which will be used in Chapter 8. The mean-field approximation considers that each node is equal (a homogeneous system) both in respect to state and degree in the network. For a network with average degree $\langle k \rangle$ we can approximate the adjacency matrix by $\mathbf{A}_{ij} \approx k_i k_j / N \langle k \rangle \approx \langle k \rangle / N$. This states that each node is connected to every other node in the network but with a weighting $\langle k \rangle / N < 1$. Substituting this approximation into (6.19) we get

$$\frac{dx}{dt} = f(x) + kg(x).$$

The mean-field approximation provides a reasonable first approximation when k/N and N are sufficiently large, and for networks with degree distribution well represented by the mean. Erdős-Rényi networks [168] have binomial distribution which in the large N limit is approximated by a normal distribution which is unimodal. In contrast, Barabási-Albert networks [169] have power-law (or *scale free*) distribution where the mean is much greater than the median and therefore dynamics on these networks are not well captured by a mean-field approach.

6.3 Discussion

In this chapter we introduced ways to describe stochastic processes through the master equation, and how to make approximations to arrive at analytically tractable

equations. These approximations cumulate in either a partial differential equation (for the Fokker-Planck equation) or an ordinary differential equation (for mean-field approaches) which can be solved or analysed using standard procedures. These procedures are not given here but can be found in any textbook on the subject.

When dealing with stochastic processes, the master equation provides an exact description of the system, but it is difficult to work with analytically. We can however make successively better approximations to the exact dynamics by using more terms of the Kramers-Moyal expansion, at the cost of complexity. Usually only two terms are sufficient to capture the system dynamics including fluctuation driven phenomena. This truncated expansion is the well known Fokker-Planck equation.

We also saw how the exponential scaling of the state space with system size rendered the master equation approach infeasible when describing dynamical processes on networks. To make progress with these systems we model the evolution of macroscopic quantities of instead state space probabilities. There has however been progress in more advanced methods such as pair approximation [170] and moment closure [171]. The macroscopic descriptions provide reasonable results when assumptions on the network hold true as we see in Chapter 8.

7

The 2q-Voter Model with Zealotry

Individual-based (or more commonly used agent-based) models have been used to describe collective behaviour in the past half century [172, 173]. These models have been particularly useful in explaining many social phenomena without the need for complex, high-dimensional models and modern computational simulation methods. One of the most famous of these models is the voter model (VM) [174], popular due to it being one of the rare exactly solvable models in statistical physics. Although originally described in continuous time, for clarity we will describe the VM which evolves in discrete time on a network structure, and take the appropriate limits in order to recover the equations most commonly used to describe the dynamics.

The classical voter model considers a network of nodes holding opinion (or spins) ± 1 . At each discrete time step a node chosen randomly adopts the opinion of one of its neighbouring nodes, also chosen at random¹. Finite systems evolve until the

¹ Note that on heterogeneous networks the ‘direction’ of opinion transfer has a major effect on the

nodes necessarily reach a consensus (all holding either the +1 or -1 opinion), and the dynamics cease.

The VM is in fact a particular case of the classical Ising model [175, 176], used as a model of ferromagnetism in physics. The widely studied Ising model describes a collection of atomic spins and their associated magnetic moments under the effect of temperature and an external magnetic field². Naturally, an important macroscopic quantity to measure in the Ising model is the magnetisation, the average spin direction, which is conserved on regular structures [177–179]. From a social perspective the magnetisation can be thought of as an ‘average opinion’.

There are two basic properties of the VM that have been widely studied. The exit probability $E_+(\rho_0)$ describes the probability of the system reaching a consensus of +1 given that the initial density of +1 voters is ρ_0 . For regular lattices this is given simply by $E_+(\rho_0) = \rho_0$ [180], however it is non-trivial on more complex structures. Another property of interest is the mean time to reach consensus, T_N . On lattice structures T_N scales as N^2 in one dimension, $N \ln N$ in two dimensions, and as N for all dimensions greater than two. For heterogeneous networks $T_N \sim N\mu_1^2/\mu_2$ where μ_k is the k th moment of the degree distribution [181, 182]. This scaling includes the assumption that the network is free from degree-degree correlations, and while this is often far from realistic, however it still provides a reasonable estimate when such correlations are present.

There are many variants of the voter model, motivated from both physical and social problems. A natural extension to the VM is to consider an opinion switching rate dependent on two or more neighbours [183, 184]. This extension brings the VM closer to reality as social scientists have established that conformity by imitation, an important mechanism for collective actions, is observed only when the group (neighbourhood) size is sufficiently large [185, 186]. For one such model, the non-linear q -voter model (q VM) [184], the dynamical update is given by:

1. Choose a (focus) node at random.

model dynamics due to the presence of high degree (hub) nodes.

² The VM is exactly an Ising model at zero temperature and in the absence of an external field.

2. Choose q neighbours of the focus node at random (here repetition of neighbours is allowed).
3. If all q share the same opinion (are in consensus) then the focus node adopts the opinion of the q neighbours. Otherwise, the focus node changes opinion with probability ϵ .

The q VM exhibits much of the same behaviour as the original VM, and eventual consensus of opinion is guaranteed. The difference however is in the time to consensus which varies with the value of q .

Another popular extension to the VM is the inclusion of voters who do not change their opinion, coined inflexible *zealots*³. The inclusion of such voters has a number of effects on the dynamics of the VM. The most prominent effect arises when there exist zealots of both opinion. In this regime consensus of opinion is never achieved in the whole population as by definition there will always be voters of each type. It is therefore no longer suitable to calculate quantities such as the the consensus time. Instead one can consider the evolution and long time distribution of the magnetisation [187, 188].

There are of course many other variations of voter-like dynamics which aim to replicate more realistic assumptions into the model as well as exploring how small changes in the microscopic dynamics of the model can lead to different types of collective phenomena. These include, but are not limited to employing the model on an adaptive network [189], and the inclusion multiple opinions for each node [190]. For the purposes of this chapter, we will require only the notion of zealots and the adoption mechanism of the q VM. These two mechanisms have been recently combined in the q -voter model with zealotry (q VMZ) [191]. This model offers the first major behavioural change from previous voter models. For later comparison the main features of the q VMZ are:

- The model satisfies detailed balance and is therefore time reversible. This also means that a stationary distribution is readily attainable.
- The model has a critical value of zealotry density, z_c , above which the model behaves like the q VM with fluctuations about a central fixed point, and above

³ Earlier work terms zealots as voters who only favoured one opinion over the other [187, 188].

which the long time distribution of opinion is bimodal. In this case the average opinion switches between two fixed points on a characteristic timescale.

- The model is extremely sensitive to small biases in zealotry densities.

Opinion Dynamics In Social Media

Social media is not purely used for developing friendships but has seen many opinions voiced, particularly political [192–194]. Opinion can represent a number of things: it can be the positive or negative opinion of a product, brand, or person; the preference of one thing over another such as Pepsi or Coca Cola, Manchester United or Liverpool; or the political party an individual chooses to vote for.

There have been a number of attempts to measure opinion on social media, predominantly through word usage [195] or sentiment analysis [196]. These studies however have many caveats. Data collected from social media are not necessarily representative of a general population it can be particularly noisy, and the quality of tools such of sentiment analysis is questionable on typically short social media messages. We therefore look to suitable mathematical models to understand the possible drivers of opinion formation in social networks.

One aspect of opinion spreading we want to capture is that the population is heterogeneous in the resolution of their opinion. Earlier studies [185, 186] have shown that individuals respond differently to social pressure from groups of varying size. In the context of social media, this can take the form of how many friends an individual may have (and the distribution of their opinions), or the number of opinionated content pieces shared on the network that an individual may view before changing their own opinion.

Chapter Outline

In Section 7.1 we introduce the $2q$ -voter model of opinion dynamics. We give a full description of the model through the master equation and show that detailed balance is not satisfied for all but the most trivial parameters. Following this, in

Section 7.2 we consider continuum descriptions of the model and compare analytical results with simulations. Furthermore in Section 7.3 we show that the linear Gaussian approximation is well suited to the analysis of the non-equilibrium steady state of the model for moderate system sizes. In Section 7.4 we attempt to understand the switching behaviour exhibited by the model by considering the mean switching time. In Section 7.5 we consider the model under asymmetric parameter schemes before concluding in Section 7.6.

7.1 The 2q-Voter Model with Zealotry

The 2q-voter model with zealotry (2qVZ) [2, 1] provides the simplest generalisation to the qVMZ [191] which considers a heterogeneous population of voters. By making this extension we can capture the differing resolves of voters in the population, with some voters requiring more convincing than others to change their opinion than others.

7.1.1 Model Outline

The 2qVZ consists of a population of N voters who hold one of two opinions, denoted ± 1 . A fraction of the population are inflexible in their opinion and, once initialised, never changes. These are labelled zealots [187, 188], and the corresponding size of their populations are denoted by Z_{\pm} . The remaining population consists of S swing voters, split further into two types: q_1 - and q_2 -susceptibles, who have population sizes S_1 and S_2 respectively (with $S_1 + S_2 = S$). In this model the behaviour of each voter is fixed, so that Z_{\pm} and $S_{1,2}$ are conserved. Also, no voters leave or enter the system, conserving the total population size $S_1 + S_2 + Z_+ + Z_- = N$.

Illustrated in Figure 7.1, the model updates discretely in time according to the following mechanism:

- (a) A (focal) voter is chosen at random from the population.

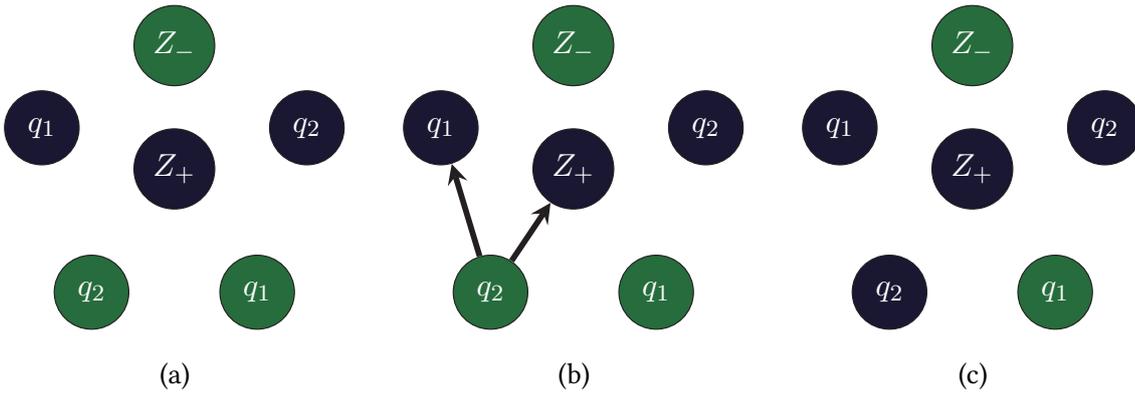


Figure 7.1: An illustration of the $2qVZ$ dynamics. (a) The system consists of q_1 - and q_2 -susceptibles as well as zealots, each holding one of two opinions (blue/green). (b) At each time step a node is randomly selected. If a zealot is selected, then nothing happens. If the node is a q_i -susceptible it chooses q_i neighbours at random and records their opinions. Here, $q_2 = 2$. (c) If those neighbours are in consensus then the original node adopts that state, otherwise no change occurs.

- (b) If the chosen voter is a zealot then no further action is taken and step (a) is repeated. If the voter is a q_i -susceptible then the opinions of a random sample of q_i neighbours are collected (note that repetition is allowed).
- (c) If the q_i neighbours all have the same opinion but opposite to that of the focal voter then the latter adopts the opinion of the neighbours. If there is no consensus between the neighbours then no opinion change occurs.

For the sake of simplicity and tractability we consider a well-mixed population where each voter can be considered as a node of a complete graph of size N .

7.1.2 Master Equation Formulation

Without explicit spatial structure the configuration space is reduced to a discrete set of $S_1 \times S_2$ points with the state of the system completely specified by the pair $\mathbf{n} = (n_1, n_2)$ where n_i is the number of susceptible voters of type i who hold the $+1$ opinion. At each update attempt there are four possible state changes as only one susceptible can change their opinion. The allowed changes at each update are therefore in the set $\{\pm e_1, \pm e_2\}$

with $\mathbf{e}_1 := (1, 0)$ and $\mathbf{e}_2 := (0, 1)$. The system may also stay in the same state if a zealot is chosen or there is not a consensus within the sampled neighbours. In this sense we can consider the evolution of the 2qVZ to be a two-dimensional random walk with inhomogeneous and biased transition rates. The full description of such a stochastic process is best described by a master equation (ME) for the evolution of $P(\mathbf{n}, T)$, the probability of finding the system in state \mathbf{n} after T discrete updates from an initial condition \mathbf{n}_0 . The main interest of the model is in the stationary distribution which is independent of the initial configuration, hence we suppress and reference to \mathbf{n}_0 in further analysis.

For any discrete Markov process such as this, the ME can be written as

$$P(\mathbf{n}', T + 1) = \sum_{\mathbf{n}} \mathcal{G}(\mathbf{n}', \mathbf{n}) P(\mathbf{n}, T) \quad (7.1)$$

where \mathcal{G} is the evolution operator, or transition matrix, describing the transition probabilities between states. For a system with $S_1 \times S_2$ possible states, \mathcal{G} can take up to $[(S_1 + 1) \times (S_2 + 1)]^2$ different values making it difficult to prescribe. However in this case, as \mathbf{n}' is restricted to the set $\{\mathbf{n}, \mathbf{n} \pm \mathbf{e}_1, \mathbf{n} \pm \mathbf{e}_2\}$, \mathcal{G} can be written explicitly as

$$\begin{aligned} \mathcal{G}(\mathbf{n}', \mathbf{n}) &= \delta(n'_1, n_1) \delta(n'_2, n_2) W^0(\mathbf{n}) \\ &+ \sum_{i=1,2, j \neq i} \delta(n'_i, n_i + 1) \delta(n'_j, n_j) W_i^+(\mathbf{n}) \\ &+ \sum_{i=1,2, j \neq i} \delta(n'_i, n_i - 1) \delta(n'_j, n_j) W_i^-(\mathbf{n}). \end{aligned} \quad (7.2)$$

Here $W^0(\mathbf{n})$ represents the probability for the system to remain unchanged while $W_1^\pm(\mathbf{n})$ and $W_2^\pm(\mathbf{n})$ are the probabilities associated with the transitions $\mathbf{n} \rightarrow \mathbf{n} \pm \mathbf{e}_1$ and $\mathbf{n} \rightarrow \mathbf{n} \pm \mathbf{e}_2$ respectively. These probabilities can be formulated by simply considering the probability of picking a voter of type q_i and subsequently picking q_i other voters from the population who hold the opposite opinion (allowing for replacement). Explicitly they are given by

$$\begin{aligned} W_i^+(\mathbf{n}) &= \frac{S_i - n_i}{N} \left(\frac{M}{N-1} \right)^{q_i}, \\ W_i^-(\mathbf{n}) &= \frac{n_i}{N} \left(\frac{N-M}{N-1} \right)^{q_i}, \\ W^0(\mathbf{n}) &= 1 - W_1^+(\mathbf{n}) - W_1^-(\mathbf{n}) - W_2^+(\mathbf{n}) - W_2^-(\mathbf{n}), \end{aligned} \quad (7.3)$$

where $M = Z_+ + n_1 + n_2$ is the total number of +1 voters.

From the ME we can derive other quantities of interest. The joint probability

$$\mathcal{P}(\mathbf{n}', T' | \mathbf{n}, T) = \mathcal{G}^{T'-T}(\mathbf{n}', \mathbf{n})P(\mathbf{n}, T), \quad (7.4)$$

describes the probability of observing the system in state \mathbf{n}' at time T' , given that it was in state \mathbf{n} at an earlier time T ($T' > T$). Using P and \mathcal{P} we can compute physical observables such as the average number of q_i voters holding the +1 opinion at time T and the two-point correlation functions at general times:

$$\langle n_i \rangle_T = \sum_{\mathbf{n}} n_i P(\mathbf{n}, T) \quad (7.5)$$

$$\langle n'_i n_j \rangle_{T', T} = \sum_{\mathbf{n}, \mathbf{n}'} n'_i n_j \mathcal{P}(\mathbf{n}', T' | \mathbf{n}, T). \quad (7.6)$$

As will become apparent in the next section, of particular interest in the 2qVZ is the *net probability current* $\mathbf{K}(\mathbf{n}, T) = (K_1, K_2)$ with

$$K_i(\mathbf{n}, T) = W_i^+(\mathbf{n})P(\mathbf{n}, T) - W_i^-(\mathbf{n})P(\mathbf{n}', T) \quad (7.7)$$

describing the net flow of probability from \mathbf{n} to $\mathbf{n}' = \mathbf{n} + \mathbf{e}_i$ for $i = 1, 2$.

Violation of Detailed Balance

An important principle of Markov processes (and kinetic systems) is that of *detailed balance*. Satisfaction of the detailed balance condition requires that there exists a (time-independent) stationary distribution $P^*(\mathbf{n})$ such that

$$\mathcal{G}(\mathbf{n}, \mathbf{n}')P^*(\mathbf{n}) = \mathcal{G}(\mathbf{n}', \mathbf{n})P^*(\mathbf{n}')$$

for all possible \mathbf{n}, \mathbf{n}' . Alternatively we can apply the Kolmogorov criterion [197] for detailed balance which states that for any closed loop of states, the product of transition rates in both directions must be equal. To establish whether detailed balance is satisfied in the 2qVZ we apply the Kolmogorov criterion to the closed loop consisting of the four states around a square: $\mathbf{n} \leftrightarrow \mathbf{n} + \mathbf{e}_1 \leftrightarrow \mathbf{n} + \mathbf{e}_1 + \mathbf{e}_2 \leftrightarrow \mathbf{n} + \mathbf{e}_2 \leftrightarrow \mathbf{n}$, illustrated in Figure 7.2. The product of transition probabilities around this loop is

$$W_1^+(\mathbf{n})W_2^+(\mathbf{n} + \mathbf{e}_1)W_1^-(\mathbf{n} + \mathbf{e}_1 + \mathbf{e}_2)W_2^-(\mathbf{n} + \mathbf{e}_2).$$

Meanwhile, the product for the *reverse* loop is

$$W_2^+ (\mathbf{n}) W_1^+ (\mathbf{n} + \mathbf{e}_2) W_2^- (\mathbf{n} + \mathbf{e}_1 + \mathbf{e}_2) W_1^- (\mathbf{n} + \mathbf{e}_1)$$

so using the explicit expressions (7.3) the ratio of the two probabilities is

$$\left(\frac{M+1}{M} \frac{N-M-1}{N-M-2} \right)^{q_1 - q_2} \geq 1.$$

The quantity in the bracket is greater than or equal to unity, achieving equality only when $q_1 \neq q_2$. Thus, the dynamics of our 2qVZ violates the detailed balance. When $q_1 = q_2$ the model is reduced to the single population qVMZ [191] which does satisfy detailed balance.

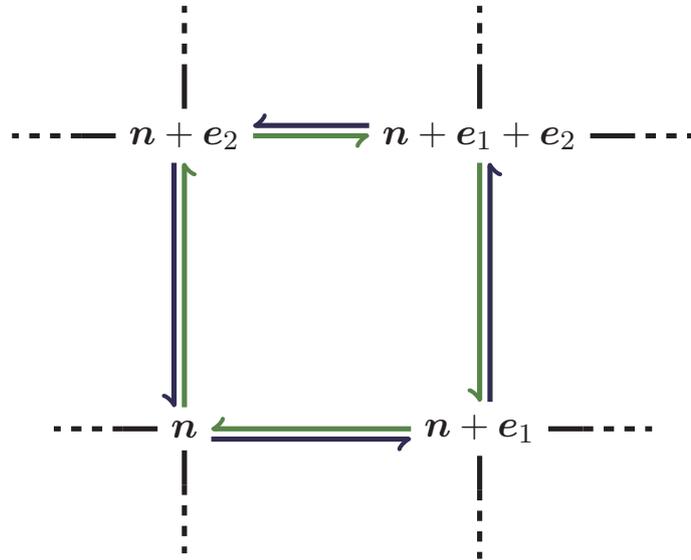


Figure 7.2: An illustration of the violation of detailed balance. The system can be represented as a two-dimensional lattice of size $(S_1 + 1) \times (S_2 + 1)$. The probability of traversing this loop clockwise (green) does not equal the probability of traversal in the opposite direction (blue). By Kolmogorov’s criterion, the detailed balance is violated.

The violation of detailed balance has a number of consequences on the behaviour of the system. Firstly this means that the Markov process is *irreversible*, that is, the system behaves differently going forwards in time to the system with time reversed. The second consequence is that the system settles into a *non-equilibrium steady state* (NESS), characterised by the presence of non-vanishing probability currents [198]. These consequences will be explored in more detail in later sections.

7.1.3 Characterisation of the NESS

As a consequence of detailed balance violation the system relaxes into a non-equilibrium steady state. Due to the system being out of equilibrium the associated stationary state probability distribution P^* is difficult to calculate [199]. However, given we have P^* , or a suitable approximation, there are a number of characteristics of the NESS that we are able to quantify.

Perhaps the most fundamental observables are the stationary state averages and two-point lagged correlations. Analogously to (7.6) we define them to be

$$\langle n_i \rangle^* = \sum_{\mathbf{n}} n_i P^*(\mathbf{n}) \quad (7.8)$$

$$\langle n'_i n_j \rangle_T^* = \sum_{\mathbf{n}, \mathbf{n}'} n'_i n_j \mathcal{P}^*(\mathbf{n}', T | \mathbf{n}, 0). \quad (7.9)$$

Furthermore we can define the stationary probability current $\mathbf{K}^*(\mathbf{n}) = (K_1^*, k_2^*)$ with

$$K_i^*(\mathbf{n}) = W_i^+(\mathbf{n})P^*(\mathbf{n}) - W_i^-(\mathbf{n} + \mathbf{e}_i)P^*(\mathbf{n} + \mathbf{e}_i) \quad (7.10)$$

which gives the net flow from \mathbf{n} to $\mathbf{n} + \mathbf{e}_i$. In the NESS, the current \mathbf{K}^* is divergence free, which on a lattice this condition reads

$$0 = K_1^*(\mathbf{n}) - K_1^*(\mathbf{n} - \mathbf{e}_1) + K_2^*(\mathbf{n}) - K_2^*(\mathbf{n} - \mathbf{e}_2). \quad (7.11)$$

What this means is that the currents must form closed loops around the system. In particular, analogously with fluid dynamics, the probability flow satisfies a continuity equation ($\nabla \cdot \mathbf{K} = 0$) which ensures that probability is not created or destroyed. This also means that \mathbf{K}^* can be represented as the curl of another quantity.

Following the literature for the incompressible fluids we can introduce the concepts of *vorticity* ($\omega = \nabla \times \mathbf{K}$), the *stream function* ($\mathbf{K} = \nabla \times \psi$), and *total angular momentum* ($\mathbf{L} = \int_{\mathbf{r}} \mathbf{r} \times \mathbf{K}(\mathbf{n})$). Explicitly for our discrete system we can define the vorticity as

$$\omega^*(\mathbf{n}) = K_1^*(\mathbf{n}) + K_2^*(\mathbf{n} + \mathbf{e}_1) - K_1^*(\mathbf{n} + \mathbf{e}_2) - K_2^*(\mathbf{n}). \quad (7.12)$$

Here the vorticity is associated with a plaquette (square) centred on the half integers $(n_1 + 1/2, n_2 + 1/2)$ [200] however we will continue to denote this as \mathbf{n} for convenience.

The stream function satisfies

$$K_i^*(\mathbf{n}) = \varepsilon_{ij} [\psi^*(\mathbf{n}) - \psi^*(\mathbf{n} - \mathbf{e}_j)] \quad (7.13)$$

where ε is the two-dimensional Levi-Civita symbol⁴ and again we associate the quantity with a plaquette. Setting the arbitrary constant in ψ^* to be zero outside the $S_1 \times S_2$ rectangle then we find

$$\psi^*(\mathbf{n}) = \sum_{\ell=0}^{n_2} K_1^*(n_1, \ell) = - \sum_{\ell=0}^{n_1} K_2^*(\ell, n_2). \quad (7.14)$$

Finally, the sum

$$\langle \mathcal{L} \rangle^* \equiv \sum_{\mathbf{n}} \varepsilon_{ij} n_i K_j^*(\mathbf{n}). \quad (7.15)$$

plays the role of the discrete probability angular momentum in the NESS. Since $\mathbf{K} \propto P^*$ the above sum can be thought of as a form of statistical average. In particular we label it as the average total probability angular momentum, written as

$$\langle \mathcal{L} \rangle^* = \langle \varepsilon_{ij} n_i V_j \rangle^*, \quad (7.16)$$

where $V_j = W_j^+ - W_j^-$.

In subsequent sections we will calculate these quantities numerically and find analytical approximations for them in the large system limit.

7.2 Continuum Descriptions and Exact Results

The master equation formulation of the $2qVZ$ provides an exact description. However, as is often the case, it is analytically intractable for all but the smallest of system sizes.

⁴ The Levi-Civita symbol is explicitly given by

$$\varepsilon_{ij} = \begin{cases} 1 & \text{if } (i, j) = (1, 2) \\ -1 & \text{if } (i, j) = (2, 1) \\ 0 & \text{otherwise.} \end{cases}$$

Typically, Monte Carlo simulations are required to explore the model behaviour and parameter dependence. Fortunately, insight can be gained by considering the model in the large system size continuum limit where demographic fluctuations are accountable and assumed small. In this limit (also known as the thermodynamic limit in statistical physics), $Z_{\pm}, S_i, N \rightarrow \infty$ while maintaining fixed densities $z_{\pm} = Z_{\pm}/N$ and $s_i = S_i/N$. For large but finite systems the correlations and fluctuations can be modelled using the Fokker-Planck equation (FPE) [158, 160, 201] (see Chapter 6).

Conversely we can solve the ME for small systems exactly by numerical methods. This allows us to study the NESS exactly and allows us to compare continuum descriptions to the ‘ground truth’.

7.2.1 Fokker-Planck Equation

In this limit of large but finite systems ($N \gg 1$) the configuration space of densities $x_i = n_i/N$ approaches a continuum within a rectangle $(x_1, x_2) \in [0, s_1] \times [0, s_2]$. In this limit we also rescale time such that $\tau = T/N$ is a continuous variable. This equates one Monte Carlo step (MCS) in continuous time to N iterations of the microscopic model, i.e. on average each voter is picked once per MCS.

Following the standard and general procedures [161] (Chp. 6) to obtain the continuum limit of the ME (7.1) we arrive at the FPE for the probability density $P(\mathbf{x}, \tau)$:

$$\frac{\partial}{\partial \tau} P(\mathbf{x}, \tau) = \sum_{i=1,2} \frac{\partial}{\partial x_i} \left[\frac{\partial}{\partial x_i} u_i(\mathbf{x}) P(\mathbf{x}, \tau) - v_i(\mathbf{x}) P(\mathbf{x}, \tau) \right] \quad (7.17)$$

where in this case $u_i \equiv (w_i^+ + w_i^-)/2N$ and $v_i \equiv w_i^+ - w_i^-$, and where $w_i^+ = (s_i - x_i)(z_+ + x_1 + x_2)^{q_i}$ and $w_i^- = x_i(z_+ + x_1 + x_2)^{q_i}$ are the continuum counterparts of W_i^{\pm} .

From the definition (7.7), the right hand side can be identified as the divergence of the probability current density, i.e.,

$$\frac{\partial}{\partial \tau} P(\mathbf{x}, \tau) = \nabla \cdot \mathbf{K}(\mathbf{x}, \tau).$$

In the NESS the system loses the dependence on time and hence all time derivatives are

also zero. The stationary FPE then becomes

$$\sum_{i=1,2} \frac{\partial}{\partial x_i} \left[\frac{\partial}{\partial x_i} u_i(\mathbf{x}) P^*(\mathbf{x}) - v_i(\mathbf{x}) P^*(\mathbf{x}) \right] = 0 \quad (7.18)$$

Therefore to find an analytical expression for the stationary state $P^*(\mathbf{x})$ we must solve (7.18) with non-trivial boundary conditions.

7.2.2 Mean-Field Analysis

The next level of approximation to the FPE is to consider the evolution of the averages of key quantities rather than the full probability distributions of each. The averages of interest are of the density of +1 voters in each population given by $\langle x_i \rangle(\tau) = \int x_i P(\mathbf{x}, \tau) d\mathbf{x}$. The evolution of these quantities is readily computed by multiplying the FPE (7.17) by x_i and integrating over \mathbf{x} . This gives

$$\frac{d}{d\tau} \langle x_i \rangle = \int x_i \frac{\partial}{\partial \tau} P(\mathbf{x}, \tau) d\mathbf{x} \stackrel{\text{(IBP)}}{=} \int K_i(\mathbf{x}, \tau) d\mathbf{x}$$

since the surface contributions from the integration by parts involve the normal components of \mathbf{K} and vanish. Noting that $\int \partial_i [u_i P] d\mathbf{x}$ is a surface term, we know that it is not necessarily zero but vanishes at the highest order for large N . Hence as $N \rightarrow \infty$, the only contributions from the right hand side of (7.17) are:

$$\begin{aligned} \frac{d}{d\tau} \langle x_i \rangle &= \langle v_i(\mathbf{x}) \rangle \\ &= \langle (s_i - x_i)(z_+ + x_1 + x_2)^{q_i} \rangle - \langle x_i(z_+ + x_1 + x_2)^{q_i} \rangle. \end{aligned} \quad (7.19)$$

To make further progress we invoke the mean-field approximation (MFA) which assumes that higher moments can be factored in terms of averages (e.g. $\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle$). Using this approach all fluctuations and correlations between variables are ignored. Applying the MFA to (7.19) we arrive at the mean-field rate equations (REs):

$$\frac{d}{d\tau} x_i = (s_i - x_i) \mu^{q_i} - x_i (1 - \mu)^{q_i} \quad (7.20)$$

where $\mu = M/N = z_+ + x_1 + x_2$ is the total density of voters holding the +1 opinion, and angled brackets have been dropped.

Reduction of the system to the mean-field equations yields an autonomous system of coupled non-linear ordinary differential equations (ODEs) which are amenable to the typical analysis of finding fixed points and checking their stability. Setting $\frac{d}{d\tau}x_i = 0$, the system admits fixed points of the implicit form

$$x_i^* = \frac{s_i}{1 + \rho^{q_i}}, \quad (7.21)$$

where $\rho = (1 - \mu^*)/\mu^*$, with $\mu^* = z_+ + x_1^* + x_2^*$ (dependent on x_i^*). The ratio ρ has meaning in being the ratio of voters holding the -1 opinion to those holding the $+1$ opinion in the steady state. It satisfies

$$\frac{1}{1 + \rho} = z_+ + \sum_{i=1,2} \frac{s_i}{1 + \rho^{q_i}} \quad (7.22)$$

since the left hand side is $\mu^* = z_+ + x_1^* + x_2^*$, which equals the right hand side (using (7.21)). At this point it is useful to introduce the variables

$$\theta_i = \frac{1}{q_i} \ln \left(\frac{s_i}{x_i} - 1 \right) \quad (7.23)$$

with $x_i = s_i/(1 + e^{q_i\theta_i})$ and the fixed points are instead given by $\theta_i = \ln \rho$. It is worth noting that while the natural variables (x_1, x_2) were restricted to the rectangle $[0, s_1] \times [0, s_2]$, the θ_i variables occupy the entire plane, i.e., $\theta_i \in (-\infty, \infty)$.

7.2.3 Linear Stability and Phases

For the remainder of this chapter, aside from Section 7.5, we will focus on the particularly interesting case of symmetric zealotry ($z_+ = z_- = z$) for which the 2qVZ exhibits a continuous phase transition. For any value of z the Equation (7.20) always admits one solution, $\rho = 1$ ($\mu^* = 1/2$). This corresponds to a central FP $\mathbf{x}^* = \mathbf{x}^{(0)} \equiv (s_1/2, s_2/2)$ where the opinion within each population is split equally. As z is decreased below a critical value z_c , Equation (7.20) admits two further solutions, $\mathbf{x}^{(\pm)}$ with $x_i^- < x_i^{(0)} < x_i^+$. For the specific case of $(q_1, q_2) = (1, 2)$ and $s_1 = s_2 = s$ these are given explicitly by

$$\mathbf{x}^{(\pm)} = \left(\frac{s}{2(1-s)} \left[1 - s \pm \sqrt{s(4-3s)-1} \right], \frac{s \pm \sqrt{s(4-3s)-1}}{2} \right).$$

To check the stability of a FP we linearise the rate Equations (7.20) about the FP and find the linear stability matrix, $-\partial\dot{x}_i/\partial x_j|_{\mathbf{x}=\mathbf{x}^*}$. This takes the form

$$\mathbb{F}(\mathbf{x}^*) = \begin{pmatrix} Y_{1\mu} - X_{1\mu} & -X_{1\mu} \\ -X_{2\mu} & Y_{2\mu} - X_{2\mu} \end{pmatrix}, \quad (7.24)$$

where

$$\begin{aligned} Y_{i\mu} &= \mu^{*q_i}(1 + \rho^{q_i}), \\ X_{i\mu} &= q_i x_i^* (1 - \mu^*)^{q_i-1} (1 + \rho). \end{aligned}$$

Evaluating $\det \mathbb{F}$ at $\mathbf{x}^* = \mathbf{x}^{(0)}$ gives the simple result

$$\det \mathbb{F}(\mathbf{x}^{(0)}) = 2^{2-q_1-q_2} [1 - q_2 s_2 - q_1 s_1]. \quad (7.25)$$

Stable FPs require that $\det \mathbb{F}(\mathbf{x}^*) > 0$, so this implies the central FP is stable whenever

$$1 > s_1 q_1 + s_2 q_2. \quad (7.26)$$

When (7.26) is not satisfied $\mathbf{x}^{(0)}$ becomes unstable, which also coincides with the emergence of the two other FPs $\mathbf{x}^{(\pm)}$ which can be shown to be stable. This means the system exhibits a pitchfork bifurcation at z_c . In Figure 7.3 we show the phase diagram with the two regimes separated by the critical line $s_1 q_1 + s_2 q_2 = 1$.

Since $1 = 2z + s_1 + s_2$, we can express the critical z_c line as a function of the qs and the asymmetry in population densities $\Delta s \equiv s_1 - s_2$. This is given by

$$z_c = \frac{\bar{q} - 1}{2\bar{q}} + \frac{q_1 - q_2}{q_1 + q_2} \frac{\Delta s}{2}, \quad (7.27)$$

where \bar{q} is the average $(q_1 + q_2)/2$ ⁵.

In Figure 7.4 we show typical steady state time series of the $2qVZ$ in both regimes. In this example, $\Delta s = 0$ and $(q_1, q_2) = (1, 2)$, leading to a critical zealotry of $z_c = 1/6$. In the low zealotry regime ($z < z_c$) the system fluctuates between the two FPs $\mathbf{x}^{(\pm)}$, whereas in the high zealotry regime ($z > z_c$) the system fluctuates around the central FP, $\mathbf{x}^{(0)}$. The mean-field approximation accurately predicts the location of the FPs in

⁵ The method applied here is powerful enough for us to generalise to a population of any number of groups of q_i -susceptibles, in which case the critical line is given by $1 = \sum_i s_i q_i$.

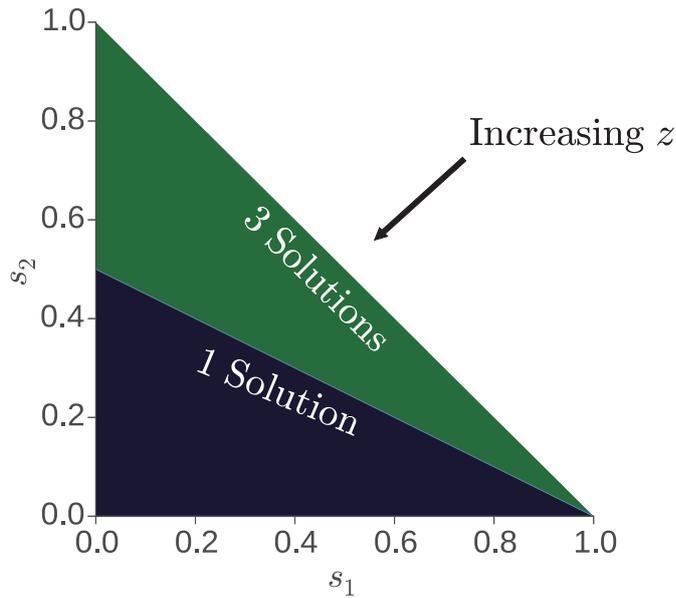


Figure 7.3: Criticality in the symmetric $2qVZ$ ($z_{\pm} = z$). Criticality at z_c occurs on the interface between the blue and green regimes, prescribed by $s_1 q_1 + s_2 q_2 = 1$ where the corresponding critical zealotry is $z_c = (1 - s_1 - s_2)/2$. In the blue regime the mean-field Equations (7.20) have a single central fixed point, whereas in the green regime they admit three fixed points. Here $(q_1, q_2) = (1, 2)$.

both regimes, however it can not tell us anything about the demographic fluctuations around (or between) the FPs.

Finally, setting $q_1 = q_2$ we are reduced to a homogeneous population as of the $qVMZ$ [191] with $z_c = (q - 1)/(2q)$. In this spirit, we may introduce an *effective* q_{eff} for our heterogeneous $2qVZ$:

$$q_{\text{eff}} = \frac{s_1 q_1 + s_2 q_2}{s_1 + s_2} \quad (7.28)$$

in the sense that the number of FPs in the $2qVZ$ are the same as those in the homogeneous $qVMZ$ with this q_{eff} ⁶.

⁶ As with the critical line, this can be generalised to $q_{\text{eff}} = \sum_i s_i q_i / \sum_i s_i$ for populations of any composition.

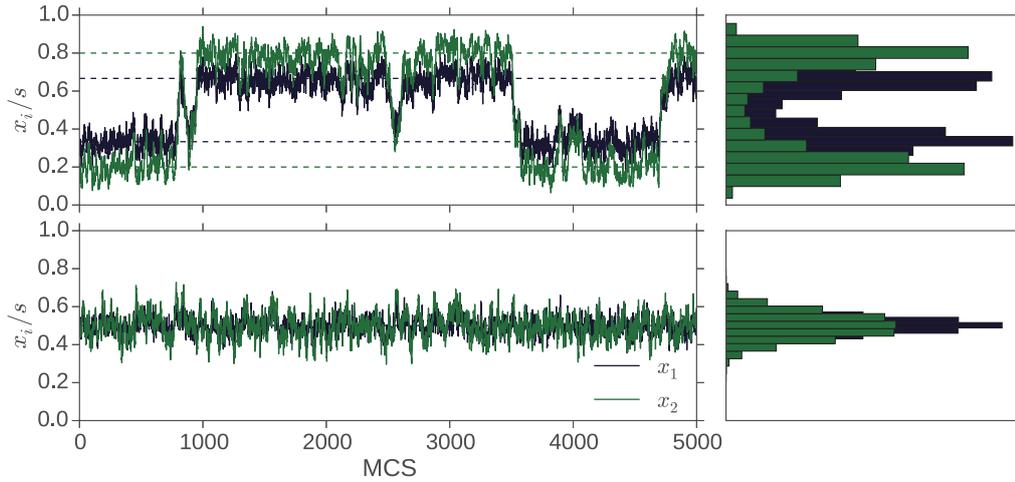


Figure 7.4: Typical time series of the $2qVZ$. (Top) time series of the fraction of +1 voters within each population (x_i/s) in the low zealotry regime ($z = 1/7$). The process switches between two fixed points, indicated by dashed coloured lines, on a characteristic timescale. The distribution of opinion levels (right) show that little time is spent in between the two fixed points. (Bottom) the $2qVZ$ in the high zealotry regime ($z = 2/9$). Both populations oscillate around $x_i/s = 0.5$. The other parameters of the model are $(S, q_1, q_2) = (250, 1, 2)$. One MCS equates to N iterations of the model.

7.2.4 Numerically Exact Results for Small Systems

As we have seen for systems which do not reach equilibrium but instead settle into a NESS it is difficult to attain exact analytical expressions for the stationary state $P^*(\mathbf{n})$. For small systems however we are able to attain a numerically exact solution for $P^*(\mathbf{n})$ (to within a prescribed accuracy). This can be achieved by numerically iterating the ME (7.1) in matrix form, i.e.,

$$\mathbf{P}(T+1) = \mathcal{G}\mathbf{P}(T)$$

where \mathcal{G} is the $(S_1 + 1)(S_2 + 1) \times (S_1 + 1)(S_2 + 1)$ transition matrix and $\mathbf{P}(T)$ is a suitably ordered probability vector whose elements are $P(\mathbf{n}, T)$.

To find \mathbf{P}^* we exploit the matrix relation $\mathcal{G}\mathbf{P}(0) = \mathbf{P}(\infty) = \mathbf{P}^*$, independently of the initial condition $\mathbf{P}(0)$. In practice we compute $\mathbf{P}^* = \mathcal{G}\mathbf{P}(0)$ by iterating $\mathcal{G}^{2k} = \mathcal{G}^k \mathcal{G}^k$ until the desired accuracy is reached. In particular for a system with $S_1 = S_2 = 100$, we find $\mathbf{P}^*(\mathbf{n})$ accurate up to 10^{-15} with 64 iterations (i.e. \mathcal{G}^{264}). In this case \mathcal{G} has

10^8 elements, however only 5×10^4 are non-zero due to there being only five possible transitions from each state. The transition matrix is therefore sparse. Subsequent powers of \mathcal{G} quickly become dense which in turn becomes problematic for the storage of the matrix in memory. The calculation of the stationary distribution for larger systems requires a trade-off between the storage of higher powers of \mathcal{G} and the computational cost of repeated multiplication of \mathbf{P} by some power of \mathcal{G} to reach the desired level of accuracy as quickly as possible. For systems up to $S = 100$ this can be done on a modern computer without issue. For systems larger than $S = 100$ it is possible to attain accurate values for $P^*(\mathbf{n})$ by utilising a method which uses expanded $P^*(\mathbf{n})$'s for smaller systems and interpolating to get an initial condition close to the stationary distribution for the larger system.

With the stationary state known, we are able to compute all the other quantities of interest such as the stationary probability current $\mathbf{K}^*(\mathbf{n})$, vorticity $\omega^*(\mathbf{n})$, and stream function $\psi^*(\mathbf{n})$. In Figure 7.5 we show the stationary state P^* and probability currents \mathbf{K}^* for the model in the high and low zealotry regimes, described in the previous section. The stationary states (a) and (b) display two peaks and a single peak respectively (in agreement with the mean-field predictions). The probability current fields (c) and (d) both show counter-clockwise “whirls” around each peak. These whirls imply correlation of the dynamic properties of the two populations, indicating the tendency of q_2 -susceptibles to “follow” q_1 -susceptibles. This is confirmed by our calculation of the correlation functions in Section 7.3.

In Figure 7.6 we fully characterise the NESS in the low zealotry regime with the stationary distribution (a), probability currents (b), vorticity (c), and stream function (d). These are plotted in θ -space (7.23) where all three FPs lie on the line $\theta_1 = \theta_2$. Although $P^*(\mathbf{n})$ looks symmetrical under this transformation it is in fact not, and the “ridge” that runs from one peak to the other through the saddle lies close to, but not on, the line $\theta_1 = \theta_2$. Panel (c) shows the vorticity field, which as expected, is positive near the peaks, corresponding to the counter-clockwise whirls of \mathbf{K}^* . Less obvious is the presence of counter rotations ($\omega^* < 0$) away from the peaks. Lastly, the stream function ψ^* (panel (d)) has appearance similar to P^* ; they are in fact proportional to each other.

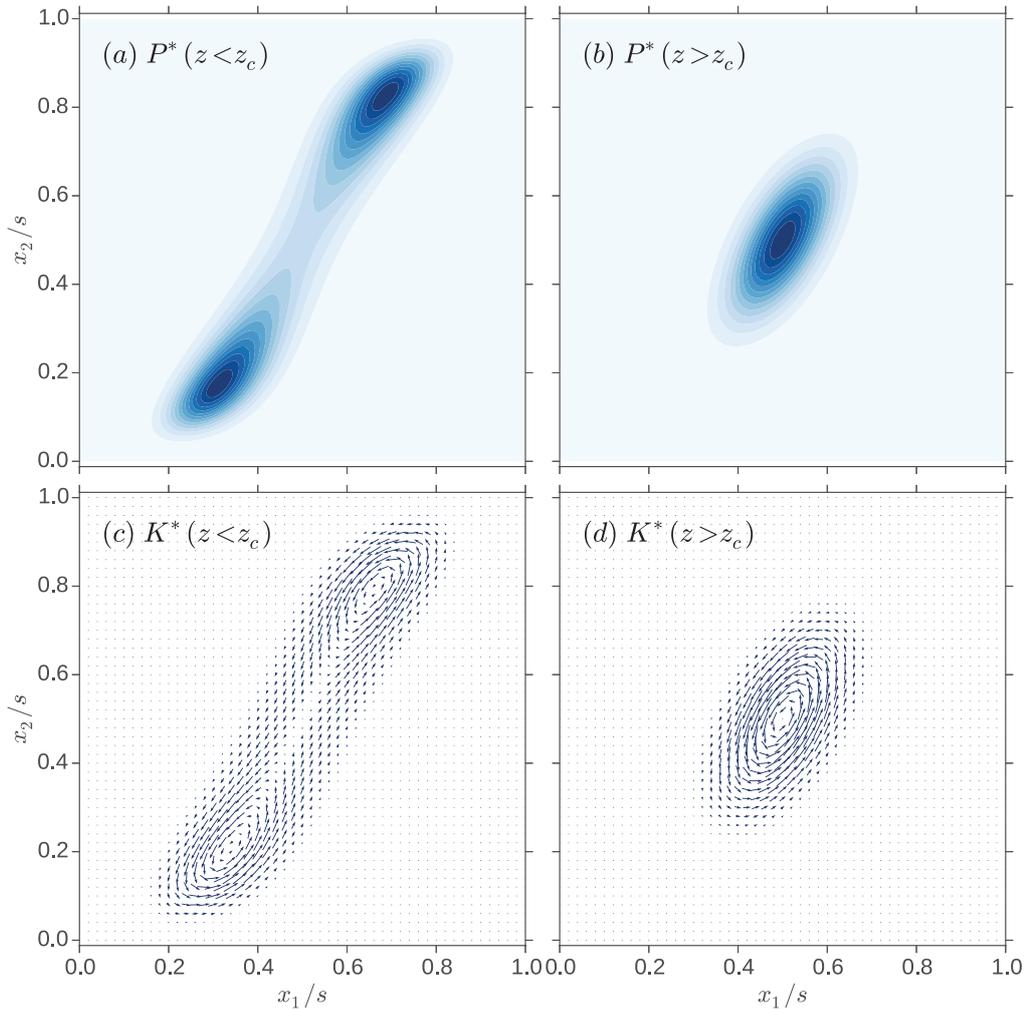


Figure 7.5: Numerically exact stationary distributions (top) and probability currents (bottom) in both the low zealotry regime, $z < z_c$ (left), and high zealotry regime, $z > z_c$ (right). Parameter values are $(S, q_1, q_2) = (100, 1, 2)$ and $Z = 40$ and 80 in the low and high zealotry regimes respectively.

7.3 The Linear Gaussian Approximation (LGA)

To describe the fluctuations in the NESS we examine the FPE beyond the lowest order in $1/N$. This procedure provides a scaling analysis of typical non-critical fluctuations in systems of large, but finite N . As with our consideration of linear stability, we consider small deviations around a stable fixed point \mathbf{x}^* : $\boldsymbol{\xi} = \mathbf{x} - \mathbf{x}^*$. The LGA consists of linearising the drift term and evaluating the diffusion term at the fixed point, resulting

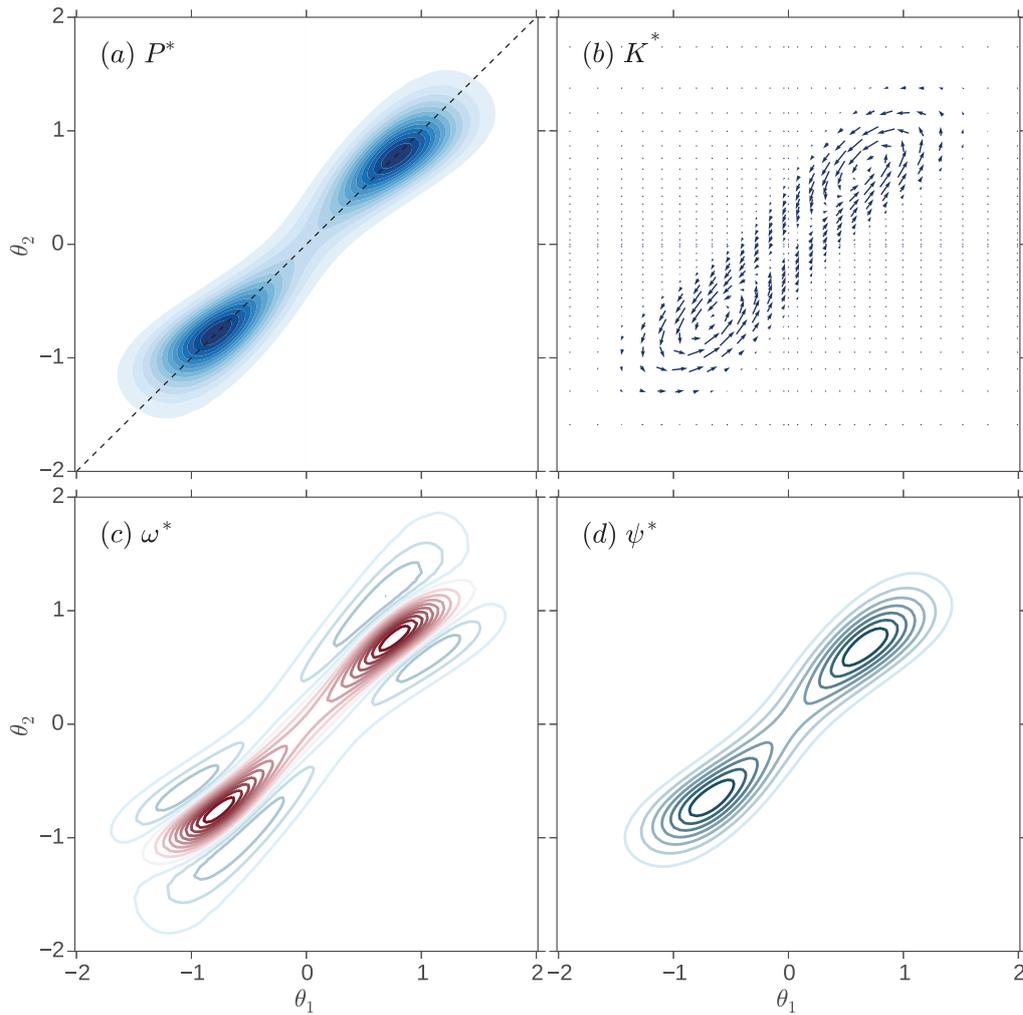


Figure 7.6: Numerically exact properties of the 2qVZ NESS with $(N, S, Z, q_1, q_2) = (280, 100, 40, 1, 2)$ (low zealotry regime). (a) Heat maps of the stationary distribution P^* in θ -space (7.23). The peaks of P^* fall on the dotted line $\theta_1 = \theta_2$ however the distribution is *not* symmetric about this line. Areas of higher probability are more darkly shaded. (b) Stationary probability currents \mathbf{K}^* . (c) Stationary vorticity ω^* . (d) Stationary stream function ψ^* .

in the LGA FPE for the stationary distribution:

$$\nabla \cdot [\mathbb{F}\boldsymbol{\xi} + \mathbb{D}\nabla] P^*(\boldsymbol{\xi}) = 0. \quad (7.29)$$

Here $\mathbb{D} = \mathbb{D}(\mathbf{x}^*)$ is the diffusion matrix given by

$$\mathbb{D}(\mathbf{x}^*) = \frac{1}{N} \begin{pmatrix} x_1^*(1 - \mu^*)^{q_1} & 0 \\ 0 & x_2^*(1 - \mu^*)^{q_2} \end{pmatrix} \quad (7.30)$$

and \mathbb{F} is the drift matrix defined by (7.24). By means of taking a Fourier transform of (7.29) and using a Gaussian ansatz we see that the LGA FPE has solution

$$P^*(\boldsymbol{\xi}) = \frac{1}{2\pi\sqrt{\mathbb{C}}} \exp \left[-\frac{1}{2} \boldsymbol{\xi}^T \mathbb{C}^{-1} \boldsymbol{\xi} \right], \quad (7.31)$$

where \mathbb{C} is the covariance matrix, with elements

$$C_{ij} = \langle \xi_i \xi_j \rangle^*. \quad (7.32)$$

The covariance matrix can be obtained as a function of \mathbb{F} and \mathbb{D} by solving the Lyapunov equation $\mathbb{F}\mathbb{C} + \mathbb{C}\mathbb{F}^T = 2\mathbb{D}$ which arises as a necessary condition for the solution given. From this equation we can also see that, as \mathbb{F} is $\mathcal{O}(1)$ and \mathbb{D} is $\mathcal{O}(1/N)$ then \mathbb{C} must also be $\mathcal{O}(1/N)$.

7.3.1 Correlations and Probability Flows

Having established an approximation of P^* within the LGA we can find LGA expressions for the observables of interest. The probability currents $\mathbf{K}^* = -[\mathbb{D}\nabla + \mathbb{F}\boldsymbol{\xi}] P^*$ can be expressed in two ways. By noting $\nabla P^* \propto -\mathbb{C}^{-1}\boldsymbol{\xi} P^*$ (using (7.31)), the first expression

$$\mathbf{K}^*(\boldsymbol{\xi}) = \{\mathbb{D}\mathbb{C}^{-1} - \mathbb{F}\} \boldsymbol{\xi} P^*(\boldsymbol{\xi}) \quad (7.33)$$

confirms the linear relationship between \mathbf{K}^* and P^* . The other, using $\mathbb{D} = [\mathbb{F}\mathbb{C} + \mathbb{C}\mathbb{F}^T] / 2$ is

$$\mathbf{K}^*(\boldsymbol{\xi}) = \frac{\mathbb{F}\mathbb{C} - \mathbb{C}\mathbb{F}^T}{2} \nabla P^*(\boldsymbol{\xi}), \quad (7.34)$$

which explicitly shows that \mathbf{K}^* is divergence free.

The key observation here is that the matrix $\mathbb{F}\mathbb{C} - \mathbb{C}\mathbb{F}^T$ is antisymmetric which leads us to define

$$\mathbb{L} \equiv \begin{pmatrix} 0 & L_{12} \\ -L_{12} & 0 \end{pmatrix} = \mathbb{F}\mathbb{C} - \mathbb{C}\mathbb{F}^T. \quad (7.35)$$

Then, rewriting (7.34) in terms of \mathbb{L} , we see that $K_i^* = \varepsilon_{ij} \partial_j (L_{12} P^* / 2)$. This takes the form of (7.14) so we can identify the stream function $\psi^* = L_{12} P^* / 2$. Furthermore, from

the continuum version of (7.15) we see that, ignoring surface terms in the integration by parts,

$$\langle \mathcal{L} \rangle^* = \int \varepsilon_{ij} \xi_i K_j^* d\xi = L_{12} \quad (7.36)$$

Finally we look to find the continuum version of the two-point correlation function of (7.9), namely

$$\langle \xi_i' \xi_j \rangle_\tau^* = \int \xi_i' \xi_j \mathcal{P}^*(\xi', \tau | \xi, 0) d\xi' d\xi.$$

Within the realms of the LGA it is easier to use the corresponding solution to the Langevin equation: $\xi(\tau) = e^{-\mathbb{F}\tau} \xi(0)$ (+ noise). As the noise is uncorrelated in time,

$$\langle \xi_i(0) \xi_j(\tau) \rangle^* = \sum_k e^{-F_{jk}\tau} \langle \xi_i(0) \xi_k(0) \rangle^* = \sum_k C_{ik} e^{-F_{jk}\tau}$$

The antisymmetric part of this correlation is necessarily odd in τ and is non-vanishing for systems in a NESS.

We can also define the time-independent quantity

$$\tilde{C}(\tau) \equiv \langle \xi_1 \xi_2 \rangle_\tau^* - \langle \xi_2 \xi_1 \rangle_\tau^* \quad (7.37)$$

which is the 1-2 element of the antisymmetric matrix

$$\tilde{\mathbb{C}}(\tau) = \mathbb{C} e^{-\mathbb{F}^T \tau} - e^{-\mathbb{F} \tau} \mathbb{C}. \quad (7.38)$$

Furthermore we can exploit Sylvester's formula [202] to write

$$e^{-\mathbb{F}\tau} = \left(\frac{\lambda_+ e^{-\lambda_- \tau} - \lambda_- e^{-\lambda_+ \tau}}{\lambda_+ - \lambda_-} \right) \mathbb{I} + \left(\frac{e^{-\lambda_+ \tau} - e^{-\lambda_- \tau}}{\lambda_+ - \lambda_-} \right) \mathbb{F},$$

where \mathbb{I} is the identity matrix and $\lambda_\pm = \left(\text{Tr} \mathbb{F} \pm \sqrt{(\text{Tr} \mathbb{F})^2 - 4 \det \mathbb{F}} \right) / 2$ are the eigenvalues of \mathbb{F} . Hence we can rewrite (7.38) as

$$\tilde{\mathbb{C}}(\tau) = \left(\frac{e^{-\lambda_- \tau} - e^{-\lambda_+ \tau}}{\lambda_+ - \lambda_-} \right) \underbrace{[\mathbb{F}\mathbb{C} - \mathbb{C}\mathbb{F}^T]}_{=\mathbb{L}}. \quad (7.39)$$

Or, in terms of its only independent quantity,

$$\tilde{C}(\tau) \equiv \langle L \rangle^* \left(\frac{e^{-\lambda_- \tau} - e^{-\lambda_+ \tau}}{\lambda_+ - \lambda_-} \right).$$

Due to the form of $\tilde{C}(\tau)$ we can see that $\tilde{C}(\tau) > 0$ for all τ , and we expect $\tilde{C}(\tau)$ to find its maximum value at $\tau^* = \ln(\lambda_+/\lambda_-)/(\lambda_+ - \lambda_-)$.

7.3.2 Simulation Results and Assessment of Fit

The equations (7.33)-(7.39) derived previously offer significant information on the behaviour of the $2qVZ$ in the realms of linear approximation about the fixed points. While this analysis holds for any values of $s_{1,2}$ and $q_{1,2}$, in this section we compare the LGA results with simulation in the specific case of $s_{1,2} = s, z_{\pm} = 1/2 - s$, and $q_2 = 2q_1 = 2$. In doing so we assess the validity of the LGA and quantify the level of fit for a specific case.

Our Monte Carlo simulations cover a range of system sizes up to $N = 14400$ and are run for 10^5 MCS. By compiling histograms from the model trajectories we find a Monte Carlo estimate of $P^*(\mathbf{n})$. We find that in the high zealotry regime the LGA prediction (7.31) is an excellent approximation, i.e. $P^*(\boldsymbol{\xi}) \simeq NP^*(\mathbf{n})$, and captures the Gaussian peak around the fixed point $\mathbf{x}^{(0)}$. In the low zealotry regime the stationary distribution is bimodal as expected with two peaks around $\mathbf{x}^{(\pm)}$. However the distribution is not fully Gaussian but is in fact skewed and more sharply peaked than the Gaussian approximation. For systems smaller than $N \sim 10^3$ there are clear visible deviations between the simulations and LGA approximation yet for larger systems the agreement is reasonable in both regimes.

To quantify this we calculate the excess kurtosis and skewness of the one-dimensional projections of $P^*(\boldsymbol{\xi})$ onto each axis, given by

$$P^*(\xi_1) = \int_0^1 P^*(\xi_1, \xi_2) d\xi_2$$

and with a corresponding expression for $P^*(\xi_2)$. For the smallest system with $S = 250$ in the low zealotry regime the kurtosis was $(-0.242, 0.750)$ for the ξ_1 and ξ_2 projections respectively, while the skewness was $(0.321, 0.969)$. By contrast, in the high zealotry regime the kurtosis was $(-0.030, -0.130)$ and skewness $(0.004, -0.007)$. This confirms that the LGA provides a better approximation in the high zealotry regime.

As the system size increases the kurtosis and skewness approach zero in both regimes. In Figure 7.7 we see the improvement of the LGA with N , this time in terms of the correlations $\langle \xi_1 \xi_2 \rangle^*$ in the NESS. This confirms that $C_{ij} \propto 1/N$, i.e. the fluctuations scale as $N^{-1/2}$ far from criticality. As with the skew and kurtosis, the LGA provides a

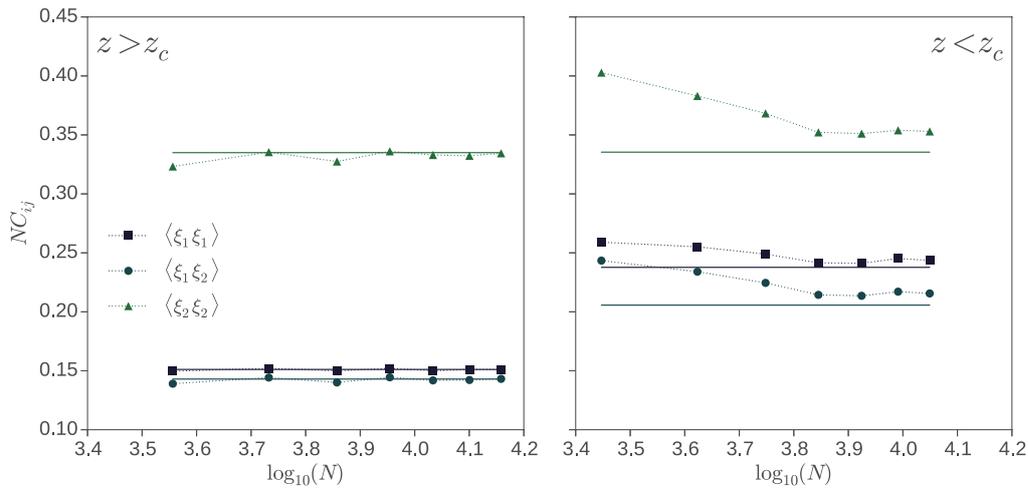


Figure 7.7: $C_{ij} = \langle \xi_i \xi_j \rangle$ as a function of system size N . Comparison of the LGA predictions (7.32) (solid) against results of stochastic simulation (marked), averaged over at least 10^5 MCS. The scaling $C_{ij} \propto 1/N$ is confirmed and the quantitative agreement improves as N increases. Here in the low zealotry regime (left) $(s, z, q_1, q_2) = (5/14, 1/7, 1, 2)$. By comparison the parameters in the high zealotry regime are $(s, z, q_1, q_2) = (5/18, 2/9, 1, 2)$, and the critical value of z being $z_c = 1/6$.

good quantitative fit in the high zealotry regime for $N \gg 10^3$, however much larger system sizes are required to achieve similar levels of quantitative agreement in the low zealotry regime. This is more likely due to the skewness in the peaks of P^* which is present for smaller systems.

We now turn to study the genuine non-equilibrium observables. Using (7.10), we can compute the probability current exactly and compare the results to the current obtained from (7.33), as shown in Figure 7.8. In this comparison there are two things of note. Firstly the mean field fixed point and the peaks of the distribution (right) are not fully aligned. This is a finite size effect (as the MF fixed point is derived for an infinite population system) and is less pronounced as the system becomes larger. Secondly the probability currents in the LGA are qualitatively similar to the exact solution; they are in agreement with the counter-clockwise whirls seen in Figure 7.8 (left) and Figure 7.5 also.

Finally we look at the antisymmetric two-point correlation function $\tilde{C}(\tau)$. This is

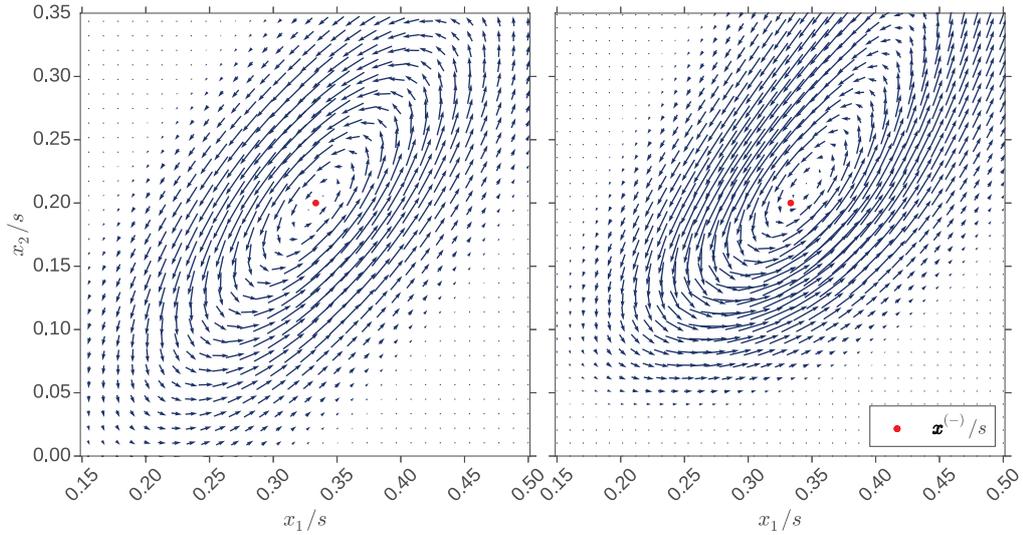


Figure 7.8: Stationary probability currents. Comparison of the LGA predictions (7.33) (left) with the numerically exact counterpart (7.7) (right) near the fixed point $\mathbf{x}^{(-)}$ (red dot) in the low zealotry phase. Parameters here are $(N, S, Z, q_1, q_2) = (280, 100, 40, 1, 2)$.

calculated from simulation by considering the lagged correlation

$$\frac{1}{R - \tau} \sum_{\tau=0}^{R-\tau} x_1(\tau)x_2(\tau + \tau) - x_2(\tau)x_1(\tau + \tau)$$

where R is the length of the run (typically 10^5 MCS). Figure 7.9 shows that the $2qVZ$ is characterised by $\tilde{C} > 0$ with a single peak expected from Equation (7.39). This means that the q_i -susceptibles are correlated in such a way that $\langle x'_1 x_2 \rangle_{\tau}^* > \langle x'_2 x_1 \rangle_{\tau}^*$, indicative of the q_2 -susceptibles ‘following’ the q_1 -susceptibles on a finite timescale ($\tau \approx 40$) with a peak at around 3 MCS. This peak is accurately captured by the LGA.

7.4 Fluctuation-driven Switching Dynamics

As with previous work on the $qVMZ$, and for stochastic processes with two favourable states in general, it is interesting to study the time taken to switch from one fixed point to another. In Figure 7.10 (left) we see that the long time behaviour of the $2qVZ$ in the low zealotry regime is characterised by so-called “swing-state dynamics”. This behaviour is encapsulated by long periods of oscillation about one of $\mathbf{x}^{(\pm)}$ followed by sudden

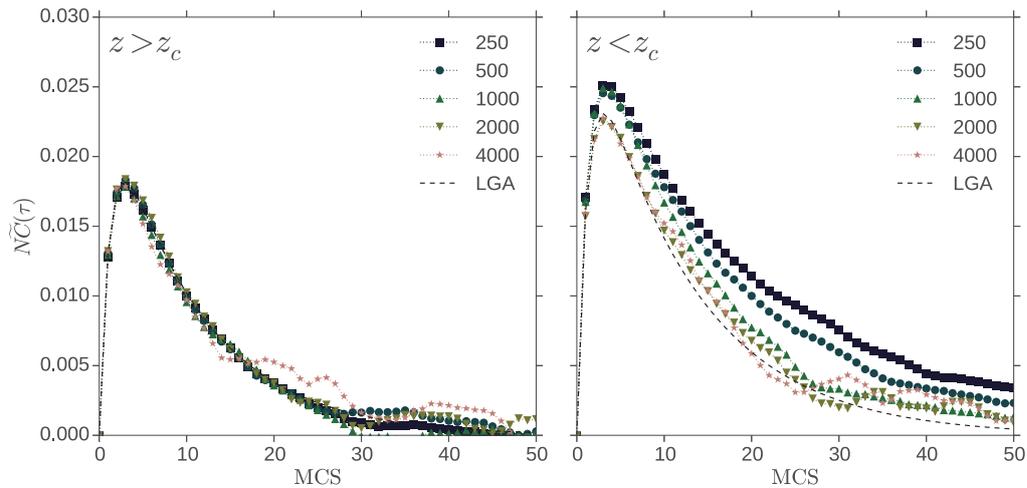


Figure 7.9: The LGA prediction of the two-point correlation function $\tilde{C}(\tau)$ (dashed) compared to simulations of varying susceptible population size in the high (left) and low (right) zealotry regimes with $s_{1,2} = s, z_{\pm} = z$ and $(q_1, q_2) = (1, 2)$. In the high zealotry regime ($z = 2/9$) the LGA captures the behaviour of $\tilde{C}(\tau)$ for all values of system size considered. However, in the low zealotry regime ($z = 1/7$) the LGA is only qualitatively accurate for $S \geq 2000$. Fluctuations for MCS greater than 20 are due to a low sampling rate. The LGA values for the peak τ^* are 2.80 and 3.05 for the high and low zealotry regimes respectively, accurately predicting the peak in the data which occurs at 3 in both cases.

switches or jumps to the opposite state on a much shorter timescale. Figure 7.10 (right) shows a typical jump from $\mathbf{x}^{(+)}$ to $\mathbf{x}^{(-)}$. As shown in the previous section, the q_2 -susceptibles “follow” the q_1 -susceptibles across the switch, however this behaviour is subtle and the lag between the switching of both populations is negligible to first order.

As the mean-field equations average out the fluctuations which drive this switching behaviour we rely on the FPE to study this phenomenon. In particular we look to measure the *mean switching time* τ_s , which measures the average time to switch from one fixed point to another⁷. Finding τ_s can be formulated as a first-passage problem [159]. This problem is readily solved in the one-dimensional case [191] (see Chapter 6), however

⁷ As this process is symmetric we can average over both $\mathbf{x}^{(+)} \rightarrow \mathbf{x}^{(-)}$ and $\mathbf{x}^{(-)} \rightarrow \mathbf{x}^{(+)}$ as they are equal.

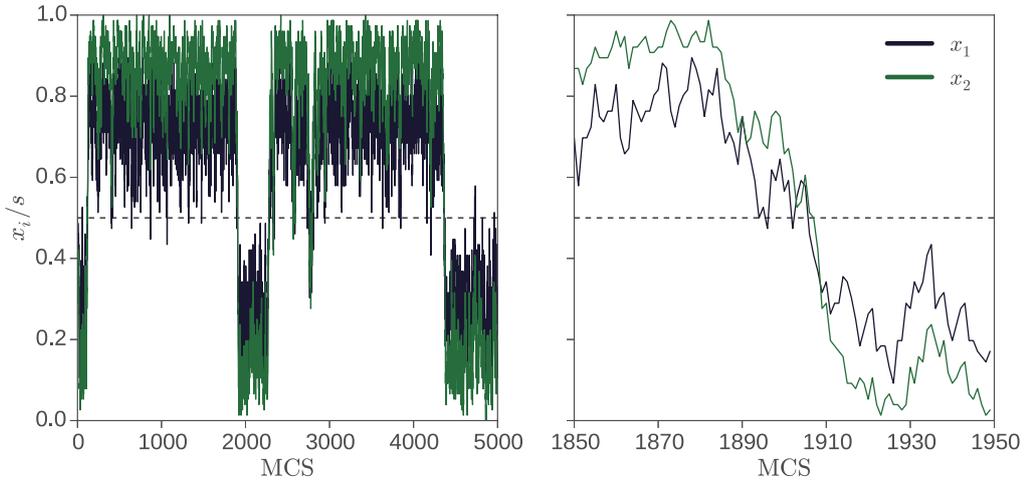


Figure 7.10: Switching behaviour of the $2qVZ$ ($z < z_c$). Time series of the total fraction of +1 susceptibles within each populations (x_i/s). (Left) a sample of 5000 MCS and (right) a closer examination of a typical switching event over 100 MCS. 1 MCS equates to N iterations of the model. Here the parameters are $(N, s, z, q_1, q_2) = (200, 0.38, 0.12, 1, 2)$

in two or more dimensions and with the violation of detailed balance the current theory breaks down. To make progress with estimating the mean switching time we exploit the mapping of the $2qVZ$ onto the equivalent $qVMZ$ such that they share the same value of z_c . For $z \lesssim z_c$ we expect that switching dynamics of the $2qVZ$ with (q_1, q_2) to be well approximated by the $qVMZ$ with a population of q_{eff} -susceptibles, with q_{eff} given by (7.28).

Within the q_{eff} approximation we can utilise the methodology used in Chapter 6 to yield an expression for τ_s . In particular we use the framework of the backward Fokker-Planck equation (bFPE) [159, 160, 158]. The infinitesimal generator of the bFPE is given by

$$\mathcal{G}_b^{\text{eff}}(x) = [\tilde{w}^+(x) - \tilde{w}^-(x)] \partial_x + \frac{1}{2N} [\tilde{w}^+(x) + \tilde{w}^-(x)] \partial_x^2 \quad (7.40)$$

where $\tilde{w}^+(x) = (s_1 + s_2 - x)(x + z)^{q_{\text{eff}}}$ and $\tilde{w}^-(x) = x(s_1 + s_2 + z - x)^{q_{\text{eff}}}$. The mean switching time can be computed by solving $\mathcal{G}_b^{\text{eff}}(x)\tau_s(x) = -1$ with the reflective and absorbing boundary conditions

$$\frac{d\tau_s}{dx}(x^{(-)}) = 0 = \tau_s(x^{(+)}).$$

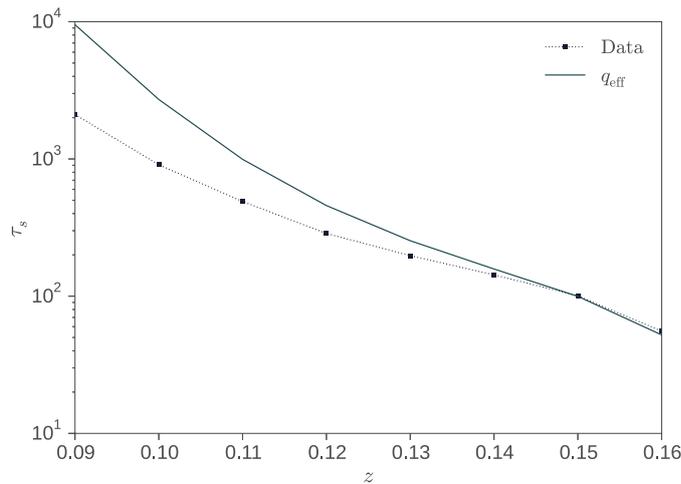


Figure 7.11: The mean switching time τ_s between the two stable fixed points for systems of size $N = 100$ with $(q_1, q_2) = (1, 2)$. Results obtained from stochastic simulations (blue \square) are compared with approximations (7.41). Simulation data are averaged over 10^6 MCS.

where $x^{(\pm)}$ are the fixed points of the q VMZ. This yields

$$\tau_s = 2N \int_{x^{(-)}}^{x^{(+)}} dy e^{-N\phi(y)} \int_{x^{(-)}}^y \frac{e^{N\phi(v)} dv}{\tilde{w}^+(v) + \tilde{w}^-(v)}, \quad (7.41)$$

where $\phi(v) = 2 \int_{x^{(-)}}^v \left\{ \frac{\tilde{w}^+(v) - \tilde{w}^-(v)}{\tilde{w}^+(v) + \tilde{w}^-(v)} \right\}$. Using Kramers' formula [203–205, 191], we then have

$$\ln \tau_s \simeq 2N \int_{x^{(-)}}^{x^{(+)}} \frac{\tilde{w}^-(y) - \tilde{w}^+(y)}{\tilde{w}^-(x^{(-)}) + \tilde{w}^+(x^{(-)})} dy \quad (7.42)$$

This predicts that the mean switching time grows exponentially with N , which is confirmed by simulation.

Figure 7.11 shows that this approximation is accurate just below z_c ($z_c = 1/6$ in this case) and overestimates by nearly an order of magnitude far below z_c . This overestimation is interesting in itself and suggests that the presence of probability currents, absent from consideration in the approximation, are likely candidates for the speed up of the switching dynamics.

7.5 Asymmetric Zealotry

So far in this chapter we have considered the case of symmetric zealotry $z_+ = z_- = z$. The $2qVZ$ with asymmetric zealotry $z_+ > z_-$ shares many of the same features as its $qVMZ$ counterpart [191]. There are high and low zealotry regimes characterised by a unimodal and bimodal stationary probability distribution respectively. However now both distributions are highly asymmetric, even for small asymmetries in zealotry. Let $z_{\pm} = (1 \pm \delta)z$ where $\delta \in [-1, 1]$ is a measure of the asymmetry.

In Figure 7.12 (bottom left) we see that even for an asymmetry of $\delta \approx 0.034$ the peak at $\mathbf{x}^{(+)}$ is much more pronounced than the peak at $\mathbf{x}^{(-)}$ in the low zealotry regime. For high zealotry the “central” fixed point $\mathbf{x}^{(0)}$ is now skewed towards states with a +1 majority of voters (not shown). As in the symmetric case, detailed balance is violated and hence there exists non-zero probability currents in the NESS (Fig. 7.12 (bottom right)). Similar to before the probability current flows in an anti-clockwise fashion around the fixed points, although the currents around $\mathbf{x}^{(+)}$ are now significantly stronger than those around $\mathbf{x}^{(-)}$.

The model in the asymmetric regime is analysed readily using the techniques in sections 7.2 and 7.3, and more specifically we can use the FPE and LGA provided that $N \gg 1$. In particular, using the LGA we can find the stationary probability density $P^*(\boldsymbol{\xi})$ around each fixed point as well as LGA expressions for the probability currents $\mathbf{K}^*(\boldsymbol{\xi})$, vorticity $\omega^*(\boldsymbol{\xi})$ and correlation functions $C_{ij} = \langle \xi_i \xi_j \rangle^*$. For low zealotry the bimodal probability distribution is dominated by the peak at $\mathbf{x}^{(+)}$ which suggests that the LGA will be more accurate than the symmetric case provided that the fixed point $\mathbf{x}^{(+)}$ is sufficiently far from the boundary. This is due to a reduction in the skewness of the peak about $\mathbf{x}^{(+)}$ which is caused by the presence of the peak at $\mathbf{x}^{(-)}$.

We may also consider the switching behaviour of the process. Figure 7.12 (top) highlights that the system is characterised by the metastability of $\mathbf{x}^{(-)}$: while in principle the continuous switching between states can be observed, the switch from $\mathbf{x}^{(+)} \rightarrow \mathbf{x}^{(-)}$ occurs with extreme rarity, and the system spends a significant amount of time fluctuating about $\mathbf{x}^{(+)}$. Using the effective q mapping (Sec. 7.4) we can study the mean

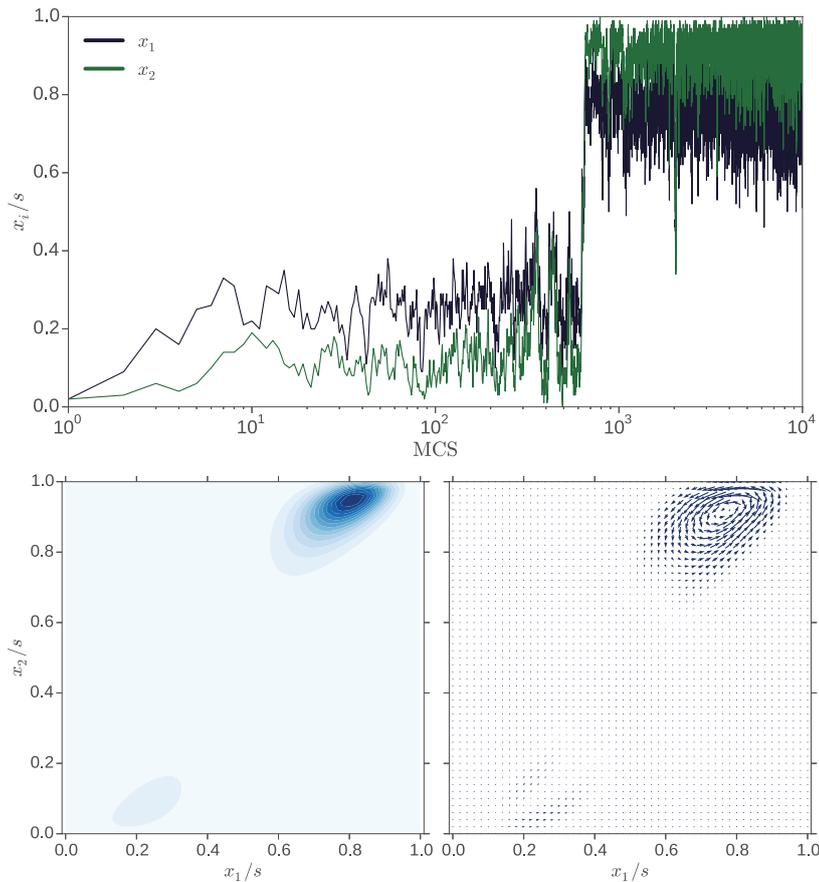


Figure 7.12: A summary of the low zealotry regime with asymmetric zealotry with parameters $(N, S, Z_+, Z_-, q_1, q_2) = (258, 100, 30, 28, 1, 2)$ and consequently $\delta = 0.01$. (Top) a typical time series for the first 10^4 MCS. Systems initialised with few +1 opinions soon reach the upper steady state where they stay for long times. (Bottom left) the stationary distribution P^* in the NESS. Darker regions are more probable. (Bottom right) the stationary probability current \mathbf{K}^* .

switching times $\tau_s^{-\rightarrow+}$ and $\tau_s^{+\rightarrow-}$ and, following the same analytic steps, we see that the mean switching times also grow exponentially with N .

7.6 Discussion

In this chapter we introduced the 2qVZ, a heterogeneous out-of-equilibrium non-linear voter model for a finite and well-mixed population. The 2qVZ is a direct but non-trivial generalisation of the qVMZ. Both models aim to replicate the concepts of group-size

dependence on conformity and the interplay between independence and conformity in opinion formation. The $2qVZ$ extends this notion by allowing individuals to require differing levels of group consensus for them to change their opinion, creating subpopulations of voters with varying degrees of conformity.

From a mathematical viewpoint the $2qVZ$ is radically different from its predecessor in that it violates detailed balance and is hence a genuine non-equilibrium system. Ignoring fluctuations and considering the mean-field behaviour, both systems display a high and low zealotry phase, separated by a critical value of z , z_c , where the system undergoes a pitchfork bifurcation. However, the $2qVZ$ settles to a non-equilibrium steady state which we have analysed in detail. In particular we have characterised the NESS in terms of the stationary probability distribution P^* , the stationary probability currents \mathbf{K}^* and other derivative features. These features have been studied for small systems by exact numerical calculation and also for larger systems by employing a linear Gaussian approximation to systems far from criticality. This has allowed us to understand the probability flows around the fixed points but has also cemented the LGA as a useful tool in the analysis of non-equilibrium social systems.

Furthermore we have investigated the asymmetric zealotry regime as well as the switching dynamics for systems with low zealotry. In particular, by mapping the model onto the one-dimensional $qVMZ$ which satisfies detailed balance we can see that the favourable flow of probability current dramatically reduces the mean switching time between states for systems far from criticality.

This work has allowed us to study in depth a genuine non-equilibrium system and has raised questions about the role and interpretation of closed loops of probability current in the NESS: are these related to the presence of “leaders” and “followers” in society?

Finally it is important to place this study into context. There are many instances of binary opinions which are shared and communicated. Twitter and Facebook have been quoted as being incredibly influential in the recent US presidential election, particularly as fake, robotic accounts have been used to swell the sizes of the respective populations of voters [206]. There are also many instances where public opinion has switched back and forth over time, such as in fashion, and attitudes to facial hair [207]. Applying the model to

complex and/or temporal networks would be an interesting extension which would be relevant from a social dynamics perspective and would help bring the model closer to real observable systems. Furthermore, we might question the Markovian assumptions of the model. Do the agents in real systems base their opinions solely on the current state of the system? While incorporating non-Markovian features into the model would drastically decrease the degree of analysis possible an empirical study of whether or not this assumption holds true would be enlightening for the 2qVZ and for voter models in general.

8

The LISA Model of Innovation Diffusion

The study of *innovation diffusion* aims to understand how new, previously unseen ideas, products, or behaviours spread throughout a society [208]. The spread of innovation can occur through various media, be that word-of-mouth, advertisement, or by direct observation. In fact the term “innovation diffusion” refers to a number of different social models and mechanisms such as contagion, imitation, and social learning¹ [210–212].

Traditionally, models of innovation diffusion have been based on a *well-mixed* population, that is, an infinitely-sized population where all possible agent-agent interactions are allowed [213]. This class of model are referred to as aggregate models, or mean-field models (see Chapter 6). Perhaps the most influential model of this class is that introduced by Bass in 1969 [214, 215, 72, 216]. In this model innovation can spread when a previous adopter of an innovation interacts with an agent susceptible to

¹ Somewhat confusingly, innovation diffusion rarely refers to the physical interpretation of diffusion described by a diffusion equation [209].

adopting (contagion), or by spontaneous adoption of the innovation through an external force (advertisement and mass media). The Bass model shares its primary mechanism with that of the susceptible-infected (SI) model of epidemic spreading [217]. This has led to many extensions of the model being studied in parallel in both fields, albeit with a different focus, and the shared use of terminology in society such as the notion of ‘viral advertising’.

The characteristic behaviour of the Bass model is exhibited by the sigmoidal shape of the fraction of adopters in the population over time. Adoption is initially slow to start however after a latency period complete adoption is quickly achieved. The main draws of the Bass model are its simplicity (having only two parameters) and that it readily yields an analytic solution. Due to this the Bass model has been successful in fitting historical data [218], in particular in economics where parameter values have been calculated for various classes of new products and innovations such as televisions, mobile phones, etc. The conclusions derived from such studies should of course be met with scepticism. The fitting of historical data does not validate the model, and examples are often chosen which agree with the model (confirmation bias).

The Bass model, as a model for real life application also has the following limitations:

- (a) The model is based on an infinitely large, homogeneous population [219, 160] and does not account for fluctuation-driven phenomena or heterogeneities in the population.
- (b) The model accounts only for contagion, omitting other possible mechanisms of innovation spread that result from social reinforcement and social pressure [220–222].
- (c) The model assumes that the entire population will eventually adopt, a so-called “pro-innovation bias”, however complete adoption is rarely (if ever) observed in real data [223–225, 208].

The last of these limitations is perhaps the most interesting. For incomplete adoption to occur in this setting either the innovation must not reach the entire population and/or

individuals must reject or oppose the innovation. This behaviour has been observed in society, for example 90% of Americans own a mobile phone as of 2014, and 68% had a smartphone in 2015 [226] however their use is accompanied by health and safety concerns [227]. Likewise the coverage for the measles, mumps, and rubella (MMR) vaccine in the United Kingdom reached only 92.7% in 2013-2014, below the government target of 95% required for herd immunity. This incomplete adoption may be attributed to scepticism surrounding the safety of the vaccination following bad press coverage which was promoted heavily by anti-vaccination movements [228].

The mechanisms for incomplete adoption will be the focus of this chapter, and will be introduced by means of incorporating a network structure to the dynamics and furthermore introducing individuals who may chose not to adopt the innovation. We predict that the network structure plays a significant role in the spread of innovation as non-adopters may block potential pathways of innovation spread if the network topology is restrictive.

Innovation Spreading In Social Media

The total spending on social media advertising in 2015 was \$23.68 billion globally [229], an increase of 33% on the previous year, with similar growth expected in subsequent years. This is indicative of the (perceived) power that social contacts and contagion has on influencing users to adopt new ideas and products. The precise mechanisms that drive social adoption are poorly understood which is reflected in the diversity of advertising campaigns; from targeted high-volume adverts, to attempts to personify the brand and converse with the user base.

There have been a number of studies aiming to model and understand the spread of innovation on social media, with 39% drawing from ideas of the diffusion of innovation [49]. There is however no universal model, and predictive power is limited. Another issue faced is the inference of adoption from social media data alone. While researchers are able to access a wealth of data from social media there have been few controlled studies. Instead, the sentiment [230] towards innovations is often calculated and used as a proxy.

Another closely related fact is that social media is increasingly being used as a medium for users to share news stories. This draws resemblance to the spread of ideas and innovation and is often easier to measure due to the unique identifier of a URL for a news article. One major difference is that innovations are beneficial to society, however news may be detrimental if the news is incorrect or strongly biased as has been seen in the recent US presidential election [29].

Chapter Outline

In Section 8.1 we introduce the LISA model of innovation spread which is based upon the Bass model of innovation diffusion and more generally on the compartment-based models used in epidemiology. We subsequently study the basic properties of the model in the mean-field limit and on complete graphs in Section 8.2, finding two regimes of behaviour dependent parameter values. In particular we look to understand how negative opinion affects the final level of adoption in the steady state and the speed of adoption up to that point. In Section 8.3 we study the model on random networks and one-dimensional lattices. In each case we compare the analysis with simulation results and highlight instances where the analytical approximations used break down and the results differ significantly. Finally we conclude with a brief discussion on how the model can inform strategies for the proliferation of potentially divisive innovations.

8.1 The LISA Model

To better understand the LISA model it is first helpful to further review the two-state Bass Model and its output. The Bass model consists of two types of agent: susceptibles S (yet to adopt), and adopters \mathcal{A} . The model allows susceptibles to become adopters by one of two mechanisms:

- (a) **Contagion-driven Spread:** a susceptible may become an adopter after interaction with another adopter, represented by the two-body interaction $S + \mathcal{A} \rightarrow \mathcal{A} + \mathcal{A}$.

- (b) **Spontaneous Adoption:** a single susceptible becomes an adopter $\mathcal{S} \rightarrow \mathcal{A}$, with a typically small rate.

The key feature of the Bass model is that the density of adopters over time is sigmoidal in shape, that is, the time derivative of this density has a single sharp peak (corresponding to an inflection point in the density itself). For all parameter values the model always reaches complete adoption eventually [208, 72, 215, 231, 232].

8.1.1 Model Outline

The LISA model [3] is a four-state model, consisting of a population of N agents that can be in one of the states of Luddite (\mathcal{L}), ignorant (\mathcal{I}), susceptible (\mathcal{S}), or adopter (\mathcal{A}). The mechanisms of the model (Figure 8.1) show how ignorant agents may either be persuaded to become susceptible and then subsequently reach the adopter state, or may convert to the Luddite state and permanently oppose the innovation. Specifically, the

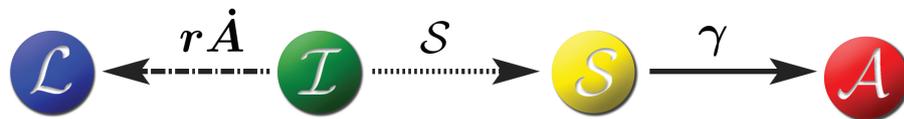


Figure 8.1: Schematic depiction of the LISA model. An ignorant \mathcal{I} can become a Luddite \mathcal{L} with rate $r\dot{A}$ (in a mean-field setting); an ignorant can also become a susceptible \mathcal{S} by contagion with rate proportional to the susceptible density. A susceptible spontaneously becomes an adopter at rate γ .

mechanisms which drive the LISA model are:

- (a) **Contagion-driven Spread:** an ignorant agent becomes susceptible upon interaction with another susceptible agent, i.e., $\mathcal{I} + \mathcal{S} \rightarrow \mathcal{S} + \mathcal{S}$ with rate 1.
- (b) **Spontaneous Adoption:** a single susceptible spontaneously becomes and adopter, $\mathcal{S} \rightarrow \mathcal{A}$, with rate γ^2 .

² Adoption could also occur by contagion according to the rule $\mathcal{S} + \mathcal{A} \rightarrow \mathcal{A} + \mathcal{A}$. Investigations into

- (c) **Luddism:** ignorant agents may spontaneously and permanently reject the innovation by becoming a Luddite, $\mathcal{I} \rightarrow \mathcal{L}$, with a rate proportional to the change in density of adopters in their neighbourhood.

The novel Luddism mechanism above incorporates two aspects of negative behaviour towards innovation. Firstly, agents have no predisposition towards the innovation and may choose to oppose the innovation simply due to its sudden increase in popularity. This ingrains a level of *non-conformity* [233, 234] in the population. The second represents a fear of the innovation or concern of its consequences on society as was apparent in the model's namesake, the 18th century Luddism movement, where the introduction of labour-saving machinery caused fear and anger over job security. Such themes are reoccurring in the modern era with the advent of sophisticated artificial intelligence and automation threatening the security of low-skilled labour. To encapsulate this behaviour in the model we define the rate at which the Luddite density increases to be proportional to the adoption rate, with constant of proportionality denoted by r (the Luddism parameter). Where agent interactions are structured (on a network, for example) this interaction is localised such that any agent can only measure the rate of adoption in its neighbourhood. We consider the scheme $r > 0$ such that the rapid introduction and adoption of an innovation will prompt a strong Luddism response.

The LISA model also incorporates a type of social reinforcement mechanism whereby in the multi-stage progression $\mathcal{I} \rightarrow \mathcal{S} \rightarrow \mathcal{A}$ a successful adoption follows a number of prompts from neighbouring agents [220, 222, 235, 232]. It is worth noting that this multi-stage progression is required to produce non-trivial results. With only three states the model produces a polarised community of Luddites and adopters where the proportion of Luddites to adopters is dependent only on r . The need for such a progression is confirmed in other relevant models [236] which find that the negative response to innovation, due to high levels of advertising, cannot be replicated with fewer than four states. In this sense the LISA model represents the most simple generalisation

the two-body mechanisms yielded similar features to the LISA model, however it was more technically difficult to analyse and was hence abandoned in favour of simplicity.

of the Bass model that gives a non-trivial final state where the adoption of the innovation is incomplete.

8.2 Mean-field Analysis

We first consider the LISA model in the mean-field limit, where agents are perfectly mixed. The densities of each type of agent are given by $(L, I, S, A) = (N_{\mathcal{L}}, N_{\mathcal{I}}, N_{\mathcal{S}}, N_{\mathcal{A}})/N$, where N_X is the number of agents of type $X \in \{\mathcal{L}, \mathcal{I}, \mathcal{S}, \mathcal{A}\}$, and N is the total number of agents. We consider the limit $N \rightarrow \infty$, so that all densities are continuous variables and all fluctuations are negligible. In this setting, the evolution of the agent densities is described by the rate equations:

$$\begin{aligned}\dot{L} &= rAI \equiv (\alpha - 1)SI, \\ \dot{I} &= -(1 + \gamma r)SI \equiv -\alpha SI, \\ \dot{S} &= S(I - \gamma), \\ \dot{A} &= \gamma S,\end{aligned}\tag{8.1}$$

where the dot denotes the time derivative and we define $\alpha \equiv 1 + \gamma r$. Since the total density is conserved, i.e. $L + I + S + A = 1$, the sum of these rate equations equals zero. A natural initial condition is a population that consists of a small density of susceptible agents that initiate the dynamics, while all other agents are ignorant; that is, $I(0) = 1 - S(0) = I_0$ and $L(0) = A(0) = 0$.

To solve these rate equations, it is useful to introduce the modified time variable $d\tau = S(t) dt$, which linearise the rate equations to

$$\begin{aligned}L' &= (\alpha - 1)I, \\ I' &= -\alpha I, \\ S' &= I - \gamma, \\ A' &= \gamma,\end{aligned}\tag{8.2}$$

with solution

$$\begin{aligned}
 L &= \frac{\alpha - 1}{\alpha} I_0 (1 - e^{-\alpha\tau}), \\
 I &= I_0 e^{-\alpha\tau}, \\
 S &= \frac{I_0}{\alpha} (1 - e^{-\alpha\tau}) + 1 - I_0 - \gamma\tau, \\
 A &= \gamma\tau.
 \end{aligned} \tag{8.3}$$

There are two basic regimes of behaviour that are controlled by the adoption rate γ , as illustrated in Figure 8.2:

- (a) **Gradual but extensive adoption.** When $\gamma < I_0$, the density of susceptibles S varies non-monotonically in time and reaches a maximum value S_{inc} at an “inception time” t_{inc} , after which S decays to 0. This non-monotonicity leads to sigmoidal curve for the adopter density, with A increasing rapidly for $t \gtrsim t_{\text{inc}}$. The rescaled inception time τ_{inc} is determined by the criterion $S' = 0$, or equivalently, $I(\tau_{\text{inc}}) = \gamma$. This gives

$$\tau_{\text{inc}} = \frac{1}{\alpha} \ln(I_0/\gamma). \tag{8.4}$$

- (b) **Rapid but sparse adoption.** When $\gamma > I_0$, the susceptibles quickly become adopters, leaving behind a substantial static population of ignorants and a small fraction of adopters, as well as Luddites.

Numerical simulations of the LISA model on a large complete graph and numerical integration of the rate equations (8.1), illustrated in Figure 8.2, give results that are virtually indistinguishable.

We can express the densities in terms of the physical time t by inverting $d\tau = S(t) dt$ to give $t = \int_0^\tau d\tau' / S(\tau')$. Substituting $S(\tau)$ from the third of Equations (8.3) and taking the limits of low adoption, $\gamma \ll 1$ and $\alpha \approx 1$, we have³

$$t = \int_0^\tau \frac{d\tau'}{1 - I_0 e^{-\tau'}} \approx \tau + \ln[1 - I_0 e^{-\tau}]. \tag{8.5}$$

³ Here the term $-\gamma\tau'$ has been neglected. This approximation is legitimate since τ' is integrated from 0 to $\tau \ll \tau_\infty \approx 1/\gamma$ and therefore $\gamma\tau' \ll 1$ in the regime being considered. A similar reasoning, with $\gamma\tau \leq \gamma \ln(I_0/\gamma) \ll 1$, leads to (8.6) when $\gamma \ll 1$.

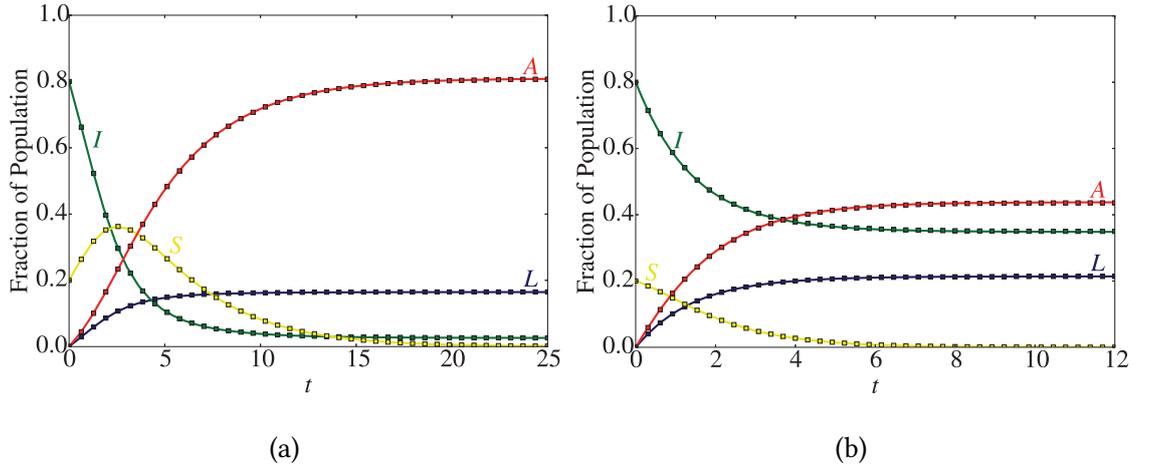


Figure 8.2: Evolution of a realisation of the LISA model on a complete graph of 10^6 nodes with $I_0 = 0.8$ and Luddism parameter $r = 0.9$. (a) $\gamma = 0.3$ (extensive adoption) (b) $\gamma = 1$ (sparse adoption). Evenly distributed samples of the stochastic simulation (\square) are indistinguishable from the solution of Equation (8.1) (solid line). The completion times for (a) and (b) are 60 and 17 respectively.

In particular, the physical inception time t_{inc} is,

$$t_{\text{inc}} \approx \int_0^{\ln(I_0/\gamma)} \frac{d\tau'}{1 - I_0 e^{-\tau'}} \approx \ln \left[\frac{I_0}{(1 - I_0)\gamma} \right] \quad (8.6)$$

and therefore grows as $\ln(1/\gamma)$.

The stationary state is reached when all susceptibles disappear, so that no further reactions can occur. This state occurs at the time τ_∞ for which $S(\tau_\infty) = 0$. This requires solving

$$\frac{I_0}{\alpha}(1 - e^{-\alpha\tau_\infty}) + 1 - I_0 - \gamma\tau_\infty = 0$$

or equivalently

$$e^{-\alpha\tau_\infty} = -\frac{\alpha\lambda}{I_0}\tau_\infty + \frac{\alpha(1 - I_0) + I_0}{I_0}.$$

This equation is of the form

$$p^{ax+b} = cx + d \quad (8.7)$$

which, using the transformation $-t = ax + ad/c$ to give $tp^t = R := -ap^{b-ad/c}/c$, gives $t = W(R \log(p))/\log(p)$ where $W(z)$ is the principal branch of the Lambert function,

defined as the solution of $z = W(z)e^{W(z)}$. The solution of (8.7) is therefore

$$x = \frac{-W\left(-a \log(p)p^{b-\frac{ad}{c}}/c\right)}{a \log(p)} - \frac{d}{c}$$

and upon the substitutions $a = -\alpha$, $b = 0$, $c = -\alpha\gamma/I_0$, and $d = [\alpha(1 - I_0) - I_0]/I_0$ we obtain

$$\tau_\infty = \frac{1}{\gamma} - \frac{I_0 r}{\alpha} + \frac{1}{\alpha} W_0\left(-\frac{I_0}{\gamma} e^{I_0 r - \alpha/\gamma}\right). \quad (8.8)$$

Here τ_∞ is a decreasing function of the adoption rate γ , with $\tau_\infty \sim 1/\gamma$ in the high and low adoption rate regimes.

We now determine the final densities by substituting τ_∞ into Equations (8.3). For a small adoption rate ($\gamma \ll 1$) this gives

$$\begin{aligned} A_\infty &= 1 - \mathcal{O}(\gamma), \\ I_\infty &\rightarrow 0, \\ L_\infty &\approx (\alpha - 1)I_0 = \mathcal{O}(\gamma). \end{aligned}$$

Similarly, the densities at the inception time are obtained by substituting τ_{inc} into Equations (8.3). This yields $A(\tau_{\text{inc}}) + S(\tau_{\text{inc}}) = 1 - [(\alpha - 1)I_0 + \gamma]/\alpha$. Since $(\alpha - 1)I_0 \sim \mathcal{O}(\gamma)$, when $\gamma \ll 1$ and r is finite, here the stationary density of adopters approximately equals the sum of the adopter and susceptible densities at the inception time, $A_\infty \approx A(\tau_{\text{inc}}) + S(\tau_{\text{inc}})$. Hence, in the low adoption rate regime (when r is finite), we can infer the final level of adoption from the adopter and susceptible densities at the inception time, i.e. well before the stationary state.

The dependence of the final densities for different parameter ranges is shown in Figure 8.3. Again simulation results for the complete graph are indistinguishable from numerical integration of the rate equations.

Interestingly, L_∞ varies non-monotonically on γ when the initial state consists mostly of ignorants and the fixed rate of Luddism r is not too high, as in Figure 8.3 (top).

Another observation is that the final fraction of Luddites is non-monotonic in γ . To see this we need to show that

$$\frac{dL_\infty}{d\gamma} = \frac{I_0}{\alpha^2} \left[r(1 - e^{-\alpha\tau_\infty}) + \alpha(\alpha - 1) \left(r\tau_\infty + \alpha \frac{d\tau_\infty}{d\gamma} e^{-\alpha\tau_\infty} \right) \right]$$

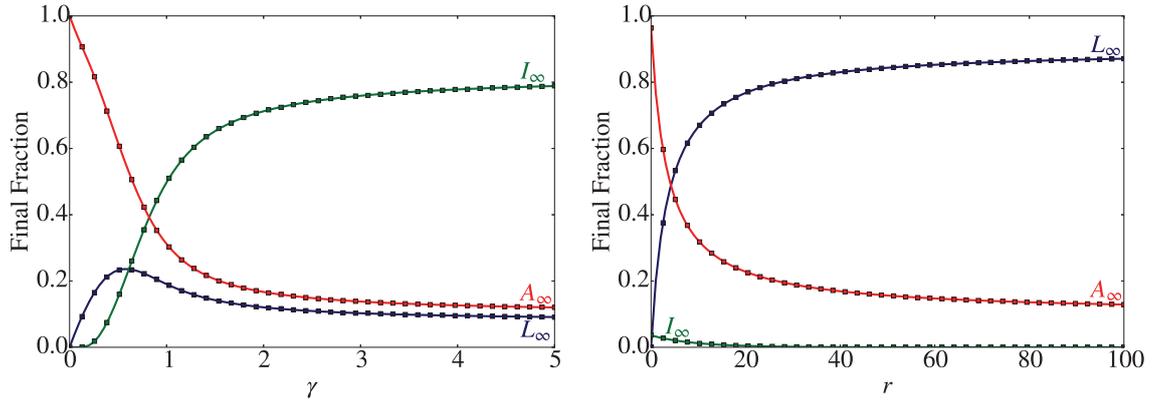


Figure 8.3: Dependences of the final-state densities L_∞ , I_∞ and A_∞ for a complete graph of 10^4 nodes and $I_0 = 0.9$. In the top panel $r = 0.9$ while γ varies, whereas in the bottom panel $\gamma = 0.3$ while r varies. Simulations (\square) in complete agreement with (8.3) with substitution (8.8) (solid line).

takes both positive and negative values for $\gamma > 0$. The derivative $d\tau_\infty/d\gamma$ is given by

$$\frac{d\tau_\infty}{d\gamma} = -\frac{1}{\gamma^2} - \frac{r}{\alpha^2} W(z) + \frac{W(z)}{\alpha z(1+W(z))} + \frac{I_0 r^2}{\alpha^2}$$

where $z = -\frac{I_0}{\gamma} e^{-(\alpha/\gamma) + I_0 r}$. For $r = 0$, $\gamma = 1$ and $I_0 = 1$, z achieves its minimum of $-1/e$ where $W(z)$ is not differentiable. However since we enforce $S_0 > 0$ then $I_0 < 1$ for all time and so $W(z)$ is differentiable in the regime of interest. Clearly $z \leq 0$ which bounds the Lambert function, $-1 < W(z) < 0$. By considering the individual terms it is apparent that

$$\frac{d\tau_\infty}{d\gamma} = -\frac{1}{\gamma^2} + \frac{1}{\alpha^2} \left[\frac{1}{z(1+W(z))} \right] \left[\alpha - rz(1+W(z)) + \frac{I_0 r^2 z(1+W(z))}{W(z)} \right] < 0,$$

which shows that the stationary time τ_∞ is a decreasing function of γ . For both $\gamma \ll 1$ and $\gamma \gg 1$, $\tau_\infty \sim 1/\gamma$ provided that $r \ll 1/\gamma$. Similarly, $d\tau_\infty/d\gamma \sim -1/\gamma^2$ in the same regimes. Consequently, τ_∞ is large when γ is small, and is small when γ is large, remaining finite for intermediate values. This gives

$$\frac{dL_\infty}{d\gamma} \sim \begin{cases} r(1 - e^{-1/\gamma}) & \text{for } \gamma \ll 1 \\ -\frac{e^{-1/\gamma}}{\gamma^2} & \text{for } \gamma \gg 1 \end{cases}$$

and hence there exists a γ^* which satisfies

$$r(1 - e^{-\alpha\tau_\infty}) + \alpha(\alpha - 1) \left(r\tau_\infty + \alpha \frac{d\tau_\infty}{d\gamma} e^{-\alpha\tau_\infty} \right) = 0$$

such that $L_\infty(\gamma^*)$ is a global maximum.

It is also worth noting that in the absence of Luddites, complete adoption is almost, but not completely, achieved since the final densities of adopters and ignorants are $A_\infty \approx 1 - I_\infty$ and $I_\infty \approx e^{-1/\gamma}$, see Figure 8.3 (bottom).

To assess the role of finite- N fluctuations on the dynamics, we simulate the LISA model on complete graphs of N nodes using the Gillespie algorithm [237]. At long times we find that the densities of each species, N_X/N , fluctuates around the corresponding mean-field density with a root-mean-square fluctuation of amplitude $\sim N^{-1/2}$, as expected from general properties of this class of reaction processes [158, 160]. We also find that the probability distribution of N_X/N is a Gaussian of width of order $N^{-1/2}$ that is centred on the mean-field density. We also estimate the completion time T_C for the system to reach its final state by the physical criterion that $S(t = T_C) = 1/N$. That is, completion is defined by the presence of a single susceptible remaining in the population [222, 235, 232]. Linearising the rate equations (8.1) with $S = 0 + \epsilon\hat{S}$ and $I = I_\infty + \epsilon\hat{I}$ gives

$$\epsilon\dot{\hat{S}} = \epsilon(I_\infty - \gamma)\hat{S} + \mathcal{O}(\epsilon^2).$$

Hence the density of susceptibles vanishes as $S(t) \sim e^{-(\gamma - I_\infty)t}$. For $t > t_{\text{inc}}$ we know that $\gamma > I(t)$. The mean completion time can then be estimated as $T_C \approx \ln(N)/(\gamma - I_\infty)$. This approximation highlights two features of the model; the time for the adoption to spread scales like $\ln(N)$, and the time for the system to reach its stationary state is fast when γ is large, and slow when γ is small. This gives a trade-off between an adoption which is slow and widespread, or an adoption which penetrates the system quickly but results in a much smaller uptake.

8.3 Random Graphs and Lattices

The mean-field analysis of the model is insightful, however real social systems are rarely well-mixed. We now consider how the LISA model behaves on different topologies including Erdős-Rényi random graphs [168] and one-dimensional lattices. In particular we look to uncover, quantify, and explain genuine non mean-field effects.

A graph with N nodes is best represented by the $N \times N$ adjacency matrix $\mathbf{A} = [A_{ij}]$, where $A_{ij} = 1$ if nodes are connected and 0 otherwise. The graphs that we consider in this section are undirected meaning edges are reciprocal, i.e. $A_{ij} = 1 \iff A_{ji} = 1$. To implement the LISA model on a graph we use a modified version of the Gillespie algorithm [237]. This approach considers the propensity for each node to participate in any of the transformations $\mathcal{I} \rightarrow \mathcal{L}$, $\mathcal{I} \rightarrow \mathcal{S}$, $\mathcal{S} \rightarrow \mathcal{A}$, resulting in $3N$ possible updates to the system at any given point.

The propensity for a susceptible to adopt is γ , independent of the state of the local environment. The propensity for an ignorant node i to become a susceptible is s_i/N where s_i is the number of susceptible neighbours of i . Finally the propensity for an ignorant node i to become a Luddite is $r\gamma s_i/k_i$ where $k_i = \sum_j A_{ij}$ is the degree of node i . The propensity of i to become a Luddite is thus proportional to the total propensity of its neighbours propensities to adopt at any given time. In this sense we encode node i 's knowledge of the local adoption rate. Trivially the propensity for any other state change is zero.

8.3.1 Erdős-Rényi Random Graphs

We first study the dynamics of the LISA model on the class of Erdős-Rényi (ER) graphs. For an ER graph an edge is formed between any two nodes with probability p , resulting in a binomial degree distribution in which each node has on average $k = p(N - 1)$ neighbours [7, 238]. We make the assumption that there are no correlations between the degrees of neighbouring nodes⁴. Under this assumption the adjacency matrix can be written as $A_{ij} \approx k_i k_j / Nk \approx k/N$. This means that the dynamics of the LISA model on ER graphs can be approximately described by a suitable generalisation of the mean-field theory, namely that connections are still all-to-all but are weighted by a factor $k/N < 1$. In particular if we consider the model from an individual-based perspective and let S_i be the probability that a node i is susceptible and let I_j be the probability that a node j

⁴ For any graph with heterogeneous degree distribution nodes of any degree are, on average, connected to nodes of higher degree.

is ignorant, then the density of susceptibles S evolves as

$$\dot{S}_i = S_i \left[\sum_j (A_{ij}/N) I_j - \gamma \right] \approx S [(k/N)I - \gamma],$$

since in this approximation all nodes are homogeneous ($S_i = S$) and interact with k/N neighbours on average.

As the only mechanism which involves two nodes, the two-body contagion $\mathcal{I} + \mathcal{S} \rightarrow \mathcal{S} + \mathcal{S}$ has an effective rate of k/N (which was previously 1 in the mean-field model). The rates for the Luddism and adoption mechanisms remain unchanged. Intuitively the effective rate equations become

$$\begin{aligned} \dot{L} &= \gamma r S I \equiv \left(\beta - \frac{k}{N} \right) S I, \\ \dot{I} &= - \left(\gamma r + \frac{k}{N} \right) S I \equiv -\beta S I, \\ \dot{S} &= S \left(\frac{k}{N} I - \gamma \right), \\ \dot{A} &= \gamma S, \end{aligned} \tag{8.9}$$

where we define $\beta = \gamma r + (k/N)$ for convenience.

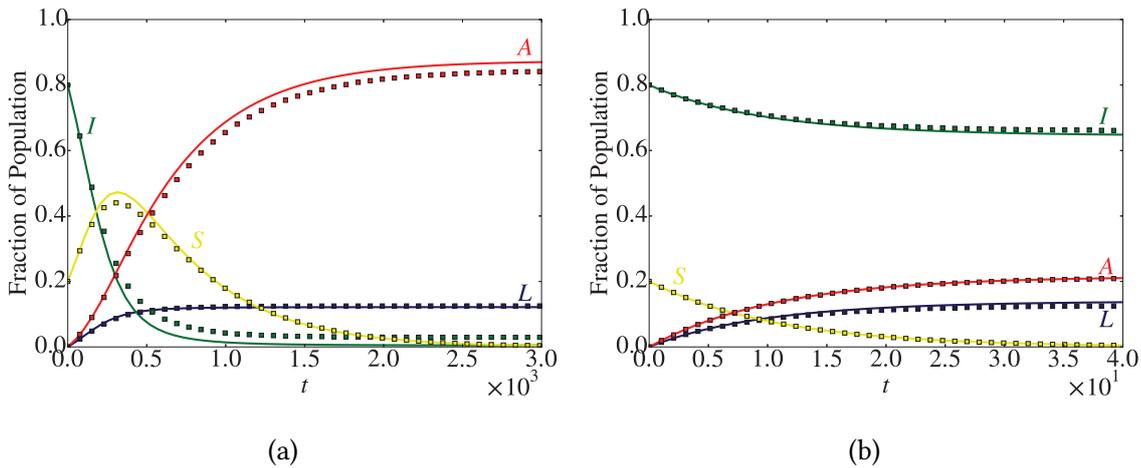


Figure 8.4: The evolution, averaged over 100 realisations, of the LISA model on an ER graph with $N = 10^3$ nodes, $k = 10$, and $I_0 = 0.8$. (a) $\gamma = 0.002$, such that $\gamma < (k/N)I_0$ and (b) $\gamma = 0.1$ such that $\gamma > (k/N)I_0$. Shown are the evenly distributed samples of the stochastic simulation (\square) and the solution of Equation (8.9) (solid line). The Luddism parameter $r = 0.9$.

As with the mean-field model, the rate equations predict two regimes of behaviour (shown in Figure 8.4):

- (a) **Slow but extensive adoption** ($\gamma < kI_0/N$). Here the density of S s peaks at an inception time $t_{\text{inc}} \sim \ln(1/\gamma)$ before vanishing. The densities of adopters and Luddites also display the characteristic sigmoidal dependence on time.
- (b) **Rapid but sparse adoption** ($\gamma > kI_0/N$). The density of S s vanishes quickly so that the density of adopters and Luddites quickly reach their steady-state values.

However, the critical value of γ between the two regimes, γ^* , is now $kI_0/N < I_0$. This suggests a lower tolerance for adoption as the local neighbourhood of a node gets smaller.

The simulation results presented in Figure 8.4 show that for $k/N = 0.1$ the mean-field approximation of (8.9) captures the main qualitative features of the model, and correctly predicts the final state of the system with reasonable accuracy. The scale of the t -axis indicates that extensive adoption (a) happens multiple orders of time slower than the rapid adoption (b), and in both cases the adoption rate is slower than in the mean-field model (Fig. 8.2).

The stationary state can be determined in the same fashion as in Section 8.2, noting that (8.9) becomes linear under the change of variable $\tau = \int_0^t S(t') dt'$. Explicitly the steady state is given by

$$\begin{aligned} I_\infty &= I_0 e^{-\beta\tau_\infty} \\ L_\infty &= \frac{\beta - k/N}{\beta} (I_0 - I_\infty) \\ A_\infty &= \gamma\tau_\infty, \end{aligned} \tag{8.10}$$

where now

$$\tau_\infty = \frac{k}{N\gamma\beta} + (1 - I_0)\frac{r}{\beta} + \frac{1}{\beta}W_0\left(-\frac{kI_0}{N\gamma}e^{-(1-I_0)r-k/(N\gamma)}\right). \tag{8.11}$$

Figure 8.5 shows the simulation results for the stationary densities as a function of the mean degree, compared to the approximate predictions of (8.10). The mean-field

predictions correctly capture the qualitative behaviour of the functional dependence of the station densities on k . The predictions are only quantitatively accurate when k/N is sufficiently large. When $k/N \ll 1$ the neighbourhood of a node is small and only represents a fraction of the entire graph, resulting in large demographic fluctuations which invalidate the underlying assumptions of (8.10). The average degree has a minimal effect on the stationary density of Luddites although this relationship is not monotonic. By contrast the adopter density is an increasing function of k , suggesting that a well connected population helps spread innovation and reduce the level of negative response. The model dependence on γ and r are qualitatively similar to that on the complete graph.

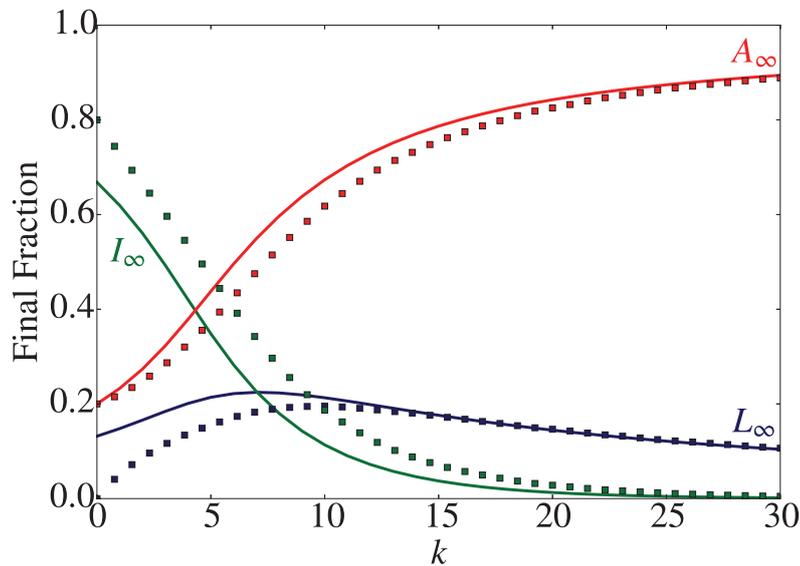


Figure 8.5: Dependence of the final densities L_∞ , I_∞ and A_∞ on the average degree for ER graphs with $N = 10^3$ nodes. The simulation (\square) represents an average over 40 model realisations for 30 randomly generated networks. Parameters are $\gamma = 0.005$, $r = 0.9$, and $I_0 = 0.9$. The mean-field predictions (8.10) (solid line) match the simulation for $k \gtrsim 20$ (see main text).

One heuristic to assess the demographic fluctuations of the system is to view the ER graph of mean degree k as a meta-population of N/k patches of well-mixed populations of size k . Under this heuristic, when $N \gg k \gg 1$ the number of agents in each component fluctuates like $k^{1/2}$ about the average value. Assuming these components

are independent (and hence the fluctuations are also independent) the total noise in the full population should have amplitude $\sim (N/k)^{1/2}k^{1/2} = N^{1/2}$, leading to fluctuations in the densities to the order of $N^{-1/2}$. This prediction is confirmed by our simulations (see Figure 8.6) which find that $N_A(\infty)/N$ has a Gaussian probability distribution over an ensemble of simulations, with mean A_∞ and a width that scales as $N^{-1/2}$. Similar distributions are found for L_∞ and I_∞ however not for S_∞ given that $S_\infty = 0$ is the condition required for the dynamics to cease.

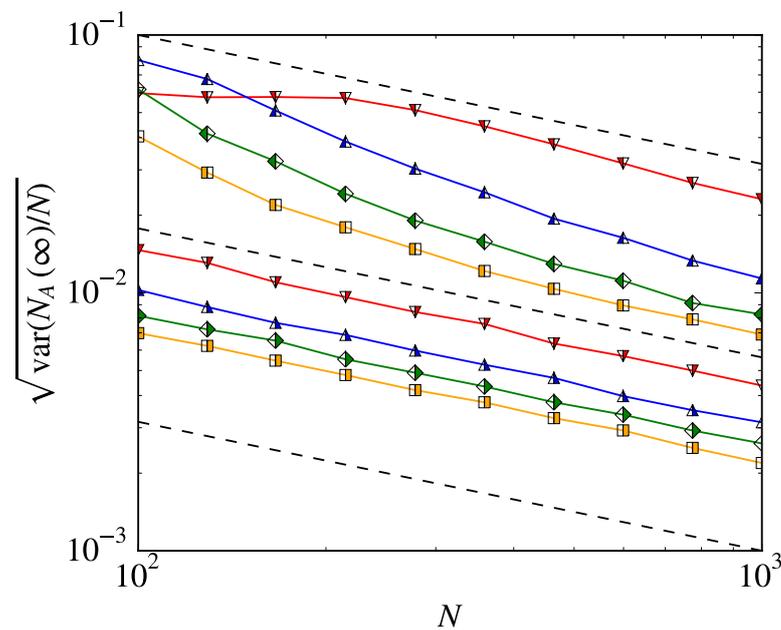


Figure 8.6: Variance in the stationary state of the density of adopters, $N_A(\infty)/N$, obtained over an ensemble of 40 realisations of the model on 75 randomly generated networks for $N = 100 - 1000$. The figure shows the standard deviation around the mean value A_∞ for $k/N = 0.01/0.1$ (left/right filled ∇), $k/N = 0.02/0.2$ (left/right filled \triangle), $k/N = 0.03/0.3$ (left/right filled \diamond), and $k/N = 0.04/0.4$ (left/right filled \square). As guides for the eyes, the dashed lines are $\propto N^{-0.5}$. The parameters used are $\gamma = 0.005$, $r = 0.9$, and $I_0 = 0.9$.

Finally it is useful to understand the model outcomes (particularly the steady state predictions (8.10)) from a marketing perspective, in particular determining whether the spread of the innovation was deemed successful. There are six possible outcomes labelled with Roman numerals in Figure 8.7, partitioning the (γ, r) parameter space,

with the boundaries defined by the level sets satisfying $A_\infty = I_\infty$, $L_\infty = I_\infty$, and $L_\infty = A_\infty$. In this context, the most desirable outcome is in region (I) where the adopter population forms the largest group in the stationary state, and the Luddite population is the smallest. This leaves a significant proportion of ignorants in the population who could potentially be targeted for subsequent marketing strategies or the spread of a new similar innovation. Region (II) can be deemed as a ‘controversial success’. The largest population is still the adopters, however the second largest group are the Luddites - the innovation has spread to a majority but has been divisive and polarised the entire population. The remaining scenarios correspond to failures of varying degree; from not reaching a majority of the population (V, VI), or by having a significant negative reaction which eclipses the adoption of the innovation (III, IV).

In summary, we have shown that the dynamics and stationary state of the LISA model can be approximated accurately using mean-field assumptions provided that the size of a neighbourhood around a node is a sufficiently large fraction of the graph.

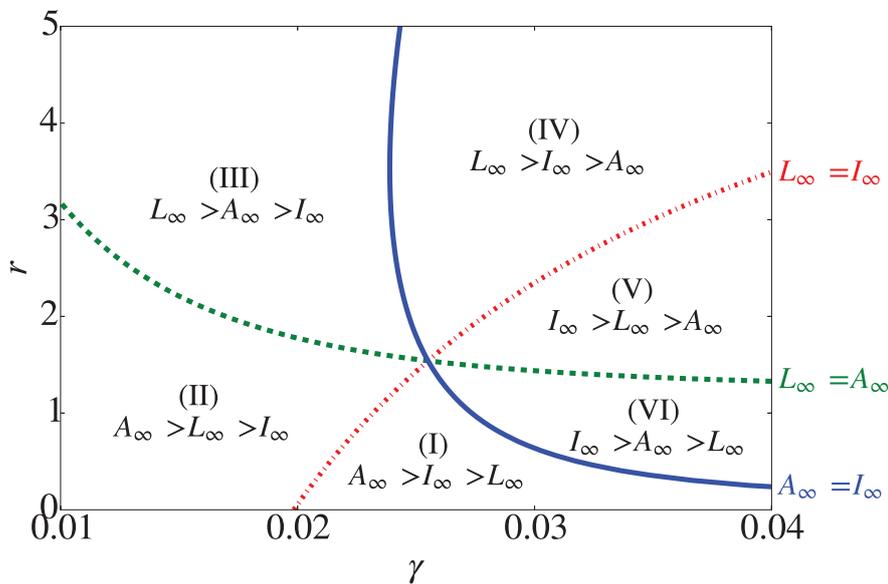


Figure 8.7: The mean-field steady state predictions (8.10) over the parameter space (γ, r) for $k/N = 0.025$ and $I_0 = 0.9$. The contours $L_\infty = I_\infty$, $L_\infty = A_\infty$, and $I_\infty = A_\infty$ split the domain into six regions which characterise the innovation (see main text).

8.3.2 One-dimensional Lattices

Recent controlled experiments have shown that innovation may spread more efficiently on clustered graphs and lattices than on random networks [225]. To understand the effect of regular and restrictive topology on the spread of an innovation, and where the mean-field approximation breaks down, we investigate the LISA dynamics on one-dimensional lattices. In particular we are interested in the ‘barriers to innovation’ that Luddites pose and the ability of Luddites to block crucial paths through the graph which allow the innovation to spread.

We consider a one-dimensional lattice of N nodes with periodic boundary conditions. In this setup, nodes are homogeneous with each node having two neighbours. Applying

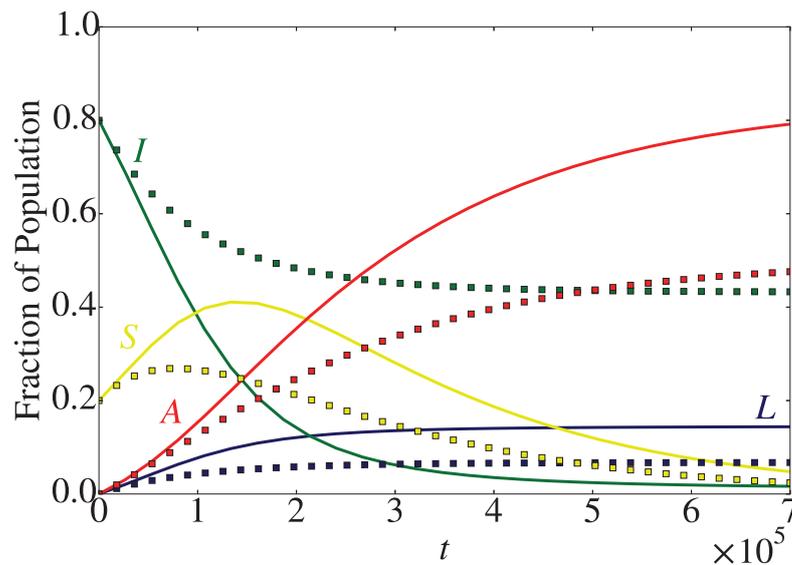


Figure 8.8: Time dependence of the densities in each state for a one-dimensional lattice of size $N = 10^5$ averaged over 100 realisations. The corresponding mean-field predictions from Equation (8.9) with $k = 2$ (solid line) deviate dramatically from the simulation samples (\square). The parameters are $\gamma = 0.005$, $r = 0.9$, and $I_0 = 0.8$.

the mean-field theory used in Section 8.3.1 to the LISA model on one-dimensional lattices provides a poor approximation to the results of simulations (Figure 8.8). The mean-field approximation systematically deviates from the simulation in that it will always overestimate the density of adopters and Luddites at any point in time, and conversely always underestimate the ignorant density. We expect the mean-field

predictions to perform poorly as our underlying assumption of a locally well-mixed population is incorrect but also because of the constraints on paths through the graph due to the restrictive topology of the lattice.

Despite quantitative inaccuracy we still observe two regimes of behaviour in the mean-field approximation. Specifically for $k = 2$ we observe slow and complete adoption for $\gamma < (2/N)I_0$ and rapid but restricted adoption for $\gamma > (2/N)I_0$. From simulations, illustrated in Figure 8.9, we observe the following three regimes:

- (A) When $\gamma \ll 2I_0/N$, there is slow adoption as well as a time-scale separation. First, almost all \mathcal{I} s are converted to \mathcal{S} 's [239, 159, 240] in a time of the order of N^2 . When the lattice consists almost entirely of \mathcal{S} 's, these become adopters after a mean time of the order of γ^{-1} . As a consequence, when $\gamma \ll N^{-1}$ the size of the adopter domains grows abruptly after a time of order $\sim N^2 + \gamma^{-1}$, when all ignorants have disappeared and the entire lattice is covered with adopters.
- (B) When $\gamma \sim 2I_0/N$, the domains of adopters grow initially almost linearly in time, whereas the average size of \mathcal{I} clusters remains approximately constant and of a comparable size to \mathcal{A} domains.
- (C) When $\gamma \gg (2/N)I_0$, adoption occurs quickly and the final state is reached in a time of order $\mathcal{O}(1/\gamma)$. The final adopter density is limited by the formation of Luddites at the ends of ignorant domains which prevent further conversion within each domain.

To study the densities of each population in the stationary state we require a different approach to model the evolution of the system.

Analysis of ignorant domains

Initially, the nodes on the one-dimensional lattice are either ignorant, with probability I_0 , or susceptible, with probability $S_0 = 1 - I_0$. Thus the initial configuration consists of connected domains of ignorant nodes bordered by susceptibles. Moreover, since ignorants can only become susceptible if a neighbour is susceptible, domains of

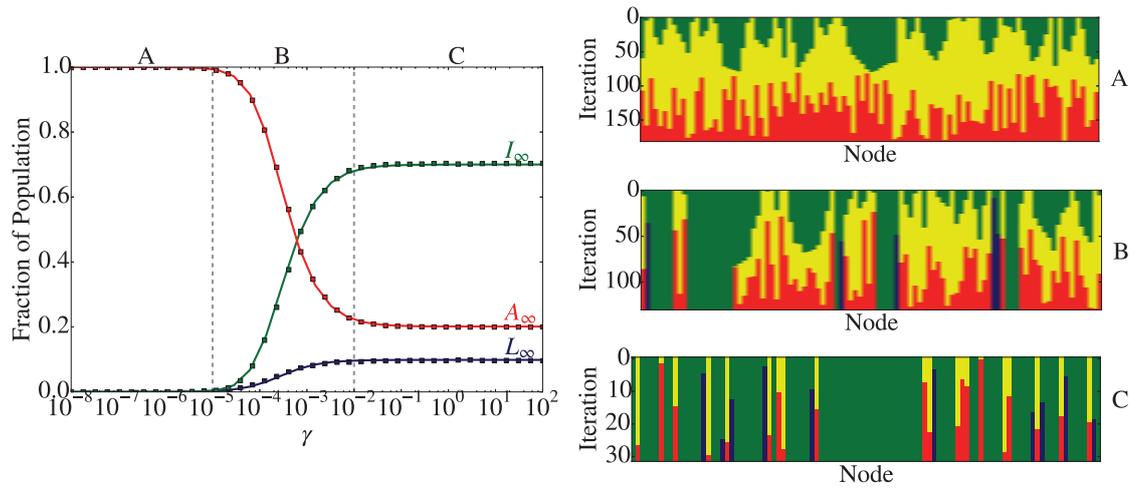


Figure 8.9: Final simulated average proportions of adopters (red/gray \square), ignorants (green/dark gray \square) and Luddites (blue/black \square) for varying values of γ , averaged over 100 simulations. Theoretical predictions using ignorant domain length are overlaid (solid line). Parameters are $N = 1000$, $r = 0.5$. Initially ignorants and susceptibles are randomly distributed, with densities $I_0 = 0.8$ and $S_0 = 0.2$. The three regimes discussed in the text are separated by dashed lines corresponding to regions where $(2/N)I_0 \ll \gamma$ and $(2/N)I_0 \gg \gamma$. Typical realisations of the model for $N = 100$ in each of the three regimes are given (right). On the vertical axis the iteration corresponds to a single step of the Gillespie algorithm, with one reaction taking place per iteration.

ignorants only evolve at their ignorant-susceptible interfaces. We will refer to these as “active interfaces”. At an active interface one of three events can occur:

- The ignorant node becomes susceptible, thus reducing the domain length by one, with probability

$$p_S = \frac{1/N}{1/N + r\gamma/2 + \gamma}.$$

- The ignorant node becomes a Luddite, thus reducing the length of the domain by one and causing the interface to become inactive, with probability

$$p_L = \frac{r\gamma/2}{1/N + r\gamma/2 + \gamma}.$$

- The susceptible node becomes an adopter, thereby terminating the interface

evolution, with probability

$$p_A = \frac{\gamma}{1/N + r\gamma/2 + \gamma}.$$

For an isolated ignorant node with two susceptible neighbours, these probabilities respectively become

$$\begin{aligned}\hat{p}_S &= \frac{2/N}{2/N + r\gamma + \gamma}, \\ \hat{p}_L &= \frac{r\gamma}{2/N + r\gamma + \gamma}, \\ \hat{p}_A &= \frac{\gamma}{2/N + r\gamma + \gamma}.\end{aligned}$$

Let $Q_n(m)$ be the probability that a domain of ignorants of initial length n with a *single* ignorant-susceptible interface has a *final* length $n - m$, with $0 \leq m \leq n$. We can determine $Q_n(m)$ as follows: if the final length of ignorants is $n - m$, with $0 < m < n$, then either m ignorant nodes must become susceptible before a susceptible node at the interface adopts, or $m - 1$ ignorant nodes must become susceptible before an ignorant node at the interface becomes a Luddite. These events occur with probabilities $p_A p_S^m$ and $p_L p_S^{m-1}$ respectively. Using similar reasoning for the cases $m = 0$ and $m = n$, we thus find

$$Q_n(m) = \begin{cases} p_A & \text{if } m = 0 \\ p_A p_S^m + p_L p_S^{m-1} & \text{if } 0 < m < n \\ p_S^n + p_L p_S^{n-1} & \text{if } m = n \end{cases} \quad (8.12)$$

By summing over m , it can be shown that $Q_n(m)$ is normalised.

We now consider the case where a connected region of n ignorant nodes initially has two ignorant-susceptible interfaces. The probability $P_n(m)$ that a region of ignorants of initial length n with two active interfaces has final length $n - m$ is given by the recursion relation

$$\begin{aligned}P_n(m) &= Q_n(m)p_A + Q_{n-1}(m-1)p_L \\ &\quad + P_{n-1}(m-1)p_S,\end{aligned} \quad (8.13)$$

where the terms $Q_n(m)$ are given by (8.12). Equation (8.13) captures the three possible events that can occur at the interface. If a susceptible node at the interface adopts, which

occurs with probability p_A , then the region of ignorants only has one remaining active interface left and there will be $n - m$ remaining ignorants with probability $Q_n(m)$, as given in (8.12). If an ignorant node at the interface becomes a Luddite, which occurs with probability p_L , then again the region of ignorants will only have one active interface. Since there will be one ignorant less the probability there will be $n - m$ remaining ignorants is $Q_{n-1}(m - 1)$. Finally, if an ignorant node at the boundary becomes susceptible, which occurs with probability p_S , then the probability that there are $n - m$ ignorants remaining is the same as if we had started with $n - 1$ ignorant nodes, i.e. $P_{n-1}(m - 1)$.

To solve the recursion relation (8.13) we need $P_n(0)$ and $P_1(1)$. The probability that a region of ignorants of initial length n remains of length n is given by

$$P_n(0) = \begin{cases} p_A \hat{p}_A & \text{if } n = 1 \\ p_A^2 & \text{if } n > 1 \end{cases}.$$

Also, the probability that a single ignorant node that initially has two susceptible neighbours becomes a susceptible or Luddite is given by

$$P_1(1) = \hat{p}_A(p_L + p_S) + \hat{p}_L + \hat{p}_S.$$

Thus the solution to the recursion relation (8.13) for $0 < m < n - 1$ is given by

$$P_n(m) = (m + 1)p_A^2 p_S^m + 2mp_A p_L p_S^{m-1} + (m - 1)p_L^2 p_S^{m-2}.$$

For $m = n - 1$ we have

$$P_n(n - 1) = p_A [\hat{p}_A + (n - 1)p_A] p_S^{n-1} + 2(n - 1)p_A p_L p_S^{n-2} + (n - 2)p_L^2 p_S^{n-3},$$

and for $m = n$ we have

$$P_n(n) = [\hat{p}_A(p_L + p_S) + \hat{p}_L + \hat{p}_S] p_S^{n-1} + (n - 1) (p_A p_S^n + 2p_L p_S^{n-1} + p_L^2 p_S^{n-2}).$$

Again it is possible to check, by summing (8.13) over m and solving the resulting recursion relation, that $P_n(m)$ is normalised.

We can use $P_n(m)$ to calculate the expected final length of ignorant domains $\langle x \rangle$. First note that since I_0 is the initial probability of being ignorant, the probability that a domain of ignorants initially has length $n > 0$ is given by $p_0(n) = I_0^{n-1} S_0$ for large N . Thus we find that

$$\langle x \rangle = \sum_{n=0}^N n p_0(n) - \sum_{n=0}^N p_0(n) \sum_{l=0}^n l P_n(l).$$

In principle, we may use the above to obtain an explicit expression for $\langle x \rangle$. In practice, however, we use the solutions to (8.13) to calculate $\langle x \rangle$ numerically.

Calculation of population densities

Initially, the mean number of ignorants is given by $I_0 N$ and so dividing by the mean length of ignorant domains, $1/(1 - I_0)$, yields the expected number of ignorant domains, $(1 - I_0) I_0 N$. Thus the final density of ignorants is

$$I_\infty = (1 - I_0) I_0 \langle x \rangle.$$

The probability that an ignorant domain survives is

$$q = 1 - \sum_{n=0}^{\infty} p_0(n) P_n(n).$$

Surviving ignorant domains have two interfaces, which are either ignorant-adopter or ignorant-Luddite, with probabilities $p_A/(p_L + p_A)$ and $p_L/(p_L + p_A)$, respectively. Thus the expected number of Luddites at the interfaces of non-vanishing ignorant domains is given by

$$\eta_+ = \frac{2p_L}{p_L + p_A} q (1 - I_0) I_0 N. \quad (8.14)$$

It is also possible for Luddites to arise when a domain vanishes. By identifying the terms in $P_n(n)$ that result in Luddites, it is possible to determine that the expected number of Luddites that arise when a domain of initial size $n > 1$ vanishes is given by

$$l_n = (\hat{p}_{APL} + \hat{p}_L) p_S^{n-1} + (n - 1) (2p_L p_S^{n-1} + p_L^2 p_S^{n-2})$$

and $l_1 = \hat{p}_{APL} + \hat{p}_L$. Thus the expected number of Luddites that arise from domains of ignorants that vanish is

$$\eta_0 = (1 - I_0)I_0N \sum_{n=0}^{\infty} p_0(n)l_n. \quad (8.15)$$

Summing equations (8.14) and (8.15) and dividing by N we arrive at the final density of Luddites

$$L_{\infty} = I_0(1 - I_0) \left(\frac{2p_L}{p_L + p_A}q + \sum_{n=0}^{\infty} p_0(n)l_n \right).$$

Since the dynamics cease when $S = 0$, the number of adopters can be found using the conservation law $A_{\infty} = 1 - L_{\infty} - I_{\infty}$. These results are plotted as a solid line in Figure 8.9 and show excellent agreement with the simulation results when N is large enough that fluctuations are negligible.

By conducting a one-dimensional analysis of the ignorant domains we successfully recovered the stationary state densities of the LISA model on the lattice, where the mean-field approximation broke down. As Figure 8.9 shows, the Luddites act as blockades to innovation when the topology is restricted preventing the possibility of ignorant nodes becoming susceptible to the innovation. A similar effect was observed in two-dimensional lattice structures however the effect was not as pronounced as in one dimension.

8.4 Discussion

In this chapter we introduced the LISA model of innovation diffusion. This model is among the simplest non-trivial extensions to the Bass model which does not lead to complete adoption of the innovation. The main addition to the model is the presence of Luddites who permanently oppose the innovation in response to the rapid proliferation of the innovation within their neighbourhood. This novel Luddism mechanism incorporates the rate of change of neighbouring states rather than the states themselves. The dynamics has two regimes; quick and partial adoption of innovation and slow, broader adoption. The two regimes are dependent on the initial density of

ignorant agents, the size of each agent's local neighbourhood, and the rate of adoption (or marketing pressure).

We also conducted a detailed analysis of the model on complete graphs, Erdős-Rényi random graphs, and on one-dimensional lattices. For complete graphs and ER random graphs we showed that simple mean-field approximations help significantly illuminate the model's features and predict the stationary state densities well for suitable parameters. For one-dimensional lattices, careful considerations of the populations of ignorant agents allowed us to successfully predict the stationary state densities of each agent type. We saw that on restrictive network topologies (in particular the one-dimensional lattice), that Luddites can act as barriers to innovation and can potentially block the innovation from spreading to parts of the network.

The model aims to replicate the fact that innovations are often met with controversy or fear, regardless of their benefit to society as a whole, such as the invention of labour-saving machinery or new vaccinations. Consequently full adoption of an innovation is rarely observed. The inclusion of this effect, and the application of the model on different topologies help explain how innovation may spread through a society that is resistant to change.

This type of model, and in particular, diagrams such as Figure 8.7 may help marketers in informing their strategy for a possible product launch. A large advertising campaign and rapid adoption may disgruntle potential adopters and potentially polarise the population, whereas a targeted and less forceful campaign may lead to a higher adopting percentage over time. The level of marketing also comes with some cost and so the problem could also be phrased as an optimisation problem. The model needs to be calibrated and validated using real data, which first involves being able to classify agents into one of \mathcal{L} , \mathcal{I} , \mathcal{S} , or \mathcal{A} before observing these agents over the course of an innovation launch. Ultimately this model may be used to forecast adoption levels based on early telemetry, or by using calibrated parameters for similar innovations.

One extension to the model would be to include an evolving or temporal graph structure which may help or hinder the spread of innovation. Another application would be to use content posted to social media and analyse the sentiment surrounding new innovations

to infer which states ($\mathcal{L}, \mathcal{I}, \mathcal{S}, \mathcal{A}$) users belong to, and track the rates of change between the states over time.

9

Summary and Outlook

9.1 Summary

Modelling social networks as temporal networks is a difficult task. However, as we have seen in this thesis the inclusion of temporal information can lead to differences in our conclusions compared to considering static networks alone. Returning our original thesis, we have shown that by considering the temporal network as a sequence of events (without any temporal aggregation) we are able to develop algorithms and representations which allow us to identify key nodes and decompose the network into natural components. Consequently this means that researchers and advertisers alike are able to analyse the ‘importance’ nodes and components of the network effectively.

In Chapter 3 we analysed a state-of-the-art centrality measure, communicability, which is currently used in industry to find influential users. We found that by aggregating

the temporal network into a sequence of static networks we are unable to freely choose the weighting parameter of the centrality and, depending on the level of temporal aggregation, the measure had different interpretations. This meant that it was difficult to attach a meaning to the centrality and maintain consistency across studies. Our introduction of a new efficient algorithm to calculate communicability centrality removes any restriction on parameter selection, avoids ambiguity in what the algorithm calculates while also being less computationally complex than current algorithms.

Continuing to model social networks as a sequence of temporal events we introduced a new network representation in Chapter 4 called the temporal event graph (TEG). The TEG builds upon previous event-based network representations and the concept of temporal motifs. Unlike other representations, the TEG uniquely defines a temporal network, meaning that no information is lost when considering temporal networks in this way. The TEG provides a unique perspective on temporal networks as they are described in terms of the behaviour of nodes (inter-event times and motifs), rather than the events themselves. Using the TEG we are able to naturally decompose the temporal network into smaller components and assess the structure and composition of the network through the temporal barcode and the statistical properties of each component.

In Chapter 5 we saw the impact of the TEG when applied to data collected from the social network, Twitter. Here we were able to highlight the differences in structure across different temporal networks, assess the temporal dependence of each network, and classify the behaviour within conversations, as well as identifying the conversation topic. This meant we were easily able to identify spam accounts and find users more likely to get involved in discussion.

The latter part of this thesis was devoted to mathematical models of social systems.

In Chapter 7 we introduced the $2q$ -voter model with zealotry ($2qVZ$) [2, 1]. The model added to the work on the the popular voter model (and variants) by introducing a heterogeneous population of voters with varying resolution in their opinion. The $2qVZ$ is one of the first non-equilibrium models of social systems for which we were able to characterise the non-equilibrium steady state (NESS) by using a linear Gaussian

approximation. In general, there is no simple way to calculate the NESS, however in this case the linear Gaussian approximation provided a suitable method and we were able to quantify the probability currents present at stationarity. From a social perspective these currents suggest the presence of ‘leaders’ and ‘followers’ in the population where changes in the average population opinion are driven by shifts in the opinion of the leader population.

In an attempt to understand the spread of innovation on social networks, and the influence of nodes opposed to innovation we introduced the LISA model [3] in Chapter 8. Extending the popular Bass model, we introduced a novel network mechanism where node states could change depending on the rate of change of the states of neighbouring nodes. The mechanism was used to model anti-bandwagon behaviour where nodes may oppose an innovation if the local rate of adoption is too quick. The model showed that rapid forcing of adoption can lead to incomplete adoption and a polarised population, especially on restrictive network topologies.

9.2 Outlook

In this section we look at the wider impact of this research and further research avenues.

The TEG provides a general framework to study temporal networks. While in this thesis we have focused on social networks, the TEG can also be used to study a sequence of temporal events where the relationships between events can be characterised. Possible other use cases include other digital communication, user interaction patterns with websites, mobility data gathered from mobile phones, and trades or transactions from financial institutions or online marketplaces.

There is also much that can be done to refine the statistical analysis of the TEG, for example, identifying the features of temporal components which provide the largest variance for classification problems. There are a number of interesting mathematical questions we can ask about the TEG. When we considered the random temporal network in Chapter 4, the induced aggregate graphs of the temporal components were random in nature but due to the complexity of the Δt -connected condition, are not Erdős-Rényi

random graphs. A suitable question is how can we characterise these graphs and what is their dependence on the parameter Δt ? Similarly we can ask what the structure of the TEG looks like with a temporal network null model. Unlike static networks where a number of random reference models have been proposed there is not a temporal network model that has been universally adopted, with studies instead using shuffled and time-reversed versions of the original network. The TEG potentially offers a way to generate temporal networks with a prescribed distribution of motifs. Generating these networks is challenging however; two-event motifs can be generated with a single chain of events. The difficulty lies with generating these networks with the correct number of nodes and node activity levels.

The TEG may also provide efficiency savings for algorithms where the time between events is critical, in the running dynamic communicability centrality, for example. As the communicability matrix decays exponentially with time, one may impose a hard cut-off after some time. This would allow the communicability centrality to be calculated on a component-by-component basis using the Δt -TEG, where Δt is appropriately chosen. Similarly the dynamics of a process on the temporal network (like an epidemic model) may be restricted to the components of the Δt -TEG which allow the original problem to be decomposed into smaller and simpler problems.

For the remainder of this section, we discuss in more depth two more potential uses of the TEG.

9.2.1 Using the Temporal Event Graph to Study Dynamical Processes on Networks

An aspect of the TEG not studied in this thesis is the inter-event time (IET) distribution conditional on the motif type, i.e. given that the relationship between two events is of a particular motif (say ABBC), what is the distribution of times between such events? One can hypothesise that the timescales for passing on or broadcasting a message (ABBC and ABAC respectively) may be different from replying to a message (ABBA) which would require time to compose a response. This concept can also be generalised to

larger motifs where one could consider the IETs of multiple event pairs in the motif. For example, we can study successive IET times between two nodes sending messages in turn (the ABBAAB motif using previous notation) and ask whether successive IETs are independent, and if not, find their dependence.

A novel source of temporal network data is from the recording of dynamical processes on networks. To help illustrate this we use the susceptible-infected (SI) model on a static network. During the evolution of the model we record the events “A has infected B at time t ”. This results in a sequence of temporal events which follow the spread of infection. Now we have captured the process evolution as a temporal network, we can apply a range of temporal network methods including the TEG. The motif distribution of the TEG will be characterised by containing a mixture of only the ABAC and ABBC motif as once a node is infected it can not be reinfected. Furthermore, the TEG will consist of as many components as there are initially infected nodes.

If we instead consider a model where nodes can recover from infection after some time like the susceptible-infected-susceptible (SIS) model, then there is the possibility of all motifs occurring. What characterises this process is the motif-conditional IETs. The ABBC and ABAC motifs will still have exponentially distributed IETs however the remaining four motifs will have an IET distribution that is dependent on both the infection rate and recovery mechanism, as seen in Figure 9.1. Combining the motif and IET information it may be possible to recover both the mechanisms and parameters of the model, or at the very least distinguish the dynamics from other processes.

In this thesis we have considered both ways to monitor and model temporal networks. The extraction of temporal networks from dynamical processes can unite these two aspects and offer new tools to analyse and classify dynamical processes.

9.2.2 Network Filtering with the Temporal Event Graph

In the case studies of Chapter 5 we saw how the selection of messages by keyword presented a bias towards a certain behaviour (retweeting). As data is collected by sampling tweets which contain a particular keyword (or keywords) if a tweet contains

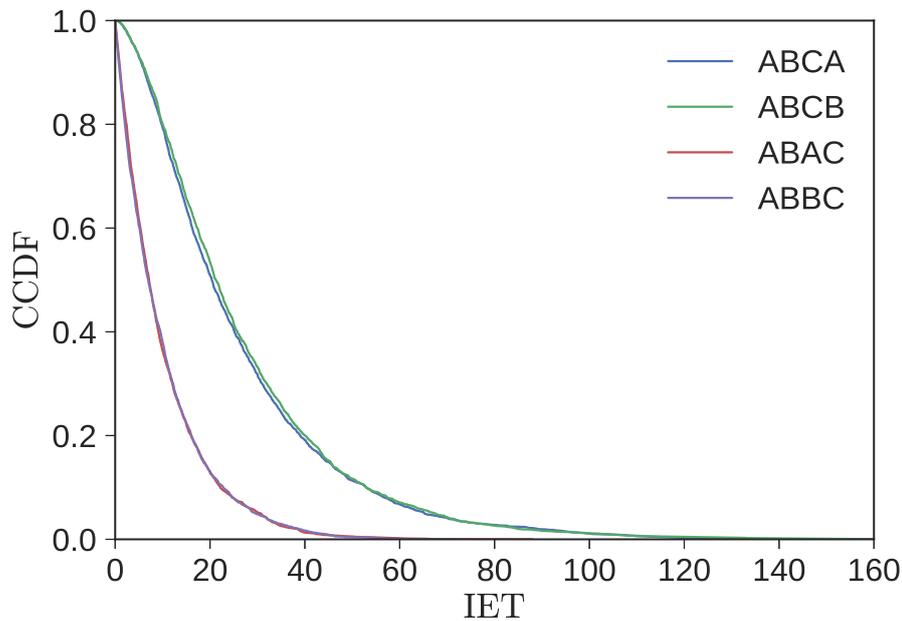


Figure 9.1: The CCDFs of the IET distributions for the SIS temporal network with constant probability of infection from each neighbour and constant recovery rate (once infected). The IET distributions are conditioned on the motif type formed between the two events. Here the ABCA and ABCB motif IETs are on average longer than those of the ABAC and ABBC motif due to the former motifs requiring a node to recover before they can appear.

that keyword then by necessity, all retweets of that tweet will also contain that keyword. By contrast a discussion around a topic may not contain the keyword in every tweet and instead be referred to by a pronoun or not mentioned at all. This leads to the sample potentially missing many conversations which are of genuine discussion around a topic.

We propose that this issue can be addressed using the TEG. The Twitter firehose is a stream of all tweets that occur on the network, currently in the order of 500 million tweets per day. While this is a manageable quantity with modern computing power there still remains the issue of sampling relevant tweets to address a research or business question.

The premise of this method is to sample components of the TEG (or Δt -TEG) instead of individual tweets. The steps of the method are as follows (illustrated in Figure 9.2):

1. **Build the TEG.** The TEG for Twitter can be constructed iteratively in real-time.

By storing the last event that each node has participated in events can be added to the TEG in an $\mathcal{O}(1)$ operation.

2. **Sample components.** Components of the TEG (or Δt -TEG) can be sampled according to a number of criteria. Possibilities include the presence of a keyword, a critical density of keyword usage within the component, or a vocabulary similar to vector of keywords. Further sampling can be used to find components of a certain size, word diversity, or with a particular motif or IET distribution.

Using this method tweets are sampled not only if they fulfil the keyword criteria but also if they are connected to events that do. Therefore these samples of the Twitter network are larger than by keyword search alone but should ultimately capture the entire conversation surrounding a particular topic. One drawback of this method is that ‘unrelated’ tweets may be included if the users participate in multiple conversations on differing topics within a short period of time, however this can be controlled by varying the parameter Δt and the component sampling criteria.

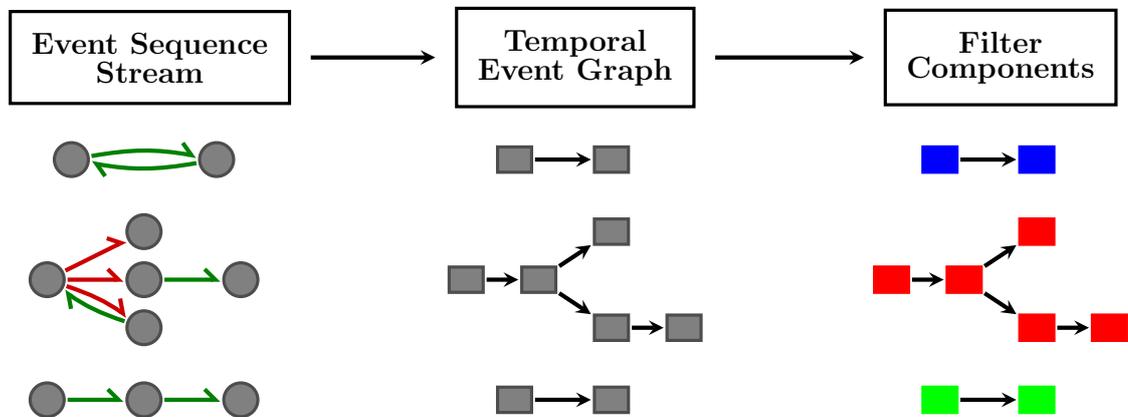


Figure 9.2: Using the TEG to filter the Twitter firehose.

9.3 Final Remarks

In summary this thesis has used a variety of techniques to examine social networks from different perspectives and at different scales. The new methods to filter and analyse temporal social network data could potentially have significant impact in industry, and

the new temporal network representation offers a new complementary approach for the temporal network community. Further exploration of the TEG and its properties should help develop our understanding of temporal networks and provide new ways to decompose and process temporal data (from social networks or elsewhere). The models introduced in the later sections of this thesis also provide insight into the behaviour of social systems and highlight new ways to analyse non-equilibrium processes, which are not amenable to standard methods.

Our ultimate goal is to develop methods to analyse social network data in real-time and build up a picture of individual behaviour in order to be able to predict user interactions, drive conversations, and effectively spread ideas through the network. This thesis takes one step towards that goal, however there are many more avenues of research and sources of data to explore.

Bibliography

- [1] A. Mellor, M. Mobilia, and R. K. P. Zia, “Heterogeneous out-of-equilibrium nonlinear q -voter model with zealotry,” *Physical Review E*, vol. 95, p. 012104, 2017.
- [2] —, “Characterization of the nonequilibrium steady state of a heterogeneous nonlinear q -voter model with zealotry,” *EPL (Europhysics Letters)*, vol. 113, no. 4, p. 48001, 2016.
- [3] A. Mellor, M. Mobilia, S. Redner, A. M. Rucklidge, and J. A. Ward, “Influence of Luddism on innovation diffusion,” *Physical Review E*, vol. 92, p. 012806, 2015.
- [4] W. W. Zachary, “An information flow model for conflict and fission in small groups,” *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1977.
- [5] J. F. Padgett and C. K. Ansell, “Robust action and the rise of the Medici, 1400-1434,” *American Journal of Sociology*, vol. 98, no. 6, pp. 1259–1319, 1993.
- [6] R. L. Breiger and P. E. Pattison, “Cumulated social roles: The duality of persons and their algebras,” *Social Networks*, vol. 8, no. 3, pp. 215–256, 1986.
- [7] M. Newman, *Networks: An Introduction*. Oxford University Press, 2010.
- [8] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Reviews of Modern Physics*, vol. 74, no. 1, p. 47, 2002.
- [9] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and dynamics,” *Physics Reports*, vol. 424, no. 4, pp. 175–308, 2006.

-
- [10] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994, vol. 8.
- [11] L. Euler, “Solutio problematis ad geometriam situs pertinentis,” *Eneström*, vol. 53, 1741.
- [12] B. Hopkins and R. Wilson, “The truth about Königsberg,” *The College Mathematics Journal*, vol. 35, no. 3, p. 198, 2004.
- [13] J. L. Moreno and H. H. Jennings, “Statistics of social configurations,” *Sociometry*, pp. 342–374, 1938.
- [14] P. Bonacich, “Power and centrality: A family of measures,” *American Journal of Sociology*, pp. 1170–1182, 1987.
- [15] M. Bianchini, M. Gori, and F. Scarselli, “Inside PageRank,” *ACM Transactions on Internet Technology (TOIT)*, vol. 5, no. 1, pp. 92–128, 2005.
- [16] E. Estrada and N. Hatano, “Communicability in complex networks,” *Physical Review E*, vol. 77, no. 3, p. 036111, 2008.
- [17] L. Katz, “A new status index derived from sociometric analysis,” *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.
- [18] M. E. Newman, “A measure of betweenness centrality based on random walks,” *Social Networks*, vol. 27, no. 1, pp. 39–54, 2005.
- [19] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, “Identification of influential spreaders in complex networks,” *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.
- [20] L. C. Freeman, “A set of measures of centrality based on betweenness,” *Sociometry*, pp. 35–41, 1977.
- [21] M. Kivela, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, “Multilayer networks,” *Journal of Complex Networks*, vol. 2, no. 3, pp. 203–271, 2014.

- [22] S. Milgram, "The small world problem," *Psychology Today*, vol. 2, no. 1, pp. 60–67, 1967.
- [23] Facebook statistics. (Accessed 21/02/2017). [Online]. Available: <http://newsroom.fb.com/company-info/>
- [24] About Twitter. (Accessed 21/02/2017). [Online]. Available: about.twitter.com
- [25] A. Lenhart, K. Purcell, A. Smith, and K. Zickuhr, "Social media & mobile internet use among teens and young adults." *Pew Internet & American Life Project*, 2010.
- [26] C. Shirky, "The political power of social media: Technology, the public sphere, and political change," *Foreign Affairs*, pp. 28–41, 2011.
- [27] H. Gil de Zúñiga, N. Jung, and S. Valenzuela, "Social media use for news and individuals' social capital, civic engagement and political participation," *Journal of Computer-Mediated Communication*, vol. 17, no. 3, pp. 319–336, 2012.
- [28] C. Llewellyn and L. Cram, "Brexit? Analyzing opinion on the UK-EU referendum within Twitter." in *ICWSM*, 2016, pp. 760–761.
- [29] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," National Bureau of Economic Research, Tech. Rep., 2017.
- [30] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, "Tastes, ties, and time: A new social network dataset using Facebook.com," *Social Networks*, vol. 30, no. 4, pp. 330–342, 2008.
- [31] A. L. Traud, P. J. Mucha, and M. A. Porter, "Social structure of Facebook networks," *Physica A: Statistical Mechanics and its Applications*, 2012.
- [32] Facebook's three and a half degrees of separation. (Accessed 21/02/2017). [Online]. Available: <https://research.fb.com/three-and-a-half-degrees-of-separation/>
- [33] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *Proceedings of the 19th International Conference on World Wide Web*. ACM, 2010, pp. 591–600.

- [34] D. Antoniadou and C. Dovrolis, "Co-evolutionary dynamics in social networks: A case study of Twitter," *Computational Social Networks*, vol. 2, no. 1, p. 14, 2015.
- [35] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.
- [36] J. Bryden, S. Funk, and V. A. Jansen, "Word usage mirrors community structure in the online social network Twitter," *EPJ Data Science*, vol. 2, no. 1, pp. 1–9, 2013.
- [37] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi, "Measuring user influence in Twitter: The million follower fallacy." *ICWSM*, vol. 10, pp. 10–17, 2010.
- [38] H. Fani, F. Zarrinkalam, E. Bagheri, and W. Du, "Time-sensitive topic-based communities on Twitter," in *Canadian Conference on Artificial Intelligence*. Springer International Publishing, 2016, pp. 192–204.
- [39] M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network," *Expert Systems with Applications: An International Journal*, vol. 40, no. 16, pp. 6266–6282, 2013.
- [40] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," *Journal of the American society for Information Science and Technology*, vol. 60, no. 11, pp. 2169–2188, 2009.
- [41] A. N. Smith, E. Fischer, and C. Yongjian, "How does brand-related user-generated content differ across YouTube, Facebook, and Twitter?" *Journal of Interactive Marketing*, vol. 26, no. 2, pp. 102–113, 2012.
- [42] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welp, "Predicting elections with Twitter: What 140 Characters reveal about political sentiment." *ICWSM*, vol. 10, pp. 178–185, 2010.
- [43] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Everyone's an influencer: Quantifying influence on Twitter," in *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*. ACM, 2011, pp. 65–74.

- [44] P. E. Brown and J. Feng, “Measuring user influence on Twitter using modified k-shell decomposition,” in *5th International AAAI Conference on Weblogs and Social Media*, 2011.
- [45] C. Castillo, M. Mendoza, and B. Poblete, “Information credibility on Twitter,” in *Proceedings of the 20th International Conference on World Wide Web*. ACM, 2011, pp. 675–684.
- [46] A. Pak and P. Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining,” in *LREc*, vol. 10, no. 2010, 2010.
- [47] T. Sakaki, M. Okazaki, and Y. Matsuo, “Earthquake shakes Twitter users: Real-time event detection by social sensors,” in *Proceedings of the 19th International Conference on World Wide Web*. ACM, 2010, pp. 851–860.
- [48] C. Buntain and J. Golbeck, “Identifying social roles in reddit using network structure,” in *Proceedings of the 23rd International Conference on World Wide Web*. ACM, 2014, pp. 615–620.
- [49] A. S. Kümpel, V. Karnowski, and T. Keyling, “News sharing in social media: A review of current research on news sharing users, content, and networks,” *Social Media and Society*, vol. 1, no. 2, p. 2056305115610141, 2015.
- [50] A. Panisson, L. Gauvin, M. Quagiotto, and C. Cattuto, “Mining concurrent topical activity in microblog streams,” in *Proceedings of the the 4th Workshop on Making Sense of Microposts co-located with the 23rd International World Wide Web Conference (WWW 2014)*, 2014.
- [51] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts, “Who says what to whom on Twitter,” in *Proceedings of the 20th International Conference on World Wide Web*. ACM, 2011, pp. 705–714.
- [52] B. A. Huberman, D. M. Romero, and F. Wu, “Social networks that matter: Twitter under the microscope,” *SSRN*, 2008.

- [53] Changing trends in the UK advertising industry. (Accessed 21/02/2017). [Online]. Available: https://marketing.conference-services.net/resources/327/2342/pdf/AM2011_0186.pdf
- [54] Reaching the right audience: How brands are using audience targeting in digital advertising, Forbes Insights. (Accessed 21/02/2017). [Online]. Available: https://info.quantcast.com/rs/quantcast/images/Quantcast_F-Insights_Report-36_Final-Web.pdf
- [55] Creative industries economic estimates 2016 report. (Accessed 21/02/2017). [Online]. Available: <https://www.gov.uk/government/statistics/creative-industries-economic-estimates-january-2016>
- [56] P. Laflin, A. V. Mantzaris, F. Ainley, A. Otley, P. Grindrod, and D. J. Higham, "Dynamic targeting in an online social medium," in *Social Informatics*. Springer, 2012, pp. 82–95.
- [57] P. Laflin, A. V. Mantzaris, D. J. Higham, P. Grindrod, F. Ainley, and A. Otley, "Twitter's big hitters," *Digital Futures 2012*, 2012.
- [58] P. Laflin, A. V. Mantzaris, F. Ainley, A. Otley, P. Grindrod, and D. J. Higham, "Discovering and validating influence in a dynamic online social network," *Social Network Analysis and Mining*, vol. 3, no. 4, pp. 1311–1323, 2013.
- [59] D. Higham, A. V. Mantzaris, P. Grindrod, A. Otley, and P. Laflin, "Anticipating activity in social media spikes," in *9th International AAAI Conference on Web and Social Media*, 2015.
- [60] P. Grindrod, D. J. Higham, P. Laflin, A. Otley, and J. A. Ward, "Inverse network sampling to explore online brand allegiance," *European Journal of Applied Mathematics*, pp. 1–13, 2016.
- [61] P. Grindrod, M. C. Parsons, D. J. Higham, and E. Estrada, "Communicability across evolving networks," *Physical Review E*, vol. 83, no. 4, p. 046120, 2011.
- [62] Bloom case studies. (Accessed 21/02/2017). [Online]. Available: <http://www.bloomagency.co.uk/work/>

- [63] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Reviews of Modern Physics*, vol. 87, no. 3, p. 925, 2015.
- [64] M. A. Porter and J. P. Gleeson, “Dynamical systems on networks,” *Frontiers in Applied Dynamical Systems: Reviews and Tutorials*, vol. 4, 2016.
- [65] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008, vol. 1.
- [66] Deloitte Global prediction on photosharing. (Accessed 21/02/2017). [Online]. Available: <https://www2.deloitte.com/global/en/pages/technology-media-and-telecommunications/articles/tmt-pred16-telecomm-photo-sharing-trillions-and-rising.html>
- [67] F. Black and M. Scholes, “The pricing of options and corporate liabilities,” *Journal of Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.
- [68] Creative industries strategy 2013-2016, Innovate UK. (Accessed 21/02/2017). [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/362253/Creative_Industries_Strategy_2013-16.pdf
- [69] F. Liljeros, C. R. Edling, and L. A. N. Amaral, “Sexual networks: Implications for the transmission of sexually transmitted infections,” *Microbes and Infection*, vol. 5, no. 2, pp. 189–196, 2003.
- [70] C. S. Riolo, J. S. Koopman, and S. E. Chick, “Methods and measures for the description of epidemiologic contact networks,” *Journal of Urban Health*, vol. 78, no. 3, pp. 446–457, 2001.
- [71] L. Kovanen, M. Karsai, K. Kaski, J. Kertész, and J. Saramäki, “Temporal motifs in time-dependent networks,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2011, no. 11, p. P11005, 2011.
- [72] F. M. Bass, “A new product growth for model consumer durables,” *Management Science*, vol. 15, no. 5, pp. 215–227, 1969.

- [73] M. C. González, H. J. Herrmann, J. Kertész, and T. Vicsek, “Community structure and ethnic preferences in school friendship networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 379, no. 1, pp. 307–316, 2007.
- [74] J. Fournet and A. Barrat, “Epidemic risk from friendship network data: An equivalence with a non-uniform sampling of contact networks,” *Scientific Reports*, vol. 6, no. 24593, 2016.
- [75] G. Kossinets and D. J. Watts, “Empirical analysis of an evolving social network,” *Science*, vol. 311, no. 5757, pp. 88–90, 2006.
- [76] R. Kumar, J. Novak, and A. Tomkins, “Structure and evolution of online social networks,” in *Link Mining: Models, Algorithms, and Applications*. Springer, 2010, pp. 337–357.
- [77] B. Meeder, B. Karrer, A. Sayedi, R. Ravi, C. Borgs, and J. Chayes, “We know who you followed last summer: Inferring social link creation times in Twitter,” in *Proceedings of the 20th International Conference on World Wide Web*. ACM, 2011, pp. 517–526.
- [78] P. Holme and J. Saramäki, “Temporal networks,” *Physics Reports*, vol. 519, no. 3, pp. 97–125, 2012.
- [79] —, *Temporal Networks*. Springer, 2013.
- [80] D. Kempe, J. Kleinberg, and A. Kumar, “Connectivity and inference problems for temporal networks,” in *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*. ACM, 2000, pp. 504–513.
- [81] B. B. Xuan, A. Ferreira, and A. Jarry, “Computing shortest, fastest, and foremost journeys in dynamic networks,” *International Journal of Foundations of Computer Science*, vol. 14, no. 02, pp. 267–285, 2003.
- [82] E. Cheng, J. W. Grossman, and M. J. Lipman, “Time-stamped graphs and their associated influence digraphs,” *Discrete Applied Mathematics*, vol. 128, no. 2, pp. 317–335, 2003.

- [83] A.-L. Barabasi, “The origin of bursts and heavy tails in human dynamics,” *Nature*, vol. 435, no. 7039, pp. 207–211, 2005.
- [84] K.-I. Goh and A.-L. Barabási, “Burstiness and memory in complex systems,” *EPL (Europhysics Letters)*, vol. 81, no. 4, p. 48002, 2008.
- [85] M. Kivela and M. A. Porter, “Estimating interevent time distributions from finite observation periods in communication networks,” *Physical Review E*, vol. 92, no. 5, p. 052813, 2015.
- [86] K. Wehmuth, A. Ziviani, and E. Fleury, “A unifying model for representing time-varying graphs,” in *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, no. 36678. IEEE, 2015, pp. 1–10.
- [87] P. Holme, “Network reachability of real-world contact sequences,” *Physical Review E*, vol. 71, no. 4, p. 046119, 2005.
- [88] J. Tang, M. Musolesi, C. Mascolo, and V. Latora, “Characterising temporal distance and reachability in mobile and online social networks,” *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 1, pp. 118–124, 2010.
- [89] R. Guimera, S. Mossa, A. Turtchi, and L. N. Amaral, “The worldwide air transportation network: Anomalous centrality, community structure, and cities’ global roles,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 22, pp. 7794–7799, 2005.
- [90] L. C. Freeman, “Centrality in social networks conceptual clarification,” *Social Networks*, vol. 1, no. 3, pp. 215–239, 1979.
- [91] L. Page, S. Brin, R. Motwani, and T. Winograd, “The PageRank citation ranking: Bringing order to the web.” Technical Report 1999-66, 1999.
- [92] J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia, “Analysing information flows and key mediators through temporal centrality metrics,” in *Proceedings of the 3rd Workshop on Social Network Systems*. ACM, 2010, p. 3.

- [93] J. Tang, S. Scellato, M. Musolesi, C. Mascolo, and V. Latora, "Small-world behavior in time-varying graphs," *Physical Review E*, vol. 81, no. 5, p. 055101, 2010.
- [94] P. Grindrod and D. J. Higham, "A dynamical systems view of network centrality," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, vol. 470, no. 2165, p. 20130835, 2014.
- [95] R. K. Pan and J. Saramäki, "Path lengths, correlations, and centrality in temporal networks," *Physical Review E*, vol. 84, no. 1, p. 016105, 2011.
- [96] E. Estrada, N. Hatano, and M. Benzi, "The physics of communicability in complex networks," *Physics Reports*, vol. 514, no. 3, pp. 89–119, 2012.
- [97] J. J. Crofts and D. J. Higham, "A weighted communicability measure applied to complex brain networks," *Journal of the Royal Society Interface*, vol. 6, no. 33, pp. 411–414, 2009.
- [98] P. Grindrod, Z. V. Stoyanov, G. M. Smith, and J. D. Saddy, "Primary evolving networks and the comparative analysis of robust and fragile structures," *Journal of Complex Networks*, p. cnt015, 2013.
- [99] A. V. Mantzaris and D. J. Higham, "Dynamic communicability predicts infectiousness," in *Temporal Networks*. Springer, 2013, pp. 283–294.
- [100] I. Chen, M. Benzi, H. H. Chang, and V. S. Hertzberg, "Dynamic communicability and epidemic spread: A case study on an empirical dynamic contact network," *Journal of Complex Networks*, p. cnw017, 2016.
- [101] J. Tang, M. Musolesi, C. Mascolo, and V. Latora, "Temporal distance metrics for social network analysis," in *Proceedings of the 2nd ACM Workshop on Online Social Networks*. ACM, 2009, pp. 31–36.
- [102] V. Nicosia, J. Tang, M. Musolesi, G. Russo, C. Mascolo, and V. Latora, "Components in time-varying graphs," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 2, p. 023101, 2012.

- [103] A. Vázquez, J. G. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabási, “Modeling bursts and heavy tails in human dynamics,” *Physical Review E*, vol. 73, no. 3, p. 036127, 2006.
- [104] H.-H. Jo, M. Karsai, J. Kertész, and K. Kaski, “Circadian pattern and burstiness in mobile phone communication,” *New Journal of Physics*, vol. 14, no. 1, p. 013055, 2012.
- [105] R. D. Malmgren, J. M. Hofman, L. A. Amaral, and D. J. Watts, “Characterizing individual communication patterns,” in *Proceedings of the 15th ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*. ACM, 2009, pp. 607–616.
- [106] Z. Dezsö, E. Almaas, A. Lukács, B. Rácz, I. Szakadát, and A.-L. Barabási, “Dynamics of information access on the web,” *Physical Review E*, vol. 73, no. 6, p. 066132, 2006.
- [107] J. G. Oliveira and A.-L. Barabási, “Human dynamics: Darwin and Einstein correspondence patterns,” *Nature*, vol. 437, no. 7063, pp. 1251–1251, 2005.
- [108] E.-K. Kim and H.-H. Jo, “Measuring burstiness for finite event sequences,” *Physical Review E*, vol. 94, no. 3, p. 032311, 2016.
- [109] A. Vazquez, B. Racz, A. Lukacs, and A.-L. Barabasi, “Impact of non-poissonian activity patterns on spreading processes,” *Physical Review Letters*, vol. 98, no. 15, p. 158702, 2007.
- [110] M. Karsai, M. Kivela, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, “Small but slow world: How network topology and burstiness slow down spreading,” *Physical Review E*, vol. 83, no. 2, p. 025102, 2011.
- [111] G. Miritello, E. Moro, and R. Lara, “Dynamical strength of social ties in information spreading,” *Physical Review E*, vol. 83, no. 4, p. 045102, 2011.
- [112] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F. Pinton, N. Khanafer, W. Van den Broeck *et al.*, “Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees,” *BMC Medicine*, vol. 9, no. 1, p. 87, 2011.

- [113] J. L. Iribarren and E. Moro, “Impact of human activity patterns on the dynamics of information diffusion,” *Physical Review Letters*, vol. 103, no. 3, p. 038702, 2009.
- [114] R. Lambiotte, L. Tabourier, and J.-C. Delvenne, “Burstiness and spreading on temporal networks,” *The European Physical Journal B*, vol. 86, no. 7, pp. 1–4, 2013.
- [115] T. Takaguchi, N. Masuda, and P. Holme, “Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics,” *PLoS ONE*, vol. 8, no. 7, p. e68629, 2013.
- [116] L. E. Rocha, F. Liljeros, and P. Holme, “Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts,” *PLoS Computational Biology*, vol. 7, no. 3, p. e1001109, 2011.
- [117] E. L. Kaplan and P. Meier, “Nonparametric estimation from incomplete observations,” *Journal of the American Statistical Association*, vol. 53, no. 282, pp. 457–481, 1958.
- [118] C. Tantipathananandh, T. Berger-Wolf, and D. Kempe, “A framework for community identification in dynamic social networks,” in *Proceedings of the 13th ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*. ACM, 2007, pp. 717–726.
- [119] F. Guo, S. Hanneke, W. Fu, and E. P. Xing, “Recovering temporally rewiring networks: A model-based approach,” in *Proceedings of the 24th International Conference on Machine Learning*. ACM, 2007, pp. 321–328.
- [120] J. Stehlé, A. Barrat, and G. Bianconi, “Dynamical and bursty interactions in social networks,” *Physical Review E*, vol. 81, no. 3, p. 035101, 2010.
- [121] H.-H. Jo, R. K. Pan, and K. Kaski, “Emergence of bursts and communities in evolving weighted networks,” *PLoS ONE*, vol. 6, no. 8, p. e22687, 2011.
- [122] T. Gross and H. Sayama, *Adaptive Networks*. Springer-Verlag Berlin Heidelberg, 2009.

- [123] P. Crucitti, V. Latora, and S. Porta, “Centrality measures in spatial networks of urban streets,” *Physical Review E*, vol. 73, no. 3, p. 036125, 2006.
- [124] M. Girvan and M. E. Newman, “Community structure in social and biological networks,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [125] L. C. Freeman, S. P. Borgatti, and D. R. White, “Centrality in valued graphs: A measure of betweenness based on network flow,” *Social Networks*, vol. 13, no. 2, pp. 141–154, 1991.
- [126] E. Estrada and N. Hatano, “Communicability angle and the spatial efficiency of networks,” *SIAM Review*, vol. 58, no. 4, pp. 692–715, 2016.
- [127] E. Estrada, E. Vargas-Estrada, and H. Ando, “Communicability angles reveal critical edges for network consensus dynamics,” *Physical Review E*, vol. 92, no. 5, p. 052809, 2015.
- [128] S. Motegi and N. Masuda, “A network-based dynamical ranking system for competitive sports,” *Scientific Reports*, vol. 2, no. 904, 2012.
- [129] Internet statistics of Twitter volumes. (Accessed 21/02/2017). [Online]. Available: <http://www.internetlivestats.com/twitter-statistics/>
- [130] M. Aprahamian, D. J. Higham, and N. J. Higham, “Matching exponential-based and resolvent-based centrality measures,” *Journal of Complex Networks*, vol. 157, no. 176, p. 4, 2016.
- [131] P. Grindrod and D. J. Higham, “A matrix iteration for dynamic network summaries,” *SIAM Review*, vol. 55, no. 1, pp. 118–128, 2013.
- [132] T. Rogers, “Null models for dynamic centrality in temporal networks,” *Journal of Complex Networks*, vol. 3, no. 1, pp. 113–125, 2015.
- [133] D. M. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs: Theory and Application*. Academic Pr, 1980, vol. 87.

- [134] A. Johansen, “Probing human response times,” *Physica A: Statistical Mechanics and its Applications*, vol. 338, no. 1, pp. 286–291, 2004.
- [135] N. M. Navaroli and P. Smyth, “Modeling response time in digital human communication.” in *ICWSM*, 2015, pp. 278–287.
- [136] M. A. Bender, J. T. Fineman, and S. Gilbert, “A new approach to incremental topological ordering,” in *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2009, pp. 1108–1115.
- [137] B. Haeupler, T. Kavitha, R. Mathew, S. Sen, and R. E. Tarjan, “Incremental cycle detection, topological ordering, and strong component maintenance,” *ACM Transactions on Algorithms (TALG)*, vol. 8, no. 1, p. 3, 2012.
- [138] M. Morris and M. Kretzschmar, “Concurrent partnerships and transmission dynamics in networks,” *Social Networks*, vol. 17, no. 3-4, pp. 299–318, 1995.
- [139] A. Barrat, C. Cattuto, A. Tozzi, P. Vanhems, and N. Voirin, “Measuring contact patterns with wearable sensors: Methods, data characteristics and applications to data-driven simulations of infectious diseases,” *Clinical Microbiology and Infection*, vol. 20, no. 1, pp. 10–16, 2014.
- [140] T. H. Cormen, *Introduction to Algorithms*. MIT press, 2009.
- [141] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, “Network motifs in the transcriptional regulation network of escherichia coli,” *Nature Genetics*, vol. 31, no. 1, pp. 64–68, 2002.
- [142] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, “Network motifs: Simple building blocks of complex networks,” *Science*, vol. 298, no. 5594, pp. 824–827, 2002.
- [143] Y. Artzy-Randrup, S. J. Fleishman, N. Ben-Tal, and L. Stone, “Comment on ‘Network motifs: Simple building blocks of complex networks’ and ‘Superfamilies of evolved and designed networks,’” *Science*, vol. 305, no. 5687, pp. 1107–1107, 2004.

- [144] P. Bajardi, A. Barrat, F. Natale, L. Savini, and V. Colizza, “Dynamical patterns of cattle trade movements,” *PLoS ONE*, vol. 6, no. 5, p. e19869, 2011.
- [145] N. Drissi, T. Chonavel, and J. M. Boucher, “Generalized cumulative residual entropy for distributions with unrestricted supports,” *Research Letters in Signal Processing*, vol. 2008, p. 11, 2008.
- [146] M. Rao, Y. Chen, B. C. Vemuri, and F. Wang, “Cumulative residual entropy: A new measure of information,” *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1220–1228, 2004.
- [147] S. Scellato, I. Leontiadis, C. Mascolo, P. Basu, and M. Zafer, “Evaluating temporal robustness of mobile networks,” *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 105–117, 2013.
- [148] A. Barrat, B. Fernandez, K. K. Lin, and L.-S. Young, “Modeling temporal networks using random itineraries,” *Physical Review Letters*, vol. 110, no. 15, p. 158702, 2013.
- [149] M. Sarzynska, E. A. Leicht, G. Chowell, and M. A. Porter, “Null models for community detection in spatially embedded, temporal networks,” *Journal of Complex Networks*, vol. 4, no. 3, pp. 363–406, 2016.
- [150] Twitter report on second screen usage. (Accessed 21/02/2017). [Online]. Available: <https://www.accenture.com/us-en/~media/Accenture/Conversion-Assets/Microsites/Documents17/Accenture-Digital-Video-Connected-Consumer.pdf>
- [151] Twitter blog post on user engagement. (Accessed 21/02/2017). [Online]. Available: <https://blog.twitter.com/2014/four-insights-about-millennials-on-twitter>
- [152] Twitter annual report 2015. (Accessed 21/02/2017). [Online]. Available: <https://investor.twitterinc.com/annuals-proxies.cfm>
- [153] Twitter development API. (Accessed 21/02/2017). [Online]. Available: <https://dev.twitter.com/overview/api>
- [154] B. Krishnamurthy, P. Gill, and M. Arlitt, “A few chirps about Twitter,” in *Proceedings of the 1st Workshop on Online Social Networks*. ACM, 2008, pp. 19–24.

- [155] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: Finding topic-sensitive influential twitterers," in *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*. ACM, 2010, pp. 261–270.
- [156] C. J. Hutto, S. Yardi, and E. Gilbert, "A longitudinal study of follow predictors on Twitter," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013, pp. 821–830.
- [157] Weekly viewing summary, Broadcasters' Audience Research Board. (Accessed 21/02/2017). [Online]. Available: <http://www.barb.co.uk/viewing-data/weekly-top-30/>
- [158] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*. Elsevier, 1992, vol. 1.
- [159] S. Redner, *A Guide to First-passage Processes*. Cambridge University Press, 2001.
- [160] C. W. Gardiner *et al.*, *Handbook of Stochastic Methods*. Springer Berlin, 1985, vol. 3.
- [161] M. Lax, "Classical noise IV: Langevin methods," *Review of Modern Physics*, vol. 38, no. 3, p. 541, 1966.
- [162] D. J. Higham, "Modeling and simulating chemical reactions," *SIAM Review*, vol. 50, no. 2, pp. 347–368, 2008.
- [163] R. Erban, J. Chapman, and P. Maini, "A practical guide to stochastic simulations of reaction-diffusion processes," *arXiv preprint arXiv:0704.1908*, 2007.
- [164] J. P. Gleeson, S. Melnik, J. A. Ward, M. A. Porter, and P. J. Mucha, "Accuracy of mean-field theory for dynamics on real-world networks," *Physical Review E*, vol. 85, no. 2, p. 026106, 2012.
- [165] A. Vespignani, "Modelling dynamical processes in complex socio-technical systems," *Nature Physics*, vol. 8, no. 1, pp. 32–39, 2011.
- [166] C. S. Gillespie, "Moment-closure approximations for mass-action models," *IET Systems Biology*, vol. 3, no. 1, pp. 52–58, 2009.

- [167] L. E. Rocha and N. Masuda, “Individual-based approach to epidemic processes on arbitrary dynamic contact networks,” *Scientific Reports*, vol. 6, 2016.
- [168] P. Erdős and A. Rényi, “On random graphs I,” *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.
- [169] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [170] E. Pugliese and C. Castellano, “Heterogeneous pair approximation for voter models on networks,” *EPL (Europhysics Letters)*, vol. 88, no. 5, p. 58004, 2009.
- [171] G. Demirel, F. Vazquez, G. Böhme, and T. Gross, “Moment-closure approximations for discrete adaptive networks,” *Physica D: Nonlinear Phenomena*, vol. 267, pp. 68–80, 2014.
- [172] T. C. Schelling, *Micromotives and Macrobehavior*. WW Norton & Company, 1978.
- [173] —, “Dynamic models of segregation,” *Journal of Mathematical Sociology*, vol. 1, no. 2, pp. 143–186, 1971.
- [174] P. Clifford and A. Sudbury, “A model for spatial conflict,” *Biometrika*, vol. 60, no. 3, pp. 581–588, 1973.
- [175] E. Ising, “Beitrag zur theorie des ferromagnetismus,” *Zeitschrift für Physik A Hadrons and Nuclei*, vol. 31, no. 1, pp. 253–258, 1925.
- [176] B. A. Cipra, “An introduction to the Ising model,” *American Mathematical Monthly*, vol. 94, no. 10, pp. 937–959, 1987.
- [177] T. M. Liggett, *Interacting Particle Systems*. Springer, 1985.
- [178] K. Suchecki, V. M. Eguiluz, and M. San Miguel, “Conservation laws for the voter model in complex networks,” *EPL (Europhysics Letters)*, vol. 69, no. 2, p. 228, 2005.
- [179] J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models*. Cambridge University Press, 2005.

- [180] T. M. Liggett, *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*. Springer Science & Business Media, 2013, vol. 324.
- [181] F. Vazquez and V. Eguiluz, “Analytical solution of the voter model on uncorrelated networks,” *New Journal of Physics*, vol. 10, no. 6, p. 063011, 2008.
- [182] V. Sood and S. Redner, “Voter model on heterogeneous graphs,” *Physical Review Letters*, vol. 94, no. 17, p. 178701, 2005.
- [183] R. Lambiotte and S. Redner, “Dynamics of non-conservative voters,” *EPL (Europhysics Letters)*, vol. 82, no. 1, p. 18007, 2008.
- [184] C. Castellano, M. A. Muñoz, and R. Pastor-Satorras, “Nonlinear q-voter model,” *Physical Review E*, vol. 80, no. 4, p. 041129, 2009.
- [185] S. E. Asch, “Opinions and social pressure,” *Scientific American*, vol. 193, no. 5, pp. 31–35, 1955.
- [186] S. Milgram, L. Bickman, and L. Berkowitz, “Note on the drawing power of crowds of different size.” *Journal of Personality and Social Psychology*, vol. 13, no. 2, p. 79, 1969.
- [187] M. Mobilia, “Does a single zealot affect an infinite group of voters?” *Physical Review Letters*, vol. 91, no. 2, p. 028701, 2003.
- [188] M. Mobilia, A. Petersen, and S. Redner, “On the role of zealotry in the voter model,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2007, no. 08, p. P08029, 2007.
- [189] T. Rogers and T. Gross, “Consensus time and conformity in the adaptive voter model,” *Physical Review E*, vol. 88, no. 3, p. 030102, 2013.
- [190] R. M. Axelrod, *The Complexity of Cooperation: Agent-based Models of Competition and Collaboration*. Princeton University Press, 1997.
- [191] M. Mobilia, “Nonlinear q-voter model with inflexible zealots,” *Physical Review E*, vol. 92, no. 1, p. 012803, 2015.

- [192] M. A. Bekafigo and A. McBride, "Who tweets about politics? political participation of twitter users during the 2011 gubernatorial elections," *Social Science Computer Review*, vol. 31, no. 5, pp. 625–643, 2013.
- [193] J. H. Parmelee and S. L. Bichard, *Politics and the Twitter Revolution: How Tweets Influence the Relationship Between Political Leaders and the Public*. Lexington Books, 2011.
- [194] C. S. Park, "Does Twitter motivate involvement in politics? Tweeting, opinion leadership, and political engagement," *Computers in Human Behavior*, vol. 29, no. 4, pp. 1641–1648, 2013.
- [195] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welp, "Election forecasts with Twitter: How 140 characters reflect the political landscape," *Social Science Computer Review*, vol. 29, no. 4, pp. 402–418, 2011.
- [196] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, "A system for real-time Twitter sentiment analysis of 2012 US presidential election cycle," in *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, 2012, pp. 115–120.
- [197] A. Kolmogorov, "Zur theorie der markoffschen ketten," *Mathematische Annalen*, vol. 112, no. 1, pp. 155–160, 1936.
- [198] R. Zia and B. Schmittmann, "Probability currents as principal characteristics in the statistical mechanics of non-equilibrium steady states," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2007, no. 07, p. P07012, 2007.
- [199] T. L. Hill, "Studies in irreversible thermodynamics IV. Diagrammatic representation of steady state fluxes for unimolecular systems," *Journal of Theoretical Biology*, vol. 10, no. 3, pp. 442–459, 1966.
- [200] A. Mellor, M. Mobilia, and R. K. P. Zia, *Figshare*, 2016, (Supplementary Material). [Online]. Available: <https://dx.doi.org/10.6084/m9.figshare.2060595>
- [201] H. Risken, *The Fokker-Planck Equation*. Springer, 1984.

- [202] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.
- [203] H. Eyring, "The activated complex in chemical reactions," *Journal of Chemical Physics*, vol. 3, no. 2, pp. 107–115, 1935.
- [204] H. A. Kramers, "Brownian motion in a field of force and the diffusion model of chemical reactions," *Physica*, vol. 7, no. 4, pp. 284–304, 1940.
- [205] P. Hänggi, P. Talkner, and M. Borkovec, "Reaction-rate theory: Fifty years after Kramers," *Reviews of Modern Physics*, vol. 62, no. 2, p. 251, 1990.
- [206] B. Kollanyi, P. N. Howard, and S. C. Woolley., "Bots and automation over Twitter during the US election," *Data Memo 2016 Oxford, UK*. [Online]. Available: www.politicalbots.org.
- [207] C. Mackay, *Extraordinary Popular Delusions and the Madness of Crowds*. Richard Bentley, London, 1841.
- [208] E. M. Rogers, *Diffusion of Innovations*. Simon and Schuster, 2010.
- [209] R. Brown, "A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies," *Philosophical Magazine Series 2*, vol. 4, no. 21, pp. 161–173, 1828.
- [210] J. Coleman, E. Katz, and H. Menzel, "The diffusion of an innovation among physicians," *Sociometry*, vol. 20, no. 4, pp. 253–270, 1957.
- [211] D. L. Kincaid, "From innovation to social norm: Bounded normative influence," *Journal of Health Communication*, vol. 9, no. S1, pp. 37–57, 2004.
- [212] B. Ryan and N. C. Gross, "The diffusion of hybrid seed corn in two iowa communities," *Rural Sociology*, vol. 8, no. 1, p. 15, 1943.
- [213] L. A. Fourt and J. W. Woodlock, "Early prediction of market success for new grocery products," *The Journal of Marketing*, pp. 31–38, 1960.

- [214] F. M. Bass, “Comments on ‘A new product growth for model consumer durables the Bass model’,” *Management Science*, vol. 50, no. 12, pp. 1833–1840, 2004.
- [215] —, “The relationship between diffusion rates, experience curves, and demand elasticities for consumer durable technological innovations,” *Journal of Business*, pp. S51–S67, 1980.
- [216] V. Mahajan, E. Muller, and R. K. Srivastava, “Determination of adopter categories by using innovation diffusion models,” *Journal of Marketing Research*, pp. 37–50, 1990.
- [217] W. O. Kermack and A. G. McKendrick, “A contribution to the mathematical theory of epidemics,” in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 115, no. 772. The Royal Society, 1927, pp. 700–721.
- [218] F. Sultan, J. U. Farley, and D. R. Lehmann, “A meta-analysis of applications of diffusion models,” *Journal of Marketing Research*, pp. 70–77, 1990.
- [219] E. Kiesling, M. Günther, C. Stummer, and L. M. Wakolbinger, “Agent-based simulation of innovation diffusion: A review,” *Central European Journal of Operations Research*, vol. 20, no. 2, pp. 183–230, 2012.
- [220] D. Centola, V. M. Eguíluz, and M. W. Macy, “Cascade dynamics of complex propagation,” *Physica A: Statistical Mechanics and its Applications*, vol. 374, no. 1, pp. 449–456, 2007.
- [221] D. Centola, R. Willer, and M. Macy, “The emperor’s dilemma: A computational model of self-enforcing norms,” *American Journal of Sociology*, vol. 110, no. 4, pp. 1009–1040, 2005.
- [222] P. S. Dodds and D. J. Watts, “Universal behavior in a generalized model of contagion,” *Physical Review Letters*, vol. 92, no. 21, p. 218701, 2004.
- [223] J. Goldenberg, B. Libai, S. Solomon, N. Jan, and D. Stauffer, “Marketing percolation,” *Physica A: Statistical Mechanics and its Applications*, vol. 284, no. 1, pp. 335–347, 2000.

- [224] D. Strang and M. W. Macy, "In search of excellence: Fads, success stories, and adaptive emulation," *American Journal of Sociology*, vol. 107, no. 1, pp. 147–182, 2001.
- [225] D. Centola, "The spread of behavior in an online social network experiment," *Science*, vol. 329, no. 5996, pp. 1194–1197, 2010.
- [226] Device ownership, Pew research center. (Accessed 21/02/2017). [Online]. Available: <http://www.pewinternet.org/data-trend/mobile/device-ownership/>
- [227] The interphone study. (Accessed 21/02/2017). [Online]. Available: http://interphone.iarc.fr/interphone_results.php
- [228] R. M. Wolfe and L. K. Sharp, "Anti-vaccinationists past and present," *British Medical Journal*, vol. 325, no. 7361, p. 430, 2002.
- [229] Social network advertising expenditure, Emarketer. (Accessed 21/02/2017). [Online]. Available: <https://www.emarketer.com/Article/Social-Network-Ad-Spending-Hit-2368-Billion-Worldwide-2015/1012357>
- [230] B. Pang, L. Lee *et al.*, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
- [231] V. Mahajan, E. Muller, and F. M. Bass, "New product diffusion models in marketing: A review and directions for research," in *Diffusion of Technologies and Social Behavior*. Springer, 1991, pp. 125–177.
- [232] H. P. Young, "Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning," *The American Economic Review*, vol. 99, no. 5, pp. 1899–1924, 2009.
- [233] E. Abrahamson and L. Rosenkopf, "Institutional and competitive bandwagons: Using mathematical modeling as a tool to explore innovation diffusion," *Academy of Management Review*, vol. 18, no. 3, pp. 487–517, 1993.

- [234] P. S. Tolbert and L. G. Zucker, “Institutional sources of change in the formal structure of organizations: The diffusion of civil service reform, 1880-1935,” *Administrative Science Quarterly*, pp. 22–39, 1983.
- [235] P. L. Krapivsky, S. Redner, and D. Volovik, “Reinforcement-driven spread of innovations and fads,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2011, no. 12, p. P12003, 2011.
- [236] S. Moldovan and J. Goldenberg, “Cellular automata modeling of resistance to innovations: Effects and solutions,” *Technological Forecasting and Social Change*, vol. 71, no. 5, pp. 425–442, 2004.
- [237] D. T. Gillespie, “Exact stochastic simulation of coupled chemical reactions,” *The Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.
- [238] M. O. Jackson *et al.*, *Social and Economic Networks*. Princeton University Press, 2008, vol. 3.
- [239] V. Privman, *Nonequilibrium Statistical Mechanics in One Dimension*. Cambridge University Press, 2005.
- [240] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics*. Cambridge University Press, 2010.