

---

# CAUSAL STRUCTURE IN NETWORKS

---

by  
James R. Clough

A thesis presented for the degree of  
Doctor of Philosophy of Imperial College London

Imperial College London  
Department of Physics  
Blackett Laboratory  
Prince Consort Road  
London SW7 2AZ  
United Kingdom

May 2017



## DEDICATION

This thesis is dedicated to my parents for always supporting and encouraging me in whatever I chose to do.

#### **DECLARATION OF ORIGINALITY**

I hereby declare that all work presented in this thesis that is not my own has been appropriately acknowledged.

James Clough

#### COPYRIGHT DECLARATION

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.



## ABSTRACT

The ‘network geometry’ approach in network science has in recent years had success in describing complex network topologies using simple geometric models. Previous work has focussed on using Riemannian spaces such as flat Euclidean space or curved Hyperbolic space to describe network structure.

Here, the geometry of Lorentzian spacetime is used to model and describe the structure of a special class of networks, directed acyclic graphs. These networks share important features, such as causal structure, with this geometry making this approach a natural extension of the network geometry programme. By recognising the relationship between these networks and this family of geometries, techniques from physical theories of discrete spacetime may be brought into the domain of network science allowing new methods, models, and analyses to be developed.

Using network datasets which form directed acyclic graphs, primarily citation networks, as illustrations, I show how characterising a network by its causal structure aids traditional analysis, how networks can be associated with spacetimes of a specific dimension and curvature, and how they may be embedded in a spacetime. Numerous applications are discussed relevant to both citation networks and directed acyclic graphs more generally, and computational implementations of the methods discussed are made available.



# Table of Contents

<b>List of Tables</b>	<b>11</b>
<b>List of Figures</b>	<b>12</b>
<b>1 Introduction</b>	<b>19</b>
1.1 Graphs . . . . .	21
1.2 Network Science . . . . .	27
1.3 Directed Acyclic Graphs . . . . .	31
1.4 Outline . . . . .	35
1.5 Datasets . . . . .	36
<b>2 Network Geometry</b>	<b>41</b>
2.1 Spatial Networks . . . . .	41
2.2 Latent Space Models . . . . .	44
2.3 Random Geometric Graphs . . . . .	46
2.4 Hyperbolic Network Geometry . . . . .	49
2.5 Summary . . . . .	54
<b>3 Causal Structure and Causal Sets</b>	<b>55</b>
3.1 Partial Orders . . . . .	56
3.2 Lorentzian Geometry . . . . .	58
3.3 Causal Set Theory . . . . .	64
3.4 Generating causal sets by sprinkling . . . . .	72
3.5 Summary . . . . .	74
<b>4 Transitive Reduction</b>	<b>76</b>
4.1 Effect on network structure . . . . .	79
4.2 Effect on individual nodes . . . . .	86
4.3 Summary . . . . .	88
<b>5 Measuring dimension</b>	<b>89</b>
5.1 Order Dimension . . . . .	90
5.2 Geodesics and Longest paths . . . . .	93
5.3 Midpoint-Scaling Dimension . . . . .	96
5.4 Myrheim-Meyer Dimension . . . . .	98
5.5 Results . . . . .	103
5.6 Scale Dependent Dimension . . . . .	107
5.7 Discussion and Interpretation . . . . .	109

5.8 Summary . . . . .	111
<b>6 Measuring curvature</b>	<b>113</b>
6.1 De Sitter Spacetime . . . . .	114
6.2 De Sitter Causal Sets . . . . .	115
6.3 Inferring Curvature . . . . .	121
6.4 Results . . . . .	123
6.5 Summary . . . . .	127
<b>7 Embedding networks</b>	<b>128</b>
7.1 Embedding in Euclidean Space . . . . .	129
7.2 Estimating spacelike separation . . . . .	131
7.3 Multidimensional Scaling . . . . .	132
7.4 Generalised MDS . . . . .	134
7.5 Results . . . . .	143
7.6 Discussion and Applications . . . . .	150
7.7 Summary . . . . .	151
<b>8 Conclusions and Outlook</b>	<b>153</b>
8.1 Future Work . . . . .	157
8.2 Available software . . . . .	159
<b>Bibliography</b>	<b>160</b>
<b>A Degree after Transitive Reduction</b>	<b>178</b>
<b>B Transitive Reduction plots</b>	<b>181</b>
<b>C Myrheim-Meyer Dimension for <math>k &gt; 2</math></b>	<b>182</b>

# List of Tables

1.1	Table of datasets.	40
4.1	Table of degrees before and after transitive reduction.	79
5.1	Table of values for Myrheim-Meyer dimension estimate.	100
5.2	Table of approximate estimated dimensions for real networks.	103

# List of Figures

1.1	Example of a graph. . . . .	21
1.2	Phase transitions in Erdős Renyi graphs. . . . .	25
1.3	Definitions of the degree, diameter and clustering in a simple graph. . . . .	29
1.4	Examples of Directed Acyclic Graphs. . . . .	32
1.5	DAG from the citation network of US Supreme Court judgements. . . . .	33
1.6	DAG from the technology tree of FreeCiv. . . . .	34
2.1	Global flightpaths: an example of a spatial network. . . . .	43
2.2	Latent space model of voting in the US house of representatives. . . . .	45
2.3	An example of a Random Geometric Graph. . . . .	47
2.4	Phase transition in the size of the largest connected component, and clustering coefficient in an RGG. . . . .	48
2.5	Example of curved spaces. . . . .	49
2.6	Positive and negative curvature on 2D surfaces. . . . .	50
2.7	Hyperbolic tiling in the Poincaré disk. . . . .	51
2.8	A random geometric graph in the hyperbolic plane. . . . .	53
3.1	Lorentz boosted vectors in Minkowski spacetime. . . . .	61
3.2	Diagram of a causal set and an element's future and past. . . . .	66
3.3	Diagram an interval in a causal set. . . . .	67
3.4	Causal set sprinkling into Minkowski spacetime. . . . .	69
3.5	Lorentz invariance of a Poisson point process compared to that of a regular lattice. . . . .	71
4.1	Transitive reduction and closure . . . . .	77
4.2	Transitive reduction of the hep-th citation network. . . . .	81
4.3	Transitive reduction of SCOTUS citation network. . . . .	82
4.4	Transitive reduction of US patent citation network. . . . .	83
4.5	Universality in citation network degree distributions before and after transitive reduction. . . . .	85
4.6	Effect of transitive reduction on degrees of nodes in hep-th citation network. . . . .	86
5.1	Examples of order dimension. . . . .	91
5.2	Example of chains in causal sets. . . . .	94
5.3	Volume and distance scaling in causal set graphs. . . . .	95
5.4	Data collapse of volume and distance scaling in causal set graphs. . . . .	96
5.5	Example calculation of the Midpoint-scaling dimension estimate. . . . .	98

5.6	Estimated fractions of k-chains used in the Myrheim-Meyer dimension estimator. . . . .	100
5.7	Myrheim-Meyer and midpoint-scaling dimension estimator on causal set graph. . . . .	102
5.8	Dimension estimates for real networks. . . . .	106
5.9	Causal set graph with scale-dependent dimension. . . . .	108
5.10	Estimated dimension for a toroidal causal set graph. . . . .	109
6.1	1 + 1 De Sitter spacetime embedded in 2 + 1 Minkowski spacetime. . . . .	115
6.2	Causal set graphs in intervals of de Sitter spacetime. . . . .	117
6.3	Expected number of chains in de Sitter spacetime causal set graphs. . . . .	120
6.4	Dimension estimates on de Sitter causal set graphs. . . . .	120
6.5	Estimated curvature for de Sitter causal set graphs. . . . .	122
6.6	Curvature estimates for real networks. . . . .	126
7.1	Graph embedding in Euclidean space using MDS. . . . .	130
7.2	Example of the naive spacelike distance estimator. . . . .	132
7.3	MDS eigenvalues of points in Minkowski spacetime. . . . .	138
7.4	MDS eigenvalues for causal set graphs and random DAG. . . . .	139
7.5	MDS eigenvalues for real networks. . . . .	141
7.6	Embedding of a small causal set graph. . . . .	143
7.7	Embedding of most cited papers in the hep-ph citation network. . . . .	146
7.8	Embedding of most cited papers in the hep-th citation network. . . . .	147
7.9	Sensitivity and specificity of network embeddings. . . . .	148
7.10	AUC values for network embeddings. . . . .	149
C.1	Myrheim-Meyer dimension estimates for real networks, for $k = 2, 3, 4$ . .	184



### ACKNOWLEDGMENTS

I would not have been able to even begin this research, let alone complete this thesis without the help of many colleagues and friends at Imperial College.

The Complexity and Networks group has provided the perfect environment for me to learn to think, read, and write about all sorts of things. Professors Henrik Jensen and Kim Christensen deserve thanks for constantly reminding our research group that, however broad and diverse we think our research interests are, there is a lot more out there. I doubt there are many places one could study where the weekly seminars feature mathematicians and artists, physicists and musicians, biologists, engineers, bankers, computer scientists and more or less everything else in between, with each of them getting a fair hearing and thorough scrutiny. Whoever was responsible for securing our office on the top of the EEE building ensured that there was always a good view to stare at while sitting and thinking about some problem or other. Thanks to Kim especially for helping to convince me to study here and securing the funding allowing me to do so.

Thanks to my fellow students in the group. Those who have already led the way ahead of me, Drs Kishan Manani, Alvis Tang, Kris Broga, Duccio Piovani and Eduardo Viegas. For the company during lab demonstrations, the raised eyebrow during a seminar reassuring me I was not the only one who didn't understand, for encouraging me to do things I would have regretted not doing, I am very grateful. I'll remember the trips to Lucca and La Herradura especially; they helped keep everything else in important context.

And thanks to those fellow students who surely clear this hurdle soon, Michael Kitromilidis, Vaiva Vasiliauskaite, Katharina Brinck, Saoirse Amarteifio, Qing Yao and Andrea Calandruccio. For all of the questions and answers, new projects, and for making sure that if whatever I was supposed to be doing had become too tedious or difficult that there was always something else to work on instead.

Thanks to those who always had a good word of advice, Drs Jelena Gruic, Paul Expert, Gunnar Pruessner, and the assessors who helped me during my time here, Professors Fay Dowker and Nick Jones.

Finally and most importantly, thanks to my primary supervisor on this project, Dr. Tim Evans. I owe him real debt of gratitude for taking me on as a student in the first place, for being willing to give me his good ideas to work on as well as being willing to entertain my own bad ones. I think I also owe him approximately 400 cups of coffee which I will endeavour to slowly repay in the coming years. The many afternoons spent attempting to prove things on the whiteboard in your Huxley office were some of the most enjoyable days of work during my time here.



## List of Publications

JC, J. G. Gollings, T. V. Loach, T. S. Evans.  
Transitive Reduction of Citation Networks,  
*Journal of Complex Networks*, 3(2):189-203, 2015

JC, T. S. Evans.  
What is the Dimension of Citation Space?  
*Physica A: Statistical Mechanics and its Applications*, 147(4):758-778, 2016

JC, T. S. Evans.  
Embedding networks in Lorentzian Spacetime  
Awaiting publication,  
Preprint: <http://arxiv.org/abs/1602.03103>, 2016

# Definitions and Notation

$\mathcal{G}$	A graph, consisting of a set of vertices and a set of edges.
$\mathcal{V}(\mathcal{G})$	The vertex set of $\mathcal{G}$ .
$\mathcal{E}(\mathcal{G})$	The edge set of $\mathcal{G}$ .
<b>A</b>	The adjacency matrix of a graph.
<b>RGG</b>	Random Geometric Graph.
<b>DAG</b>	Directed Acyclic Graph.
$\mathcal{P}$	A partially ordered set, or poset.
$\mathcal{C}$	A causal set, or causet.
$\prec$	The binary relationship in a poset or causet, called ‘precedes’.
$\mathbb{E}^D$	D-dimensional Euclidean space.
$\mathbb{H}^D$	D-dimensional Hyperbolic space.
$\mathbb{M}^{d+1}$	$(d+1)$ -dimensional Minkowski spacetime.
$d\mathbb{S}^{d+1}$	$(d+1)$ -dimensional de-Sitter spacetime.
$  \mathbf{u}  _{\mathbb{M}}^2$	Pseudo-inner product in Minkowski spacetime.
<b>TR</b>	Transitive reduction.
<b>TC</b>	Transitive closure.
<b>MMD</b>	Myrheim-Meyer dimension.
<b>MPSD</b>	Midpoint-scaling dimension.
<b>MDS</b>	Multidimensional scaling.
<b>AUC</b>	Area under receiver-operator curve.

# Chapter 1

## Introduction

The scientist is not a person who gives the right answers, he's one who asks the right questions.

---

Claude Levi-Strauss

Physics is traditionally reductionist in its approach. Its goal is to describe the underlying nature of reality, and in attempting to do that scientists seek to find explanations for the phenomena they observe in terms of something simpler. This ultimately leads to descriptions in purely mathematical terms, with as little else left as possible. Experiments must remove, or at least account for outside factors so that one process can be observed and understood without others interfering. Any success in this task is a great achievement. We've no right to expect that there are even underlying simple rules that are possible to deduce from what we're capable of observing. Nonetheless, most of the history of physics has focussed on answering this question: 'how can we deduce simple rules from observing the complicated world?'

It is easy to forget though that even if we know the underlying simple mathematical laws (we certainly do not know all of them yet, if they exist at all), the world we observe is still complicated. This tension between the simple rules and the complicated observations goes both ways. So a second question remains: 'how does the complicated world emerge from those simple laws?'

I think that second question is just as fundamental as the first. These complicated

things are the ones we want to know about. Complex systems like the human brain, the Internet, cities, or the global scientific process; these are all things that are worth studying because they are important to how we live our lives, but to leave it at that is to undersell them. Knowing how such complexity arises from simple rules is just as interesting for its own sake as the underlying mathematics and fundamental physics are.

There's a sense in which the first question, that of deriving the simple laws, is always going to be easier to answer, since if the type of answer you're looking for is 'simple' then the space of possible answers is smaller than for 'complicated'. In the case of complex systems, and the understanding of their emergent properties, it's not even clear what the type of answer we're seeking is. Maybe there will be elegant and simple explanations that tell us how complex brain function emerges from simpler underlying processes, and maybe there won't be. We are probably still at the prior stage of that scientific process, that of working out which observations are going to be useful, and cataloguing the universalities and differences of various systems of interest. Understandably then, widely applicable explanatory theories for complex systems can be hard to find. This is nothing to be upset about though, in my opinion at least. These exploratory steps in the dark, where we neither know what kind of things we are looking for or what we might find can be the most exciting, even if the lack of overarching conclusions can be frustrating.

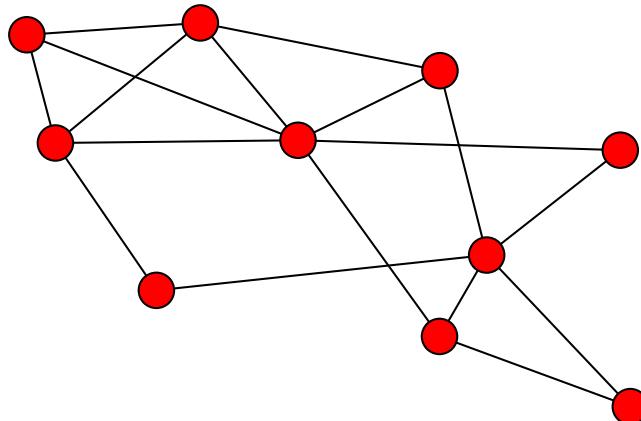
Studying complex systems is an inherently interdisciplinary task because complexity arises in so many different domains. Interdisciplinarity provides an opportunity for progress because by drawing together techniques and tools from mathematical and physical theories, and data from fields where those theories have not been previously applied, we can make new observations and learn more about both the tools and the data. This is the approach taken in the coming chapters. The data, as discussed later in this chapter, is primarily from social systems of interaction between individual people, in the form of communication and citation networks. The mathematical theory is that of graphs, networks, spacetime geometry and causal sets.

## 1.1 Graphs

A variety of approaches have been developed in the study of complex systems, each with their own history and nomenclature, overlapping to varying extents. Let us begin by introducing the *network approach* to studying complex systems. The premise is simple enough; consider a complex system as a network of interacting parts, and then study the structure, or topology, of the pairwise<sup>1</sup> interactions between these smaller simpler components. This structure is mathematically described as a graph, denoted  $\mathcal{G}$  which is a collection of vertices, or nodes,  $\mathcal{V}$ , representing the small components, and a collection of edges, or links,  $\mathcal{E}$ , the pairwise relations, and so we write

$$\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}. \quad (1.1)$$

A simple example is shown in Figure 1.1. Abusing notation slightly, we will also denote the vertex set and edge set of  $\mathcal{G}$  respectively by  $\mathcal{V}(\mathcal{G})$  and  $\mathcal{E}(\mathcal{G})$ .



**Figure 1.1:** An example of a graph with 10 vertices and 16 edges. Although the graph here is drawn in a particular way, with  $x$  and  $y$  coordinates assigned to each vertex, it is important to note that this is simply for convenience in illustrating the graph. There is no special significance to the way the graph is drawn (unless stated otherwise, as discussed in later chapters) and the drawing does not change the mathematical structure of the graph which consists *only* of which vertices are neighbours with which others.

---

<sup>1</sup>Interactions of more than two components can be considered in generalisations of graphs such as hypergraphs or simplicial complexes [214], but won't be discussed here.

If a vertex  $v$  is in the graph, then  $v \in \mathcal{V}$ . If there is an edge between vertices  $u$  and  $v$  then the pair  $(u, v) \in \mathcal{E}$ , and we will call  $u$  and  $v$  neighbours. Graphs can be *undirected*, meaning that each edge  $(u, v)$  is an unordered pair, or *directed*, meaning  $(u, v)$  is ordered and so different from  $(v, u)$ . Here we will only be considering *simple graphs*, which have at most one of each edge, and in which there are no self-loops, meaning  $(u, v) \in \mathcal{E} \implies u \neq v \forall u, v \in \mathcal{V}$ .

This structure can be specified in matrix form, by the graph's *adjacency matrix*,  $\mathbf{A}$ . If  $\mathcal{G}$  has  $N$  vertices, then  $\mathbf{A}$  is an  $N \times N$  matrix. Labelling each vertex with a unique index,  $i \in \{1, 2, \dots, N\}$ , so that the index of vertex  $u$  is  $i(u)$ , we then say that  $\mathbf{A}_{i(u), i(v)} = 1$  iff  $(u, v) \in \mathcal{E}$  and 0 otherwise. It immediately follows from this definition that in undirected graphs  $\mathbf{A}$  is symmetric, since then  $(u, v) \in \mathcal{E} \implies (v, u) \in \mathcal{E}$  and so  $\mathbf{A}_{i(u), i(v)} = \mathbf{A}_{i(v), i(u)}$ . Since we are not considering graphs with self loops the elements on the diagonal of  $\mathbf{A}$  are always 0.

There are many ways of assigning these indices as labels to the vertices of a graph, and so many different adjacency matrices can represent the same graph. Furthermore, two graphs can be different (by having different vertex sets) but still have the same structure, if there is a correspondence between their edges and vertices. Two such graphs are called *isomorphic* when this is the case. More formally, two graphs  $\mathcal{G}$  and  $\mathcal{H}$  are isomorphic iff there is a bijection

$$f : \mathcal{V}(\mathcal{G}) \rightarrow \mathcal{V}(\mathcal{H}) \quad (1.2)$$

such that

$$(u, v) \in \mathcal{E}(\mathcal{G}) \iff (f(u), f(v)) \in \mathcal{E}(\mathcal{H}) \quad \forall u, v \in \mathcal{V}(\mathcal{G}). \quad (1.3)$$

This is to say that the vertices of  $\mathcal{G}$  can be relabelled with the names of the vertices of  $\mathcal{H}$  in such a way that the relabelled graph has exactly the same edge set as  $\mathcal{H}$ . The focus of this thesis is on the structure of graphs and so if two graphs are isomorphic then they are, for our purposes, effectively the same. Of course, any measurement of graph structure, including those developed in later chapters, must be the same for two

isomorphic graphs since the names of the vertices are irrelevant.

The mathematical study of graphs has a long history, with Euler's work on the bridges of Königsberg widely regarded as founding the field [39]. Famously, by considering the graph representing the islands (vertices) and bridges (edges) of the city, Euler demonstrated that it was not possible to make a journey crossing each bridge exactly once.

Since these early beginnings, graph theory has grown enormously and is now a well established mathematical field. It is used in topology, for example in Euler's formula relating the numbers of vertices, edges and faces on polyhedra [39]. Graphs are used to describe molecular structure in Chemistry [21], particle interactions in Quantum Field Theory [138], and are used in Quantum Information theory [117].

Graph theory has posed some of the most well known problems in mathematics such as the famous 'four colour problem' [16], which asks whether a plane divided into contiguous regions can have those regions coloured such that no bordering regions share a colour. This problem, first raised in the mid-19<sup>th</sup> century took more than 100 years and a controversial computer assisted proof to solve [96].

Computer scientists also commonly use graphs. Many useful data structures and algorithms are described in graphical terms [120]. Graph problems appear often in the study of computational complexity<sup>2</sup> with many famous NP-complete problems posed in the language of graph theory, such as the sub-graph isomorphism problem [195] or the graph colouring problem [208]. Indeed, the problem of determining whether two graphs are isomorphic to each other or not remains a very active question of current study [18].

Graph theory allows us to answer questions about particular graphs and their structure, but more is needed to describe complex systems. These complex networks can be extremely large and so need to be described not just microscopically but also statistically. We are not just interested in the graph that exactly represents the structure of, for example, the Internet, even if such a thing can be found in a reasonable amount

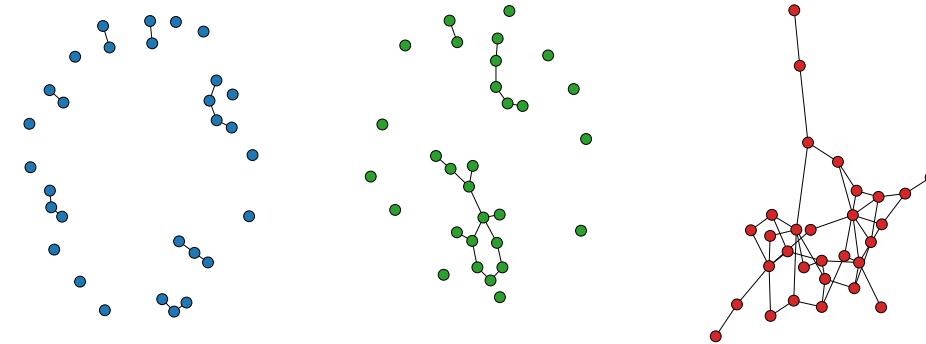
---

<sup>2</sup>Despite the similarities in their names, 'complexity science' and 'computational complexity' are distinct fields of study.

of time, or be said to exist at all. Rather we are interested in the *kinds of graph* that the Internet is like. Physicists do not attempt to describe the position and motion of every particle in a room to describe its properties, but instead describe them statistically and discuss useful observable properties of the particles (such as temperature and pressure) in terms of these statistics. *Network science* takes the same approach to studying these large graphs. We calculate statistical properties of their structure and choose interesting and meaningful statistics to observe and measure in order to characterise the system in question.

The development of random graph theory and the statistics of graphs by Erdős and Renyi begins this advancement [38]. The model they introduce is extremely useful and provides a good example for the introduction of the discussion of statistical properties of graphs. An Erdős and Renyi (ER) random graph,  $\mathcal{G}_{\text{ER}}(n, p)$  has  $n$  vertices, and each of the  $\frac{n(n-1)}{2}$  possible edges is present with a constant probability  $p$ . The ER graph is very useful as a simplest null model and if we know nothing about the structure of a graph other than how many vertices and edges it has, an ER graph is a good place to start. Exponential random graph (ERG) theory provides a rigorous method of defining the ensemble of graphs with certain average properties or constraints that otherwise maximises entropy [168]. The ER model can be thought of as the simplest ERG where only the number of nodes and expected number of edges is specified.

However, for even moderately large  $n$  it would be neither practical nor helpful to describe such a model by specifying the probability of each possible graph it could create since there are so many [5]. Instead, we can describe it statistically. We can measure *graph observables*, which are measurable properties of the graph, and characterise the model with their statistics. Of course, there are many such observables one could think up, and the difficulty lies in creating observables which are meaningful for the system the graph represents. For example whether a graph has an odd or even number of nodes might be something we can easily measure, but if the network in question is a social network, knowing the parity of a person's number of friends does not tell us how popular they are.



**Figure 1.2:** Three ER graphs with  $n = 30$  and  $p = 0.03, 0.05, 0.1$ . On the left, most pairs are not connected by an edge, and those that are form small disconnected groups, with no cycles. In the middle, we approach the critical probability,  $np = 1$  for a giant connected component of the graph, with connected vertices mainly forming one large component which contains a cycle. On the right, we are well above the critical probability and all vertices are part of one component which contains many cycles.

Let us consider some more useful graph observables. Remember,  $\mathcal{G}_{\text{ER}}(n, p)$  is not one particular graph but is a random variable that describes a whole ensemble of graphs each with an associated probability, so these observables will have some associated probability distribution. The simplest observables we can consider are the number of vertices, which we will denote as  $N(\mathcal{G}) = |\mathcal{V}(\mathcal{G})|$ , and edges,  $E(\mathcal{G}) = |\mathcal{E}(\mathcal{G})|$  in the graph, where  $|\cdot|$  denotes the cardinality of a set. It is clear that  $N(\mathcal{G}_{\text{ER}}(n, p)) = n$  with probability 1. The number of edges,  $E$  follows a binomial distribution since each of the possible edges is present independently with the same probability and so

$$E(\mathcal{G}_{\text{ER}}(n, p)) \sim \text{Binomial}\left(\binom{n}{2}, p\right). \quad (1.4)$$

Let us define a *path* of length  $\ell$  as a sequence of  $\ell + 1$  vertices,  $v_1, v_2, \dots, v_{\ell+1} \in \mathcal{V}(\mathcal{G})$  such that each of the  $\ell$  pairs  $(v_1, v_2), (v_2, v_3), \dots, (v_\ell, v_{\ell+1}) \in \mathcal{E}(\mathcal{G})$ . Such a path is a *cycle* if the first and last vertex are the same, that is  $v_1 = v_{\ell+1}$ . A vertex  $u$  is *reachable* from  $v$  if there is some path containing both of them, and a *connected component* is a set of vertices which are all mutually reachable. In a directed graph, paths must also respect the direction of the edges. A *strongly connected component* is one in which the all vertices are mutually reachable when accounting for the direction of edges, and a *weakly connected*

*component* is one in which all vertices are mutually reachable when edge direction is not accounted for.

A useful graph observable is the expected number of vertices in the largest (in the sense of containing the most vertices) connected component of the graph. We can then see in Figure 1.2 that the ER model displays a phase transition around a critical probability below which there is no large component (more precisely, the probability of a component containing a finite proportion of the vertices goes to 0 as  $n$  tends to infinity) and above which such a component emerges.

The *degree*,  $k(u)$  of a vertex  $u$  in a network is the number of neighbours of  $u$ .

$$k(u) = |\{e \in \mathcal{E} : u \in e\}|. \quad (1.5)$$

This can be calculated from the graph's adjacency matrix by summing across rows or down columns. In undirected graphs, the symmetry of  $\mathbf{A}$  means either one gives the same result, but in directed graphs this is not the case. We will refer to the number of edges in which a vertex is the first of the pair as its *out-degree*,  $k_{\text{out}}$  and the other case as its *in-degree*  $k_{\text{in}}$ . Using the graph's adjacency matrix we can define them as

$$k_{\text{out}}(u) = \sum_j \mathbf{A}_{i(u),j}, \quad (1.6a)$$

$$k_{\text{in}}(u) = \sum_j \mathbf{A}_{j,i(u)}. \quad (1.6b)$$

The distribution of degrees in a network tells us about its structure. Regular graphs are those in which all vertices have the same degree. In a typical ER graph the degree distribution is binomial. Generally we will talk about a graph's degrees in terms of a probability distribution,  $p(k)$  describing the probability of a vertex having a degree equal to  $k$ .

## 1.2 Network Science

The recent explosion of interest in network science means that the discipline is now far too large, and too rapidly evolving for a fully comprehensive review here but various textbooks such as those of Newman [146] and Barabási [22] provide well-rounded introductions. Instead I will here highlight some key results and ideas, many of which have led to separate fields of study important in their own right, which will be referenced or used in later chapters.

One of the early discoveries of network science was that many real complex networks from physical and social systems with apparently little in common such as the Internet, protein interactions and academic collaborations [147] share a *power-law* or *scale-free* degree distribution where, for large  $k$ ,  $p(k) \sim k^{-\gamma}$  and the exponent  $\gamma$  is typically around 2 or 3 [23]. In ER graphs the binomial degree distribution means that the probability of a high degree decays faster than exponentially, making it very unlikely that a vertex exists with a degree much higher than the mean degree  $\langle k \rangle$ . But in these scale-free graphs, when  $\gamma < 3$  the second moment of the degree distribution diverges representing the fact that a small number of nodes with very high degree, often called *hubs* appear. For many networks and their associated dynamics these hubs can be extremely important. In disease spreading processes for example this divergence of the degree distribution's second moment means that the probability of disease transmission required for an outbreak becomes arbitrarily small because these hubs are connected to so many other nodes that they can spread the disease process throughout a large fraction of the network should they get infected [157, 150]. In the network representing the Internet, these hubs are a small number of websites which take a significant fraction of the world's total Internet traffic. It has been shown that this scale-free topology makes the network more resilient against random attacks (removal of nodes) but more susceptible to targeted attacks against these important hubs [193].

Power-laws appear often in statistical mechanics [56] and in the study of self-organisation [20]. A simple and elegant explanation for power-laws in degree distributions is the famous *preferential attachment* [24] or *cumulative advantage* [160] mechanism.

The principle of nodes gaining edges proportional to their current degree can be explicit (such as in the BA model) or self-organising (for example in [77] where new edges appear from random walks on the graph). The ubiquity<sup>3</sup> and importance of power laws has led to numerous suggestions of other mechanisms and processes which generate them [149, 58].

Counting a vertex's degree is the simplest measure of local structure and arguably the next simplest is clustering<sup>4</sup>. Clustering measures how likely two neighbours of a given vertex are to be neighbours of each other. We define the *clustering coefficient* for node  $u$  to be:

$$c_u = \frac{|\{(v, w) \in \mathcal{E}(\mathcal{G}) : (u, v) \& (u, w) \in \mathcal{E}(\mathcal{G}), v \neq w\}|}{k(u)(k(u) - 1)} \quad (1.7)$$

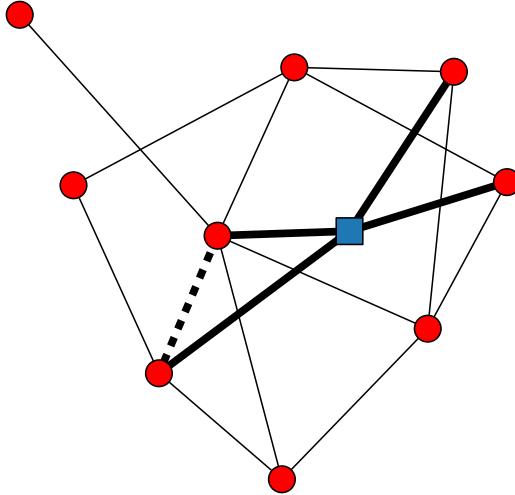
which is simply the fraction of pairs of neighbours of  $u$  which share an edge. Similarly, the *global clustering coefficient*, counts the fraction of connected triples which form a triangle in the graph. Graphs with local structure, such as some simple lattices usually have high clustering coefficients, as we'll explore in more detail in chapter 2, while ER random graphs have low clustering [146] since the probability of edges being present is independent.

The *diameter* of a graph is the longest of all the shortest paths between each pair of vertices. The *Small World* effect is the phenomenon that in many complex networks the diameter, and average shortest path length between two nodes is very small compared with the number of nodes in the network, as famously discussed in [205, 151]. ER graphs are small world graphs since their diameter grows logarithmically with the number of vertices in the graph,  $N$ . Lattices on the other hand commonly have large diameters which scale as some power of  $N$ . Many complex networks are also small world, a result which was initially surprising but has now become well known [146].

---

<sup>3</sup>It should be mentioned that there is a lot of research and dispute on methods of fitting power-law distributions [92, 58] and so one must be careful before claiming to have found another instance of this 'universality'.

<sup>4</sup>Clustering is sometimes also referred to as *transitivity* although this term will be used in a more specific way when talking about directed acyclic graphs in chapter 4 and so is avoided here.



**Figure 1.3:** The highlighted square node has a degree of four (corresponding to the four bold edges). Its four neighbours could have a total of six edges between them, of which one is present (the dashed edge). Its clustering coefficient is therefore  $1/6$ . The diameter of this graph is three, since when the shortest path between each pair is computed, the length of the longest such path is three.

This is interesting because these networks seem to have the highly clustered property of a lattice, but also the small world property of a random graph, so lie somewhere in between these two simple models.

A common task in network science is to use the structure of a network to decide which nodes are ‘important’, ‘influential’ or ‘central’ in some way. Clearly many possible approaches to this *network centrality* idea can exist depending on which definition of ‘important’ one uses. For example, if shortest paths are particularly relevant in a network, then *betweenness centrality* [45] which counts how many shortest paths travel through a node may be useful, or *closeness centrality* which measures how far a given node is from all others. If important nodes share edges with other important nodes then a self-consistent answer based on the spectral decomposition of the adjacency matrix, such as *eigencentrality* [44] of the network might be relevant.

Degree and clustering are microscopic measures defined for individual nodes, whereas measures like the diameter, or connectivity are defined for the whole graph

and so are macroscopic. Mesoscopic measures lie in between, measuring structures larger than the scale of individual vertices but smaller than the entire graph. Community structure is mesoscopic in this way and the definition and detection of communities in graphs is now a well established field. The general idea is to partition the graph's vertices into collections called communities such that edges between vertices in the same community are likely and edges spanning communities are rare. Different methods of measuring these likelihoods give rise to different methods of community detection; see the many reviewed in [80] for some examples. Furthermore some methods allow vertices to be in many communities at once [76], some provide different communities at different scales [179] and many other approaches exist with specific relevance to specific applications. In networks where we expect 'similar' nodes to share edges and 'dissimilar' nodes to not share edges, community detection methods provide a way of partitioning the nodes of the network into groups which should then share common features.

Various mathematical extensions to the network approach have been used where required by the particular features of the complex system being modelled. Edges can be associated with a *weight* representing some cost or distance and most of the methods discussed above are easily generalisable to this context [146]. Different types of edges might be represented in different layers of a network (called a *multi-layer* or *multiplex* network). Processes on multilayer networks can have richer dynamics than single-layer networks so for systems which are naturally multilayered (such as a transport network where each layer represents a different mode of transportation) this kind of modelling can be extremely useful [152, 144].

In systems which evolve in time it can be useful to consider edges as appearing and/or disappearing at certain times. These *temporal networks* [109, 134] can model air travel where edges representing flights only occur at certain times, or communication networks where different people talk to each other at certain times of the day.

All of these approaches have their uses but it is important to make sure that the right techniques are used for the right task. There is a balance to be struck between

the impressive universality of many network based approaches and the need for application-specific knowledge for the interpretation of the results. In the next section I will introduce a special class of graphs and then explain why their particular properties and constraints suggest that a geometric approach may be appropriate.

### 1.3 Directed Acyclic Graphs

The focus of this thesis is a special class of networks, those which form Directed Acyclic Graphs (DAG). These are directed graphs, in which there are no directed paths which contain the same vertex twice (a cycle)<sup>5</sup>; see Figure 1.4, Figure 1.5, and Figure 1.6 for examples. Such networks appear in many systems, such as scheduling problems [116], logic flow in spreadsheets [136], dependency in software libraries [130], artificial neural networks [177], family trees and the case we use as illustration in later chapters, citation networks [62].

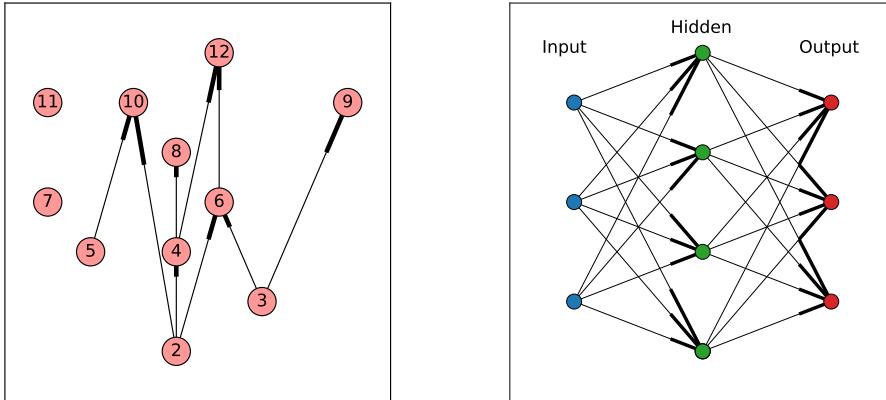
In general, a network will form a DAG when there is some kind of ordering on the nodes which the edges respect [113, 114, 188]. For example, in the case of a family tree, a parent must be older than their child. In scheduling problems tasks must be completed in some particular order. An ordered list of the nodes of the network in which edges only go from earlier nodes to later nodes (or vice versa) is called a *topological sorting* of the graph or, in the language of partially ordered sets, a *linear extension*. One graph may admit many linear extensions, and they can be found efficiently by well known algorithms [15]. This partial order<sup>6</sup> on the vertices is a very useful property and a constraint around which we will later build a lot of analysis.

The adjacency matrix  $\mathbf{A}$  of a DAG also has special properties. It may be written in upper-triangular form, by making sure that the indices of the nodes respect a valid

---

<sup>5</sup>This is the common nomenclature but it is not ideal, as ‘directed acyclic graph’ could also refer to a graph which is acyclic when edge direction is not accounted for (usually called a tree), that has been given directed edges. In [114] ‘directed ordered graphs’ is suggested as more appropriate, but the authors note that the literature has already settled on a name. From now on then, acyclic means that there are no directed cycles, but there may still be cycles in the undirected version of the graph.

<sup>6</sup>See chapter 3 for the definition and further discussion of partial orders.



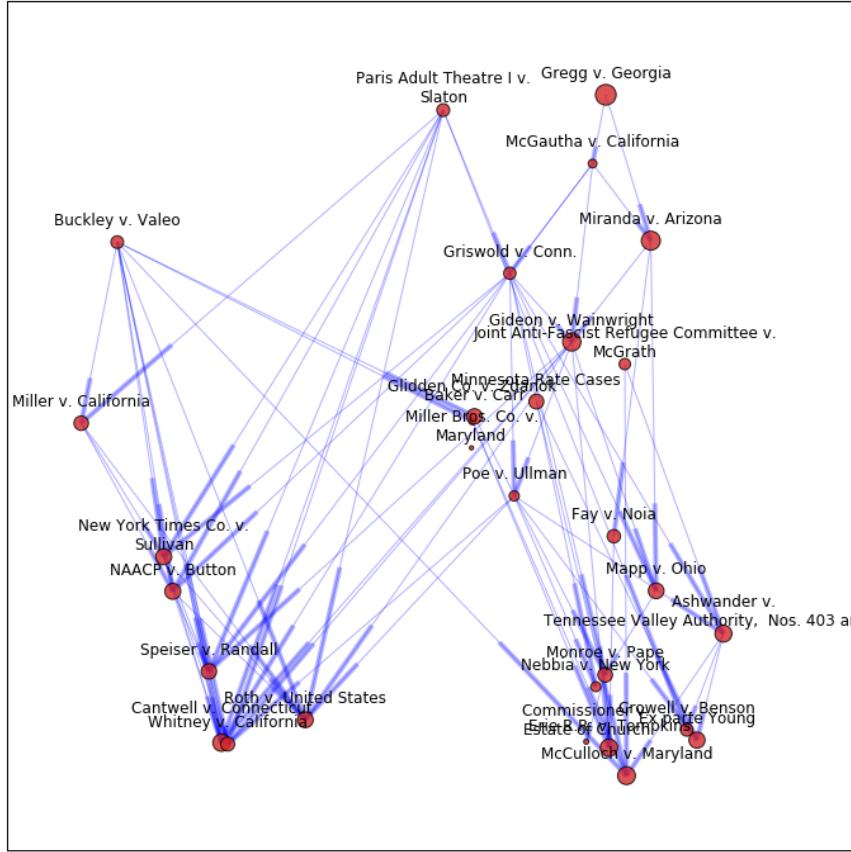
**Figure 1.4:** Examples of small directed acyclic graphs (DAGs). Edge direction is indicated by the thicker end of each edge. Left: A graph showing the factorisation relations of integers up to 12. The order on the vertices is their size, since a number can only be a factor of a larger number and this property ensures the graph is acyclic. Right: A schematic of a neural network, as in [107]. Information propagates from an input layer on the left, through a hidden layer, to an output layer on the right. The direction of this information flow determines the order on the nodes which ensures the graph is acyclic.

topological sorting of the graph.

$$i < j \implies A_{i,j} = 0 \quad (1.8)$$

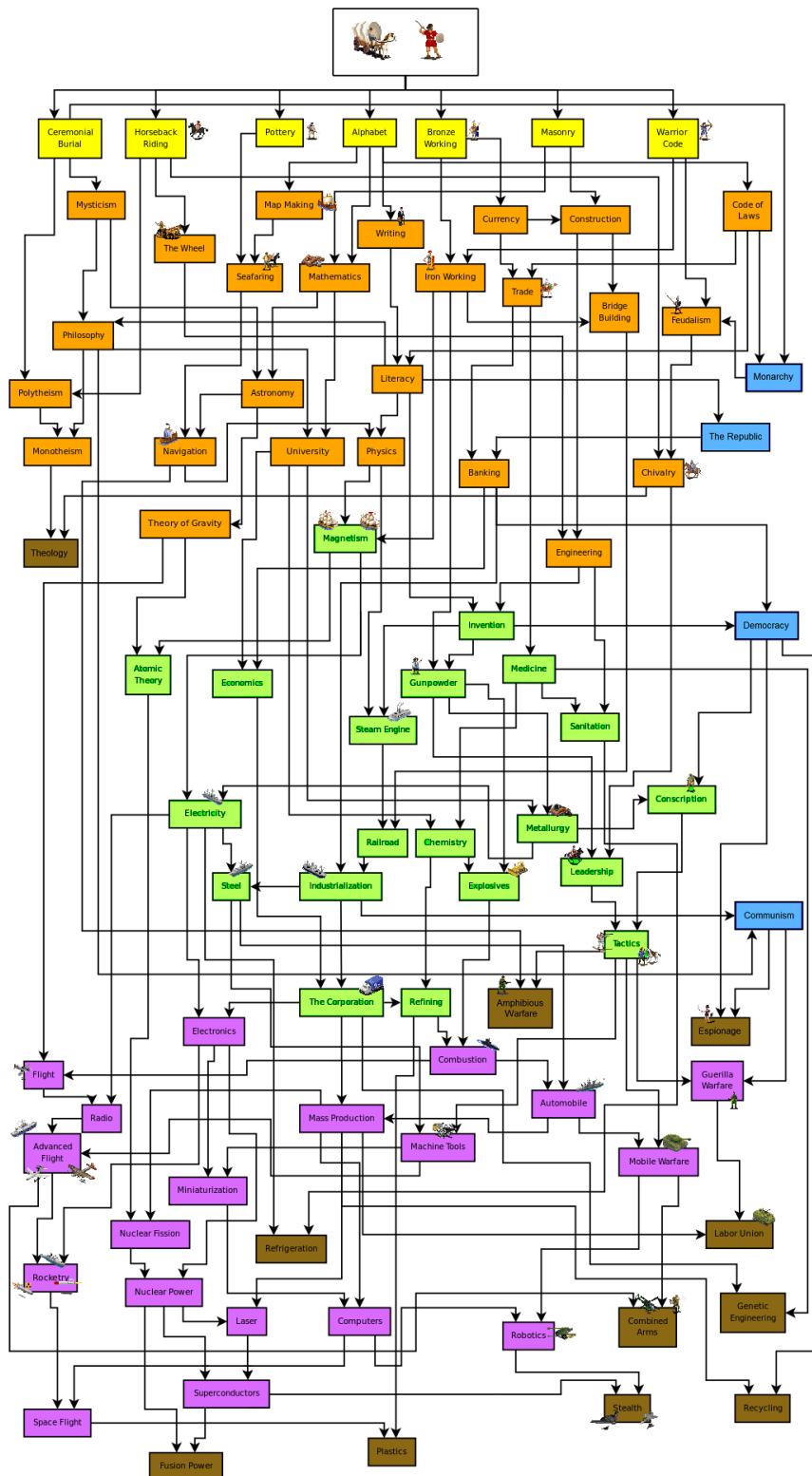
Because  $A$  is upper-triangular with 0 on the diagonal all of its eigenvalues are zero [137]. DAGs have no strongly connected components (aside from individual vertices) because if there is a directed path from  $u$  to  $v$ , we know there cannot be a directed path from  $v$  to  $u$  and so these vertices are not mutually reachable. Therefore when talking about components in a DAG, the meaning is always that of weakly connected components.

A common theme in network science is taking note of particular aspects of network structure, or constraints, and developing novel ways of characterising networks which take these into account. For example, when considering the network of phone contacts in a country, we might notice that people who live near to each other are more likely to share a phone conversation. The structure of the network is dominated by this effect of local structure. But because we know about this effect, we can develop measures which account for it, and can then observe second-order effects, such as shared language.



**Figure 1.5:** An example of a DAG. This graph represents the citation network of the top cited judgements in the history of the US supreme court. The graph is acyclic because a judgement can only cite a previous judgement and so time forms an order on the vertices which the edges must respect. By drawing the graph such that the y-axis represents time we ensure all of the edges point the same direction in the axis (in this case, down). In chapter 7 are details of the exact method used for this visualisation.

In [78] this is done with the Belgian phone contact network, and after accounting for local effects, a clear divide between French speakers in the South and Brussels, and Flemish speakers in the North is visible. In the study of temporal networks, usual measures such as reachability need to be redefined to take into account the constraint of edges only existing at certain times [109] and in multiplex networks community structure [129] and failure cascades [99] need to account for the special properties of these networks in their theoretical descriptions. The aim of this thesis is to take this approach of structurally-aware measures and apply it to DAGs.



**Figure 1.6:** An example of a DAG from the technology tree of the computer game FreeCiv. Each vertex in the graph (the boxes in this diagram) are technologies that the player can research, and edges represent dependence on prior technologies. More advanced technologies in the game require simpler, earlier technologies to have been already researched.

## 1.4 Outline

My aim in this thesis is to introduce new methods and measures for DAGs. In the same vein as the examples above, this will be done by paying attention to the particular structural properties that this class of networks has, and developing tools which take these constraints into account.

We begin in chapter 2 by reviewing the ‘network geometry’ approach in network science. We will see how geometric analysis is used in network science and how simple geometric models can give rise to a surprising wealth of features in network structure. The later chapters are an extension of this geometric approach to DAGs, with their own appropriate geometry.

Chapter 3 will discuss the geometry used to characterise the structure of DAGs, which is the geometry of Lorentzian spacetime. We will review the differences between the Riemannian geometry that has been used in the past for network analysis, and the Lorentzian geometry which I will use here. We will see that there is a natural geometric interpretation for many networks forming DAGs because the acyclic property that defines them arises naturally from the causal structure of spacetime. Many of the tools I will use in later chapters have their basis in causal set theory, which is a theory of discrete Lorentzian spacetime. Causal sets will be defined and some of their properties introduced in this chapter.

Chapter 4 begins the analysis of DAGs in earnest, starting with the simple property of transitivity. This chapter is based on work published in [62] with Tamar Loach, Jamie Gollings and my supervisor Tim Evans. Algorithms performing transitive closure and reduction will be defined, and their effect on real networks discussed. We will see that even these simple techniques can helpfully characterise network structure and distinguish otherwise similar networks, as well as highlight particular nodes within them.

Chapters 5 and 6 focuses on characterising network structure by estimating the dimension and curvature of the spacetime with the same (or closest) causal structure as a given network. They are partially based on work published in [61] done with Tim

Evans. We will introduce and analyse several methods of discrete spacetime dimension estimation for flat spacetime, and measure the spacetime dimension of more real networks. This work is then extended by considering curved spacetime, and real networks are shown to be characterised by their curvature and how it changes at different length scales.

We then tackle the problem of finding an effective embedding for a network in a spacetime, in chapter 7. This chapter is based on work available as a preprint in [60] written with Tim Evans. In this chapter a method for finding a network embedding is developed and the level of correspondence between the network and spacetime quantified.

Finally, chapter 8 concludes by discussing the scope of this approach and the possible future work that can be developed.

## 1.5 Datasets

To illustrate the methods developed in later chapters it is useful to have a variety of real networks which form DAGs, and those used here are detailed in Table 1.1. Citation networks are an excellent example and will be the primary source of datasets used here. This is because most of the large, freely available datasets of networks forming DAGs are citation networks (as evidenced by their use as examples in much of the previous literature on DAGs). This is not to say that the approaches developed here are only applicable to citation networks, simply that at the time of writing the most interesting DAG datasets tend to be of this type.

A citation network is a network of documents (the nodes) where a directed edge goes from one document to another if the first cites the second in its bibliography. The graph is directed because the citation relationship is asymmetric; being cited is not the same thing as citing someone else. The graph is acyclic because there is an order on the nodes which the direction of the edges must respect, which is time. One document can only cite another which has already been written. It can be argued that in some cases this is not technically true because a document may be published at multiple different

times, and in different versions, and so a later version may cite something written after an earlier version was written. Furthermore, two documents may be written at the same time by authors who are in contact and so may mutually cite each other. We will refer to the citations which appear to cite something in the future as *acausal edges*, but it is worth noting at this stage that such edges are, in every dataset analysed here, a small minority of the total.

We will look at citation networks from three different fields. Firstly, academic papers. Our dataset comes from the arXiv ([arxiv.org](http://arxiv.org)), and is provided by [86]. The arXiv is an online repository of pre-prints of scientific papers in physics, mathematics, computer science and other related disciplines. We will look in particular at the physics section which (at the time of writing) contains most of the arXiv's papers<sup>7</sup>. For ease of analysis and to allow greater detail in the analysis of these papers the citation network is split into the subcategories used on the arXiv, and we will investigate six of the largest, spanning high energy physics (theory and phenomenology), cosmology and general relativity, astrophysics, quantum physics and condensed matter theory.

Citation networks of academic papers have been subjected to much study in recent years [127, 162, 91, 57]. One of the earliest and best known models is that of Price in [160], who was interested in the distribution of citation counts of papers (in network terms, the in-degree distribution). The power-law distribution he found and described with the mechanism of ‘cumulative advantage’ can now be seen as part of the network phenomenon of ‘preferential attachment’. Since then, various models of citation networks have been proposed [200, 75, 93, 165, 91], with the aim of understanding the mechanism behind the citation process, and the more general process of the flow of information between researchers. These networks are interesting datasets in their own right, being large networks with complex structure. But they also shed light on the scientific process itself, viewed as a complex system of individual agents (the scientists) interacting according to their own goals and beliefs, giving rise to a wider emergent phenomenon (the body of knowledge they collectively create). A better understanding

---

<sup>7</sup>See the arXiv mirror site <http://front.math.ucdavis.edu> for a breakdown by section.

of the scientific process may help researchers of all kinds to better advance their fields, absorb and integrate the knowledge of others into their own work, and increase the impact and readership of their research.

Secondly, we use a network of court judgements (sourced from [82]) from the US Supreme Court (abbreviated as SCOTUS). When the court makes a decision, it cites its own previous decisions creating a citation network. In these court cases, precedent (the decisions previously made by the court) are very legally significant, and the structure of this citation network captures that [55, 81]. As is the case for academic papers, most court cases are rarely cited while others become far-reaching in their impact. Significant differences between this citation network and those from academic papers include the authorship, and timespan of publication. Unlike scientific papers, these court judgements are all written by essentially the same group of people, the judges on the Supreme Court at that time. Therefore, we can reasonably assume that the authors are all more or less aware of all of the previous documents in this corpus, either from having written them or from their presumably expert knowledge of past legal cases. Importantly this is not the case in academia, where most published papers are unknown to most other authors simply due to the vast number of journals and articles in them, and the fact that very new publications require time to be read by other scientists who will not be immediately aware of every new paper appearing in their field. Secondly, the timespan of these publications is unusually large, spanning more than two centuries; approximately the entire history of the USA. Clearly this means that there have been many different generations of judges on the court and so their collective citation behaviour may have changed from one era to the next. This is unlike the arXiv network which has a shorter history than the length of the typical career of a scientist.

Thirdly, we will use a network of citations between patents [100, 6]. When a new patent application is made, there is a legal obligation to cite previously existing patents so as to delineate which features are new innovations, and which already exist. Effectively each citation is giving up a claim to a potentially lucrative invention. This

is an interesting feature because inventors are unlikely to do this unless it is necessary, whereas authors of academic papers can add more citations to other papers at no cost to themselves and consequently may be willing to cite more freely. It is this kind of difference in citation dynamics that will hopefully be captured by the methods proposed in later chapters. There are still many similarities though as all of these citation networks share some common features such as having heavy-tailed degree distributions [162] (see chapter 4 for more details) and high clustering.

The second class of networks forming DAGs which we study here comes from communication networks. A network where nodes represent people, and time-associated edges represent communication between two people would usually be considered as a temporal network in the sense used in [109]. However, we can use a *line graph* transformation, as in [76] to turn such networks into DAGs. This transformation maps edges in the original graph, to nodes in the line graph. Nodes in the line graph are connected if the edges they represent in the original graph share a vertex. We will use a dataset of emails released from the recent investigation into Hillary Clinton consisting of 9255<sup>8</sup> emails between 513 different people over a period of five years. In the original network, each email represents an edge between people, and is associated with a particular time. In the line graph, each email is a node, and one email has a directed edge to another if the sender of the second is the receiver of the first, and is within some time window (we use one week). While it is not the focus of this thesis, it is worth noting that this line graph transformation, parametrised by the time window used to determine edge presence, allows any temporal network to be represented as a DAG and so opens up all of the analysis in later chapters to this important class of networks.

Finally, the FreeCiv technology tree graph in shown Figure 1.6 is also included, as an example of a DAG which is small enough to be visualised in a worthwhile manner, but large enough that the measures developed in later chapters may be meaningfully applied.

---

<sup>8</sup>The discrepancy between this value and that in Table 1.1 is because the largest weakly connected component is taken when analysing the network.

network	description	$N$	$\langle k_{in} \rangle$	Acausal edges	Source
hep-th	high-energy theory	60880	17.96	0.8%	[86]
hep-ph	high-energy phenom.	76836	17.63	0.8%	[86]
astro-ph	astrophysics	133346	14.1	0.9%	[86]
cond-mat	condensed matter	118496	6.46	1.6%	[86]
gr-qc	cosmology	27176	8.72	1.9%	[86]
quant-ph	quantum physics	36959	7.13	3.3%	[86]
SCOTUS	US Supreme Court	25378	8.52	0.2%	[81]
US Patent	US Patents 1970-1999	3774753	4.38	0.004%	[6]
Clinton	Line graph of emails	5674	18.82	0	[2]
freeciv	Freeciv tech tree	88	1.77	0	[1]

**Table 1.1:** Table of datasets analysed in this thesis. Acausal edges are those which go against an ordering of the vertices, such as a citation to a paper published in the future. These edges can be removed to ensure that the network dataset forms a DAG.

# Chapter 2

## Network Geometry

We can only see a short distance  
ahead, but we can see plenty there  
that needs to be done.

---

Alan Turing

Throughout its short history network science has had a strong connection to statistical mechanics. Research has focused on scaling [151, 176], criticality [11, 135], and dynamical processes [197, 178]. Classical problems in statistical physics such as percolation [94, 115], the Ising and Potts models [70, 163], random walkers [153, 175], Bose-Einstein condensation [33] and spreading processes [157], which are reasonably well understood for the regular lattices or Bethe lattices common in statistical mechanics were then studied on complex networks. More recently though, a complementary approach based on geometry and topology has emerged in the study of networks. This approach, usually called *network geometry* [31, 155] has led to a series of insights in recent years which will be reviewed in this chapter.

### 2.1 Spatial Networks

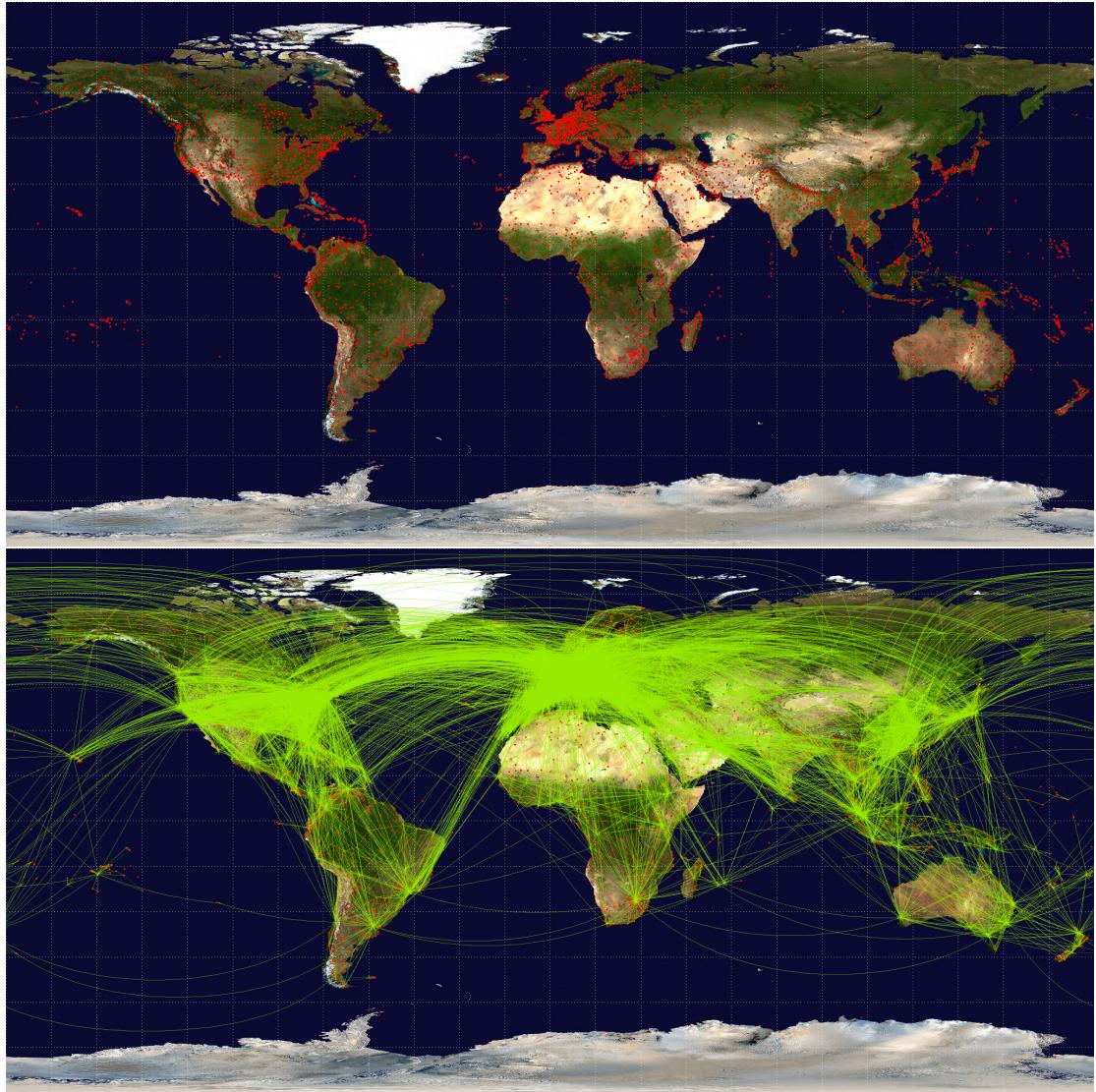
Many of the simplest random models in network science, such as the Erdős-Renyi random graph, or the configuration model [143] have in common the property that, given their degrees, no two nodes are more likely to share an edge than any other two. In

the absence of other information this may be the right thing to assume, and for some real systems it may be a reasonable approximation but for many others it is not. An interesting class of such systems are *spatial networks*, comprehensively reviewed in [27]. These are networks in which the nodes are associated with coordinates in some real space which has a metric, and the distance between two nodes in that space is related to the likelihood of finding an edge between them, with closer pairs more likely to share an edge.

This space is most often two-dimensional, and its metric Euclidean [27, 84, 198], but other examples exist [25, 17]. Well known examples include networks of transport links such as airline networks [207], roads [190], or cargo ships [112]. Many social networks can be considered spatial networks since people are more likely to be friends with those who live near them than those who live further away.

Many spatial networks are planar, meaning that their graph can be drawn on the plane such that no edges cross. This most often occurs in spatial networks as some natural consequence of the system the network represents. For example many road or rail networks are planar (or at least approximately so, accounting for bridges and tunnels). This constraint imposed by the system the network is modelling can aid analysis, since many particular results are known for planar graphs, such as the four colour theorem, Euler's formula (related the number of vertices, edges and faces) and Kuratowski's theorem (showing non-planar graphs must contain certain subgraphs). The constraints imposed by natural geometric rules on the structure of networks will be a continuing theme in later chapters.

Moving beyond the plane, spatial networks can exist in three-dimensional space. Most notably, the brain is a three-dimensional object and its neurons (nodes) and axons (edges) form a network which has been, unsurprisingly, of significant academic and medical interest [173, 54, 189]. Again, we find that nearby neurons are more likely to be connected to each other than those which are far away, which is an expected result of the fact that longer axons are more costly to grow, and signals along them take longer to propagate.



**Figure 2.1:** An example of a spatial network from global flightpaths. Airports (top) and connecting flights (bottom) in the global airline network, sourced from [4]. Strong spatial clustering is apparent in both images. Edges are most commonly between near neighbours but edges spanning longer distances are also clearly visible, such as the clear band crossing the North Atlantic.

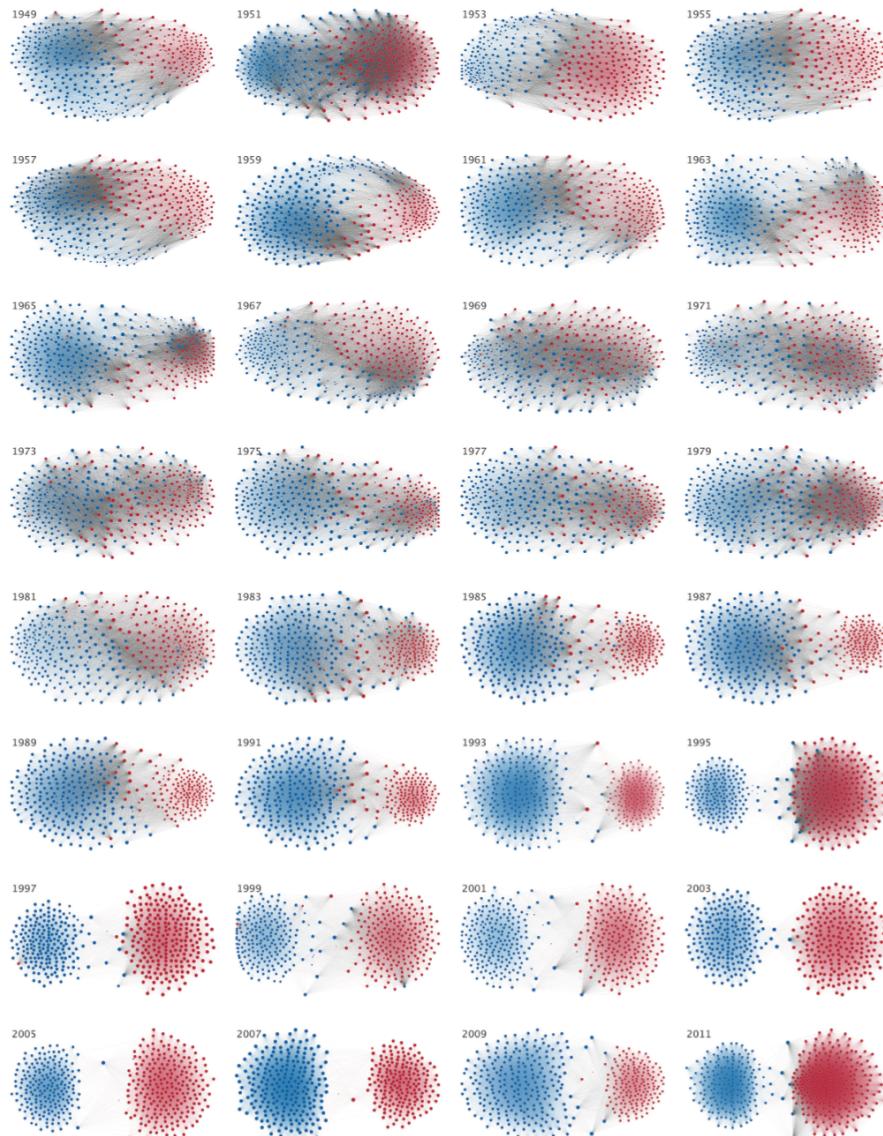
As well as the dimensionality of the space its topology can also affect the structure of a spatial network on it. The network of roads in Britain may be well modelled as a spatial network on the two dimensional plane, but the spatial network of airline flights must account for the fact that the world is a sphere. For a clear example of the effect of topology on network structure, note that the famous four-colour theorem is dependent on the topology of the surface the graph is on. A graph on the plane, or the surface of a sphere with no crossing edges requires no more than four colours such that adjacent nodes are coloured differently. But a graph on the surface of a torus can require up to seven colours for this property to hold, and in general the genus of the surface (ie. the number of holes it has) determines the number of colours required to ensure that no adjacent vertices share a colour [17].

## 2.2 Latent Space Models

When discussing spatial networks the ‘space’ in question was the physical space we live in. When considering social networks, it made sense that those living near to each other are more likely to be friends than those far away. But it is also the case that people of the same age are more likely to be friends than those of differing ages. The same is true of other similarities and differences, such as wealth and gender. It is tempting to think of these effects in the same sense as other spatial effects, placing people (in the case of age) along a one-dimensional axis, and using the distance between two people on this axis to estimate how likely or unlikely they are to be friends. Such approaches are generally (in the literature on social networks) referred to as *latent space models* [108]. The significant difference between these models and spatial networks is that in latent space models this space is hidden, or abstract, and we cannot always observe it. The challenge then is to take the structure of a network and try to deduce or predict these latent spaces, and see how well they can recreate the actual network structure.

Both correspondences and deviations between measured properties and latent spaces can yield useful information. For example, the voting patterns of politicians in parliament can be well modelled by a one-dimensional axis which roughly corres-

ponds to the political left and political right [47, 14]. If over time, the ability of a latent space model to predict party affiliation increases we might say that politics is becoming more polarised, as illustrated in Figure 2.2. On the other hand, deviation from expected voting patterns, observed as edges which would be considered unlikely in the latent space model, spanning an unusually large political distance can be used to measure particular voting patterns such as vote trading, as in [98].



**Figure 2.2:** Figure 2 from [14] (open access). Latent space model of voting in the US house of representatives. Nodes are members in the house, and edges represent voting similarity for bills in the house. Nodes are coloured according to party affiliation (blue for Democratic, red for Republican) and as time goes on the two parties become more and more distinct.

Regardless of the number of dimensions though, or whether the space the network lives in is real physical space, or some abstract space representing age, wealth, political opinion or anything else, the key fact about spatial networks and latent space models which make them different to random graphs is that nearby nodes are more likely to share edges than those far from each other. This rule allows the prediction of missing edges, and the properties of unobserved nodes [106], as well as giving us useful information about the process responsible for a network's structure. It is this idea which motivates simple mathematical models discussed in the next section.

## 2.3 Random Geometric Graphs

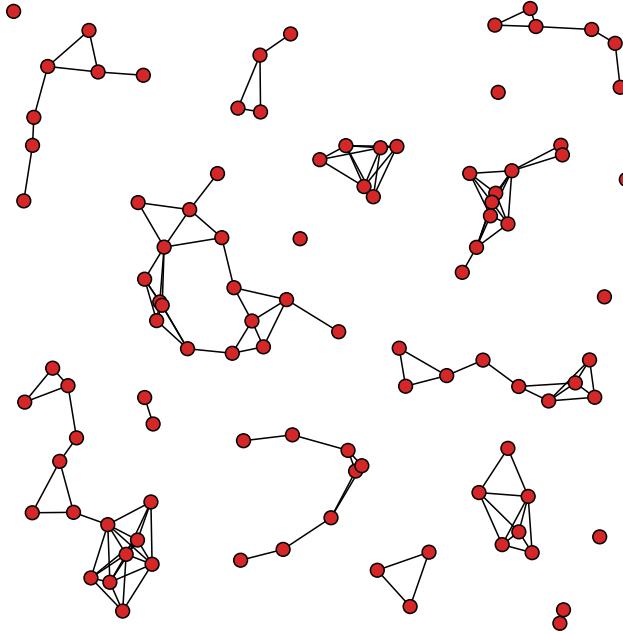
The simplest network geometric model is the random geometric graph, or RGG [67, 158] in Euclidean space, like that in Figure 2.3. Like the ER random graph, the RGG model constructs a graph as a random variable. We denote it as  $\mathcal{G}_{\text{RGG}}(N, d, r) = \{\mathcal{V}, \mathcal{E}\}$  and it is defined as follows. Each of the  $N$  vertices  $v \in \mathcal{V}$  is associated with coordinates in the  $d$ -dimensional unit cube<sup>1</sup>,  $[0, 1]^d$ . These coordinates are sampled uniformly at random, and independently. Vertices whose coordinates are within distance  $r$  share an edge.

Random geometric graphs (RGG) can be considered in the context of continuum percolation [85, 139], where each vertex is a ball of radius  $r/2$  and so an edge in the graph corresponds to two of the balls overlapping. It is unsurprising then that, as is the case for many other percolation models [56], the connectivity of an RGG shows a phase transition, with the fraction of vertices in the largest connected component equal to zero (in the limit of many vertices) when  $r$  is below a threshold value, but becoming finite above this threshold [67].

As illustrated in Figure 2.4 these phase transitions are similar to that in an ER random graph. A key difference between ER random graphs and RGGs though is in the *local structure* created by the Euclidean geometry underlying the RGG model. Edges in

---

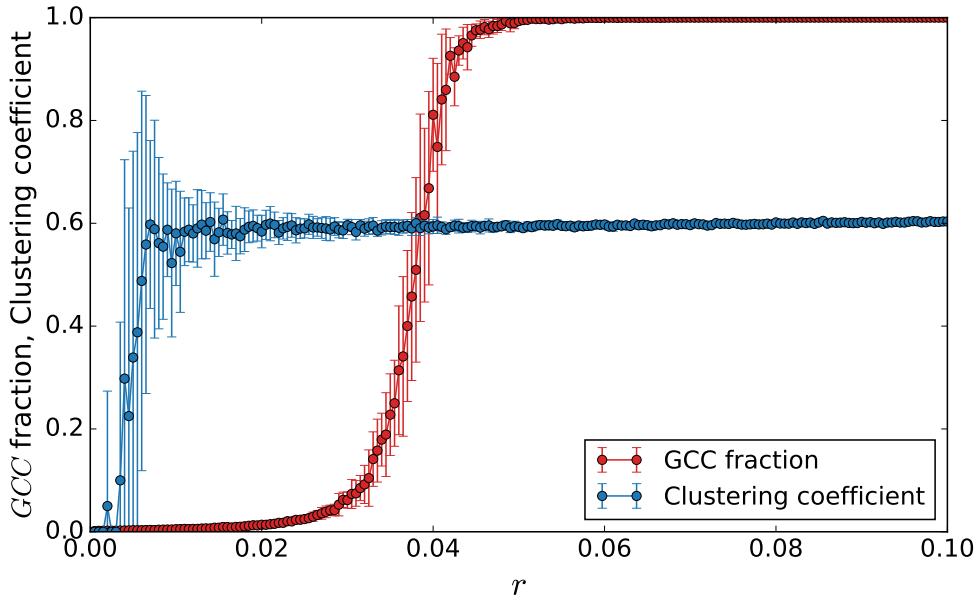
<sup>1</sup>It is of course possible to have an RGG in Euclidean space in other ways, perhaps with periodic boundary conditions, or by scattering in spaces other than the unit cube. See for example [63, 64].



**Figure 2.3:** A random geometric graph with  $d = 2$ ,  $N = 100$  and  $r = 0.1$ . This value of  $r$  is below the critical value for a giant connected component to emerge and so the graph is split into many small components. Unlike an ER random graph there are still many closed cycles in this disconnected phase due to the graph's geometric structure.

the RGG exist between vertices which are close to each other, and the triangle inequality,  $d(a, c) \leq d(a, b) + d(b, c)$  where  $d$  is a distance function, means that if there is an edge between vertices  $a$  and  $b$ , and an edge between  $b$  and  $c$ , then we can bound the distance between  $a$  and  $c$  with a value of  $2r$ . This means that  $a$  and  $c$  are much more likely to share an edge than two randomly selected vertices in an ER random graph. A consequence of this local structure induced by the model's geometry is that RGGs exhibit much higher clustering than ER graphs.

A common thread in this review of network geometry is that geometric rules can induce interesting and complex structure into random graph models. Indeed, it was recently shown that random graphs in which the average degree and clustering coefficient are fixed to a constant are, for sufficiently large clustering, equivalent (in the sense of having equivalent ensembles) to RGGs in one dimension [122]. Because of the structure emerging from geometric rules, random graphs with geometric properties can also



**Figure 2.4:** Phase transition in the size of the largest connected component, and clustering coefficient in a random geometric graph with  $D = 2$  and  $N = 1000$ . There is a clear phase transition in the fraction of vertices in the giant connected component, which is 0 below the threshold value for  $r$  but quickly jumps to 1 above this threshold value. Aside from  $r$  values so small as to make any edges unlikely the value for the clustering coefficient  $C$  is approximately constant, and high.

fit some real network datasets much better than some other random graphs. For example, in [106] RGGs are fit to networks of interacting proteins and predict network structure significantly better than equivalent ER random graphs. We will revisit the problem of fitting real networks to geometric spaces in chapter 7. Furthermore, simple geometric growth mechanisms combined with quantum statistics of the Bose-Einstein and Fermi-Dirac distributions can generate networks with a range of properties fitting real data [34, 32].

Graph observables such as the size of the largest connected component or clustering coefficient depend on the dimension of the RGG [67]. In general, we will find that changing properties of the geometric space a network is embedded in can change structural properties of that network and it is this relationship, between geometry and network structure that network geometry attempts to understand.

As was noted in chapter 1, many real world networks of interest have common properties such as an approximately power-law degree distribution, strong clustering,



**Figure 2.5:** A cartoon of triangles in curved and flat space, from Fermilab/SLAC Symmetry magazine [35]. On the left, positively curved, or closed space. In the centre, negatively curved, or open space. On the right, the familiar flat Euclidean space. Note that the sum of the internal angles of the triangles only equals  $\pi$  in flat space. Positive curvature increases this sum, and negative curvature decreases it.

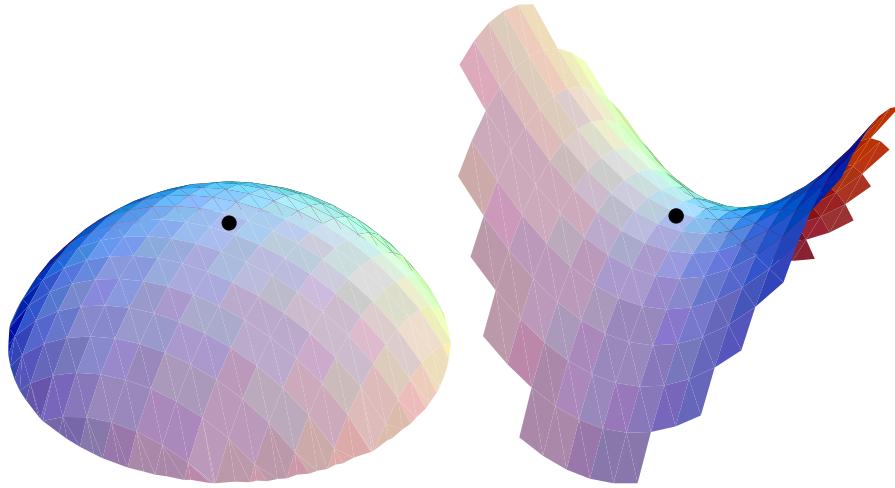
and a modular or hierarchical structure. It is then natural to ask, are there geometric spaces which can reproduce *these* properties?

## 2.4 Hyperbolic Network Geometry

It turns out that random geometric graphs in the hyperbolic plane can reproduce many common features of complex networks. Hyperbolic geometry (sometimes called Lobachevskian geometry after one of its discoverers) emerged out of various attempts by geometers to derive Euclid's parallel postulate from the other axioms in his *Elements*. This postulate states that given a straight line, and a point not on the line, there is exactly one line which can be drawn that passes through the point but never intersects the first line (remaining parallel to it). It was long thought that this postulate could be derived from the others as a theorem, the strength of this belief displayed by the number of false proofs that have appeared over the years [102]. Eventually, following attempts to prove this 'theorem' by contradiction, it was found that perfectly consistent and reasonable geometries could exist without this postulate holding at all. Since then, the geometry where there is one such other line which never meets the first is called Euclidean, and hyperbolic geometry is one in which there are many such lines, and elliptic geometry that where there are no such lines.

Figure 2.5 shows the three cases of uniformly curved space: elliptic geometry (pos-

itive, or closed curvature), hyperbolic geometry (open, or negative curvature) and Euclidean (flat, or no curvature). We will denote  $D$ -dimensional hyperbolic space as  $\mathbb{H}^D$  although in this section the primary focus will be on the hyperbolic plane,  $\mathbb{H}^2$ . Unlike positively curved elliptic geometry which can be modelled on the surface of a sphere,  $\mathbb{H}^2$  cannot be embedded in Euclidean space in a way that preserves distances and volumes. In some sense it is simply too big. The area of a circle drawn in  $\mathbb{H}^2$  grows exponentially with its radius, whereas in Euclidean space,  $\mathbb{E}^D$  this area or volume grows as the radius to the power  $D$ . For any value of  $D$ , this exponential growth will eventually outrun the polynomial growth of  $\mathbb{E}^D$  and so the hyperbolic plane can't fit in  $\mathbb{E}^D$  for any value of  $D$ . We must settle for either visualising small areas of  $\mathbb{H}^2$  as in Figure 2.6, or projecting it in some way that distorts distances, as in Figure 2.7.



**Figure 2.6:** Small sections of positively (left) and negatively (right) curved space, from the ‘Math and Art of MC Escher’ [26]. Sections of the sphere-cap shape on the left can be stitched together to form a sphere in  $\mathbb{E}^3$ . However, if we try to stitch together sections of the saddle point shape on the right, we run out of space and the surface starts overlapping with itself. This illustrates why hyperbolic space, where every point is locally saddle-like, cannot be embedded in Euclidean space.

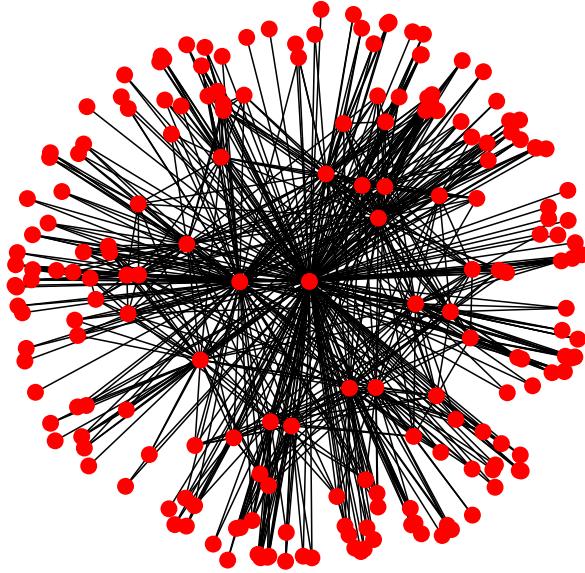
The Poincaré disk model in Figure 2.7 draws the infinite plane of  $\mathbb{H}^2$  in a unit circle. In order to fit this infinite space into a finite representation the space is warped and squashed. It is a conformal model meaning that it has the helpful property that angles between two lines in the model are the same as in the hyperbolic space. However the differences from Euclidean geometry mean that the shortest line between two points in



**Figure 2.7:** A regular hyperbolic tiling, with Schläfli symbol  $\{5, 4\}$  of an image of Imperial College's Queens tower in the  $\mathbb{H}^2$ , represented in the Poincaré disk, created using [3]. Each tower has the same hyperbolic area, and there are infinitely many, but this representation warps their apparent size, with objects appearing smaller near the edge of the disk which allows for the entire hyperbolic plane to be represented. Geodesics in this representation follow curved arcs, which can be seen here as the arcs formed by towers joined at the base and tip. Famous artworks such as MC Escher's 'Circle Limit III' woodcut show similar effects.

this model appears not as a straight line, but as a curve. Intuitively this is because the distance ‘around the edge’ of the disk is very large, and so curving towards the centre provides a shortcut.

We can now ask questions about RGGs in this geometry, and observe the properties of the graphs generated. Without exactly specifying how RGGs in hyperbolic geometry are defined (see [12] for an implementation), there are some informal reasons to suggest that such a geometry could produce networks with interesting properties. Firstly, because of the exponential growth of the space, tree graphs can be isometrically embedded (that is with each node taking up the same volume, and edge the same distance) in  $\mathbb{H}$  while they cannot in  $\mathbb{E}$ , since the number of vertices in a tree grows exponentially with its radius. This means that embedding networks whose nodes have an approximately tree-like taxonomic structure requires a negatively curved space. Such taxonomies are common in biology (consider perhaps a network of species where closely related species share an edge), but also in systems such as the Internet network where personal computers sit at the bottom of the taxonomic tree, and larger computer networks appear going up the tree with very highly connected servers sitting near the top. The path a packet of information takes on the Internet network mirrors the geodesics in hyperbolic geometry. Packets between one personal computer and another do not travel through the many small personal computers between the two, but rather first move to the large, highly connected servers at the top of the hierarchy (and in the centre of the hyperbolic space) before moving back down the hierarchy to its destination. In fact, such network geometric representations have already been used in the development of more efficient Internet routing algorithms [37]. Consider then the simplest model of uniform node density in  $\mathbb{H}^2$ . Nodes are distributed in a disk of radius  $R$  (which is a function of the number of nodes  $N$ ). The degree distribution that emerges from these graphs is a power-law, meaning that  $p(k) \sim k^{-\gamma}$  [124], the same as for the preferential attachment model [23]. By generalising the node distribution from uniform to an exponential function of the radius of the disk, and changing the strength of the negative curvature of the space, this exponent in the power-law degree distribution can



**Figure 2.8:** A random geometric graph in  $\mathbb{H}^2$  generated using the program described in [12]. Note that points near the centre of the disk have many more edges than those near the edge, creating a degree distribution with a few hubs and many nodes with a low degree. The difference between this graph, and the RGGs shown previously is only the different geometry used, and yet this is enough to completely change the typical properties of these graphs.

be tweaked, as can the average degree. See Figure 2.8 for an example of a hyperbolic random geometric graph with  $N = 200$  nodes and an average degree of  $\langle k \rangle = 5$ . This geometric model on  $\mathbb{H}^2$  can be shown to be equivalent with another well known model on  $\mathbb{S}^1$ , the circle, where the nodes also have an intrinsic fitness [156]. We can then effectively interpret the angular coordinate in the  $\mathbb{H}^2$  model, and the position on the circle in the  $\mathbb{S}^1$  model as corresponding to node *similarity*, and then the radial coordinate in the  $\mathbb{H}^2$  model, and the node fitness in the  $\mathbb{S}^1$  model as corresponding to node *popularity*. These two equivalent frameworks have sometimes been termed ‘Newtonian’ and ‘Einsteinian’ [118]. The model on the circle with node fitness is Newtonian because some nodes have higher fitness and so connect to many others (analogously to mass in Newtonian gravity). The hyperbolic model is Einsteinian because this same effect emerges by letting the curvature of space replace the node fitness, or mass, just as New-

tonian gravity emerges from the curved spacetime of General Relativity. These ideas have generated enough interest for research on efficient computational implementation [199] and fitting the model to data [36] to have since been undertaken.

## 2.5 Summary

We have seen that the network geometric approach can, despite its apparent simplicity, provide real insight into the underlying factors behind network structure. The mathematical simplicity of random geometric graphs allows for mean field, continuum approaches to be used to analytically estimate network observables. Yet the ability of geometric models to replicate relevant and interesting network properties suggests their potential for providing useful descriptions of the real systems these networks represent.

At the very least this approach can be used as a mathematical tool to simplify analysis and modelling, while apparently losing little in the way of descriptive power. At its best, it may provide a new way of considering the structure and mechanisms of growth of these networks, by making analogies between the geometric spaces used and the system in question.

In the next section this geometric approach will be extended from the Riemannian spaces considered so far to pseudo-Riemannian geometries such as Lorentzian spacetime as used in the physics of causality in relativity.

## Chapter 3

# Causal Structure and Causal Sets

Be wise - discretize.

---

Mark Kac

Previously we introduced graphs as an approach to modelling complex systems, and DAGs as a particular type of graph which has special structural constraints. The focus of this chapter is the introduction of the idea of *causal structure*. In the language of networks this simply means that we are considering networks in which the relation that a directed edge represents is a causal one. This thesis is not in the business of the philosophical discussion of what causality truly is, or how we can infer it. Rather, by a causal relation between two events  $a$  and  $b$ , we just mean that if  $b$  has occurred then we know that  $a$  must have occurred already. For example, if  $a$  represents me writing this sentence, and  $b$  represents you reading it, then it is reasonable to say that  $a$  and  $b$  are causally related in this way.

This binary relation is simple, but more complicated causal structures can be built by combining many of these relations, just as a graph is built of many simple binary edges. In this chapter we will explore sets of causally related objects, called *causal sets* and also continuous spaces where geometric relationships determine this causal relation. We begin with the discrete case by reviewing the mathematics of partial orders, then continue with the continuous case, introducing Lorentzian geometry and its role in relativistic physics. We then discuss causal sets as a physical theory but also as a

mathematical tool that can be of use in network science. The idea of taking the tools of causal set theory and of Lorentzian geometry, and applying them to help understand the structure of networks of causal relations will be the main thrust of later chapters.

### 3.1 Partial Orders

Order theory is that branch of discrete mathematics which studies binary relations which correspond to the idea of order. The basic building block of the theory is a *partially ordered set*, or *poset* denoted by  $\mathcal{P} = (P, \preceq)$ , which is an underlying set  $P$ , and a binary relation  $\preceq$  called ‘precedes’. We will write  $x \prec y$  to mean that  $x$  strictly precedes  $y$  and is not equal to it. To be a partially ordered set, the order must adhere to the following rules:

**Reflexivity:**  $x \preceq x \forall x \in P$ .

**Transitivity:** if  $x \preceq y$  and  $y \preceq z$  then  $x \preceq z \forall x, y, z \in P$ .

**Asymmetry:** if  $x \preceq y$  and  $y \preceq x$  then  $x = y \forall x, y \in P$ .

For two elements  $x, y \in P$ , we say that  $x$  and  $y$  are *related* or *comparable* iff  $x \preceq y$  or  $y \preceq x$ . If there is no such relation they are called *unrelated* or *incomparable*. An important special case of a partial order is a full, or total order, one in which every element is related to every other. The integers, ordered by the usual ‘less than or equal to’ relation  $\leq$  is an example of a full order. Some further useful definitions are:

**Chain:** A subset of  $P$  in which all elements are related. A  $k$ -chain is a chain of  $k$  elements.

**Antichain:** A subset of  $P$  in which no elements are related.

**Height:** The number of elements in the largest chain.

**Width:** The number of elements in the largest antichain.

**Extension:** An order  $\tilde{\preceq}$  is an extension of  $\preceq$  if  $x \preceq y \implies x \tilde{\preceq} y \forall x, y \in P$

**Linear Extension:** An extension of  $\preceq$  which is a total order is called a linear extension. This can be thought of as a list of elements of  $P$  which respects the partial order.

**Covering relation:**  $x$  covers  $z$  if there is no element  $y$  between them such that  $x \prec y \prec z$ . We say this relation is a covering relation.

**Maximal/minimal element:**  $x$  is maximal if there is no element  $y$  such that  $x \prec y$ , and minimal if there no  $y$  where  $y \prec x$ .

**Hasse Diagram:** A drawing of a poset in which only the covering relations are drawn. The other relations are implicit because can always be deduced from the transitivity rule.

**Descendants/future:** The descendants, or future, of an element  $x$  is the set  $J^+(x) = \{y \mid x \prec y\}$ .

**Ancestors/past:** The ancestors, or past, of an element  $x$  is the set  $J^-(x) = \{y \mid y \prec x\}$ .

Posets form the basic mathematical description of order and we will now investigate physical cases where the order relation between elements is one of causality. Note that causal relationships do intuitively meet the rules of partially ordered sets, in particular transitivity, and asymmetry and so this description is a good mathematical

candidate for discussing causal structure. Posets and DAGs are closely related mathematical objects, and we can form the *comparability graph* of a poset as in [211] as follows: if we have poset  $\mathcal{P} = \{P, \prec\}$  then its comparability graph is  $\mathcal{G}_{\mathcal{P}} = \{\mathcal{V}, \mathcal{E}\}$  where  $\mathcal{V} = P$  and for  $x, y \in P$ ,  $x \prec y \implies (x, y) \in \mathcal{E}$ . This gives us a DAG representing the relations in the poset.

## 3.2 Lorentzian Geometry

Causal structure and partial orders are linked to a particular kind of geometry, *Lorentzian geometry*, which is best known for its use in the description of spacetime in Einstein's relativity. To begin the introduction of Lorentzian geometry we will use the simplest example, which is *Minkowski spacetime*. Like Euclidean space it is flat and homogeneous, and is described by just one parameter, its dimensionality,  $D$ . We will denote  $d$  dimensional Euclidean space by  $\mathbb{E}^d$  and  $D = d + 1$  dimensional Minkowski spacetime as  $\mathbb{M}^D$ , with the reason for this distinction explained below. Both  $\mathbb{E}$  and  $\mathbb{M}$  can be thought of as a *vector space*, a collection of vectors together with an inner product which composes two vectors and returns a real number. In  $\mathbb{E}^d$  we will denote the vectors as  $\mathbf{x} = (x_1, x_2, \dots, x_d)$ , and in  $\mathbb{M}^D$  as  $\mathbf{u} = (u_0, u_1, u_2, \dots, u_d)$ . This distinction is because this first element of a vector in Minkowski spacetime,  $u_0$  is special and refers to the *time* coordinate, and the other  $d$  coordinates refer to spatial coordinates with the structure of a  $d$  dimensional Euclidean space. It is for this reason that we will often refer to a  $d + 1$  dimensional spacetime, to be explicit about the number of space and time dimensions.

In Euclidean space the inner product is the familiar dot product:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^d x_i y_i . \quad (3.1)$$

The length of a vector is given by its inner product with itself,

$$\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle = \sum_{i=1}^d x_i^2 . \quad (3.2)$$

This quantity is always non-negative, and a value of 0 means the vector  $\mathbf{x}$  is the **0**-vector. The distance between two vectors is then given by:

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| \quad (3.3)$$

which again is always positive and a value of 0 means the two vectors are identical.

In Minkowski spacetime we have an inner product-like operation,  $\eta(\mathbf{u}, \mathbf{v})$  for two vectors  $\mathbf{u}, \mathbf{v} \in \mathbb{M}^D$  which we'll call the Minkowski inner product,

$$\eta(\mathbf{u}, \mathbf{v}) = -u_0 v_0 + \sum_{i=1}^d u_i v_i , \quad (3.4)$$

where we are using the  $(- + ++)$  sign convention<sup>1</sup> for the metric, and where units are chosen such that the speed of light is equal to 1. The ‘square length’ we get from this inner product is different to Euclidean space:

$$\|\mathbf{u}\|_{\mathbb{M}}^2 = \eta(\mathbf{u}, \mathbf{u}) = -u_0^2 + \sum_{i=1}^d u_i^2 . \quad (3.5)$$

In this form it is clear that this ‘square length’ in Minkowski spacetime  $\|\mathbf{u}\|_{\mathbb{M}}^2$  is the Euclidean square length in the spatial dimensions minus the square length in the time dimension. This operation is not technically an inner product and so called a pseudo inner product because it can be zero for non-zero vectors, or negative if the contribution from the time coordinates outweighs that of the spatial coordinates. Vectors in Minkowski spacetime can then be categorised into three groups depending on the sign of  $\|\mathbf{u}\|_{\mathbb{M}}^2$ :

---

<sup>1</sup>The mostly-plus convention goes by a variety of names, including spacelike East Coast, Pauli, and GR while the mostly-minus convention by West Coast, Feynman, Landau-Lifschitz, particle physics. Although it makes no difference to any physical results, the mostly-plus convention used here hopefully makes the relationship between Euclidean and Minkowski spaces clearer since the spatial part of the Minkowski metric is now exactly the same as the Euclidean metric.

$$\begin{aligned} \|\mathbf{u}\|_{\mathbb{M}}^2 < 0 & : \text{timelike}, \\ \|\mathbf{u}\|_{\mathbb{M}}^2 = 0 & : \text{lightlike}, \\ \|\mathbf{u}\|_{\mathbb{M}}^2 > 0 & : \text{spacelike}. \end{aligned} \tag{3.6}$$

Euclidean distance is a useful and meaningful measure of spatial distance because it is invariant under rotations and translations. Intuitively the distance between two points should not depend on the direction we are looking at them from, or where we are standing. In the same way, this particular form for  $\eta$  is significant due to its invariance under a set of transformations called *Lorentz transformations*. Under the Lorentz transformations, the separation between two vectors in Minkowski spacetime is unchanged. These transformations include spatial rotations and translations, but also include changes of relative velocity, or *Lorentz boosts*. This ensures that the separation between two events in spacetime does not depend on the velocity of the observer. This encapsulates the core principle of special relativity, which is that observers in any (inertial) frame of reference must see the same laws of physics.

The equations for a Lorentz boost by velocity  $\beta$  in the  $x$  direction on coordinates  $t = u_0$  and  $x = u_1$  are<sup>2</sup>

$$x' = \gamma(x - \beta t) \tag{3.7a}$$

$$t' = \gamma(t - \beta x) \tag{3.7b}$$

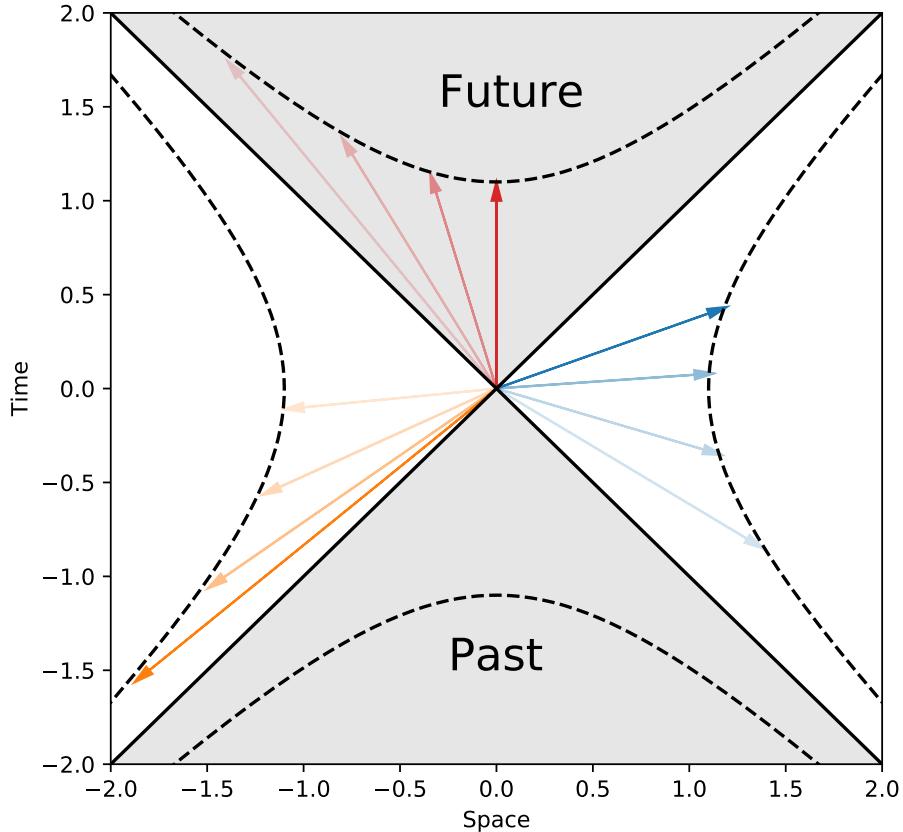
where

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}}. \tag{3.7c}$$

Under these boosts, for any  $\mathbf{u}, \mathbf{v} \in \mathbb{M}^D$  the value of  $\eta(\mathbf{u}, \mathbf{v})$  does not change. As illustrated in Figure 3.1 the set of timelike vectors has two disjoint subsets, one with a positive time coordinate and one with a negative time coordinate. We refer to these as the *future* and *past* respectively. For a given point in spacetime,  $\mathbf{u}$ , we can say that  $\mathbf{v}$  is timelike separated from  $\mathbf{u}$  if the vector  $\mathbf{u} - \mathbf{v}$  is timelike, that is if  $\|\mathbf{u} - \mathbf{v}\|_{\mathbb{M}}^2 < 0$ . We

---

<sup>2</sup>Unless specified otherwise we will always use natural units where the speed of light  $c = 1$ .



**Figure 3.1:** Lorentz boosted vectors in Minkowski spacetime with  $D = 1 + 1$ . Shaded regions show past and future lightcones of a point at the origin. Three vectors are shown, one future directed and timelike (red), and two spacelike vectors (blue and orange).

Faded vectors are boosted by successively higher velocities. Note that timelike vectors remain timelike, and spacelike vectors remain spacelike under these transformations. The dotted lines show constant  $\eta$  from the origin, and boosted vectors will move along these lines. Note that the units of space and time are chosen such that the speed of light is equal to 1.

will denote the future and past of  $\mathbf{u}$  as  $J^+(\mathbf{u})$  and  $J^-(\mathbf{u})$  respectively, and collectively the futures and pasts of all points in our spacetime makes up its *causal structure*:

$$J^+(\mathbf{u}) = \{\mathbf{v} : v_0 > u_0, \|\mathbf{u} - \mathbf{v}\|_{\mathbb{M}}^2 < 0\}, \quad (3.8a)$$

$$J^-(\mathbf{u}) = \{\mathbf{v} : v_0 < u_0, \|\mathbf{u} - \mathbf{v}\|_{\mathbb{M}}^2 < 0\}. \quad (3.8b)$$

In physics these geometric relations have particular importance. According to special relativity, if two points in spacetime (which physically mean two events occurring at a specific place and at a specific time) are timelike separated, that means that a beam of light, or in general physical information, would have the time to travel from the past event to the future event and so it is possible that the past event can causally effect the future one. If these events are spacelike separated then neither light nor anything else can travel from one event to the other quickly enough and so neither event can causally effect the other. Observers travelling at a relative velocity can disagree on which of two spacelike events occurred first, but they will always agree on which of two timelike separated events happened first, because proper<sup>3</sup> Lorentz transformations do not change the causal structure of these events.

We can now establish the connection between partial orders and Lorentzian geometry. It was not coincidence that ‘causal structure’ has been used both to refer to the relations in a poset, and between timelike separated events in Lorentzian spacetime. If we take the relationship of timelike separation to correspond to  $\prec$ , with the past event preceding the future event, then the points in a Lorentzian spacetime<sup>4</sup> form a partial order with infinitely many elements. So we will write  $u \in J^-(v) \implies u \prec v$ . This object fulfils the requirements of a partial order since the past/future relationship of spacetime is reflexive, transitive, and asymmetric. It is for this reason we use the

<sup>3</sup>Strictly speaking the symmetry group which leaves  $\eta(u, v)$  invariant includes time reversal transformations such as  $t \rightarrow t' = -t$ . Such transformations are called *improper*. In the language of posets, such time reversals simply correspond to reversing the  $\prec$  relation, or in the language of DAGs, changing the direction of the edges. We will generally only be referring to proper Lorentz transformations from now on.

<sup>4</sup>This is the case assuming the spacetime has no closed causal loops, an assumption which we will generally make throughout this thesis.

same notation for the past and future of an event in spacetime as for the ancestors and descendants of an element in a poset.

In this section we have introduced the simplest Lorentzian geometry which is Minkowski spacetime, but other, more complex spacetimes exist and will be used later in this thesis. They also have causal structure and so also correspond to partial orders. In particular, *de Sitter* spacetime,  $d\mathbb{S}$  is the simplest (in the sense of being maximally symmetric) spacetime that has uniform positive curvature. As Minkowski spacetime is analogous to Euclidean space, the geometry of de Sitter spacetime is analogous to the geometry of the sphere. This spacetime will be discussed in detail in chapter 6.

The connection between causal structure (as described by partial orders) and Lorentzian spacetimes is made complete by the important results of Hawking *et. al* [103, 104] and Malament [132]. These are the results that give us good reason to think that associating continuous spacetimes with partial orders is a sensible goal. They prove that the causal structure of a spacetime, meaning simply the set of  $J^+(x) \forall x \in \mathbb{M}$ , almost entirely determines the whole structure of that spacetime. The only missing piece is a measure of volume. The simplicity of partial orders and the fact that they are almost enough to totally describe a spacetime has motivated much of the research discussed in the next section. As noted in [71], the causal structure of a spacetime determines its chronological structure, topology, differentiable structure and its metric, and so it is tempting to want to start with the simple partial order, and derive these manifold properties from that, rather than the other way around. This idea, that partial orders can play a unifying role in describing spacetime is the basis for causal set theory as discussed in the next section.

### 3.3 Causal Set Theory

Causal set theory is a physical theory in which spacetime is fundamentally discrete, and these discrete elements can be related to each other by a notion of ‘before’ and ‘after’ [71]. This idea appears independently in the work of Myrheim in 1978 [145] and the seminal 1987 paper of Bombelli, Lee, Meyer and Sorkin [42], as an approach to solving the problem of *quantum gravity*, the name commonly given to the ongoing attempt by the physics community to reconcile two theories: general relativity (GR), and quantum mechanics (QM). Both of these theories are immensely successful in their own domains (see [142, 206] for popular textbooks); GR in describing gravity, black holes, and the large scale structure of the universe; QM in describing the stochastic mechanics of fundamental particles. They are amongst our most fundamental theories and have been experimentally verified to extraordinary precision [101].

Unfortunately these two theories are fundamentally incompatible. Because the measurements that verify GR are done on extremely large scales, and those that verify QM on extremely small scales, both theories can make validated predictions to extremely high degrees of precision in their own domains, while still appearing to fail to fit together to provide one coherent description of reality. The incompatibility between the two approaches is severe. QM describes the world stochastically, with in-built randomness (without getting into interpretations of what that means philosophically) while GR is entirely deterministic. In QM particles are not in just one place, but are described by a wavefunction over a pre-existing spacetime, and yet GR says that matter actually curves spacetime and so the theory needs to know where that matter is. Both theories treat the fundamental concept of time differently [171, 111, 13]. Due to the seriousness of these issues it is generally accepted that a satisfactory resolution can only come through fairly radical changes in how matter, space, time and energy are conceptualised in physics [154]. The causal set approach is one of the numerous approaches to reconcile these differences with others including string theory [69, 7], loop quantum gravity [172], quantum graphity [121] and quantum Regge calculus [170] to name just a few.

The aim of this thesis is, thankfully, not to attempt to resolve the problem of quantum gravity and so I will not discuss these competing approaches in any further detail. Introducing causal set theory in this section will involve some discussion of the problems in physics it seeks to solve and the phenomena it attempts to explain, but this is just to provide background and context. Whether or not causal sets are eventually part of a successful theory of quantum gravity is not important for the work in later chapters (although I wish the theory and its proponents all the best of luck). Regardless of whether the physics is correct, causal sets are useful as a discrete way of thinking about spacetime, and Lorentzian geometry [71] and that is what we need them for here.

As we saw in the previous section, the geometry of a Lorentzian spacetime is fully determined (up to a factor of volume) by its causal structure; the relationship of past and future between its points. Usually in physics, we start with the continuous spacetime and its metric and then derive this causal structure from the lightcones which tell us which points are in the past and future of which others. But we could do the opposite. Starting with the causal structure (which as causal set advocates point out, is mathematically simpler [42]) one could derive the continuous spacetime. The problem though is that this causal structure does not tell us what the volume of a region of spacetime should be. The solution is to not have a continuum with an infinite number of points in every finite volume of spacetime, but rather just a finite, albeit very large number instead. This way, the volume of a region can just be determined by the number of causal set points in that region, and the problem of determining volume reduces simply to counting. With this motivation in mind, let's begin by formally defining a causal set.

A causal set, or *causet*  $\mathcal{C} = \{C, \preceq\}$  is a partially ordered set which is *locally finite*. It is an underlying set  $C$  endowed with a relation  $\preceq$ , called *precedes* which adheres to the following axioms:

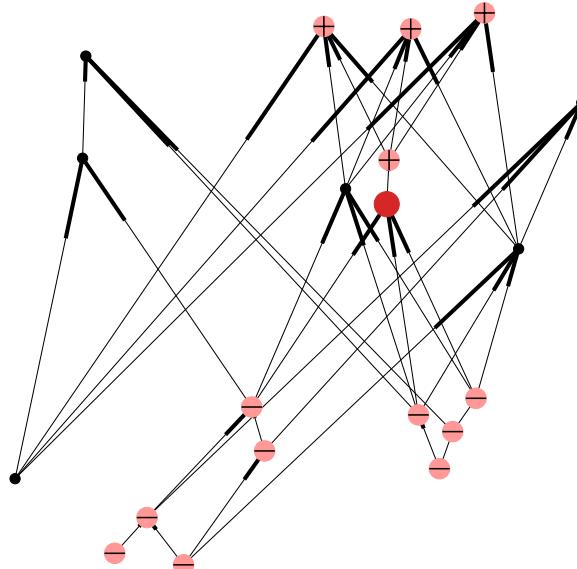
**Reflexivity:**  $x \preceq x \forall x, \in C$ .

**Transitivity:** if  $x \preceq y$  and  $y \preceq z$  then  $x \preceq z \forall x, y, z \in C$ .

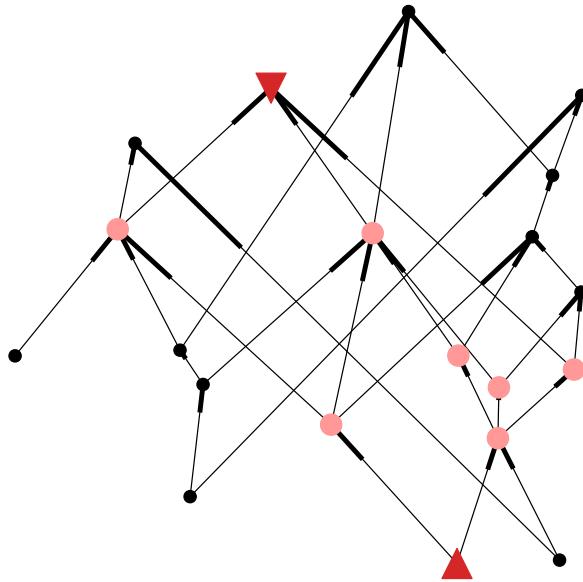
**Asymmetry:** if  $x \preceq y$  and  $y \preceq x$  then  $x = y \forall x, y \in C$ .

**Local finiteness:**  $\forall x, z \in C$ , the set  $\{y \mid x \preceq y \preceq z\}$  has a finite number of elements.

Any poset meets the first three criteria and adding local finiteness is what makes  $\mathcal{C}$  a causal set.



**Figure 3.2:** For the highlighted red node in this causal set graph, its past ( $-$ ), future ( $+$ ) and unrelated nodes (smaller). Note that only the Hasse diagram is drawn here.



**Figure 3.3:** The interval between the two triangular nodes in a causal set graph drawn as a Hasse diagram.

The causal set hypothesis says that at very small length scales, spacetime can no longer be described by a continuous manifold but instead there is only the discrete causal set from which the manifold emerges as an approximation at large length scales. The causal structure of the manifold is just that of the underlying causal set, and volume in the manifold is determined simply by the number of elements in that region, an idea captured in what is something of a slogan for causal set theory, ‘Number + Order = Geometry’ [201]. The fundamental unit of spacetime volume that would correspond to one causal set element is generally thought to be something on the order of the Planck volume, roughly  $10^{-148} \text{m}^3\text{s}$  [43]. The fact that this number is very small explains why the continuum is such an effective approximation for our everyday length scales.

An *interval*, or *Alexandrov set*  $I[x, z] = \{y \mid x \preceq y \preceq z\}$ , is the intersection of the past of  $z$  and the future of  $x$  and can be thought of as everything ‘in between’ the two, plus the two endpoints. See Figure 3.3 for an example. If  $x$  and  $z$  are unrelated, or if

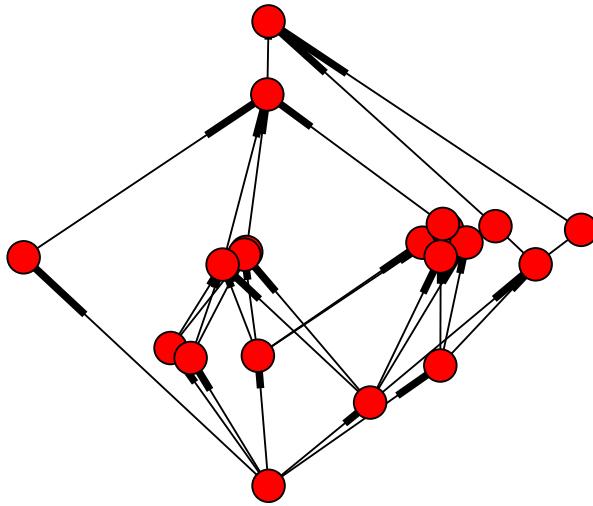
$z \prec x$  then this interval must be empty. Note that while local finiteness means that any interval must be finite in its cardinality, the causal set as a whole may have an infinite number of elements. This corresponds to the idea of a finite region of spacetime having a finite volume even though as a whole spacetime might continue indefinitely and so have an infinite volume.

A covering relation, which is usually called a *link* in the causal set literature, is a relation between two elements  $x \prec y$  for which there is no element  $z$  such that  $x \prec z \prec y$ . They can be thought of as defining ‘near neighbours’ in the causal set. Note that because of the axiom of transitivity, these link relations are sufficient to describe the whole structure of the causal set, since any relation which is not a link must be implied by some others which are links.

To bridge the gap between continuous spacetime and the discrete causal sets we will use the notions of embedding and sprinkling [72]. An *embedding* of a causal set is a mapping of its elements to points in a Lorentzian spacetime such that the order relation of the causal set is the same as that generated by the lightcone structure of the spacetime. It can be thought of as assigning coordinates in a spacetime to each element in our causet so that the order relations agree with those the spacetime would suggest. A *faithful embedding* is one in which the number of points mapped to a region of spacetime is proportional to its volume. The idea is that if a causal set is faithfully embeddable into a Lorentzian spacetime then at large scales it will physically approximate that spacetime.

Given a spacetime, how can we construct faithfully embeddable causal sets? The simplest method is sprinkling. A *sprinkling* into a Lorentzian spacetime  $\mathcal{M}$  is a Poisson point process where points of  $\mathcal{M}$  are randomly chosen, with some density, from the continuum to be elements in a causal set  $\mathcal{C}$ . These points inherit the order relation of the original spacetime as in Figure 3.4.

Since this selection process is independent and uniform, on large scales the number of points in a region is proportional to that region’s volume ensuring the resulting causet will be faithfully embeddable. In this way, we can create causal sets which ap-



**Figure 3.4:** A causal set sprinkling into an interval of  $D = 1 + 1$  dimensional Minkowski spacetime. The time dimension is vertical, and the spatial dimension horizontal.

proximate the geometry of a continuous spacetime. At this stage one could object that the entire purpose of this process was to do away with the continuous manifold and replace it with a discrete object; and yet we needed to use the manifold to generate the discrete causet with the right properties! This is a fair objection in the sense that for causal sets to be a good theory in physics its proponents must devise some other (presumably combinatorial) mechanism for generating causets which are faithfully embeddable in four-dimensional, roughly flat spacetime like the one we live in. However, this is a difficult problem [166, 8] which is not relevant to our interests here. As long as sprinkling provides a method of generating faithfully embeddable causets then it is sufficient for the purpose of providing a discrete analogue to continuous spacetime.

Why must the sprinkling be a random Poisson point process, and not, for example, a regular lattice? At first glance it would appear that both satisfy the criteria of having the number of points in a spacetime interval approximate the interval's volume, but

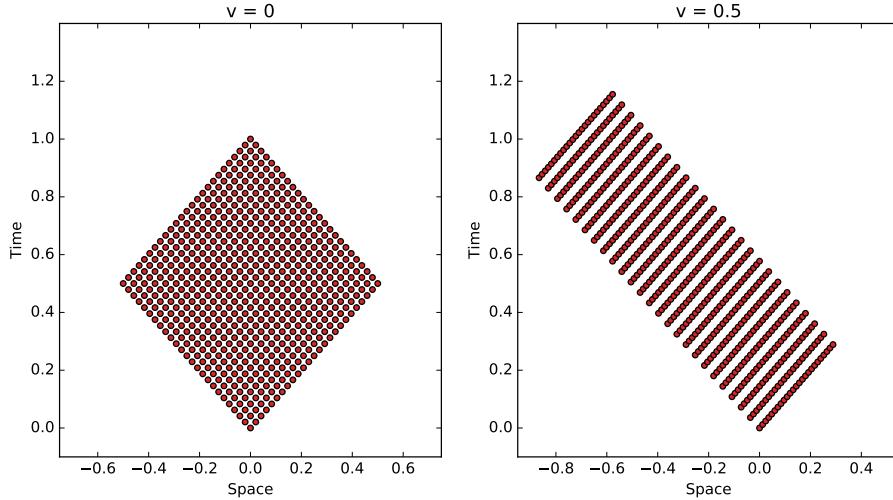
this turns out to be false. A regular lattice of points would not be Lorentz invariant because it has a preferred direction from the orientation of the lattice. After a Lorentz boost the lattice appears warped and it is clear that some spacetime intervals can be very large without containing any of the sprinkled points, as illustrated in Figure 3.5a and Figure 3.5b. A random sprinkling does not suffer from this problem and after a Lorentz boost still looks like a random sprinkling [72, 191]. This is because the expected number of points in a spacetime volume is proportional to that volume, and the spacetime volume is unchanged by a Lorentz transformation<sup>5</sup>.

There remains a number of unanswered questions in causal set theory. Firstly, what kind of process creates causal sets which approximate Lorentzian manifolds if we are not allowed to start with a manifold to sprinkle into? This is still a significant open question in and one of fundamental importance. Almost all causal sets (measured by weighting each distinct causal set equally) with a large number of elements do not look like the spacetime of our universe at all. Instead, they have a very particular three-layer structure, described by the Kleitman-Rothschild theorem in [119] and numerically investigated in [105]. Such a causal set would correspond to a universe that only lasts for three Planck times. Due to this dominance of unphysical cases, any stochastic model for generating causal sets modelling the universe's spacetime must have a way of suppressing their number. Some simple generative models (such as classical growth models) can reproduce at least some interesting properties present in the spacetime of the universe [48], but may still not approximate manifolds well enough [50].

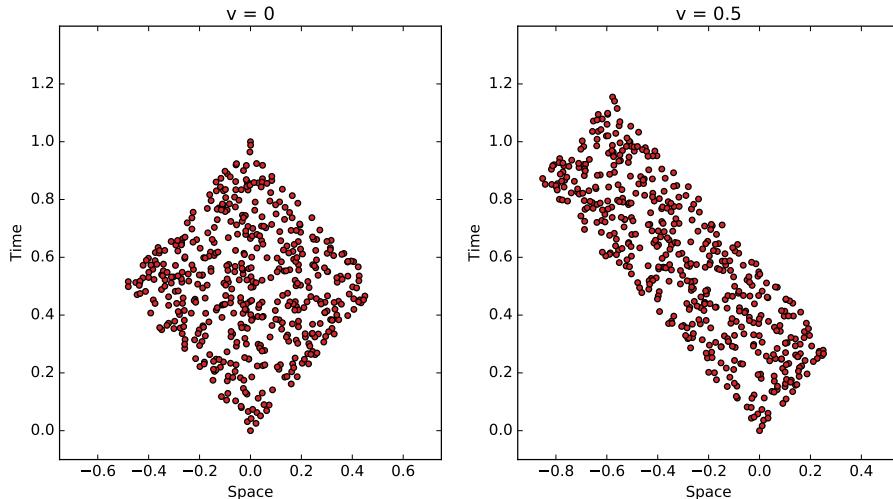
Secondly, whilst sprinkling does let us generate a causal set which approximates a manifold, if we are given a causal set to begin with how can we tell which manifolds, if any, it will approximate? For causal set theory to make sense at all, it cannot be that the one causal set making up our universe can approximate two completely different manifolds, since we are only perceiving one of them. The conjecture that the right kind of causal sets will embed faithfully into a unique spacetime is called the ‘Hauptvermutung’ of causal sets, and is another significant open question in the theory.

---

<sup>5</sup>More formally, the Jacobian for these transformations is equal to 1.



**(a)** In its rest frame (left) a lattice of points appears to work as a sprinkling into Minkowski spacetime, but after a Lorentz boost (right) clearly has a preferred direction and the average density of the points is not constant.



**(b)** The random Poisson causal set sprinkling (left) is Lorentz invariant. After a Lorentz boost (right) the shape of the interval has changed, but the density of points has not, and points still appear to be randomly placed after the boost, in contrast to an ordered lattice. The average density of the scattered points is still constant after a Lorentz boost. It is for this reason that random sprinkling is used to model causal sets rather than regular lattices.

**Figure 3.5:** A lattice of points (top) fails to be Lorentz invariant, while a random Poisson point process is, as it maintains a constant density of scattered points with no large voids.

Further work such as [192, 29, 30, 89, 90] focuses on the dynamics of particles in a causal set spacetime, and how one might define and calculate the action in a discrete combinatorial way, as well as the other open questions in the theory.

### 3.4 Generating causal sets by sprinkling

In later chapters we will need to generate causal sets approximating certain Lorentzian spacetimes by sprinkling. In this section I will briefly describe how this is computationally implemented. The simplest case is flat Minkowski spacetime,  $\mathbb{M}^D$ . In most applications, we wish to have a causal set of  $N$  elements in some spacetime interval, which for the sake of simplicity will be designated as the interval between points at coordinates  $\mathbf{u}_a = (0, 0, \dots, 0)$  and  $\mathbf{u}_b = (1, 0, \dots, 0)$ . These points are at the origin spatially, and are one unit of time apart. The challenge is to make sure that the random process we are using to generate causal set elements samples uniformly from the allowed volume of spacetime.

The simplest but inefficient approach to generating these causal sets is the trial-and-error approach. We uniformly at random generate a point in the cube  $[0, 1] \times [-0.5, 0.5]^D$ . We then check whether the point is in the interval  $[\mathbf{u}_a, \mathbf{u}_b]$  and if it is, add it to our set, discarding it otherwise. This process continues until  $N$  acceptable points have been generated. Such a process will definitely generate points uniformly within our interval, but it is inefficient in the sense that many points will be discarded. As  $D$  grows, the fraction of discarded points will approach 1 potentially making this process take a significant time if  $D$  and  $N$  are large.

A second approach is to devise a way of mapping points in a unit cube to the spacetime interval in such a way as to preserve volume. For  $\mathbb{M}^2$  this is easy, as the diamond-shaped interval in two-dimensional Minkowski spacetime can simply be rotated into a square. Given two uniform random points in  $[0, 1]$ , which we will call  $x_0, x_1$ , we find

the interval coordinates  $u_0, u_1$  with

$$u_0 = \frac{1}{\sqrt{2}}(x_0 + x_1) , \quad (3.9a)$$

$$u_1 = \frac{1}{\sqrt{2}}(x_0 - x_1) . \quad (3.9b)$$

In higher dimensions the process is somewhat more complicated. The general process involves splitting the interval into layers of constant spacetime volume, and then mapping these to layers of the unit cube with uniformly distributed random numbers so as to sprinkle evenly into each layer. The calculations will differ depending on the spacetime, but an example of  $\mathbb{M}^3$  is sketched here.

As in [181] using polar coordinates,  $(\tau, r, \theta)$  where

$$u_0 = \tau , \quad (3.10a)$$

$$u_1 = r \cos(\theta) , \quad (3.10b)$$

$$u_2 = r \sin(\theta) . \quad (3.10c)$$

Taking layers of constant  $\tau$ , we calculate the spacetime volume of the interval up to a given  $\tau$ , in the lower and upper halves of the interval. We will denote the volume of the interval as  $V(I)$ , and the volume up to time  $\tau$  as  $V(I, \tau)$  and then find

$$V(I, \tau | \tau \leq 1/2) = \int_0^\tau d\tau' \int_0^{\tau'} r dr \int_0^{2\pi} d\theta \quad (3.11a)$$

$$V(I, \tau | \tau \leq 1/2) = \frac{\pi \tau^3}{3} \quad (3.11b)$$

$$V(I, \tau | \tau > 1/2) = V(I) - \int_\tau^1 d\tau' \int_0^{\tau'} r dr \int_0^{2\pi} d\theta \quad (3.11c)$$

$$V(I, \tau | \tau > 1/2) = \frac{\pi}{12} - \frac{\pi(1-\tau)^3}{3} . \quad (3.11d)$$

Our aim then is to construct a map such that volumes are preserved. We can do this by stipulating that if a random number  $0 < x_0 < 1$  must map to the  $\tau$  which has  $x_0$  of the

volume of the interval below it, or more precisely

$$x_0 V(I) = V(I, \tau(x_0)) . \quad (3.12)$$

Which is achieved by:

$$\tau(x_0) = \begin{cases} \frac{x_0}{4}^{1/3} & \text{if } x_0 \leq 1/2 , \\ 1 - \frac{1-x_0}{4}^{1/3} & \text{if } x_0 > 1/2 . \end{cases} \quad (3.13)$$

The other coordinates are easier, as we now just need to scatter in each disk of constant  $\tau$  uniformly, given random numbers  $x_1$  and  $x_2$ . This is achieved by:

$$r(x_0, x_1) = \begin{cases} \tau(x_0) \times \sqrt{x_1} & \text{if } x_0 \leq 1/2 , \\ 1 - \tau(x_0) \times \sqrt{x_1} & \text{if } x_0 > 1/2 , \end{cases} \quad (3.14a)$$

$$\theta(x_2) = x_2 \times 2\pi . \quad (3.14b)$$

This gives us coordinates  $(\tau, r, \theta)$  which will be uniformly distributed within the unit interval in  $\mathbb{M}^3$  from random numbers uniformly distributed in  $[0, 1]^3$ . Importantly, unlike in the trial-and-error method we do not need to discard random numbers until we have  $N$  results because every generated point is guaranteed to fall inside the interval.

### 3.5 Summary

Causal set theory is by no means complete, and there is plenty of work to be done in the mathematical analysis of how causal sets correspond to continuous spacetimes, let alone how to create the wider theory needed to make testable physical predictions. The useful part of causal set theory for this thesis is the idea that discrete mathematical objects with causal structure can correspond to continuous spacetimes in a meaningful way via the idea of a faithful embedding. Since DAGs have causal structure, with each node having a set of ancestors and a set of descendants, we can use the tools of causal set theory to analyse DAGs by comparing them to continuous spacetimes in a variety

of ways.

We can think of these causal set graphs as extending the ‘network geometry’ idea discussed in the previous chapter from space (and Riemannian geometry) to spacetime (and Lorentzian geometry). The ‘edge means similar’ interpretation of network geometry makes spatial distances are useful. The ‘edge means causally related’ interpretation of DAGs makes Lorentzian spacetime the natural and appropriate embedding manifold. Given the insight that network geometry has had into the structure of many real world networks, we hope to find that developing a geometric approach for this new class of networks can prove equally fruitful.

## Chapter 4

# Transitive Reduction

The true method of knowledge is  
experiment.

---

William Blake

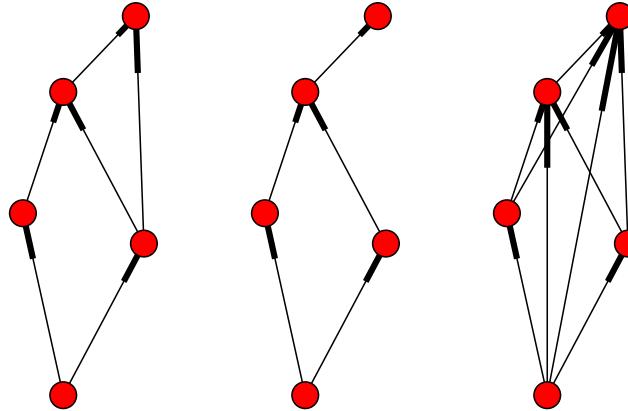
*This chapter is based on material published in:  
Transitive Reduction of Citation Networks,  
J. R. Clough, J. G. Gollings, T. V. Loach, T. S. Evans,  
Journal of Complex Networks, 2015.*

Network science concerns itself not only with abstract mathematical models but importantly with real datasets and so here we will begin to investigate real complex systems which can be represented as networks which form DAGs. As noted previously, causal structure is a key feature of this class of networks and so we seek real systems in which objects of interest, which will form the vertices of the network, are related to each other by causal relationships, which will be the directed edges of the network.

Before any other analysis there is an important distinction to be made between the structure of networks which form DAGs, and causal sets. One of the causal set axioms was *transitivity*, the restriction that if  $a \prec b$  and  $b \prec c$  then it must be the case that  $a \prec c$ . This is not necessarily the case for a network forming a DAG, if  $\prec$  represents an edge in the network. In the case of citation networks, when I write a paper, I might cite some important paper in my field but I do not have to then cite all the papers in

its bibliography (and all the papers in all of their bibliographies and so on). However in the case of a software dependency network, if an edge from  $a$  to  $b$  in that network corresponds to ‘ $a$  requires  $b$  to be installed’ then that relation clearly is transitive. To describe this difference we introduce the following notation.

A DAG,  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$  is *transitively closed* or *transitively complete* if  $\forall a, b, c \in \mathcal{V}, (a, b) \in \mathcal{E}$  and  $(b, c) \in \mathcal{E} \implies (a, c) \in \mathcal{E}$ . The *transitive closure* or *transitive completion* of a DAG is the operation which adds edges between any such  $(a, c)$  pair to make transitively closed graph [9]. In the example of a family tree network, originally an edge may represent ‘is the parent of’ but after transitive closure, an edge represents ‘is an ancestor of’. A transitively complete graph has the same causal structure as the original graph, because the new added edges cannot change which nodes are in the past, or future of others. It has as many edges as can be added without changing the causal structure since every possible pair of causally related nodes share an edge.



**Figure 4.1:** Left: A DAG,  $G$  which is neither transitively reduced or closed. Centre: The transitive reduction of  $G$ . Right: The transitive closure of  $G$ .

We can also perform the opposite operation, to find the graph with the *fewest* possible number of edges which retains the causal structure of the original. This operation is called *transitive reduction* [9]. After transitive reduction, an edge only exists between  $a$  and  $b$  if  $a \prec b$  and there is no node in the middle  $m$  such that  $a \prec m \prec b$ . If node  $m$

did exist then we would not need the edge  $(a, b)$  since it would be implied by  $(a, m)$  and  $(m, b)$ . Performing transitive reduction on a graph removes all edges except those which are necessary for its causal structure, and this causal structure is again the same as that of the original graph. When drawing the Hasse diagram of a poset, we are effectively drawing the transitive reduction of its graph [10]. Algorithms describing transitive closure and transitive reduction are given respectively in algorithm 1 and algorithm 2, using implementations from networkx,<sup>1</sup> a Python package used in this project.

---

**Algorithm 1 :** Transitive Closure

---

```

1: procedure TRANSITIVE CLOSURE( $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ )
2:   for  $u \in \mathcal{V}$  do
3:     for  $v \in$  DEPTH FIRST SEARCH( $\mathcal{G}$ , source =  $u$ ) do
4:       ADD EDGE( $\mathcal{G}, u, v$ )

```

---

**Algorithm 2 :** Transitive Reduction

---

```

1: procedure TRANSITIVE REDUCTION( $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ )
2:   for  $u \in \mathcal{V}$  do
3:     for  $v \in \mathcal{V} \mid (u, v) \in \mathcal{E}$  do
4:       for  $w \in$  DEPTH FIRST SEARCH( $\mathcal{G}$ , source =  $v$ ) do
5:         REMOVE EDGE( $\mathcal{G}, u, w$ )

```

---

DAGs formed from real datasets are not necessarily transitively reduced or complete, but somewhere in between meaning that some of the edges implied by transitivity are there, and others are missing. When analysing the causal structure of a DAG the simplest question to ask is how many of these extra transitive edges are there? What changes if we remove them? Can we identify these edges as representing a different kind of relationship to those edges that remain after transitive reduction? These are the questions we will investigate in this chapter.

---

<sup>1</sup>See <https://networkx.github.io/>.

## 4.1 Effect on network structure

To begin we will compare simple structural measures of the networks before, and after transitive reduction, on the citation networks described in chapter 1. If the network in question is not a DAG, then the result of transitive reduction is not necessarily well defined [9], and so the acausal edges mentioned in Table 1.1 are removed to ensure that this criterion is met. Table 4.1 shows the effect of transitive reduction on the average in-

	$N$	$\langle k \rangle$	$\langle k_{\text{TR}} \rangle$	$\frac{\langle k_{\text{TR}} \rangle}{\langle k \rangle}$
hep-th	60880	17.96	2.28	12.7%
hep-ph	76836	17.63	2.65	15.0%
astro-ph	133346	14.1	3.36	23.8%
cond-mat	118496	6.46	2.36	36.2%
gr-qc	27175	8.72	2.29	26.3%
quant-ph	36955	7.12	2.44	34.3%
SCOTUS	25378	8.52	2.32	27.3%
US Patent	3774753	4.38	3.13	71.6%
Clinton Emails	5674	18.81	8.41	44.7%

**Table 4.1:** Table of in-degrees for various citation networks before and after transitive reduction.

degrees of nodes in various citation networks. We can see that for the citation networks from the arXiv, and from the US Supreme Court, most edges, approximately 70 – 80% are lost in transitive reduction.

How can we interpret this result? In [185, 186, 187], Simkin and Roychowdhury investigate whether or not authors of scientific papers have read everything that they cite. By tracking typographical errors in bibliographies they estimate how often an author has copied an entry from the bibliography of another paper they have cited. If paper  $A$  cites paper  $B$  in its bibliography, it may contain an error in the name of an author, the journal in which it was published or in the title of the paper. If paper  $C$  is

then written, citing both  $A$  and  $B$ , making the exact same error as  $A$  did, Simkin and Roychowdhury assume that this is because the authors of  $C$  simply copied the information for paper  $B$  from the references of  $A$ . In the days before automatic referencing software, this suggests that these authors did not have a copy of paper  $B$  to hand to note down the authors' names, since they copied from  $A$ . As the title of Simkin and Roychowdhury's paper 'Read before you cite!' [185] implies, they believe that these copying authors may not even be reading these papers at all. Interestingly the fraction of papers that they estimate are cited in this bibliography-copying manner is around 80% meaning that only 20% are read by authors<sup>2</sup>, which is similar to the fraction of edges removed by TR in the citation networks of academic papers.

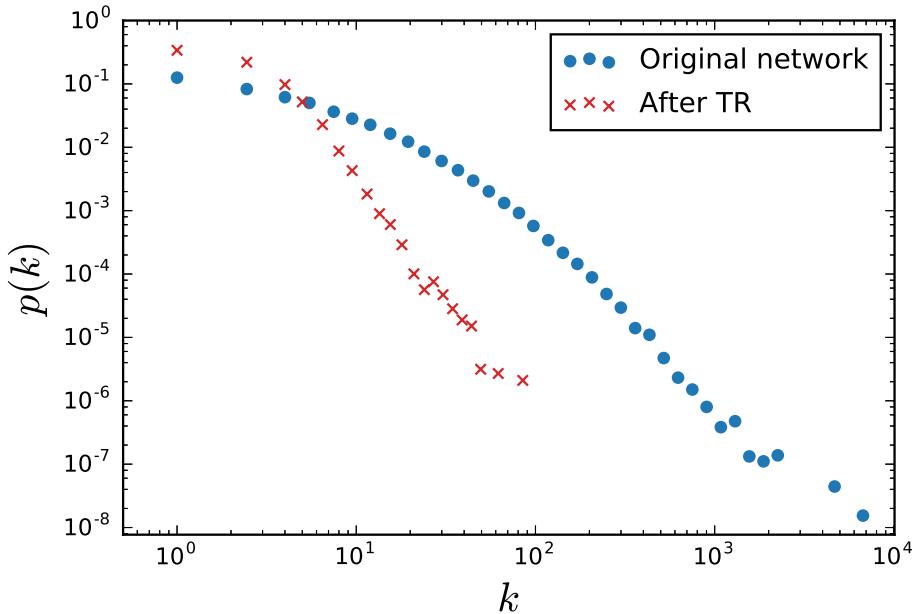
How might we identify such edges from the citation network? Well, the copied citations identified in Simkin and Roychowdhury's statistical analysis are necessarily those removed by transitive reduction, since the propagation of typographical errors can only be seen in such triangles of citing papers as described above. The edge from paper  $C$  to paper  $B$  would be removed, and it was precisely this one which may have been copied. Of course, we do not have the data to know whether or not an author genuinely found another paper useful and cited it for that reason, copied the citation of a famous paper from a bibliography of another, or made that citation for some other reason such as those discussed in [53, 131]. However, it is surely reasonable to say that a citation which is retained after transitive reduction is less likely to have been copied from elsewhere, since the paper from which it may have been copied was not also cited, and so it is more likely to have been read than otherwise.

As well as counting how many edges are removed by transitive reduction, we can observe how it changes the overall degree distribution. Figure 4.2 shows the degree distribution of the hep-th arXiv citation network before and after transitive reduction. The degree distribution for the original network is the classic fat-tailed shape commonly observed in many citation networks, such as for papers in Physical Review [162] or the SPIRES database [127]. Typically these distributions are approxim-

---

<sup>2</sup>It is up to the reader to decide whether this figure is a reasonable estimate of the fraction of papers in a bibliography that have actually been read.

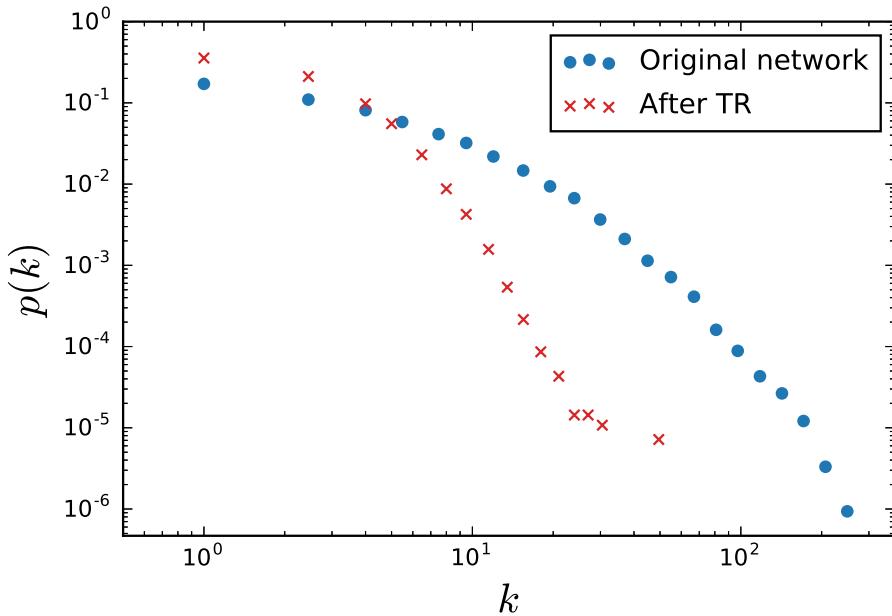
ately power-law for large  $k$ , which corresponds to a straight line in log-log plots such as Figure 4.2. The significant leftward shift of the curve after transitive reduction re-



**Figure 4.2:** Degree distribution of the hep-th citation network before and after transitive reduction. The characteristic fat-tailed shape of the original degree distribution shifts leftwards on this plot after transitive reduction as most of the network's edges are removed. Other citation networks of academic papers show similar effects.

flects the fact that most of the edges in the network have been removed by this process, and so the probability of finding high-degree nodes has significantly lessened. The steepening of the curve shows us that it is the nodes with high degrees in the original network which lose disproportionately many edges. This observation lends some further evidence to the hypothesis that edges removed by transitive reduction are more likely to be those which have been copied, in the sense ‘Read before you cite!’, since famous papers which are cited without being read are likely to have many citations, and so such papers are likely to lose many edges in transitive reduction.

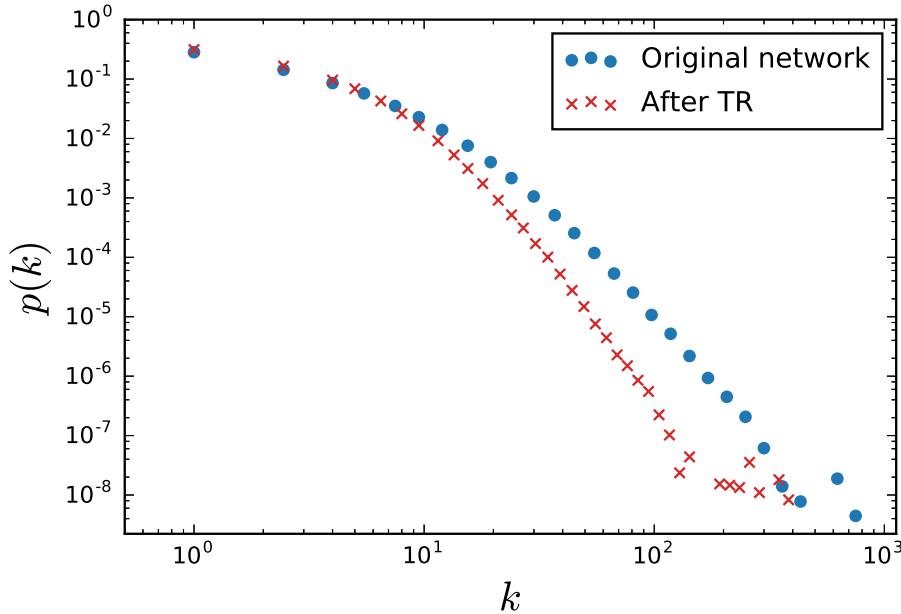
Figure 4.3 shows the same plot for the US Supreme Court (SCOTUS) citation network of court judgements. Again, we see that the curve after transitive reduction is steeper than that of the original network. This effect cannot be explained by a naive copying mechanism judges on the Supreme Court are likely to be familiar with the



**Figure 4.3:** Degree distribution of the SCOTUS citation network before and after transitive reduction. The characteristic fat-tailed shape of the original degree distribution shifts leftwards on this plot after transitive reduction as most of the network's edges are removed.

cases that they cite and to have read them. A possible explanation though is that, due to the long timespan of this citation network (more than two centuries) it seems likely that, when separated by sufficiently long times, most judgements will have some path in the citation network to most others, and so these edges will also be lost to transitive reduction.

Figure 4.4 shows the opposite case, in the US Patent citation network. Here, as indicated in Table 4.1 very few edges are lost under transitive reduction, and so there is less of a shift from the curve describing the original network to the reduced one. Clearly there is some significant difference in citation behaviour for patent-writers, as compared to academic authors or court judges. One possible explanation can be made by appealing to the differing incentives of writers in each case. Academics and judges can cite each other freely, with little cost. The more references given to a claim, the more plausible it appears and so there is a strong incentive to add citations to a piece of work to make it appear more convincing, or the authors more well read. Conversely,

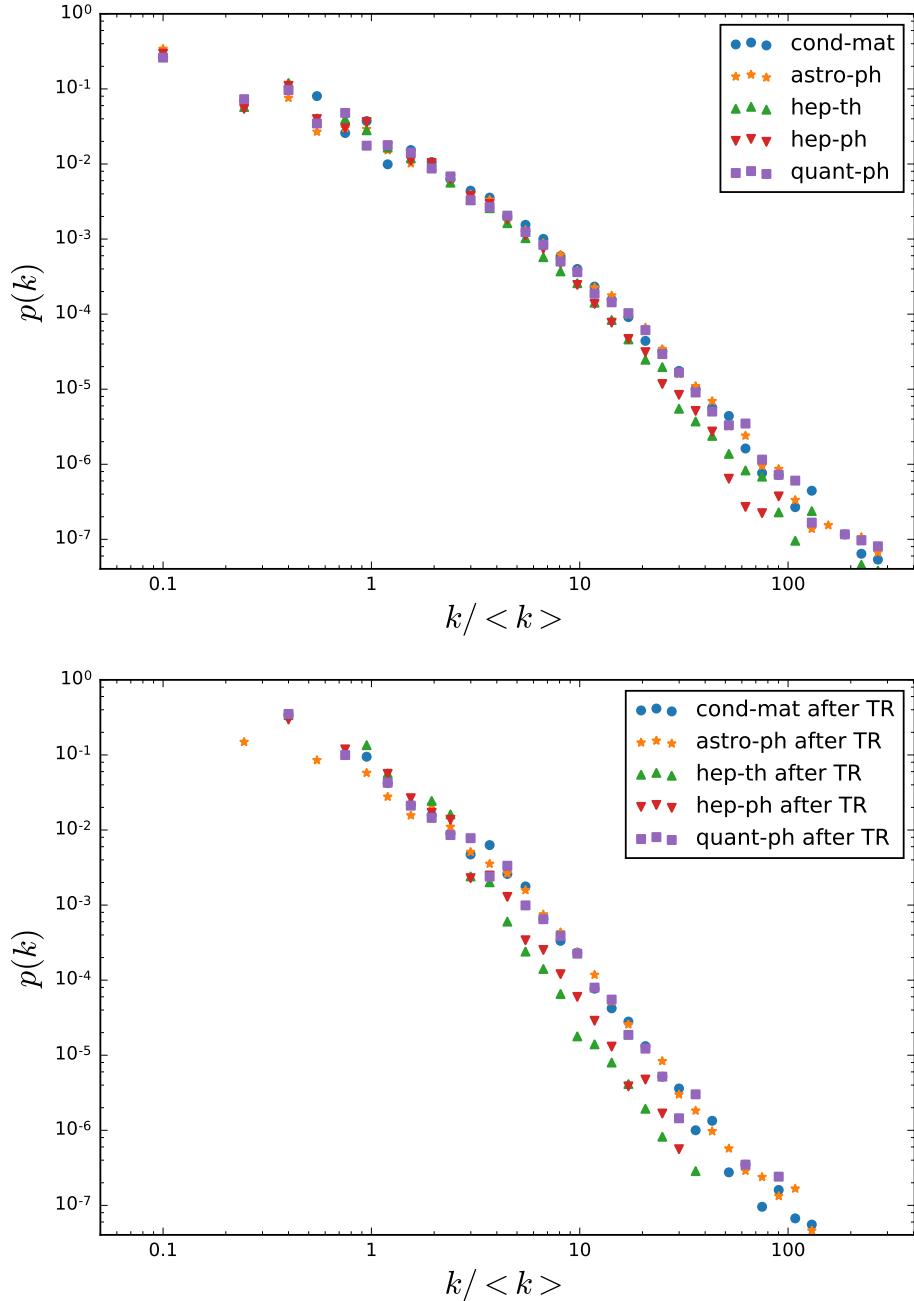


**Figure 4.4:** Degree distribution of the US patent citation network before and after transitive reduction. Few edges are lost and so the curve after transitive reduction is close to that of the original network, retaining its fat-tailed shape.

patent authors cite other patents out of legal obligation and a citation recognises work covered by a different patent which is therefore not the property of the inventor writing this one. The incentive of a self-interested patent writer is therefore to cite as little as it is possible to get away with. If edges that are removed by transitive reduction are viewed as less necessary to the network's structure, or less likely to correspond to information flow between authors then it makes sense that those authors who are incentivised to cite excessively will add such 'redundant' edges, while those who are punished for doing so will avoid it as much as possible [62].

We cannot at this stage identify with certainty the citation dynamics which drive the differing effects in these different citation networks. However, network science has had success in identifying universal features of different complex networks. Citation networks of academic papers from different fields of study appear to have similar scale-free structure, such that their degree distributions collapse on to one curve when scaled by the average degree [161]. This suggests that, when accounting for the fact that some fields just have more researchers and thus more citations than others, the

process by which new citations are made is universal. In this vein, we can ask whether the transitively reduced network appears similar across different citation networks too. In Figure 4.5 we see that there is a universal shape of the citation distributions amongst the arXiv citation networks, where each network's degree distribution is scaled by the average degree, as in [161]. Notably, after transitive reduction this still remains the case, and all of the points still fall on the same universal curve, despite a majority of the edges having been removed. This suggests that, in as much as the shape of the degree distribution of citation networks appears to be a common property amongst many different cases, the underlying skeleton of the edges surviving TR is as well. In the same sense that there appears to be universal dynamics behind the growth of citation networks, there are also another set of universal dynamics behind the growth of the reduced network over which transitive edges are then placed. Models such as that in [91] have used this idea of a reduced network overlaid with a copying process to recreate networks with these features.

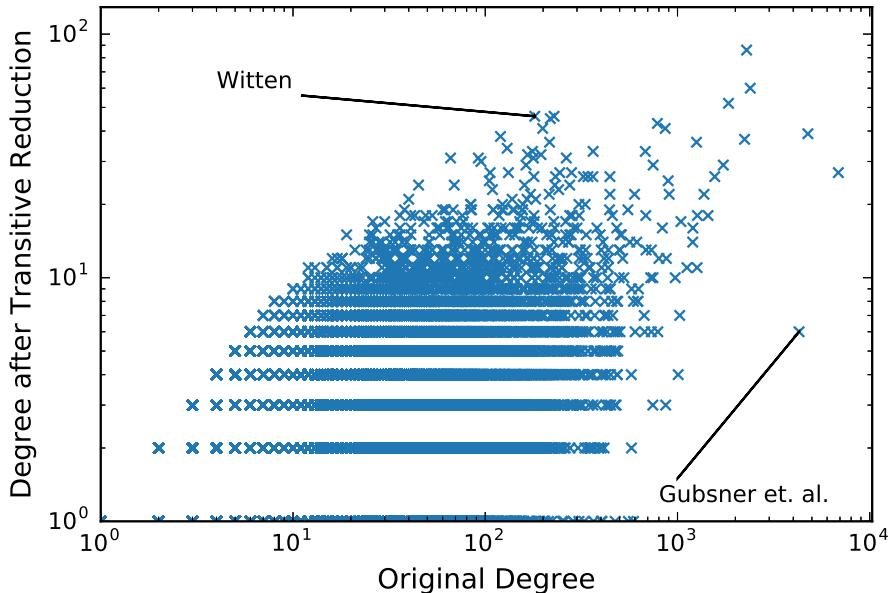


**Figure 4.5:** Degree distributions for the arXiv citations networks show a universal shape when scaled by the mean degree. Top: original citation networks. Bottom: citation networks after transitive reduction. This data collapse illustrates that the same shape degree distribution is present in all of the arXiv citation networks before, and after TR, even though in all cases most of the edges are removed.

## 4.2 Effect on individual nodes

In the previous section we saw that for citation networks from academia, transitive reduction removed most of the networks' edges and revealed a universality in structure, while for a citation network of patents very few edges were removed. Even though the degree distributions for a citation network of academic papers, and a citation network of patents looks very similar, transitive reduction reveals that the underlying causal structure of these networks is in fact very different.

Now we will look at the effect on individual nodes within a network, taking the hep-th network again as an example. Figure 4.6 shows there is a very wide range of effects on the nodes' degrees even within one citation network, an effect also present in the other academic citation networks studied here. This diversity of effect indicates a diversity of citation behaviour. For example, paper [97] ('Gauge Theory Correlators from Non-Critical String Theory', Gubser, Klebanov, Polyakov, 1999) has 4289 citations in the original citation network, making it the third most highly cited. After transitive



**Figure 4.6:** The degree before, versus the degree after transitive reduction for the hep-th citation network. The points are bounded by the line  $y = x$  since edges can only be lost, and not gained by transitive reduction. There is clearly a large spread of points indicating a diverse range of effects on individual papers. Highlighted is Gubsner *et al.* [97], which loses disproportionately many edges, and Witten [212] which loses disproportionately few.

reduction this value falls to just six citations indicating that all of the other papers citing this one were also causally related to some of these six papers. On the other hand, paper [212] ('Ground Ring Of Two Dimensional String Theory', Witten, 1993) begins with just 181 citations and but retains 46 after transitive reduction, meaning there were 46 papers citing it none of whom cited each other. This suggests a wider range of readers citing the paper, unaware of, or not interested in each others work, whereas the first paper appears to be of interest to a more cohesive group.

If there was a single process driving the growth of a citation network, we might expect to see a very strong correlation between a node's degree before and after TR, since there would be no other factor to cause them to differ[62]. The large spread of points in Figure 4.6 suggests that in fact there are competing processes since some papers gain most of their new citations from papers they already have a path to in the citation network (those which lose many edges), whereas other papers receive many more citations from papers they are otherwise unconnected to in the network (those which retain many edges).

The hypothesis that it is this *reduced citation count* which is a better indicator of real information flow would then imply that some papers have had a much higher impact than their raw citation count suggests, because they have been read and cited by many researchers independently. Even if the reduced citation count is not a superior measure of information flow in the network, it does indicate that the network's growth process is not uniform across all nodes as some gain most of their edges from new nodes with neighbours in common (those with a low reduced degree) while others gain most of their edges from new nodes which are not linked to each other (a higher reduced degree).

Finally, all of the later measures of causal structure in coming chapters are invariant to transitive reduction, only depending on the underlying skeleton of the network. Important features such as the number of ancestors and descendants a node has, which pairs of nodes are comparable, the number of chains and antichains, and the longest chains between any pair are unchanged. Transitive reduction is a useful pre-processing

step to computing these measures since the resulting reduced graph has as few edges as possible whilst still retaining the essential causal structure.

### 4.3 Summary

The main interest of this thesis is to analyse networks forming DAGs by considering their relationship with Lorentzian spacetimes and that means looking at the causal structure of the network. Transitive reduction reveals the causal skeleton of the network by removing everything else. When we perform this operation on real networks the results are not trivial and uniform but instead can reveal differences in otherwise similar networks, and differences between otherwise similar nodes in those networks. The simple idea expressed in this chapter is that looking at a network's causal structure can yield new insight into the system it represents, and the dynamics of the network's growth. In the coming chapters we will explore methods of characterising that causal structure mathematically.

# Chapter 5

## Measuring dimension

If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.

---

John Von Neumann

*This chapter is based on work presented in:*

*What is the dimension of citation space?*

*J. R. Clough, T. S. Evans, Physica A, 2016.*

Our overall aim is to analyse networks which form DAGs by associating them, or more precisely, their causal structure with Lorentzian spacetimes. To know which spacetime corresponds to which graph we must find some observable depending only on the causal structure of the graph, which we can measure to tell us about the kind of spacetime which is associated with it. In this chapter we will consider the simplest Lorentzian spacetime, which is Minkowski spacetime  $\mathbb{M}^D$ . There is only one parameter defining this spacetime which is its dimensionality,  $D$ , so this is what we will measure.

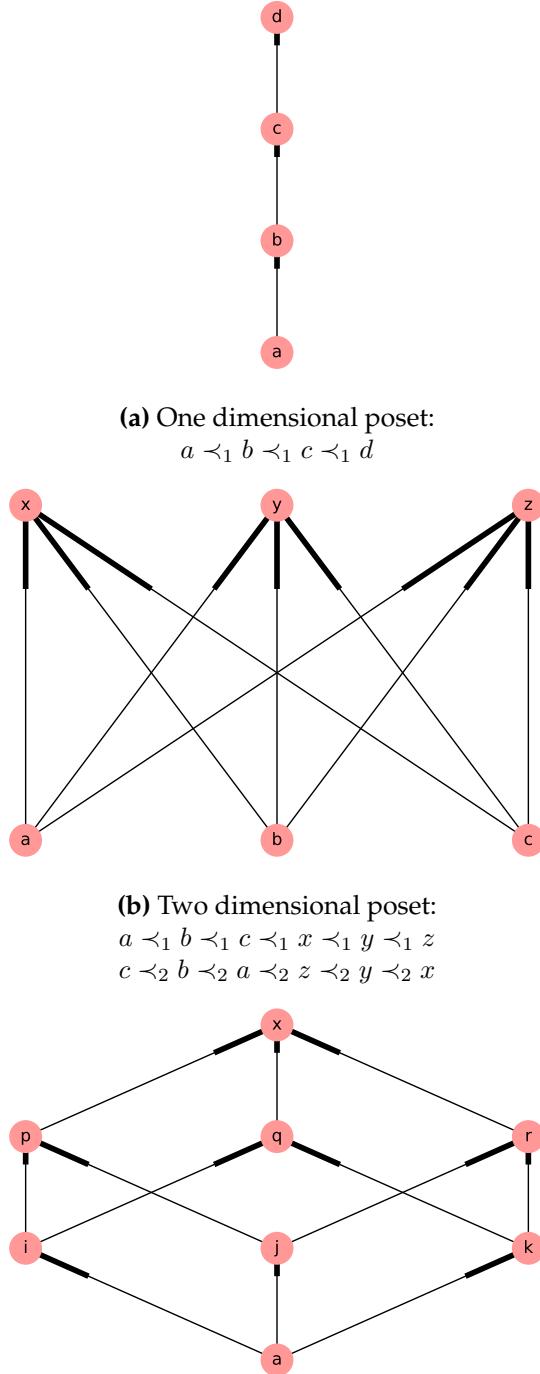
Two approaches to measuring this quantity will be presented here. Firstly, box-counting methods where we use estimates of distance and volume to find how one scales with the other, giving a measurement of dimensionality. Secondly, a chain-

counting method where we first estimate how many causally connected chains of length  $k$  we expect to see in a graph embedded in  $\mathbb{M}^D$  and then fit this number to the observed number of such  $k$ -chains. This chapter will introduce these methods in theory and demonstrate their effectiveness on causal set graphs, for which purpose they were first developed. We will then use these methods to provide an estimated Minkowski spacetime dimension for the real networks introduced previously and give an interpretation of these results in terms of citation behaviour. Furthermore, in chapter 7 we actually calculate the embedding coordinates for the nodes in these networks and calculating the dimension of that embedding spacetime is a necessary precursor to that process.

## 5.1 Order Dimension

The simplest way to associate a dimension to a poset (and so also a causal set or a DAG) is the ‘order dimension’ first discussed by Dushnik and Miller [73]. Given a partial order  $\mathcal{P} = \{P, \preceq\}$ , this dimension measures the minimum number of linear extensions of  $P$ , the intersection of which fully defines the partial order  $\preceq$ . Recall that a linear extension is a full order,  $\tilde{\preceq}$  on the elements of  $P$ , such that  $x \preceq y \implies x \tilde{\preceq} y$ . By saying that an intersection of linear extensions fully define a partial order, we mean that, for linear extensions  $\tilde{\preceq}_1, \tilde{\preceq}_2, \dots, \tilde{\preceq}_D$ , then  $x \preceq y \Leftrightarrow x \tilde{\preceq}_1 y \text{ and } x \tilde{\preceq}_2 y \text{ and } \dots \text{ and } x \tilde{\preceq}_D y$ .

Some examples of small posets with different order dimensions are given in Figure 5.1. A poset with an order dimension of 1 (eg. Figure 5.1a) is one in which all elements are related to each other, and so the partial order is in fact a total order. The poset is one-dimensional in the sense that its whole structure is determined by one ordered list of its elements. This corresponds to an embedding in  $0 + 1$  dimensional spacetime, where there is no spatial dimension so every pair of points must be time-like separated and so is related to each other. The poset in Figure 5.1b has an order dimension of 2. There are three mutually unrelated elements  $a, b, c$  which are all in the past of three more mutually unrelated elements  $x, y, z$ . To allow for unrelated elements a spatial dimension is needed so that some elements may be spatially separated from



**Figure 5.1:** Examples of small posets with varying order dimensions. If a poset has order dimension  $D$  then its structure can be fully determined by as few as  $D$  linear extensions. Such linear extensions,  $\preceq_i$  are given for each case.

each other. The poset in Figure 5.1c has an order dimension of 3 because 3 different full orders are required to fully define this partial order.

This measure of dimension has a natural geometric interpretation in terms of the *cube space* first introduced by Winkler in [211, 210]. In this model each point has spatial coordinates in  $[0, 1]^{D_{\text{cu}}}$  and one point precedes another if its coordinates are smaller in every dimension<sup>1</sup>. The smallest  $D_{\text{cu}}$  for which a given poset can be assigned coordinates in the cube space model is equal to its order dimension, and this is because the ranking of the points along each of the  $D_{\text{cu}}$  coordinates corresponds to their place in the  $D_{\text{cu}}$  linear extensions used in the order dimension.

Despite its definitional simplicity though, we will not use the order dimension to characterise the structure of DAGs, and this is for a few reasons. Firstly, it is inherently difficult to compute. Calculating the order dimension of a poset is an NP-complete problem [213]<sup>2</sup>. This makes it so costly for large complex networks as to be infeasible to use.

Secondly, the order dimension does not tell us something about the typical properties of the graph, but rather some extremal properties, because this measure requires the whole graph to be determined by  $D$  linear extensions, even if most of the graph could be determined by a much smaller number. This means that the order dimension for a large network could be completely determined by a small subgraph with a particular pathological structure meaning that a vast majority of the network may play no part in the measurement at all. We would prefer a more statistical measure where different parts of the network contribute in some proportional way.

Thirdly, the implicit geometry used in calculating the order dimension is that of a cube space, and not Minkowski spacetime. It is tempting to suggest a similar method where one defines a new order dimension based on the lowest dimensional Minkowski spacetime into which a poset can be embedded, with the poset's order arising from

<sup>1</sup>Again we use the  $[0, 1]$  interval out of convention but note that here it is only the order of the points in each dimension which is relevant so this choice makes no difference to the resulting poset.

<sup>2</sup>For the special case of determining if the order dimension is at most 2 a polynomial time algorithm does exist, but such a low order dimension is unlikely for complex networks since to have a higher order dimension only requires a subgraph such as the poset in Figure 5.1c.

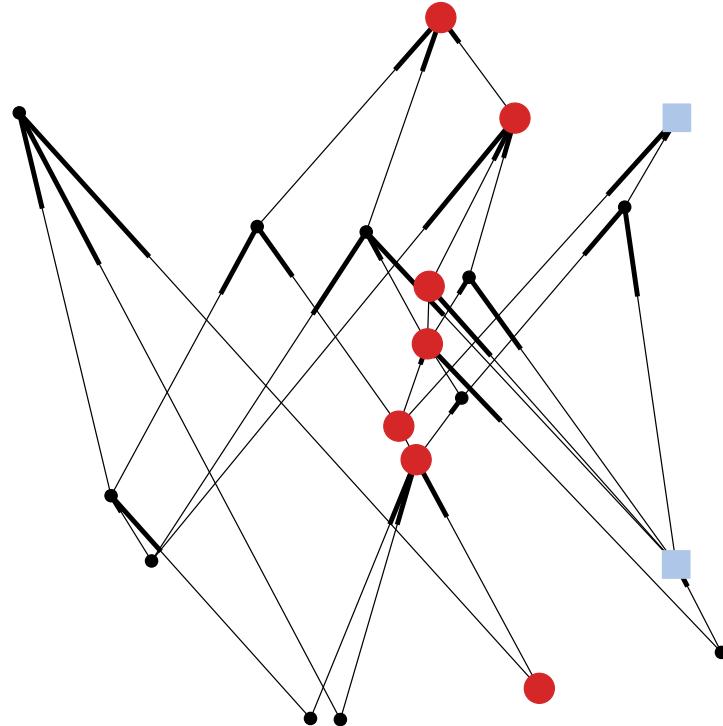
the causal structure of the spacetime. Such a method was defined in [141] where this value is called the ‘Minkowski dimension’ of a poset, but is related to previous work on ‘circle orders’ [180, 52, 95]. Unfortunately though this definition is still problematic for our intended purposes. Not only does it still suffer from the above problems but there is an additional issue which is that, surprisingly, there exist finite posets which cannot be embedded in Minkowski space of any finite dimension at all [79], and so have an undefined, or infinite Minkowski dimension. Furthermore, as noted in [49] a poset which is not embeddable in a  $D$ -dimensional Minkowski spacetime may be embeddable in a  $D$ -dimensional curved spacetime.

Instead of the order dimension, to characterise the geometric structure of our complex networks we will use statistical measures which first arose in causal set theory. This will involve calculating the expected values of structural properties of causal sets whose elements are uniformly sprinkled in Minkowski spacetime as functions of the spacetime’s dimension, and then comparing these values to those measured in a network to infer for it an estimated dimension.

## 5.2 Geodesics and Longest paths

In chapter 3 while discussing causal sets, we noted how the number of elements in an interval in the causal set corresponded to the notion of that interval’s volume in an embedding spacetime. This means that volume can be estimated just by counting the number of nodes in an interval, a very simple graph observable. What graph observable corresponds to the notion of *distance* in the embedding spacetime? For this task we will only need to consider timelike separation, and so we seek a way of estimating this quantity for related pairs of nodes in a causal set graph. As defined in chapter 3, a *chain*  $S$  is a set of vertices in the graph which are all related to each other. The elements of  $S$  form a sequence,  $x_1 \prec x_2 \prec \dots \prec x_k$ , and we will call  $S$  a  $k$ -chain when it has  $k$  elements. We will say that  $S$  is a chain from  $x$  to  $y$  if  $x$  is the first element in this sequence and  $y$  the last.

The result now known as the *Myrheim length conjecture* [145] states that when the



**Figure 5.2:** Two chains in a causal set graph. The 7-chain of circular nodes is maximal, meaning no more nodes can be added to it while it still remains a chain. The 2-chain of square nodes is not maximal since there are other nodes which could be added to this chain. Note that a maximal chain must consist only of links, but non-maximal chains, such as that on the right may have other nodes in between nodes in the chain. Furthermore, note that the chain of circular nodes, which is the longest chain between its two endpoints is a good approximation to a straight line between the endpoints, illustrating how the length of the longest chain can approximate the timelike distance between two nodes in an embedding spacetime.

causal set elements are sprinkled in a Lorentzian spacetime the timelike separation between points  $x$  and  $y$  in the continuum is closely approximated by the length of the *longest* chain from  $x$  to  $y$ . It was proven to be the case in flat spacetime [40] and the result holds for curved spacetime when the scale of the curvature is much larger than the typical distance between the causal set elements [49, 19]. Figure 5.2 gives an example illustrating that a chain approximates the path of the geodesic between the chain's two endpoints.

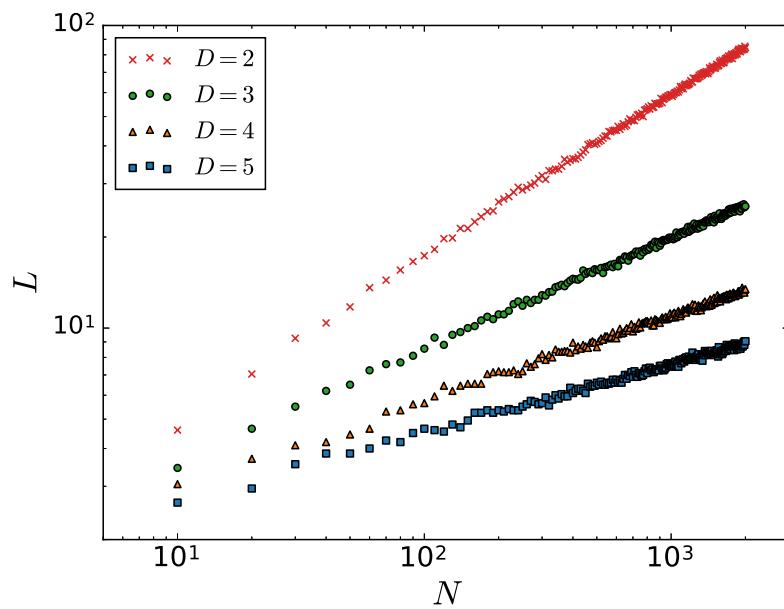
For the purposes of dimensionality estimation, timelike separations will be sufficient since we are interested in comparing the height of an interval with its volume,

and the endpoints of an interval are necessarily timelike separated. Estimating space-like separation is more complicated and will be discussed later in chapter 7.

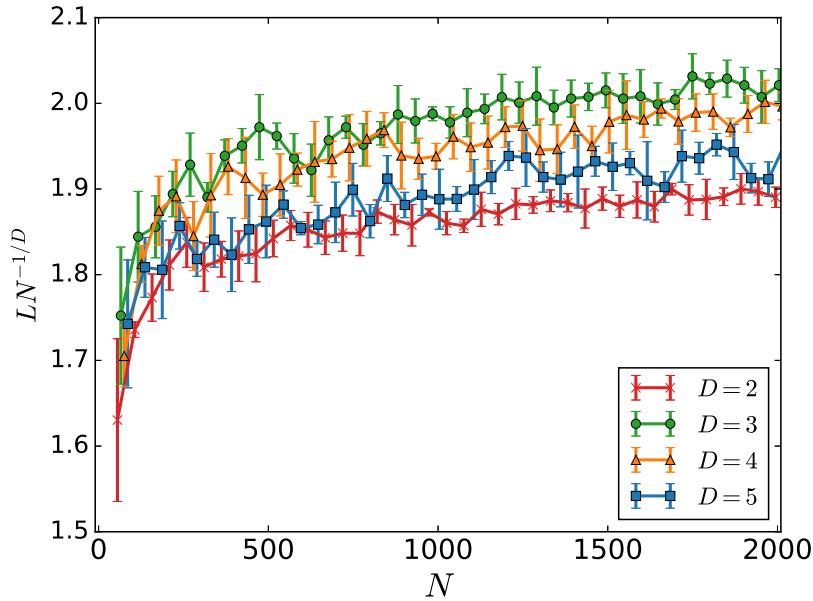
The Myrheim length conjecture tells us that for a given interval, the length of its longest chain  $L$  scales with its population  $N$  as:

$$\lim_{N \rightarrow \infty} LN^{-1/D} = m_D \quad (5.1)$$

in the limit of  $N \rightarrow \infty$ , for  $D$ -dimensional Minkowski spacetime and some constant  $m_D$  which depends on the dimension. While surprisingly little is known about the values of these constants,  $m_2 = 2$  and it is believed that no  $m_D$  is larger than  $e$  [40], but fortunately their precise values do not matter for our purposes. Using this simple equation we should then be able to estimate  $D$  by sampling many intervals and observing how  $N$  scales with  $L$ , as in Figure 5.3 and Figure 5.4.



**Figure 5.3:** Volume and distance scaling in causal set graphs. The length of the longest chain,  $L$ , which approximates timelike distance in the causal set sprinkling, scales to first order as  $L = m_D N^{1/D}$ , for some constant  $m_D$  which depends on the dimension only. The number of points in the interval,  $N$ , approximates volume. Points on the plot show measured values for causal set graphs longest chains and each point is the average of 20 random sprinklings. The axes are logarithmically scaled so straight lines on the plot correspond to power laws in the data.



**Figure 5.4:** Data collapse of volume and distance scaling in causal sets. In this figure data is binned, and longest chain lengths divided by  $N^{1/D}$ . According to the Myrheim length conjecture, these curves should each converge to a dimension dependent constant  $m_D$ , which is around a value of 2.

Note though that this method requires us to sample many intervals to get one global estimate of dimensionality. What if this estimate varies throughout different regions of the graph? It would be preferable to have some sort of local estimate, for which we can take a particular interval and obtain an estimate of its dimension without having to calibrate it against many others. The first method we'll discuss here, the midpoint-scaling dimension estimate, is such an approach.

### 5.3 Midpoint-Scaling Dimension

The midpoint-scaling dimension estimate [49] takes an interval in the graph,  $I = [u, v]$ , where  $u, v \in \mathcal{V}(\mathcal{G})$  and estimates the interval's embedding dimension by comparing how sub-volumes and sub-heights scale within that interval. Firstly, a *midpoint* element,  $p \in I$  is found by a method described below. Secondly, two sub-intervals  $I_u = [u, p]$  and  $I_v = [p, v]$  are defined. Since  $p$  is an element of  $I$  we know that  $I_u$  and  $I_v$  are non-empty; each contains at least its two endpoints. We then compare the

populations of these two sub-intervals,  $|I_u| = N_u$ ,  $|I_v| = N_v$  with the population of the total interval  $|I| = N$ . If the midpoint  $p$  is truly in the centre of the original interval's embedding spacetime we should expect  $N_u \approx N_v \approx N \times 2^{-D}$ . To have  $p$  as close to the interval's midpoint as possible, it is selected so as to maximise the population of the smaller subinterval,  $p = x \in I : \max_x(\min(N_u, N_v))$ . An example for a small causal set graph in 2-dimensional spacetime is given in Figure 5.5.

---

**Algorithm 3** : Midpoint-scaling Dimension

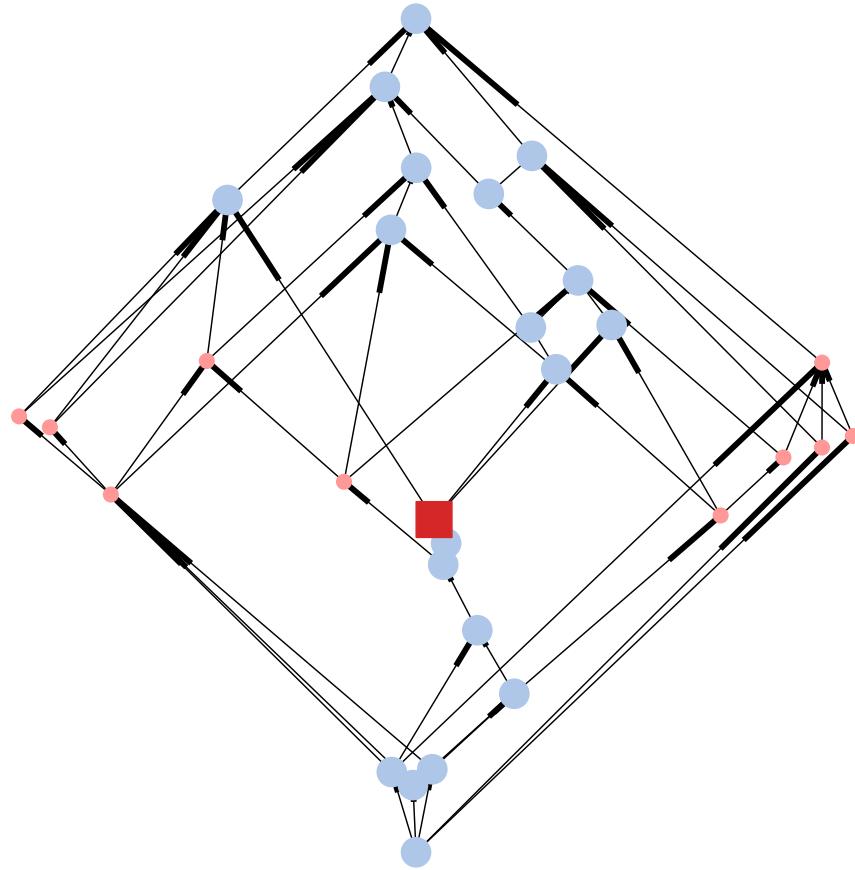
---

```

1: function MIDPOINT-SCALING DIMENSION( $\mathcal{G}, u, v$ )
2:    $N_{\max} \leftarrow 0$ 
3:    $I \leftarrow J^+(u) \cap J^-(v)$                                  $\triangleright$  interval between  $u$  and  $v$ 
4:    $N \leftarrow |I|$ 
5:   for  $p \in I$  do
6:      $I_{\text{lower}} \leftarrow J^+(u) \cap J^-(p)$                        $\triangleright$  lower interval
7:      $I_{\text{upper}} \leftarrow J^+(p) \cap J^-(v)$                        $\triangleright$  upper interval
8:      $N_{\text{midpoint}} \leftarrow \min(|I_{\text{lower}}|, |I_{\text{upper}}|)$ 
9:      $N_{\max} \leftarrow \max(N_{\max}, N_{\text{midpoint}})$ 
10:     $D \leftarrow \log_2 \frac{N}{N_{\max}}$ 
11:   return  $D$ 

```

---



**Figure 5.5:** An illustration of the midpoint-scaling dimension estimate on an instance of a  $D = 1 + 1$  causal set graph with  $N = 30$  nodes. With the square node as the midpoint, the upper and lower intervals (blue) are highlighted. The logarithm of the ratio of the size of these intervals and the size of the overall network gives the dimension estimate. The smaller interval, the lower one, has a population of 9 and in this case,  $D_{\text{MPSD}} = \log_2(30/9) \approx 1.74$ . Underestimation of the original causal set's dimension is usual for small  $N$ , as discussed below.

## 5.4 Myrheim-Meyer Dimension

A separate method of estimating a DAG's embedding dimension is the Myrheim-Meyer estimate which is based around counting the number of chains in an interval. For an intuitive approach on how this method works, consider the following extreme cases:  $\mathbb{M}^1$ , and  $\mathbb{M}^D$  for  $D \gg 1$ . In the first case,  $\mathbb{M}^1$ , there is only a time dimension and no space dimension, so the spacetime forms a one-dimensional line. Every point

is in the past or future of every other point as there is no spatial separation at all. The fraction of pairs of points which are related in a given direction (eg. in the past) is called the *ordering fraction* in the causal set literature [164]. For  $\mathbb{M}^1$  the ordering fraction is  $1/2$  since each point is, on average, in the past of half of the others. In the other extremal case,  $\mathbb{M}^{D \gg 1}$  there are many spatial dimensions contributing to make two random points spacelike separated, and only one time dimension contributing to their timelike separation, so most pairs of points will be spacelike separated and therefore the ordering fraction will be low. In general, the ordering fraction decreases monotonically as  $D$  increases, and since it can be measured it can give us an estimate of  $D$ . This idea can be generalised from considering just pairs of points to considering chains of  $k$  points<sup>3</sup>, the number of which we will denote  $S_k$ . Again, as  $D$  grows,  $S_k$  decreases. For an interval in Minkowski spacetime containing  $N$  sprinkled points, it was calculated in [140] that the expected number of  $k$ -chains is given by the expression

$$\langle S_k \rangle = N^k \frac{\Gamma\left(\frac{D}{2}\right) \Gamma(D) \Gamma(D+1)^{k-1}}{2^{k-1} k \Gamma\left(\frac{kD}{2}\right) \Gamma\left(\frac{(k+1)D}{2}\right)} \quad (5.2)$$

where  $\Gamma$  is the standard Gamma function, appearing here due to the hyper-spherical shape of the lightcones forming intervals in Minkowski spacetime. Such an expression is difficult, in general, to invert, so for use in calculations we will generally write

$$f_k(D) = \frac{\langle S_k \rangle}{N^k} \quad (5.3)$$

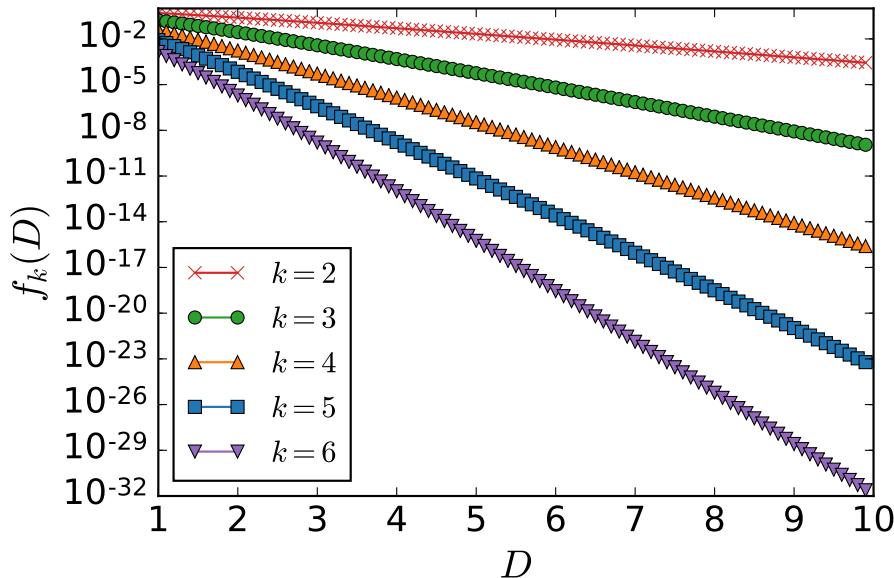
and since for any  $k$  and  $D > 1$  the function  $f_k(D)$  is monotonically decreasing, as seen in Figure 5.6, and therefore we can simply numerically invert it to provide an estimated  $D$  for a given interval. Table 5.1 gives some values of  $f_k(D)$  for integer arguments. In the simplest case,  $k = 2$  we are only measuring the ordering fraction, in which case our

---

<sup>3</sup>It is important here to recall that a chain of points here means a set of points which are all related to each other, and so in a real network which is not transitively complete, this will include points which do not share an edge so long as there is some directed path between them.

$f_k(D)$	$k = 2$	$k = 3$	$k = 4$
$D = 1$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{24}$
$D = 2$	$\frac{1}{4}$	$\frac{1}{36}$	$\frac{1}{576}$
$D = 3$	$\frac{4}{35}$	$\frac{2}{525}$	$\frac{4}{75075}$
$D = 4$	$\frac{1}{20}$	$\frac{1}{2100}$	$\frac{1}{705600}$

**Table 5.1:** Values for estimating the Myrheim-Meyer dimension, for small  $k$  and small  $D$ . In the top row, which is the one dimensional case,  $f_k(1)$  is simply the inverse of  $k!$  since all elements in the causal set are related there is only the combinatorial factor of counting the  $k!$  possible groups of  $k$  elements. As  $D$  increases,  $f_k(D)$  decreases because in larger dimensions the chance of elements being unrelated is higher.



**Figure 5.6:** Estimated fractions of  $k$ -chains used in the Myrheim-Meyer dimension estimator. The y-axis is logarithmically scaled, so the approximately straight lines on this plot indicate an exponential fall in the expected number of  $k$ -chains as dimension increases. Each curve is monotonically decreasing, so if  $k$  and  $f_k(D)$  are known,  $D$  can be inferred.

expression simplifies to

$$f_2(D) = \frac{\Gamma(D+1)\Gamma(\frac{D}{2})}{4\Gamma(\frac{3D}{2})}. \quad (5.4)$$

---

**Algorithm 4** : Myrheim-Meyer Dimension,  $k = 2$ 

---

```

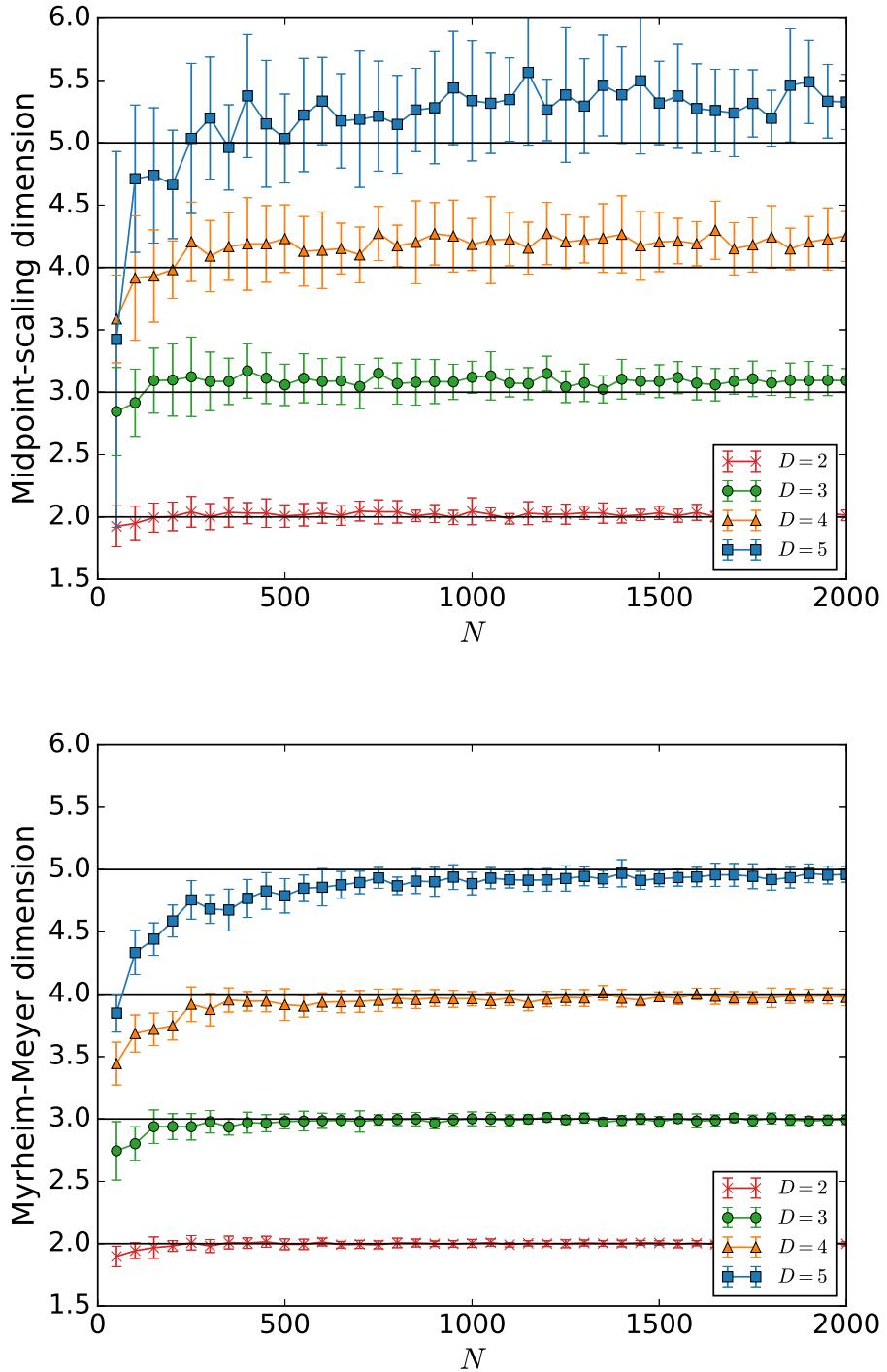
1: function MYRHEIM-MEYER DIMENSION( $\mathcal{G}, u, v$ )
2:    $I \leftarrow J^+(u) \cap J^-(v)$                                  $\triangleright$  interval between  $u$  and  $v$ 
3:    $N \leftarrow |I|$ 
4:    $S_2 \leftarrow 0$ 
5:   for  $w \in I$  do
6:     for  $z \in I$  do
7:       if  $w \prec z$  then
8:          $S_2 \leftarrow S_2 + 1$ 
9:    $f \leftarrow \frac{S_2}{N^2}$ 
10:   $D \leftarrow \text{INVERT MYRHEIM MEYER}(f)$ 
11:  return  $D$ 

```

---

Algorithm 4 describes the  $k = 2$  case but can be easily generalised for larger  $k$  by counting longer chains and using the more general chain estimate formula. Although the results shown in detail in this chapter only use the simplest version of the Myrheim-Meyer dimension where  $k = 2$ , results for  $k = 3$  and  $k = 4$  are shown in Appendix C. These further results are very similar to those with  $k = 2$ , illustrating a robustness to this method of characterising network structure.

Both of the midpoint-scaling and Myrheim-Meyer estimates are effective at recovering the embedding dimension of a causal set graph. Figure 5.7 shows numerical results of these estimates on causal set graphs of dimensions 2, 3, 4, 5. Note that in both cases, the causal set's dimension is systematically underestimated for small intervals, as also seen in previous studies in the causal set literature such as [164] and converges on the correct value of the causal set's dimension for larger intervals.



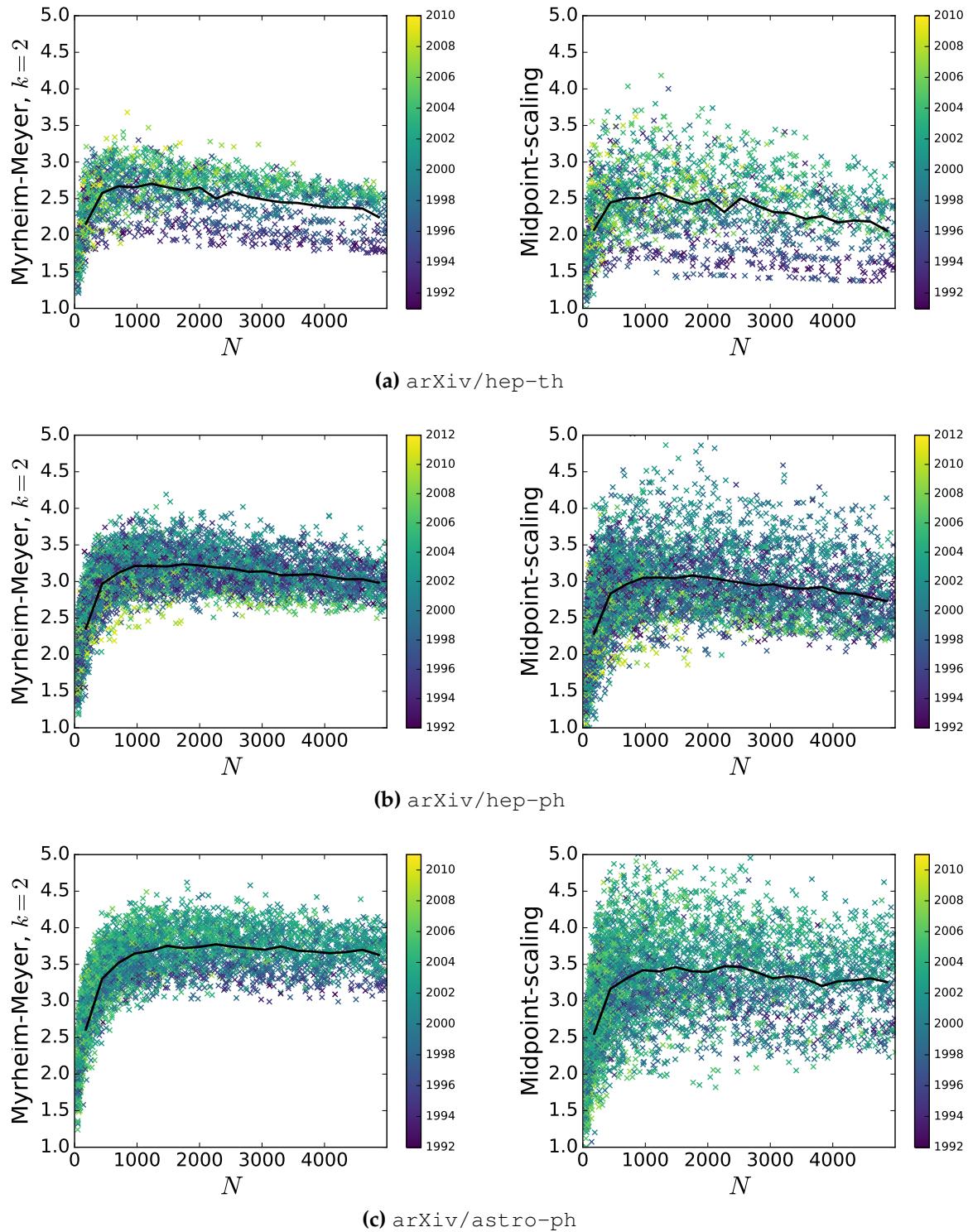
**Figure 5.7:** Myrheim-Meyer and midpoint-scaling dimension estimator on causal set graphs of dimensions  $D = 2, 3, 4, 5$ . In each case 20 graphs are generated and error bars show the standard error of the calculated estimates. As  $N \rightarrow \infty$  these dimension estimates both converge to the correct dimension of the original causal set graph. However it can be seen from this plot that for high dimensions, and particularly for the midpoint-scaling estimator, that even for  $N = 2000$  the estimate can miss the dimension of the original causal set graph.

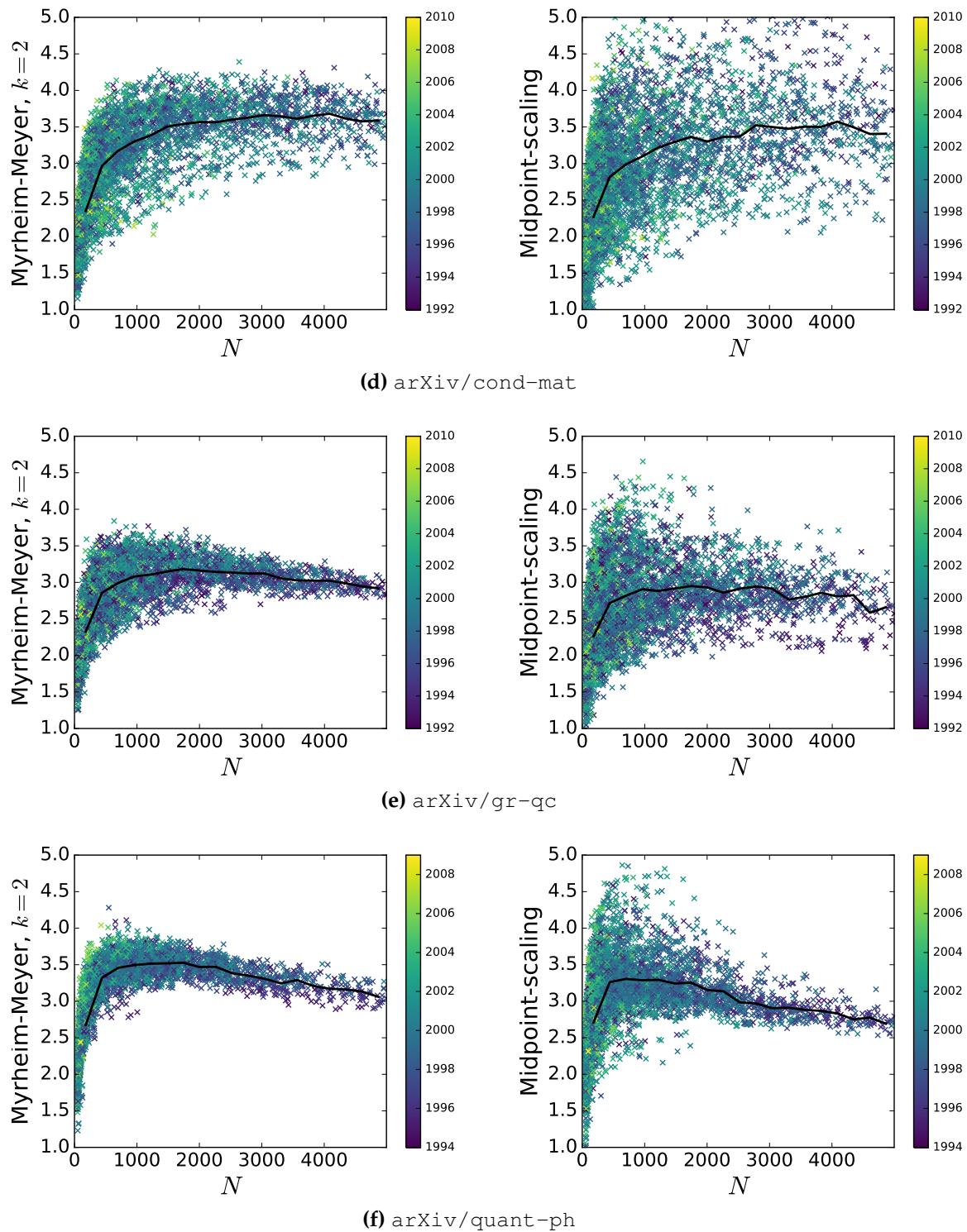
## 5.5 Results

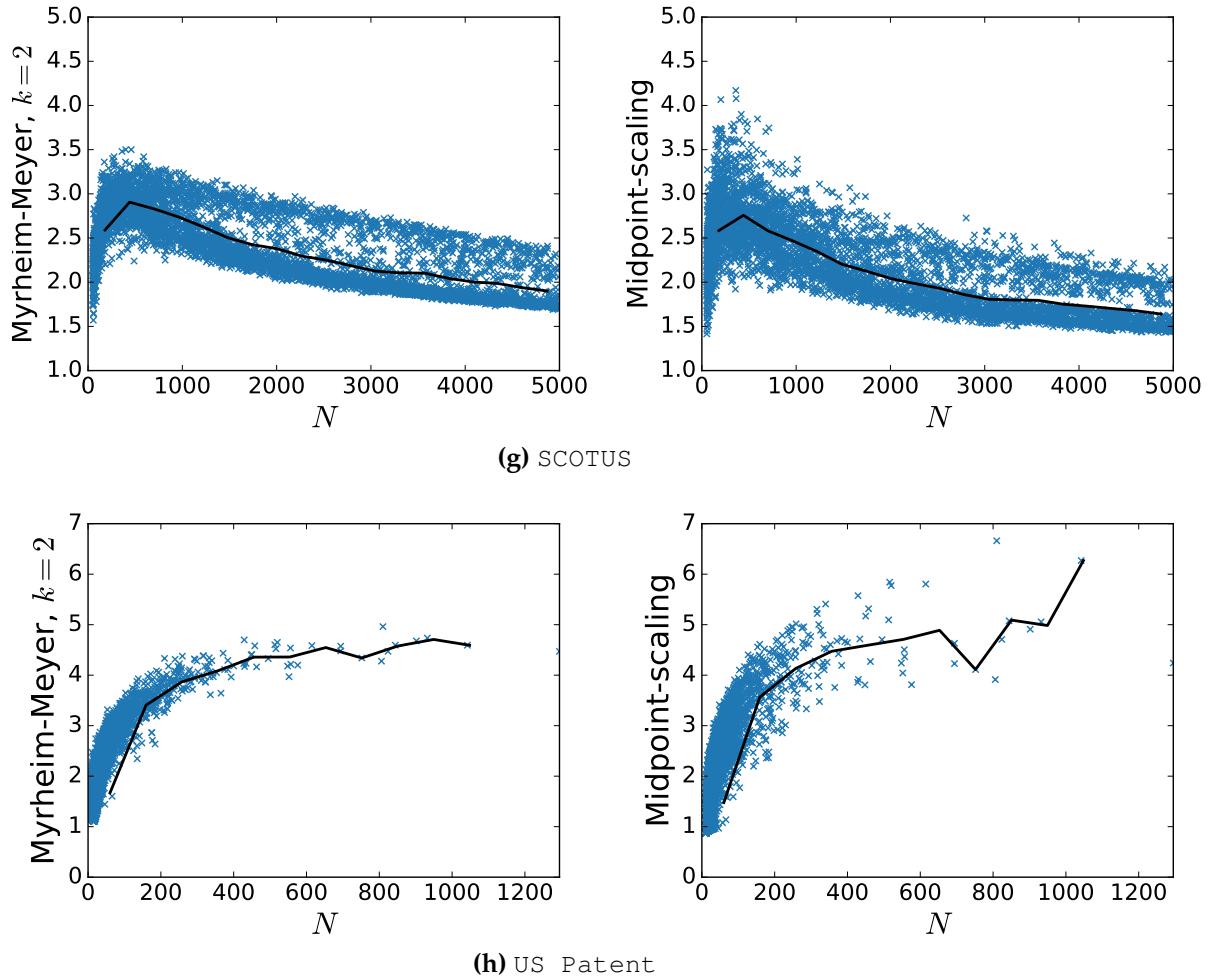
We can now use these dimension estimators on real networks. Figure 5.8 shows numerical results for some citation networks with the Myrheim-Meyer estimate on the left and the midpoint-scaling estimate on the right. Table 5.2 summarises the approximate estimated dimensions for large intervals. For each network 5000 intervals with populations between 20 and 2000 are randomly sampled and as in Figure 5.7 the plots show the dimension estimate against the size of the interval measured. Where the data was available the colour of a point represents the publication date of the first document in this interval. Note that in all cases the total populations of sampled intervals is much larger than the size of the network so the same nodes will contribute to many points in these plots, and furthermore many intervals may have very similar populations due to the random nature of the interval selection process. However, since sampling *all* of the intervals in graphs of this size would be extremely computationally costly, random sampling is the best way to get a representative view of the network's structure. The thick lines in the plots are the binned average values of the dimension measure.

	Myrheim-Meyer	Midpoint-scaling
hep-th	$\approx 2.5$	$\approx 2.5$
hep-ph	$\approx 3.0$	$\approx 3.0$
astro-ph	$\approx 3.5$	$\approx 3.5$
cond-mat	$\approx 3.5$	$\approx 3.25$
gr-qc	$\approx 3.0$	$\approx 3.0$
quant-ph	$\approx 3.5$	$\approx 3.25$
SCOTUS	$2.5 \rightarrow 2.0$	$2.5 \rightarrow 2.0$
US Patent	$> 4.0$	$> 4.0$
freeciv	1.88	1.64

**Table 5.2:** Table of approximate estimated dimensions for real networks. For small networks, the exact value is given. For larger networks an approximation from the plots above is given.







**Figure 5.8:** Dimension estimates for real networks. Left, Myrheim-Meyer estimate. Right, midpoint-scaling estimate. The dark line shows a binned average. As was the case for causal set graphs sprinkled in Minkowski spacetime, for most networks here the estimated dimension is low for small intervals and grows to reach a plateau for larger intervals. In the case of the SCOTUS citation network, the initial rise is followed not by plateau but by a drop in estimated dimension, suggesting differing dimensions on different length scales. The US Patent citation network contains very few large intervals, as indicated by the sparsity of points on the right of the plot, and so did not have a meaningful large-scale dimension in the way the other networks did.

The first obvious feature of these plots is that the left and right plots show a similar trend for each network, but with the midpoint-scaling dimension on the right having a larger spread of values. This is also the case in Figure 5.7, and so it is reasonable to say that this is simply a result of this estimator being more sensitive to noise. This sensitivity can be explained by noting that the Myrheim-Meyer estimator is counting chains and the midpoint-scaling estimator is counting nodes. There are many more 2-chains than there are nodes in these networks and so the smaller numbers present in the midpoint-scaling estimator means that statistical noise is comparatively more significant.

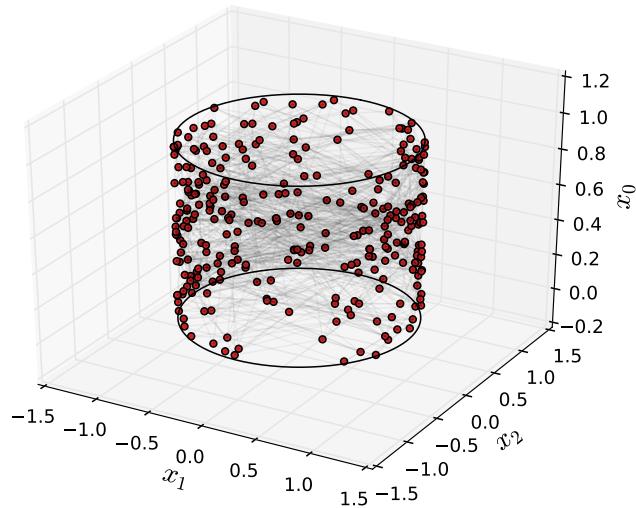
A feature present in all of these plots (except for the court judgements network) is the same rising curve reaching a plateau, as was the case in Figure 5.7. Unlike the causal set graphs though, the real citation networks show a much greater variance of dimension values. This is unsurprising as there is no reason to believe that these citation networks should have the totally homogeneous structure of the causal set graphs and so different intervals are likely to have different measured dimensions.

However despite this there are visible differences between the citation networks of different families of academic papers. For example, compare the plots on the left side of Figure 5.8a and Figure 5.8f. In the first case, almost every value for large intervals is below 3, and in the second case above 3. This means that given an interval from these citation networks we would be able to tell them apart by measuring their Myrheim-Meyer dimension even though they are very similar with regards to other network measures and come from different subsections of the wider arXiv citation network.

## 5.6 Scale Dependent Dimension

In the case of the causal set graphs it appeared that the estimated dimension always approached the original value used to create the causal sets. But for other networks the estimated dimension does not necessarily have a stable value. It can also decrease with increasing interval size. This effect can be replicated in the causal set graphs if their dimension is scale dependent. As is commonly discussed in the quantum gravity

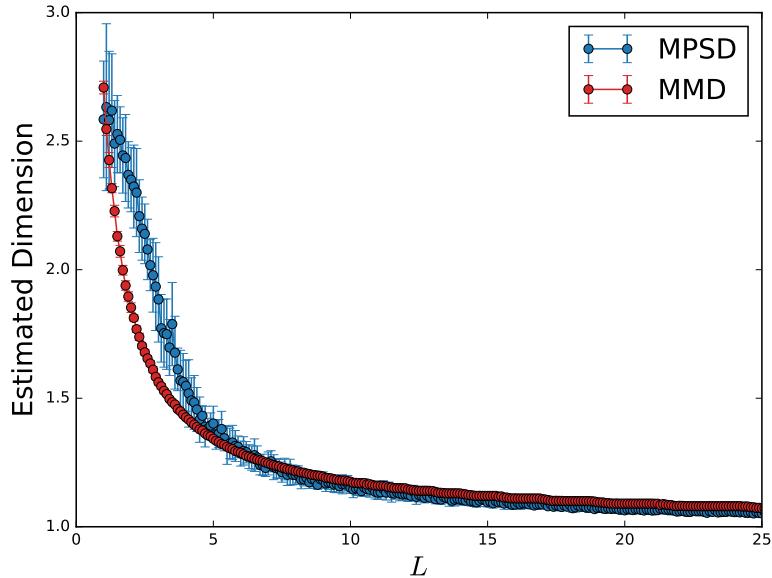
context in which causal set theory originates, some of the dimensions of the embedding spacetime can be ‘wrapped up’ at small length scales so that they are not observable at larger length scales. By creating causal sets with periodic boundary conditions in their spatial dimensions we can replicate this effect.



**Figure 5.9:** A topologically cylindrical  $D = 1 + 1$  causal set graph with 300 vertices. The surface that the elements are sprinkled on is flat Minkowski spacetime but is topologically cylindrical because it has periodic boundary conditions. At length scales much smaller than the circumference of the cylinder its surface appears two-dimensional. At length scales much larger than the circumference, the surface appears one-dimensional since almost all vertices in the graph will be related to each other. Note that while this example is drawn in three dimensions the vertices exist only on the cylinder’s surface and distances between vertices are calculated along the surface.

In Figure 5.9 an example is shown which has  $D = 2$  at small length scales and  $D = 1$  at larger length scales. In Figure 5.10 the Myrheim-Meyer and midpoint-scaling dimension estimates are shown for a causal set graph with two such wrapped-up dimensions at different length scales (the spatial dimensions are topologically toroidal), such that at small scales it appears to have  $D = 3$  but at larger length scales  $D = 1$ . The estimated dimension slides between these values, getting smaller as the ‘small dimensions’ can no longer be seen.

In terms of growing complex networks, these ‘small dimensions’ define a timescale over which the network’s growth process has a memory of recent events. At short



**Figure 5.10:** Estimated dimension for a toroidal causal set graph. On the x-axis  $L$  measures the size of the time dimension in terms of the radius of the toroidal spatial dimensions in a  $D = 2 + 1$  Minkowski spacetime. When this ratio is small the graph appears to have dimension  $D = 2 + 1$  but as the temporal length grows the graph appears to become one dimensional as the two spatial dimensions are ‘too small’ to be seen.

timescales, this wrapped up dimension affects the presence of edges but at larger timescales it does not.

## 5.7 Discussion and Interpretation

It isn’t immediately obvious how to interpret these results. For a causal set graph, we know there was some original spacetime in which the nodes were scattered, and we are estimating the dimensionality of this original hidden spacetime. But how should we interpret these results for a complex network where there was no original spacetime model? Consider the extremal cases; first, a network with a measured dimension close to 1. It must have almost all of its nodes related to almost all others. In a citation network this would mean that it would be possible to find any paper by looking back through the bibliography of a newly published paper, and through their bibliographies, and so on and so on, searching back through ‘citation ancestors’. This would be a very *narrow* field of study with just this one track of research being built upon,

and not a citation network where it was likely to find papers written on unrelated subjects. Consider then a network with a higher measured dimension. This means newly published papers are citing papers which do not cite each other or share ancestors, and many papers are causally unrelated to many others. This suggests the converse, a *diverse* academic field with many independent lines of study, and that highly cited papers are cited by researchers in unrelated fields. In this sense, we are quantifying the level of diversity in the citation behaviour of the papers in the citation network. Notably, we are doing so without regard for the content, or authors of the papers, and are using only the citation behaviour. As discussed in chapter 4 the causal structure of the network, which is the only relevant factor in both dimension estimates used here, ignores the edges which are likely to correspond to copied citations and so is more likely to correspond to citations reflecting authors reading papers.

These dimension measures quantify how narrow or diverse the behaviour growing the network is, but we can go further and think of geometric interpretations of the dimension this method is describing. When considering the ‘latent geometry’ of a network in the Riemannian case, as in [108, 183, 122], the abstract space we are considering is a similarity space, where nearby points are more similar than far away points. In the Lorentzian case, there are space and time coordinates. The model we are fitting these networks to is the causal set graph, so in this case similarity in the spatial coordinates makes edges more likely, but similarity in time coordinates make edges less likely. Recalling that in causal set theory, related elements correspond to timelike separation and the possibility of information travelling from one event to another, we can take a physical interpretation and think of information being able to travel between nodes in the network if they are comparable but not being able to reach those nodes which are incomparable. So in the case of citation networks, the information present in a newly written paper takes time to spread into different fields of study. It first reaches nearby fields, making citations from those areas more likely, and later on reaches less similar topics. The measured dimension of the network would then correspond to how quickly this similarity space of papers, court judgements, or patents grows, in terms of

the time required for information to percolate from one area to another. A low dimensional network grows slowly, whereas a high dimensional network grows quickly.

## 5.8 Summary

In this chapter I aimed to characterise the causal structure of real networks using causal set dimension estimators. This began with a review of the more well known poset, or order dimension and its relation to the cube space model, which was found to be inappropriate for this task. I then introduced two approaches to dimension estimation. Firstly, the midpoint-scaling estimator which is based on measuring how the height of an interval scales with its volume. Secondly, the Myrheim-Meyer estimator which counts the number of chains in an interval and compares it to expected values. When tested on causal set graphs both estimators recovered the correct dimension of the causets for large intervals, as is expected. When used on real data, we saw plots with similar characteristic shapes as the causal set graphs (namely small dimension for small intervals, growing to reach a plateau for large intervals). We saw a larger variance in estimated dimension for most real datasets than for the causal set graphs, explained by the greater heterogeneity of real network data than uniform null models. The important results from this chapter are that, firstly, the two dimension estimators show agreement for most datasets, which is not necessarily guaranteed. This suggests a meaningful concept of ‘embedding dimension’ for these real networks.

Secondly, we saw that otherwise similar networks from similar social systems could have significantly different estimated dimensions, suggesting that these dimension measures can see differences in underlying network structure which is otherwise hard to detect. It is therefore reasonable to believe that such measures can be a useful addition to the toolbox of network science when the network is a DAG.

Thirdly, I gave a qualitative interpretation of these methods in which they suggest how diverse or narrow the network’s growth process is, a conjecture which is open to further analysis using domain specific metadata.

Finally, an estimated embedding dimension is necessary for embedding algorithms,

as described in chapter 7 and so it is useful to provide, and demonstrate the effectiveness of such methods.

# Chapter 6

## Measuring curvature

In mathematics the art of proposing a question must be held of higher value than solving it.

---

Georg Cantor

Minkowski spacetime is the simplest spacetime to which we can fit a network's causal structure because it is uniform and flat, with only the dimension as a parameter. Fitting directed acyclic graphs to Minkowski spacetime is analogous to fitting undirected networks to random geometric graphs in Euclidean space, in that there is zero curvature in either case. Recall though that in chapter 2, when discussing network geometry, that many relevant features of real complex networks emerge naturally when fitting a network to a *curved* space, with the hyperbolic plane,  $\mathbb{H}^2$  the most popular candidate. Given the success of network geometry with uniformly curved spaces, it is immediately tempting to try to fit DAGs to uniformly curved spacetimes. These are *de Sitter* spacetime, and *anti-de Sitter* spacetime (positively and negatively curved respectively). Notably, there is a correspondence between hyperbolic space and de Sitter spacetime as described in [123] making this spacetime a particularly interesting candidate for this kind of geometric model. I will begin by discussing these spacetime geometries (focussing mainly on de Sitter spacetime both for simplicity, and because numerical results later in this chapter suggest that it is the better fit to real net-

works) and then describe a process for creating causal set graphs in this spacetime by sprinkling. Then we move on to the procedure for estimating the parameters defining the spacetime, its dimension and curvature. As before with estimators in Minkowski spacetime, this method is tested on de Sitter causal sets, and then used to estimate the curvature of real networks.

## 6.1 De Sitter Spacetime

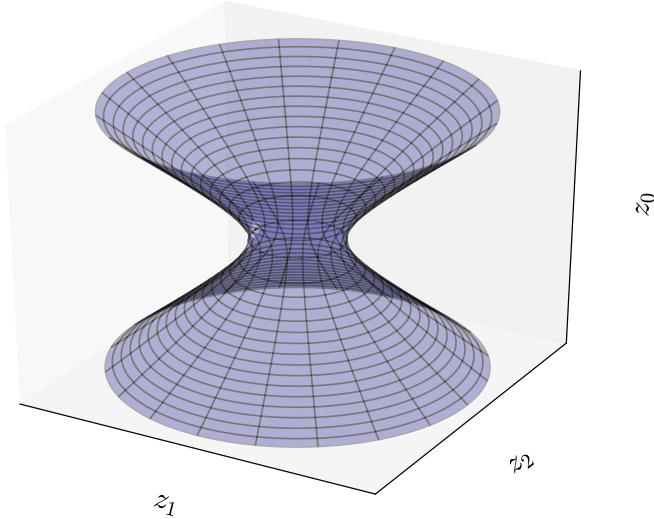
De Sitter spacetime can be constructed in a number of ways; we will begin by describing it as a submanifold, defined by a solution to an equation, in Minkowski spacetime of one higher dimension, in a similar way that a  $d$  dimensional sphere can be described as a solution to an equation in  $d + 1$  dimensional Euclidean space. We will begin with a  $d + 2$  dimensional Minkowski spacetime,  $\mathbb{M}^{d+2}$ , with coordinates  $z_0, z_1, \dots, z_{d+1}$ . Our  $d + 1$  dimensional de Sitter spacetime,  $d\mathbb{S}^{d+1}$  is then defined as the set of points lying on the one sheeted hyperboloid  $-z_0^2 + z_1^2 + \dots + z_{d+1}^2 = a^2$  where  $a$  is called the pseudoradius. The metric in these coordinates is simply the metric of the embedding Minkowski spacetime. The topology of  $d\mathbb{S}^{d+1}$  is  $\mathbb{R} \times \mathbb{S}^d$  representing one time dimension and  $d$  spatial dimensions as the surface of a  $d$  dimensional sphere.

While this coordinate system is easy to visualise (see Figure 6.1), it is less helpful for easily constructing causal set graphs in because it is hard to calculate whether two given points are spacelike or timelike separated. Instead, *conformal coordinates* make this calculation easier because in conformal coordinate systems, lightcones are straight lines like in Minkowski spacetime. Using the same conformal coordinate system as Meyer in [140] we have  $d + 1$  coordinates,  $t, x_1, x_2, \dots, x_d$  where the metric is then given by:

$$ds^2 = \frac{1}{\sigma^2}(-dt^2 + dx_1^2 + dx_2^2 + \dots + dx_d^2) \quad (6.1)$$

where

$$\sigma = 1 + \frac{K}{4}(-t^2 + x_1^2 + x_2^2 + \dots + x_d^2) \quad (6.2)$$



**Figure 6.1:** A representation of a section of  $1 + 1$  dimensional de Sitter spacetime embedded in  $2 + 1$  dimensional Minkowski spacetime. Time,  $t$  is on the vertical axis, with the one other spatial dimension  $\theta_1$  perpendicular to that.

and  $K = 1/a^2$  is called the *Gaussian curvature* of the spacetime. Since the curvature has units of inverse square distance, intervals are characterised by the dimensionless value of  $KT^2$  where  $T$  is the height of the interval (the proper time between its minimal and maximal points).

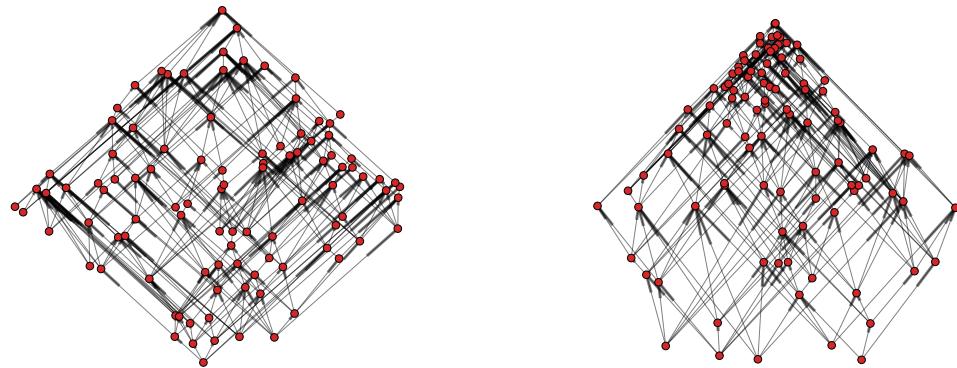
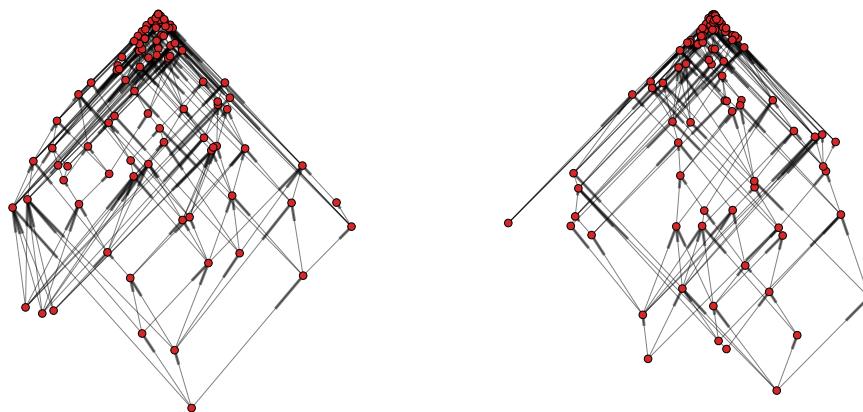
## 6.2 De Sitter Causal Sets

The benefit of using this conformal coordinate system is that it is easy to calculate the causal structure of points sprinkled within it, but the drawback is that since spacetime volumes are distorted, and sprinkling must be at a uniform density in spacetime, we must distort the sprinkling in these coordinates to make sure that the probability of finding one of the scattered points in some region is proportional to that region's true spacetime volume. Calculating causal structure is easy because the  $(-dt^2 + dx_1^2 + dx_2^2 + \dots + dx_d^2)$  term in the metric, as in Equation 6.1 means that lightcones are straight lines just as in flat spacetime.

Uniform point density is achieved by scattering points proportional to a *conformal factor* which in this case is given by  $\sigma^{-(d+1)}$ . As in [140] this can be done most simply using the *rejection method*. In this method, each scattered point is assigned a number  $w$  uniformly randomly sampled from the range  $[0, W]$ , where  $W$  is the maximum value taken by the conformal factor  $\sigma^{-(d+1)}$  in the interval we are scattering into. Then, if  $w < \sigma^{-(d+1)}$ , where  $\sigma$  is calculated using the coordinates of this point, then the point is accepted, and otherwise it is rejected. At the end of this process each point will have survived with a probability equal to the conformal factor and so those remaining will be scattered proportional to it, as is required.

As one would expect, in the case where the curvature  $KT^2 \approx 0$ , the conformal factor is just equal to 1 and all points are accepted by the rejection method meaning that they are sprinkled uniformly in the interval just as in Minkowski spacetime. We therefore recover the fact that a causal set in de Sitter spacetime with sufficiently low curvature,  $K \approx 0$ , or one where the size of an interval is very small,  $T^2 \approx 0$ , should be indistinguishable from one in Minkowski spacetime of the same dimension.

Figure 6.2 shows examples of causal set graphs made by uniformly sprinkling into  $D = 1 + 1$  de Sitter spacetime with increasing curvatures, from top left to bottom right. Note that initially, this picture is indistinguishable from similar examples in Minkowski spacetime (eg. Figure 3.2). However as the curvature increases a visible asymmetry occurs, with most of the points falling in one end of the interval. This gives the interval's causal structure measurably different statistics.

(a) Left:  $KT^2 = 0$       Right:  $KT^2 = 3.5$ (b) Left:  $KT^2 = 3.9$       Right:  $KT^2 = 3.95$ 

**Figure 6.2:** Causal sets in intervals of  $D = 1 + 1$  de Sitter spacetime, with  $N = 100$  points. Curvature is increasing from top left, to bottom right. Noticeably there is an asymmetry in the structure of the resulting graphs with more vertices clustered near the top of the interval. This results in a measurably different causal structure from which the curvature and dimension of the embedding de Sitter spacetime can be estimated.

The statistic we will use to quantify the dimension and curvature of these causal sets, and, later in this chapter, real networks, will be the number of chains of  $k$ -chains, and in that sense this approach is similar to the Myrheim-Meyer dimension described in chapter 5. The expected number of  $k$ -chains in an interval in (anti) de Sitter space-time is calculated in [140] and is given by:

$$\langle C_k \rangle = \left( \frac{V_{d-1}}{2^d} \right)^k \sum_{i_1, \dots, i_k=0}^{\infty} \left( \frac{K}{4} \right)^{i_1 + \dots + i_k} T^{k(d+1)+2(i_1+\dots+i_k)} G_d(i_1, \dots, i_k) \quad (6.3a)$$

$$G_d(i_1, \dots, i_k) = \prod_{j=1}^k \frac{1}{j(d+1) + 2(i_1 + \dots + i_j)} \frac{\Gamma(d + i_j + 1)}{\Gamma(i_j + 1)} \frac{\Gamma\left(\frac{(j-1)(d+1)}{2} + i_1 + \dots + i_j + 1\right)}{\Gamma\left(\frac{(j+1)(d+1)}{2} + i_1 + \dots + i_j\right)} \quad (6.3b)$$

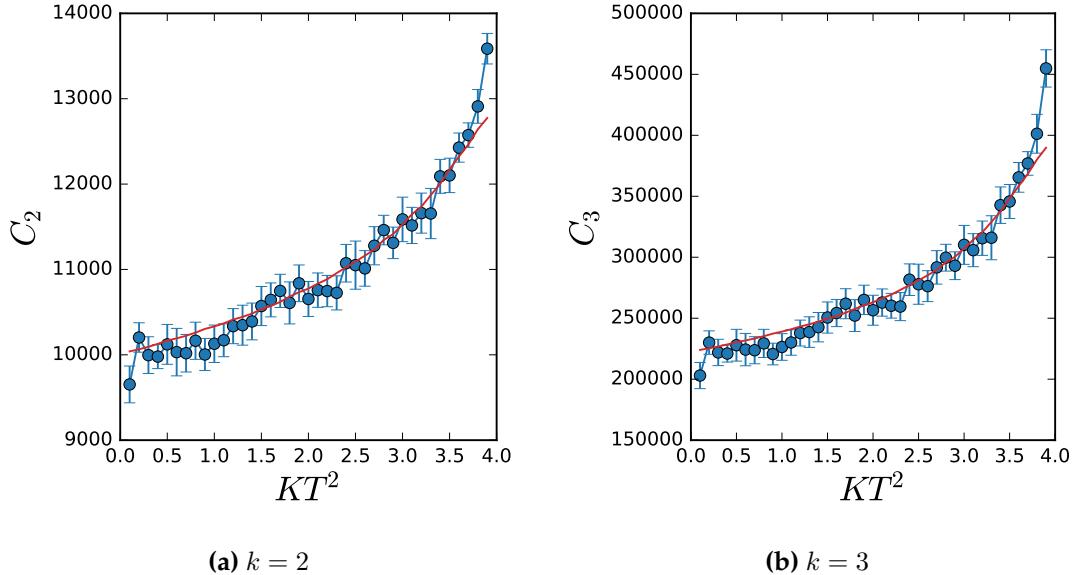
where  $d + 1$  is the dimension,  $K$  the Gaussian curvature, and  $T$  the height of the interval.  $V_{d-1}$  denotes the volume of a unit  $(d - 1)$ -sphere. These equations assume that the sprinkling density  $\rho = 1$  and so the expected number of points in an interval corresponds to its volume. There are three parameters to estimate;  $d$ ,  $K$ , and  $T$  so we need to make three measurements, of  $C_1 = N$ ,  $C_2$ , and  $C_3$ . For  $k = 1, 2, 3$  we therefore have to solve:

$$\langle C_1 \rangle = \left( \frac{V_{d-1}}{2^d} \right) \sum_{i_1=0}^{\infty} \left( \frac{K}{4} \right)^{i_1} T^{(d+1)+2i_1} \frac{1}{d+1+2i_1}, \quad (6.4a)$$

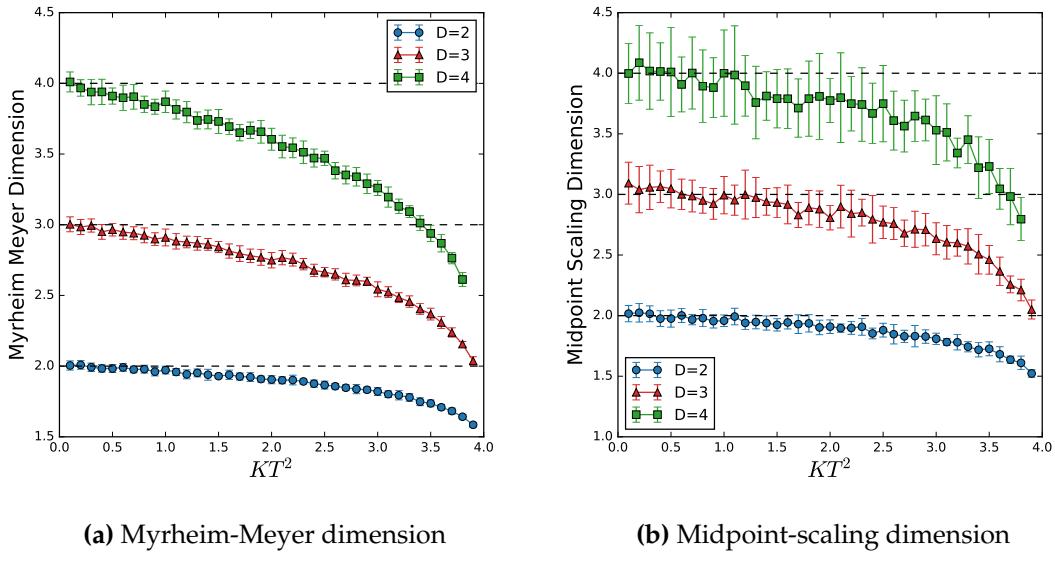
$$\begin{aligned} \langle C_2 \rangle &= \left( \frac{V_{d-1}}{2^d} \right)^2 \sum_{i_1, i_2=0}^{\infty} \left( \frac{K}{4} \right)^{i_1+i_2} T^{(d+1)+2(i_1+i_2)} \frac{1}{d+1+2i_1} \\ &\times \frac{1}{2(d+1)+2(i_1+i_2)} \frac{\Gamma(d+i_2+1)}{\Gamma(i_2+1)} \frac{\Gamma\left(\frac{d+1}{2}+i_1+i_2+1\right)}{\Gamma\left(\frac{3(d+1)}{2}+i_1+i_2\right)}, \end{aligned} \quad (6.4b)$$

$$\begin{aligned} \langle C_3 \rangle &= \left( \frac{V_{d-1}}{2^d} \right)^3 \sum_{i_1, i_2, i_3=0}^{\infty} \left( \frac{K}{4} \right)^{i_1+i_2+i_3} T^{(d+1)+2(i_1+i_2+i_3)} \frac{1}{d+1+2i_1} \\ &\times \frac{1}{2(d+1)+2(i_1+i_2)} \frac{\Gamma(d+i_2+1)}{\Gamma(i_2+1)} \frac{\Gamma\left(\frac{d+1}{2}+i_1+i_2+1\right)}{\Gamma\left(\frac{3(d+1)}{2}+i_1+i_2\right)} \\ &\times \frac{1}{3(d+1)+2(i_1+i_2+i_3)} \frac{\Gamma(d+i_3+1)}{\Gamma(i_3+1)} \frac{\Gamma(d+2+i_1+i_2+i_3)}{\Gamma(2(d+1)+i_1+i_2+i_3)}. \end{aligned} \quad (6.4c)$$

To check the consistency of our method of producing causal set graphs in de Sitter spacetime and this calculation of the expected number of chains, we count the chains in the de Sitter causal set graphs and compare these numerical results to the expected analytic values in Equation 6.4 in Figure 6.3. In Figure 6.4 we see how the dimension measures used in chapter 5 recover the correct dimension of the spacetime for low curvature, but fail as curvature increases, systematically underestimating the true dimension. These plots make it clear why multiple measurements are necessary to estimate the dimension and curvature of a causal set graph in de Sitter spacetime, as otherwise there would be no way to distinguish between two different de Sitter causal set graphs with the same measured Myrheim-Meyer or midpoint-scaling dimensions.



**Figure 6.3:** Expected number of chains in  $D = 1 + 1$  de Sitter spacetime causal sets, left,  $k = 2$ , and right,  $k = 3$ , with  $N = 200$  points. Error bars on the calculated values indicated two standard errors. The theoretical prediction (red) is a good fit for the calculated values (blue), except at very high curvatures where the two appear to diverge. This can be accounted for by noting that the analytical results is computed by summing terms in Equation 6.4, and here the first 20 terms in the sum are used. However, for the highest curvatures the terms after this point may no longer be entirely negligible.



**Figure 6.4:** Dimension measures from chapter 5 on de Sitter causal sets with  $N = 1000$  points in dimension  $D = 2, 3, 4$ , against their curvature,  $KT^2$ . In the low curvature regime, where these causal set graphs are approximately the same in structure to Minkowski space causal set graphs the dimension measures recover the correct embedding dimension as expected. As the curvature increases however the estimates diverge from the true values.

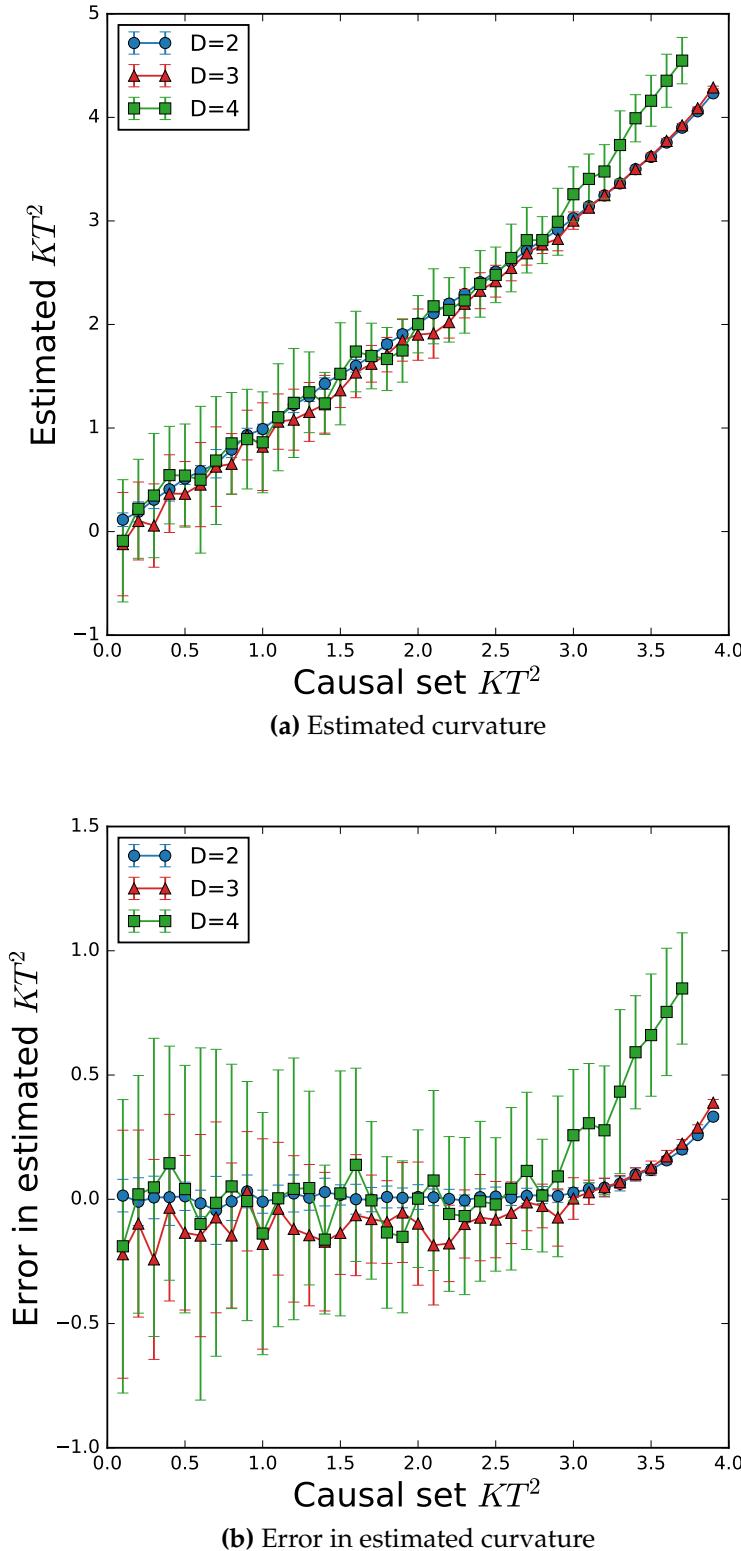
### 6.3 Inferring Curvature

To find general solutions to these highly non-linear equations analytically is not feasible. Therefore they must be solved by numerical methods. So long as the curvature is not extremely high (meaning in this coordinate system that  $KT^2$  is not too close to 4), the initial terms in the sums in Equation 6.3b are good approximations of the totals (the first 10 terms are used here). Then, once the numbers of  $k$ -chains are counted the parameters  $d$ ,  $K$  and  $T$  can be fit by any standard numerical optimisation package. Here the numerical equation solver in `scipy`<sup>1</sup> is used. The initial values used to begin the numerical optimisation were a curvature of 0, and a dimension estimated using the Myrheim-Meyer method. This implementation is available, along with examples using de Sitter causal set graphs, in [59].

We begin by testing this estimate on causal set graphs in de Sitter spacetime checking whether the curvature can be recovered from the graph's causal structure. Figure 6.5 shows the results of this estimator applied to de Sitter causal sets of known parameters and show that the estimator is very effective up to the high curvature regime where it shows some deviation from the expected values.

---

<sup>1</sup>Specifically <https://docs.scipy.org/doc/>, and the `optimize` module and its `minimize` function.

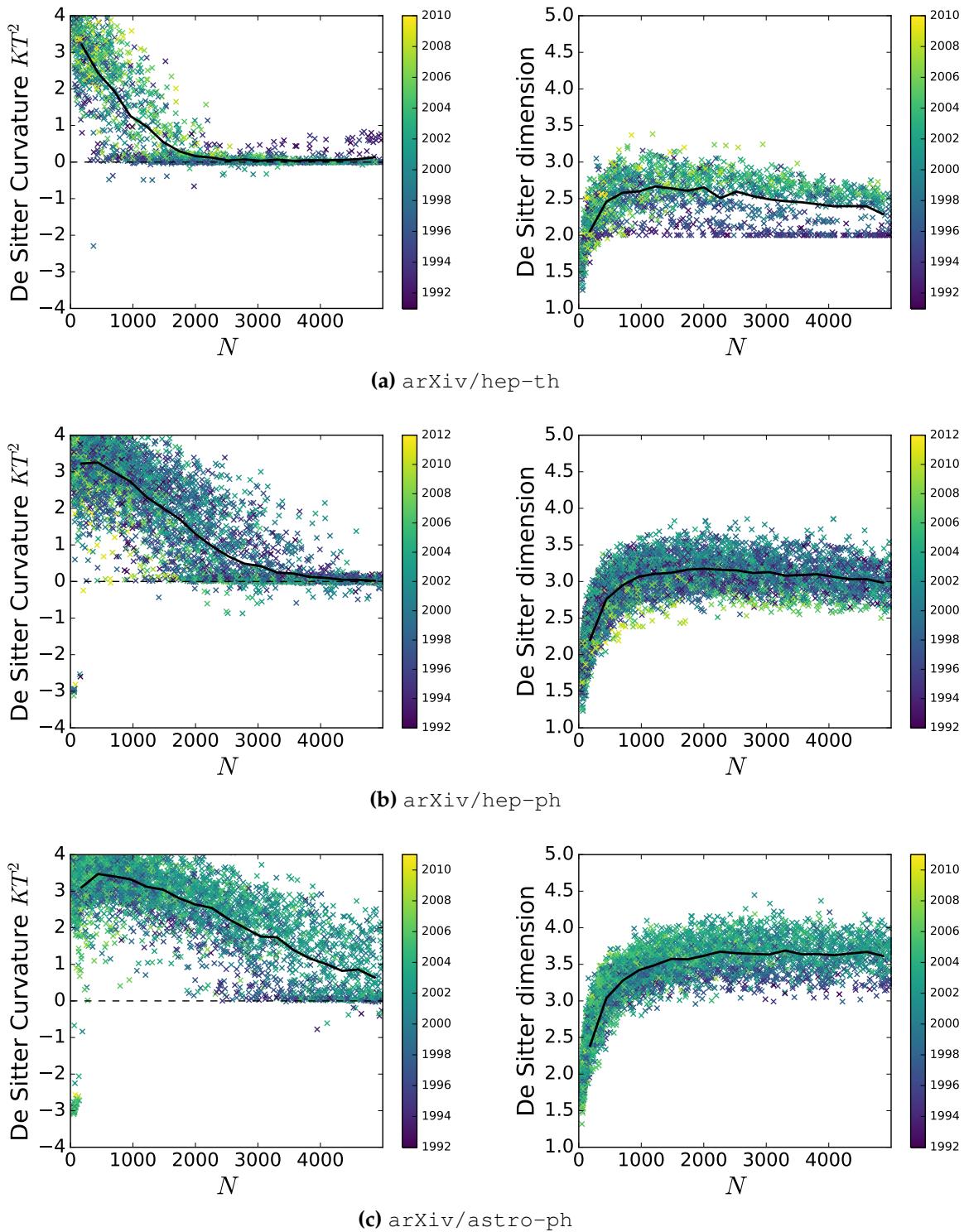


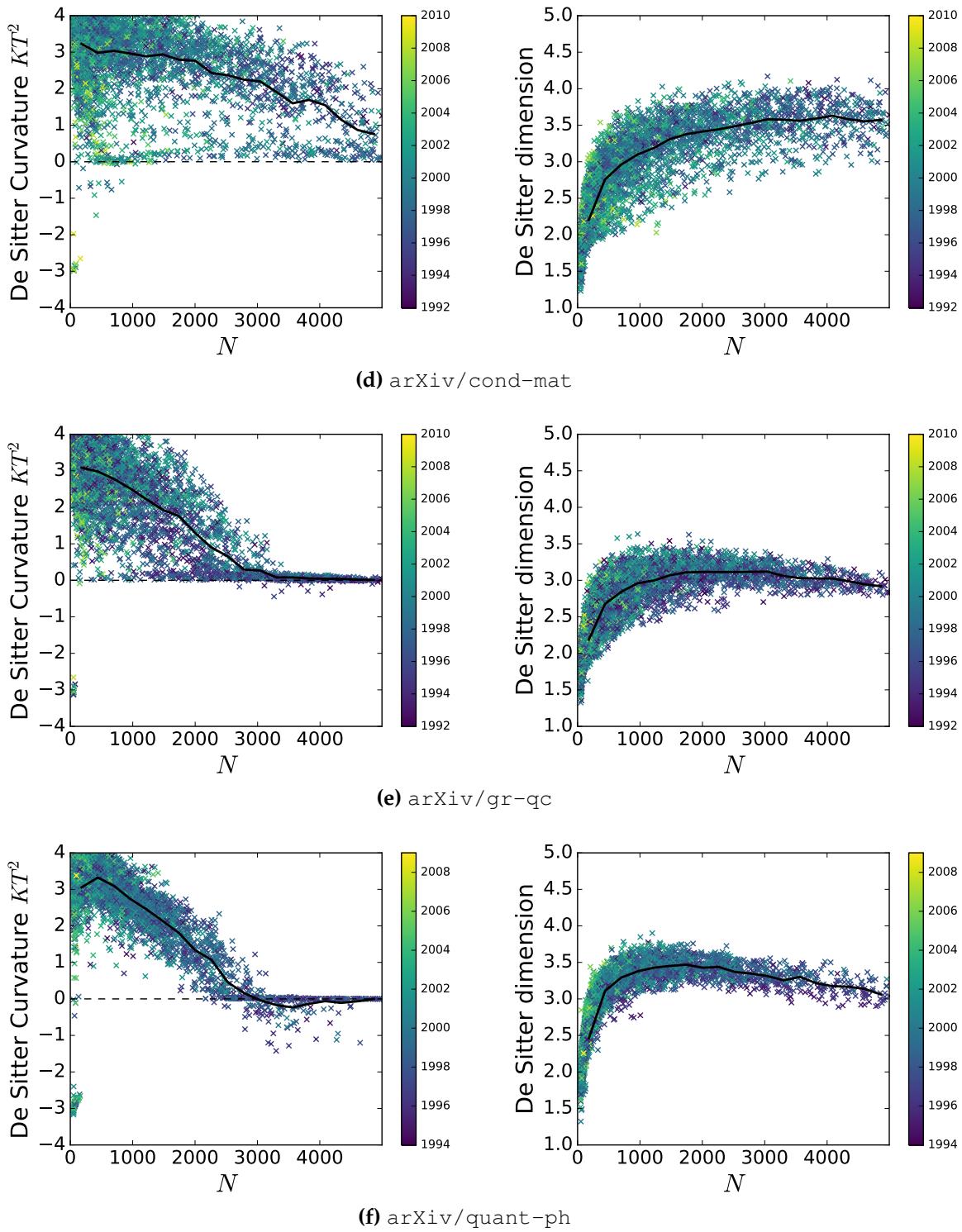
**Figure 6.5:** Estimated curvature for de Sitter causal set graphs,  $N = 1000$ , and the deviation from the values used to create these graphs. For each curvature and dimension value 20 graphs are generated and measured, with error bars showing the standard deviation of measured values. For low curvature the mean values of the errors are around 0 indicating the estimators are working well, although there is some divergence from 0 at high curvatures. As noted above this is likely due to higher order terms in Equation 6.4 which are not accounted for in the numerical approach used here. The statistical deviations shown by the error bars can be accounted for by noting that the causal set graphs are generated stochastically and so will naturally deviate slightly from the parameters used to generate them even when the estimators work perfectly. This effect is more prominent in higher dimensions as was noted in [140].

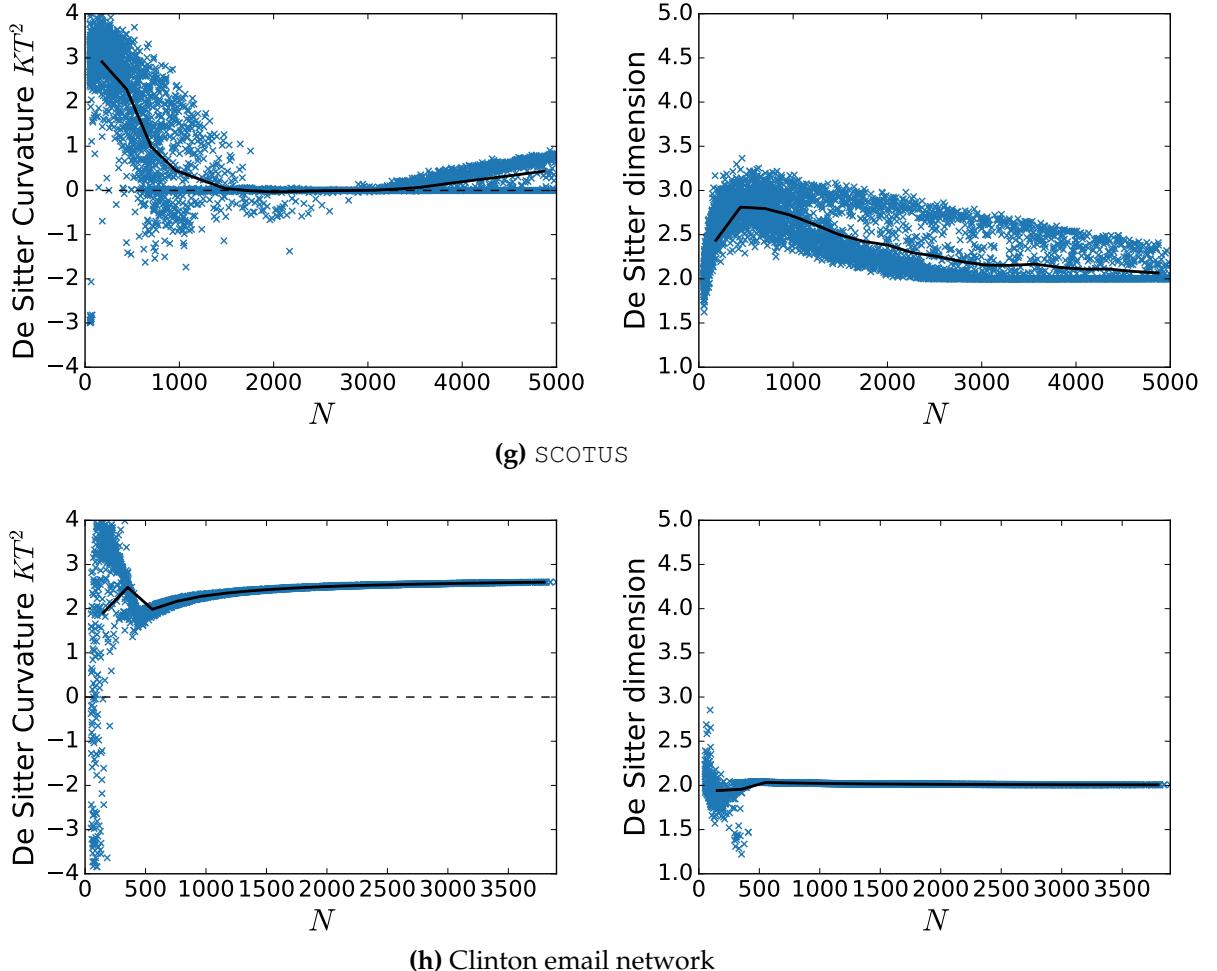
## 6.4 Results

In Figure 6.6 we can see the estimated curvature values for real networks (left) and the associated dimension of the fitted de Sitter spacetime (right). For small intervals (small  $N$  on the x-axis) we see a large spread of curvature values, which is explained by the fact that statistical fluctuations will be more significant in this regime, and due to the non-linearity in Equation 6.4 even small fluctuations in the number of chains counted can lead to large fluctuations in the eventual estimated curvature. For most of these networks (with the notable exception of the Clinton email network) the average curvature for large intervals is closer to 0 indicating approximately flat spacetime. Where there is non-zero curvature it is almost always positive corresponding to de Sitter spacetime as opposed to negatively curved anti de Sitter spacetime. Given the correspondence between de Sitter spacetime and hyperbolic space discussed previously, this is an encouraging result because it draws together the two complementary approaches of considering networks as embedded in a negatively curved Riemannian space, and as embedded in a positively curved Lorentzian spacetime.

For the arXiv citation networks, where the colour of scattered points indicated the publication date of the first paper in the interval, there is no real visible relation between this date and the estimated curvature suggesting that this property is not particularly dependent on time. However, the fact that the curvature seems to decrease as the interval size increases suggests that, while curved de Sitter spacetime describes small intervals well, large intervals appear flatter. The ‘best fitting’ spacetime may then not be the uniformly curved simplest case as assumed in this chapter and spacetime with a scale-dependent curvature may be more appropriate, although such investigations are outside the scope of the work done here. It is interesting though to note that the length scale over which the dimension measures plateau is different to (and shorter than) that over which the curvature measures do, suggesting two different length scales of interval population being relevant here.







**Figure 6.6:** Curvature estimates for real networks. On the left, the estimated curvature  $KT^2$ . On the right, the estimated dimension of the associated de Sitter spacetime. Note that almost all intervals have a positive estimated curvature, indicating a fit to de Sitter spacetime as opposed to anti de Sitter spacetime. The curvature is generally higher for small intervals but lower for large intervals, suggesting that locally, these networks do not fit well to flat Minkowski spacetime, but at large length scales they do. The estimated dimensions are generally similar to those seen in chapter 5 indicating that the measures used there are relatively robust to curvature on short length scales for these networks.

## 6.5 Summary

In this chapter we investigated a method for characterising a network's causal structure by associating it with the geometry of a curved spacetime. By making measurements of the number of chains in the network, we could fit its structure to that of a de Sitter or anti de Sitter spacetime with a given curvature scale, and dimension. Although results varied between networks, it appears that it is positively curved de Sitter spacetime which best fits small intervals in citation networks but flat space which best fits large intervals suggesting that the curvature is scale dependent. Nonetheless, the measured curvature was almost always positive indicating a de Sitter spacetime which has known correspondence to the hyperbolic space which has previously been seen to be a successful description of network structure in the network geometry literature. Therefore I think it is reasonable to believe that the network geometry approach, using curved spacetimes, could have the same kind of success in describing and explaining network structure for DAGs as has been the case for undirected networks in hyperbolic space.

# Chapter 7

## Embedding networks

A problem worthy of attack  
proves its worth by fighting back.

---

Piet Hein

*This chapter is based on work presented in:*

*Embedding Graphs in Lorentzian Spacetime,*

*J. R. Clough, T. S. Evans,*

*<http://arxiv.org/abs/1602.03103>, 2016.*

In previous chapters we have examined the problem of associating a DAG with a Lorentzian spacetime of some dimension and curvature. Macroscopic measurements allowed us to make statements like ‘this network has causal structure similar to  $\mathbb{M}^D$ ’ or ‘this network has positive spacetime curvature’. In this chapter we go one step further by actually associating coordinates in the continuous spacetime to each node in the network such that the network’s edges emerge naturally from the causal structure of the embedding spacetime.

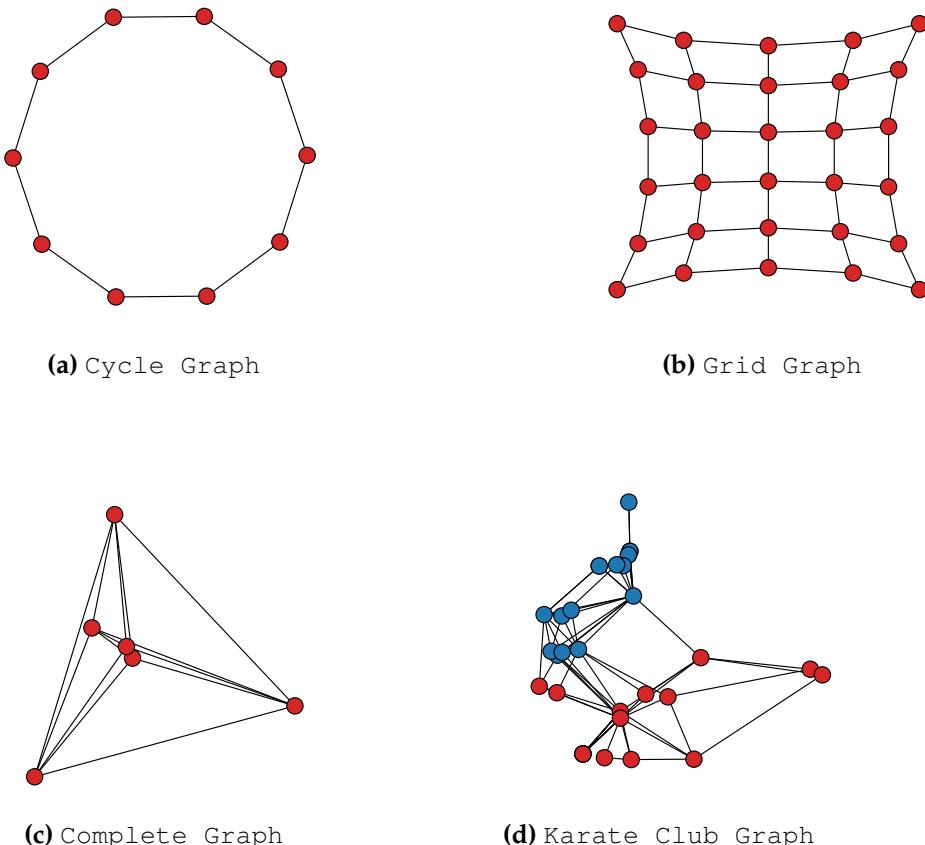
The algorithm to solve this embedding problem has two key steps. The first is to estimate the spacelike or timelike separation of each pair of nodes from the structure of the network using a null model of a causal set graph. This step primarily uses tools developed in causal set theory. The second step is to assign coordinates in a given

spacetime, in such a way as to respect, as much as is possible, the separations calculated in the first step. This second step involves using a generalised version of *multidimensional scaling* (MDS). MDS is usually only appropriate for assigning coordinates in a Riemannian space, but here it is generalised in the sense of allowing for embedding into manifolds with any metric signature [60], including Lorentzian spacetime.

## 7.1 Embedding in Euclidean Space

Let's first consider this problem in Euclidean space, and ask how we would fit a network to a random geometric graph model. We have a graph,  $\mathcal{G}$  and wish to assign  $d$  coordinates,  $\mathbf{x}_i$  to each vertex  $i$  such that the Euclidean distance between the coordinates of vertices is proportional to the distance between them in the graph as measured by shortest path length. The classical MDS algorithm (which is described in detail below) will return coordinates minimising the total square errors on these pairwise distances. Figure 7.1 shows some visualisations produced by this process.

These visualisations are optimal fits to random geometric graphs in the sense of minimising the sum of the square differences between the graph distances and distances in the embedding space. An analogous process for DAGs and Lorentzian spacetime would involve fitting the DAGs to causal set graphs, the equivalent of RGGs in spacetime. The process is more complex though. Whereas in the Euclidean case the distance in the RGG was easily estimated as the number of edges on the shortest path between two vertices, in the Lorentzian case the problem is more complicated as there are two classes of separation (spacelike and timelike) and estimating these from the graph structure is less trivial.



**Figure 7.1:** Some small graphs embedded in two-dimensional Euclidean space using MDS. Note that while some graphs such as (a) and (b) can be embedded in this way such that distance in the embedding reflects distance on the graph, this is not true of all graphs. The complete graph with six vertices cannot be embedded in this way as six mutually equidistant points, a regular six-simplex, requires five dimensions to be embedded. The karate club graph is similarly hard to embed, but it is interesting to note the approximation found by the MDS algorithm nicely recovers the two clubs in this well known social network [88], coloured here in red and blue.

## 7.2 Estimating spacelike separation

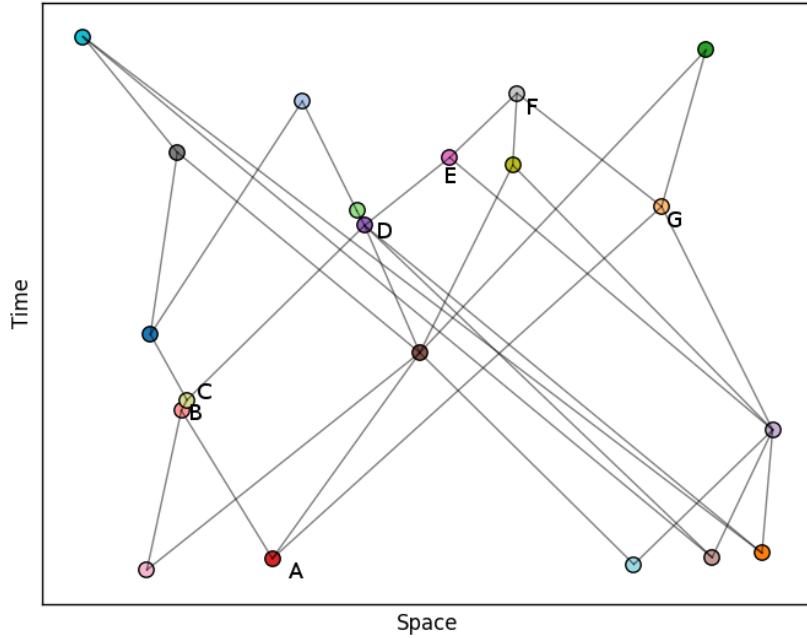
The algorithm begins by estimating the separations in the embedding spacetime between pairs of nodes using the null model of a causal set graph. For related nodes, we estimate a timelike separation, and for unrelated nodes a spacelike separation. If the network was embedded in Euclidean space like a random geometric graph the distance between two nodes could be approximated by the number of edges on the shortest path between them, as in [106]. What is an analogous step to take in for DAGs in Lorentzian geometry? We have already seen in chapter 5 that for timelike separated vertices, that separation is approximated by the length of the longest chain between them. We also need a method for estimating the spacelike separation for spacelike separated pairs, which turns out to be a more difficult problem. There is currently no method which is both as accurate and as easily calculated as the longest path is for timelike pairs [51]. Approximations are known, and we will use a very simple one, described in [49, 167] as *naive spatial distance*. This distance estimate works as follows. Suppose we have two unrelated vertices  $i$  and  $j$  in  $\mathcal{G}$  meaning they are spacelike separated. We then look for all pairs of nodes,  $k$  and  $l$ , where  $k$  is in the future of both  $i$  and  $j$  while  $l$  is in their pasts<sup>1</sup>. Of each of these pairs we then measure the length of the longest chain between them and choose as  $k^*$  and  $l^*$  the pair for which this length is minimised. The timelike separation, as measured by this longest chain, between  $k^*$  and  $l^*$  is then used as an estimate for the spacelike separation between  $i$  and  $j$ . Figure 7.2 gives a worked example.

This estimate is simple and at first appealing, but unfortunately it becomes increasingly inaccurate in more than two dimensions when  $N \rightarrow \infty$  (hence the ‘naive’ in the name). Nonetheless it is sufficient for our purposes<sup>2</sup>. This is partly because naive spacelike distance is inaccurate only for large graphs but also because in the

---

<sup>1</sup>If no such pair exists, we set the spacelike distance of  $i$  and  $j$  equal to some maximal distance which is a parameter of the algorithm. In the examples shown here, we used the length of the longest chain in the graph as this parameter.

<sup>2</sup>I have tried using the two-link method described in [167] but found that even for graphs with 2000 nodes too many cases had no two-links and so the spacelike separation couldn’t be calculated resulting in a worse embedding.



**Figure 7.2:** The timelike separation between nodes A and F is approximated as 5 units - as 5 is the number of edges in the longest direction-respecting path between them.

Nodes B and G are spacelike separated. To estimate this separation we find a pair of points in their mutual past and future. In this case, the only such pair is (A, F). The naive spatial separation between (B, G) is then given by the timelike separation between (A, F) so is also 5 units.

MDS algorithm each point's coordinates are fixed by many separations, both timelike and spacelike, which limits the effect of noise from a few poor estimations. Given a graph, these timelike and spacelike separation estimates define our separation matrix  $\mathbf{S}$  (where timelike separation has the  $-$  sign in our conventions), on which we can perform the Lorentzian MDS algorithm described below to embed it in  $\mathbb{M}$ .

### 7.3 Multidimensional Scaling

Multidimensional scaling (MDS) is a well known data analysis technique that, as its input, takes data in the form of similarity or distance between pairs of objects, and as its output returns coordinates for those objects which maximally respect those distances in the sense of minimising the sum of the squares of the errors. We begin with an introduction to classic MDS in Euclidean space, before introducing generalised Lorentzian

MDS.

Suppose we have  $N$  objects, which live in a  $d$  dimensional Euclidean space,  $\mathbb{E}^d$  and we are given the squared Euclidean distance,  $S_{ij}^{(2)}$  between each pair of objects  $i$  and  $j$ . We wish to find the coordinates of the objects, which will be  $d$  dimensional vectors,  $\mathbf{x}_i$  for each object  $i$ , such that they fit the constraint that  $|\mathbf{x}_i - \mathbf{x}_j|^2 = S_{ij}^{(2)}$ . This is the problem that the classical MDS algorithm is designed to solve. First, the  $N \times N$  matrix of square distances,  $\mathbf{S}^{(2)}$ , is defined with elements  $S_{ij}^{(2)}$ . It is then used to find the double centred matrix  $\mathbf{B} = -\frac{1}{2}\mathbf{JS}^{(2)}\mathbf{J}$  where  $\mathbf{J} = \mathbf{1} - \frac{1}{N}\mathbf{1}\mathbf{1}^T$ . It can then be shown<sup>3</sup> that

$$\mathbf{B} = \mathbf{XX}^T \quad (7.1)$$

where  $\mathbf{X}$  is an  $N \times d$  matrix of coordinate vectors  $\mathbf{x}$  which satisfy the constraint of recovering the original distances, and with the centre of mass of the coordinates at the origin.  $\mathbf{B}$  is guaranteed to be semi-positive definite (i.e. it has no negative eigenvalues). So we can then find (up to a factor of a rotation) the coordinates in  $\mathbf{X}$  by decomposing  $\mathbf{B}$  into

$$\mathbf{B} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T \quad (7.2)$$

where  $\boldsymbol{\Sigma}$  is a diagonal matrix of the eigenvalues of  $\mathbf{B}$ , and  $\mathbf{U}$  a matrix of its eigenvectors. A solution is given by

$$\mathbf{X} = \sqrt{\boldsymbol{\Sigma}}\mathbf{U}. \quad (7.3)$$

This process yields coordinates in  $N$  dimensions, but only  $d$  of the eigenvalues in  $\boldsymbol{\Sigma}$  will be non-zero. It is possible to retrieve coordinates in fewer dimensions, by using only the largest  $\hat{d}$  eigenvalues and their corresponding eigenvectors. The larger eigenvalues correspond to principle components, meaning that using them as the coordinates minimises the square difference between the original distances we started with, and those calculated from these inferred coordinates. These coordinates are in this sense the most accurate  $\hat{d}$  dimensional representation of the original data and it is in this manner that

---

<sup>3</sup>See [65] for the classical case, although the derivation given in this chapter also covers this as a special case where the matrix representing the metric is the identity matrix.

MDS can be used for dimensionality reduction. Note that the coordinates found by this MDS algorithm can differ by factors of rotation and translation while still meeting the constraint of recovering pairwise distances, and so it is a matter of choice whether to apply such a rotation or translation to the answer that the algorithm returns.

As well as this simple version of the algorithm, faster approximations also exist. Landmark MDS [68] is a two step process, where first a small number of ‘landmark’ points have their positions fixed to each other using the usual MDS method, and second, the remaining points are fixed using only their distances to the landmarks. Pivot MDS [46] provides further improvements by iteratively updating the positions of the landmarks, or pivots, using the rest of the points, and then updating the positions of the rest of the points using the pivots. Both of these approximations are easily extendible to our Lorentzian MDS method as implemented in [59].

## 7.4 Generalised MDS

Now we will consider the same problem in Minkowski spacetime  $\mathbb{M}^{d+1}$ . Consider a collection of points  $i \in \mathbb{M}^{d+1}$  with coordinates  $u_{i,\mu}$  consisting of a time coordinate,  $u_{i,0}$ , and spatial coordinates  $u_{i,k}$ , with  $k = 1, 2, \dots, d$ . The matrix  $\mathbf{M}^{(2)}$  has as its elements the separation in this spacetime between objects  $i$  and  $j$ :

$$\mathbf{M}_{ij}^{(2)} = \eta(u_i - u_j, u_i - u_j) = -(u_{i,0} - u_{j,0})^2 + \sum_{k=1}^d (u_{i,k} - u_{j,k})^2. \quad (7.4)$$

We can now ask the same question that classical Euclidean MDS poses: given pairwise separations  $\mathbf{M}^{(2)}$  can we recover coordinates which respect these separations? If you were to just try the same process as for the Euclidean case as described in the previous section, a problem occurs when taking the eigenvalues of the double centred matrix  $\mathbf{B}$ , as they are now no longer all positive and so we cannot take their square root to find the desired coordinates. Therefore we must rederive a generalised solution to this problem where the pairwise distances are now given by  $(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{G} (\mathbf{x}_i - \mathbf{x}_j)$ ,  $N$  generic vectors  $\mathbf{x}_i$  in  $D$  dimensions, and some  $D \times D$  matrix representing the metric  $\mathbf{G}$ . We will

restrict ourselves to the case where  $\mathbf{G}$  is diagonal, and, for simplicity, where the values on this diagonal are in  $\{+1, -1\}$ .

As in the classical case we first fix the coordinates' centre of mass, placing it at the origin, such that

$$\sum_{i=1}^n x_{i,k} = 0 \quad \text{for } k = 1, 2, \dots, D. \quad (7.5)$$

Then, the square separations between each point,  $S_{ij}^{(2)}$  according to our metric matrix  $\mathbf{G}$  are given by:

$$S_{ij}^{(2)} = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{G} (\mathbf{x}_i - \mathbf{x}_j). \quad (7.6)$$

In a similar fashion to the derivation of classical MDS we then find that:

$$S_{ij}^{(2)} = \mathbf{x}_i^T \mathbf{G} \mathbf{x}_i + \mathbf{x}_j^T \mathbf{G} \mathbf{x}_j - 2 \mathbf{x}_i^T \mathbf{G} \mathbf{x}_j, \quad (7.7a)$$

$$\frac{1}{n} \sum_{i=1}^n S_{ij}^{(2)} = \mathbf{x}_j^T \mathbf{G} \mathbf{x}_j + \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^T \mathbf{G} \mathbf{x}_i - \frac{2}{n} \sum_{i=1}^n \mathbf{x}_i^T \mathbf{G} \mathbf{x}_j, \quad (7.7b)$$

$$\frac{1}{n} \sum_{i=1}^n S_{ij}^{(2)} = \mathbf{x}_j^T \mathbf{G} \mathbf{x}_j + \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^T \mathbf{G} \mathbf{x}_i - \frac{2}{n} \sum_{i=1}^n \sum_{k=1}^D x_{i,k} \mathbf{G}_{kk} x_{j,k}. \quad (7.7c)$$

The last term on the right vanishes if the order of the sums is reversed due to the centre of mass of the coordinates being chosen to lie at the origin, as stipulated above. Doing the same summing over  $j$  as well gives us two equations:

$$\frac{1}{n} \sum_{i=1}^n S_{ij}^{(2)} = \mathbf{x}_j^T \mathbf{G} \mathbf{x}_j + \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^T \mathbf{G} \mathbf{x}_i, \quad (7.8a)$$

$$\frac{1}{n} \sum_{j=1}^n S_{ij}^{(2)} = \mathbf{x}_i^T \mathbf{G} \mathbf{x}_i + \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j^T \mathbf{G} \mathbf{x}_j. \quad (7.8b)$$

Summing again provides a third expression:

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(2)} = \frac{2}{n} \sum_{i=1}^n \mathbf{x}_i^T \mathbf{G} \mathbf{x}_i. \quad (7.9)$$

We can now write down each element in the matrix  $\mathbf{B} = \mathbf{X}^T \mathbf{G} \mathbf{X}$  solely in terms of  $S_{ij}^{(2)}$ .

$$B_{ij} = \mathbf{x}_i^T \mathbf{G} \mathbf{x}_j , \quad (7.10a)$$

$$B_{ij} = -\frac{1}{2} \left[ S_{ij}^{(2)} - \mathbf{x}_i^T \mathbf{G} \mathbf{x}_i - \mathbf{x}_j^T \mathbf{G} \mathbf{x}_j \right] , \quad (7.10b)$$

$$B_{ij} = -\frac{1}{2} \left[ S_{ij}^{(2)} - \frac{1}{n} \sum_{i=1}^n S_{ij}^{(2)} + \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^T \mathbf{G} \mathbf{x}_i - \frac{1}{n} \sum_{j=1}^n S_{ij}^{(2)} + \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j^T \mathbf{G} \mathbf{x}_j \right] , \quad (7.10c)$$

$$B_{ij} = -\frac{1}{2} \left[ S_{ij}^{(2)} - \frac{1}{n} \sum_{i=1}^n S_{ij}^{(2)} - \frac{1}{n} \sum_{j=1}^n S_{ij}^{(2)} + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{(2)} \right] . \quad (7.10d)$$

Which, as in the classical case, means that we can calculate  $\mathbf{B}$  from the matrix  $\mathbf{S}^{(2)}$  using the double centring matrix  $\mathbf{J}$  as:

$$\mathbf{B} = -\frac{1}{2} \mathbf{J} \mathbf{S}^{(2)} \mathbf{J} . \quad (7.11)$$

The matrix  $\mathbf{B}$  is guaranteed to be real and symmetric, from the symmetry of equation Equation 7.10d. This means it can always be decomposed into the form

$$\mathbf{B} = \mathbf{U}^T \Sigma \mathbf{U} \quad (7.12)$$

where  $\mathbf{U}$  is a matrix of orthogonal eigenvectors and  $\Sigma$  the corresponding eigenvalues. However, unlike in the previous case there is no guarantee that the eigenvalues are all positive. We are searching for the matrix of coordinates  $\mathbf{X}$  that satisfy  $\mathbf{B} = \mathbf{X}^T \mathbf{G} \mathbf{X}$  and so propose a solution:

$$\mathbf{Y} = \sqrt{|\Sigma|} \mathbf{U} . \quad (7.13)$$

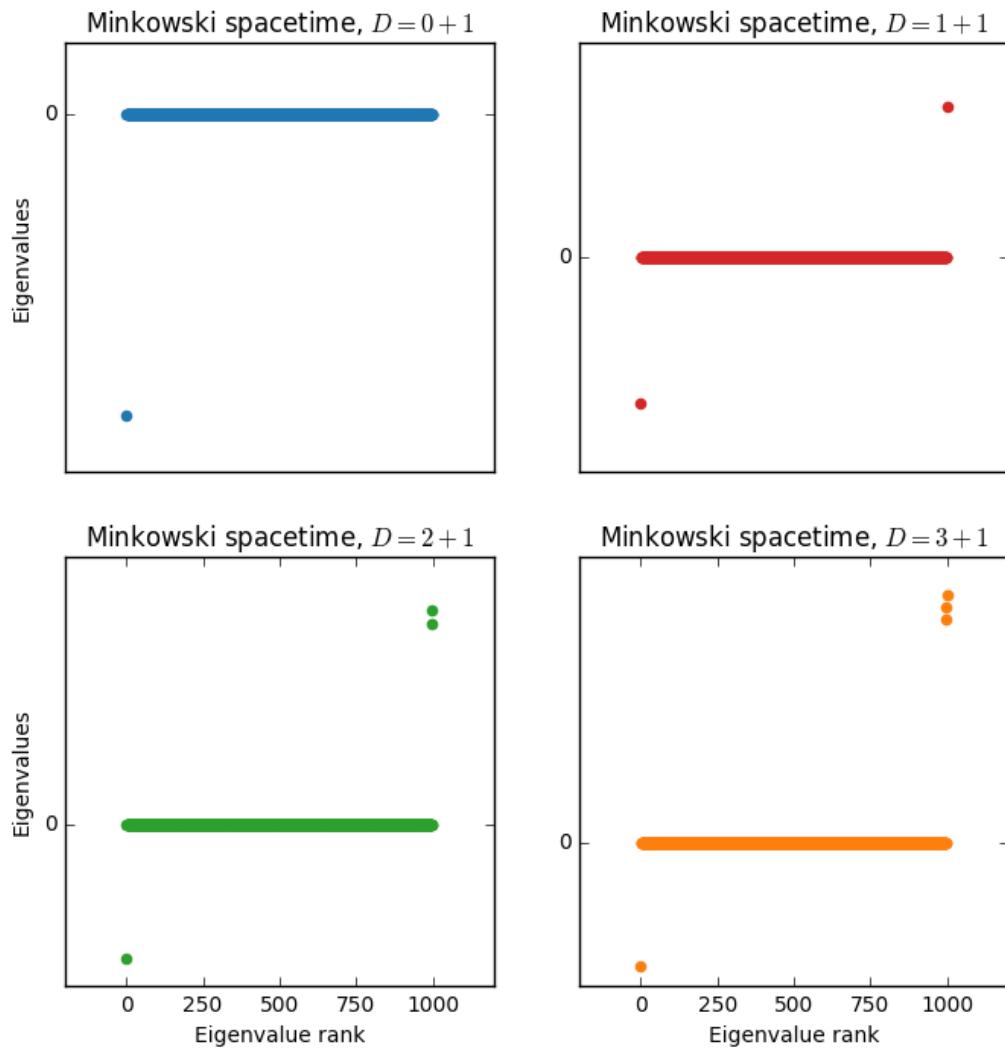
Trying  $\mathbf{Y}$  as a solution gives

$$\mathbf{Y}^T \mathbf{G} \mathbf{Y} = \mathbf{U}^T \sqrt{|\Sigma|} \mathbf{G} \sqrt{|\Sigma|} \mathbf{U} , \quad (7.14a)$$

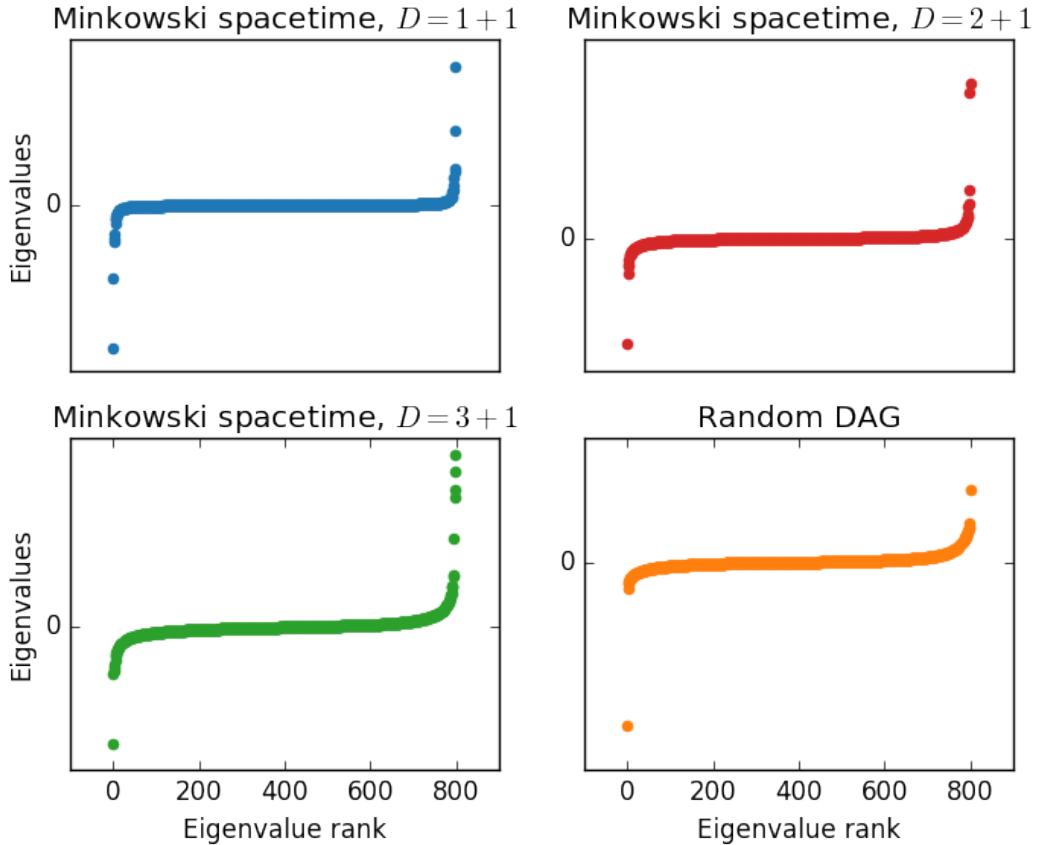
$$\mathbf{Y}^T \mathbf{G} \mathbf{Y} = \mathbf{U}^T |\Sigma| \mathbf{G} \mathbf{U} . \quad (7.14b)$$

We can now see that  $\mathbf{Y}^T \mathbf{G} \mathbf{Y}$  will be equal to  $\mathbf{B}$  only if the metric matrix  $\mathbf{G}$  is chosen to

be the signs of the eigenvalues of  $\mathbf{B}$ . This means that, whereas in the Euclidean case for MDS we chose the largest  $D$  eigenvalues to find coordinates in a  $D$  dimensional space, we now choose the largest  $D_+$  positive eigenvalues, and the largest  $D_-$  negative eigenvalues and this provides coordinates in a space with a metric of  $D_+$  plus signs, and  $D_-$  minus signs. In this manner we can pick out the most significant space (plus) and time (minus) dimensions for the embedding spacetime. The  $D$  dimensional Euclidean case is that where  $D_+ = D$  and  $D_- = 0$ , whereas  $(d+1)$  dimensional Minkowski spacetime, which is the case of relevance here, is where  $D_+ = d$  and  $D_- = 1$ . This is illustrated in Figure 7.3 where the eigenvalues of  $\mathbf{B}$  are shown for points scattered in Minkowski spacetimes of various dimensions. In Figure 7.4 the eigenvalues for the MDS embedding of causal set graphs are shown, and in Figure 7.5 those for real networks. The embedding process is outlined in pseudo-code in algorithm 5, assuming some function exists to find the longest chain between two vertices in the graph.



**Figure 7.3:** MDS eigenvalues  $\Sigma$  for exact coordinates of causal sets sprinkled in Minkowski spacetime. Eigenvalues are sorted by size. Note the large positive eigenvalues on the right of each plot, and the large negative eigenvalue on the left. In this diagram the distance matrix used is the exact distance matrix from a set of  $N = 1000$  scattered points, and so of the 1000 eigenvalues extracted, only  $d + 1$  are non-zero, with  $d$  spatial dimensions corresponding to positive eigenvalues and the 1 time dimension to the negative eigenvalue.

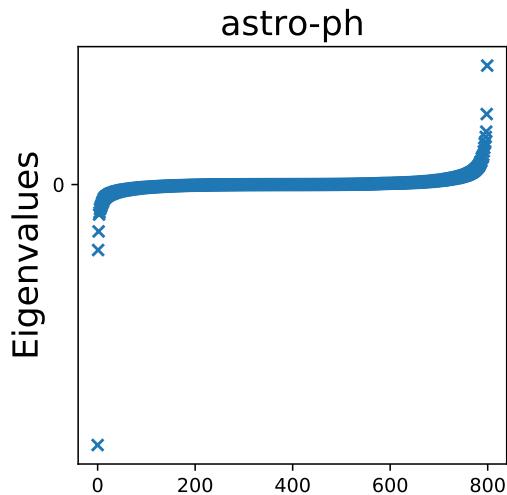


**Figure 7.4:** MDS eigenvalues  $\Sigma$  for causal set graphs and a random DAG each with  $N = 800$ . Eigenvalues are sorted by size. The large positive eigenvalues on the right of each plot correspond to spatial dimensions in the MDS embedding. The number of these large positive eigenvalues generally corresponds to the dimension of the causal set graphs in question, as would be expected if the MDS process accurately extracted the original sprinkling coordinates.

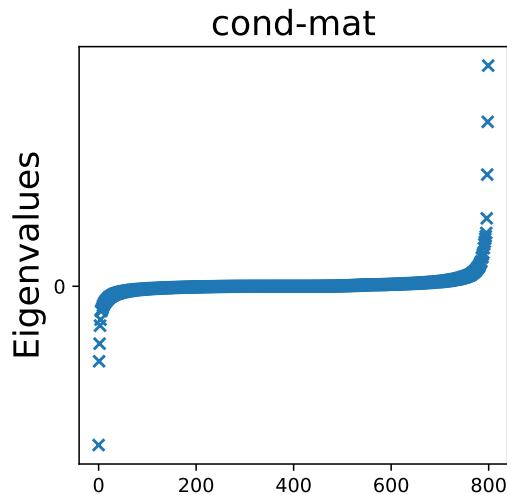
As the dimension of the causal set graphs grow, the distinction of these particularly large eigenvalues diminishes, as can be seen for the  $D = 3 + 1$  case in the lower left plot. The noise of the random sprinkling process begins to drown out the signal of the dimension of the sprinkling coordinates. The large negative eigenvalue (on the lower left of each plot) in each case corresponds to the one time dimension of the original causal set graphs. In all cases this eigenvalue is distinct and significantly larger than the others.

For the random DAG in the lower right plot, this one time dimension emerges as a consequence of the ordering of the vertices present in all DAGs.

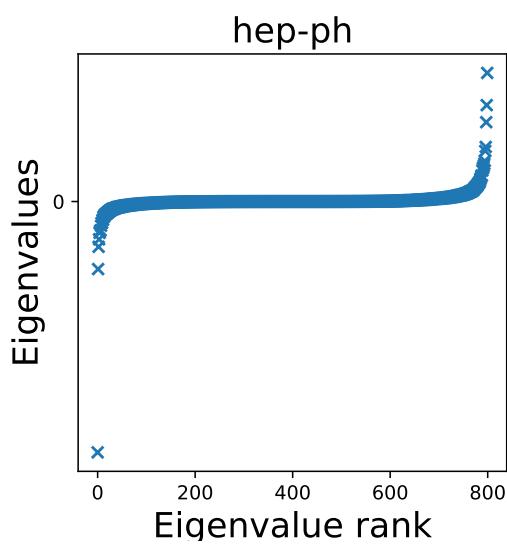
Note the symmetry in the upper-left plot between positive and negative eigenvalues. This corresponds to the symmetry between the one time and one space dimension in  $D = 1 + 1$  Minkowski spacetime, where the time and space dimensions can be interchanged while the causal set graph remains statistically invariant.



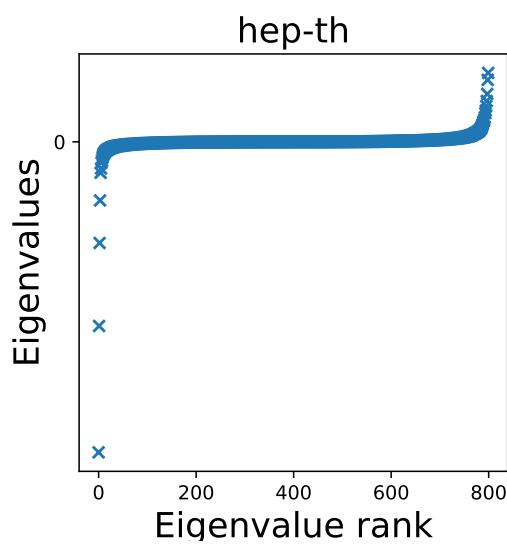
(a) arXiv/astro-ph



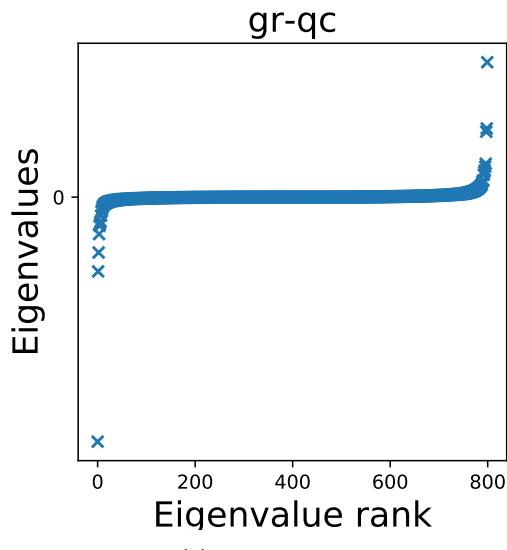
(b) arXiv/cond-mat



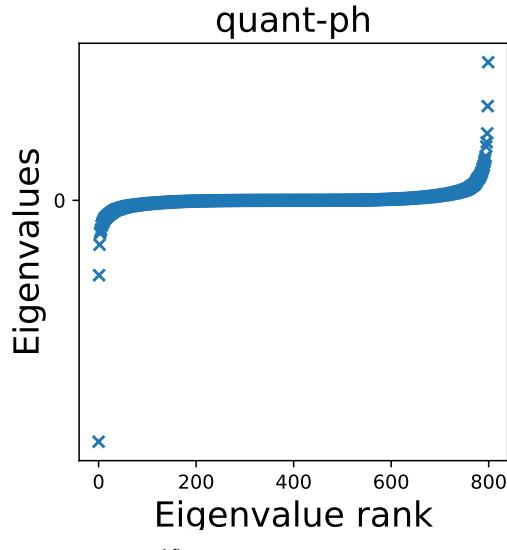
(c) arXiv/hep-ph



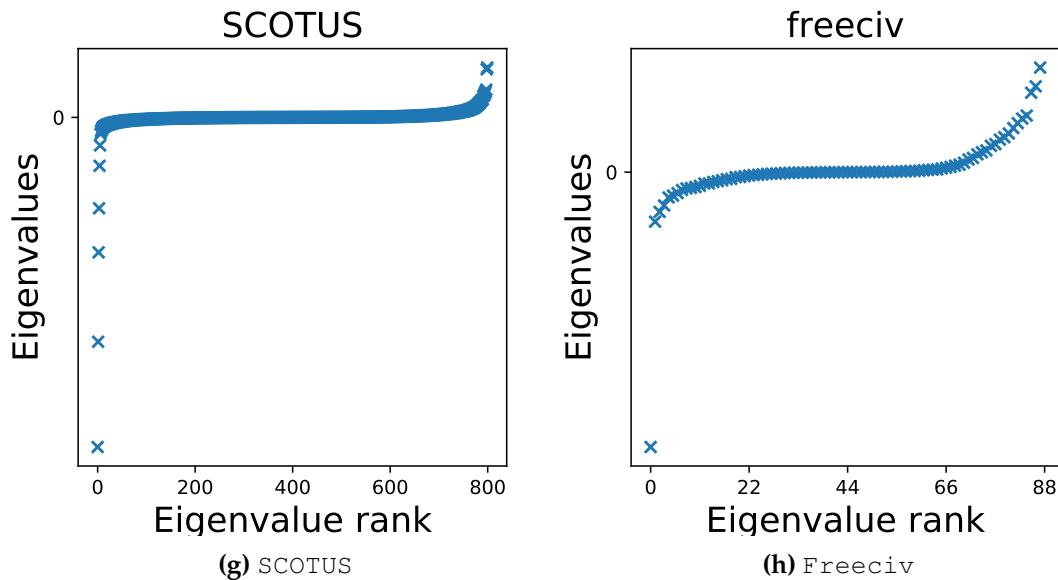
(d) arXiv/hep-th



(e) arXiv/gr-qc



(f) arXiv/quant-ph



**Figure 7.5:** MDS embedding eigenvalues estimates for real networks. As in the case of the causal set graphs, most real networks show a large negative eigenvalue corresponding to the order on the vertices, and which will correspond to the time dimension in the Minkowski space-time embedding. However, unlike the causal set graphs there is generally not a distinct number of ‘large’ positive eigenvalues suggesting that there is not an obvious number of spatial dimensions to choose.

---

**Algorithm 5** Embedding a DAG in Minkowski Spacetime

---

```

1: function EMBED GRAPH( $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}, \hat{D}$ )
2:    $M \leftarrow \text{SEPARATION MATRIX}(\mathcal{G})$ 
3:    $N \leftarrow |\mathcal{V}|$ 
4:    $J \leftarrow \mathbf{1} - \frac{1}{N}\mathbf{1}\mathbf{1}^T$ 
5:    $B \leftarrow -\frac{1}{2}JMJ$ 
6:    $U, \Sigma \leftarrow \text{DECOMPOSE}(B)$                                  $\triangleright$  Eigenvectors  $U$  and eigenvalues  $\Sigma$ 
7:    $X_0 \leftarrow \sqrt{-\Sigma_N} U_N$                                  $\triangleright \Sigma_N$  is the most negative eigenvalue
8:   for  $d \in 1, 2, \dots, \hat{D} - 1$  do
9:      $X_d \leftarrow \sqrt{\Sigma_d} U_d$                                  $\triangleright \Sigma_d$  is the  $d^{\text{th}}$  largest positive eigenvalue
10:    return  $X$                                                $\triangleright$  The embedding coordinates

```

```

1: function SEPARATION MATRIX( $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ )
2:    $M \leftarrow N \times N$  matrix of zeroes
3:   for  $u \in \mathcal{V}$  do
4:     for  $v \in \mathcal{V}$  do
5:       if  $u \prec v$  or  $v \prec u$  then                                 $\triangleright$  timelike separated
6:          $L \leftarrow \text{LONGEST CHAIN}(\mathcal{G}, u, v)$ 
7:          $M_{ij} \leftarrow -L^2$ 
8:       else                                               $\triangleright$  spacelike separated
9:          $L \leftarrow \text{NAIVE SPACELIKE DISTANCE}(\mathcal{G}, u, v)$ 
10:         $M_{ij} \leftarrow L^2$ 
11:   return  $M$ 

```

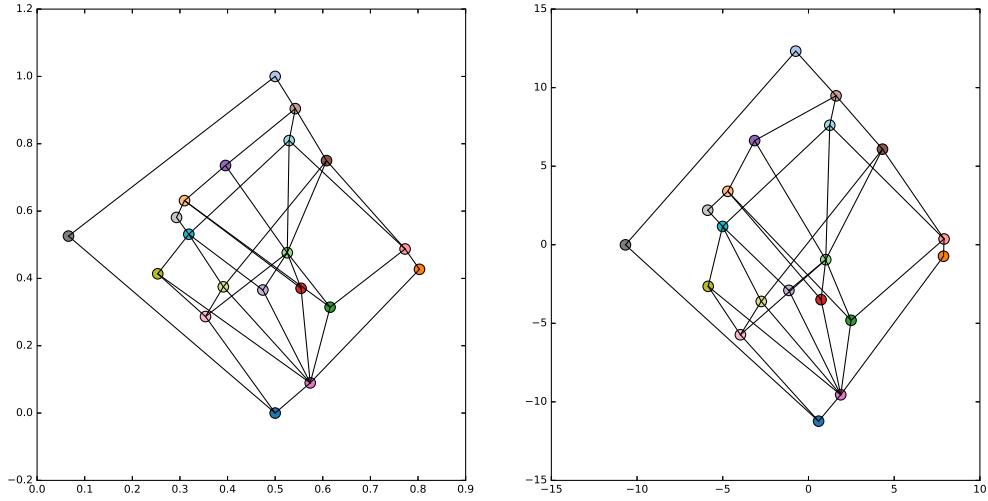
---

```

1: function NAIVE SPACELIKE DISTANCE( $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}, u, v$ )
2:    $\text{maxpath} \leftarrow 0$ 
3:   for  $w \in J^+(u) \cap J^+(v)$  do                                 $\triangleright$  intersection of futures
4:     for  $z \in J^-(u) \cap J^-(v)$  do                                 $\triangleright$  intersection of pasts
5:        $L \leftarrow \text{LONGEST CHAIN}(\mathcal{G}, w, z)$ 
6:        $\text{maxpath} \leftarrow \max(\text{maxpath}, |L|)$ 
7:   return  $\text{maxpath}$ 

```

---



**Figure 7.6:** An example of this embedding algorithm applied to a causal set graph with  $N = 200$  vertices. Of these 200 vertices, 20 have been drawn, and only the covering edges are drawn between them. On the left, the original sprinkling coordinates used to generate this causal set graph. On the right, the estimated coordinates from the embedding algorithm. The similarity between these two drawings demonstrates the algorithm’s effectiveness.

The placements of the vertices are generally very similar, but some small differences are noticeable. These occur due to the fact that many sprinklings will generate the same causal set structure and since it is only this structure that the embedding process has access to, there is no way for it to distinguish between these sprinklings. The difference in scale between the two drawings is because the density of the sprinkling does not affect the causal structure used to embed the graph and so the estimated coordinates have their own natural length scale with an approximate timelike separation of 1 between nearest neighbours.

## 7.5 Results

Figure 7.7 and Figure 7.8 give examples of a  $D = 1 + 1$  embedding for the hep-ph and hep-th citation networks. Nodes in the network are drawn using the coordinates assigned by the Lorentzian MDS algorithm and node size is proportional to citation count. Node colour is determined by the publication date of the papers. Note that this is strongly correlated to the embedded time coordinate for the papers, a feature which emerges naturally from the embedding process.

Once the Lorentzian MDS algorithm has estimated coordinates in  $\mathbb{M}^D$  for each vertex in the graph how can we assess the accuracy of the embedding? Informally, a good embedding is one where related vertices in the graph are timelike separated, and un-

related nodes are spacelike separated. When discussing the order dimension we saw that not every DAG can be embedded perfectly in Minkowski spacetime so that this relation is respected for all pairs, even for an arbitrarily large dimension. Even where a DAG can be perfectly embedded the discrete structure of the graph will introduce noise meaning that we will not necessarily recover the correct causal structure.

To quantify the effectiveness of an embedding we will take the estimated coordinates and use them to rebuild the graph by again placing edges only between timelike separated pairs. If there are edges between the same vertices in the recreated graph and in the original graph, the embedding is an accurate one, and if not then it is poor. As in [106] we will measure this using the sensitivity (the fraction of the correct edges which were predicted) and specificity (the fraction of correct non-edges which were predicted). We are effectively considering the estimated coordinates as a method of predicting edges in the original graph. To compare the sensitivities and specificities of the various embeddings we use the established method of the area under the receiver-operator curve (AUC). Varying a continuous parameter, the sensitivity and specificity of the embedding is measured, and plotted, as in Figure 7.9 and the area under this curve describes the embedding's quality.

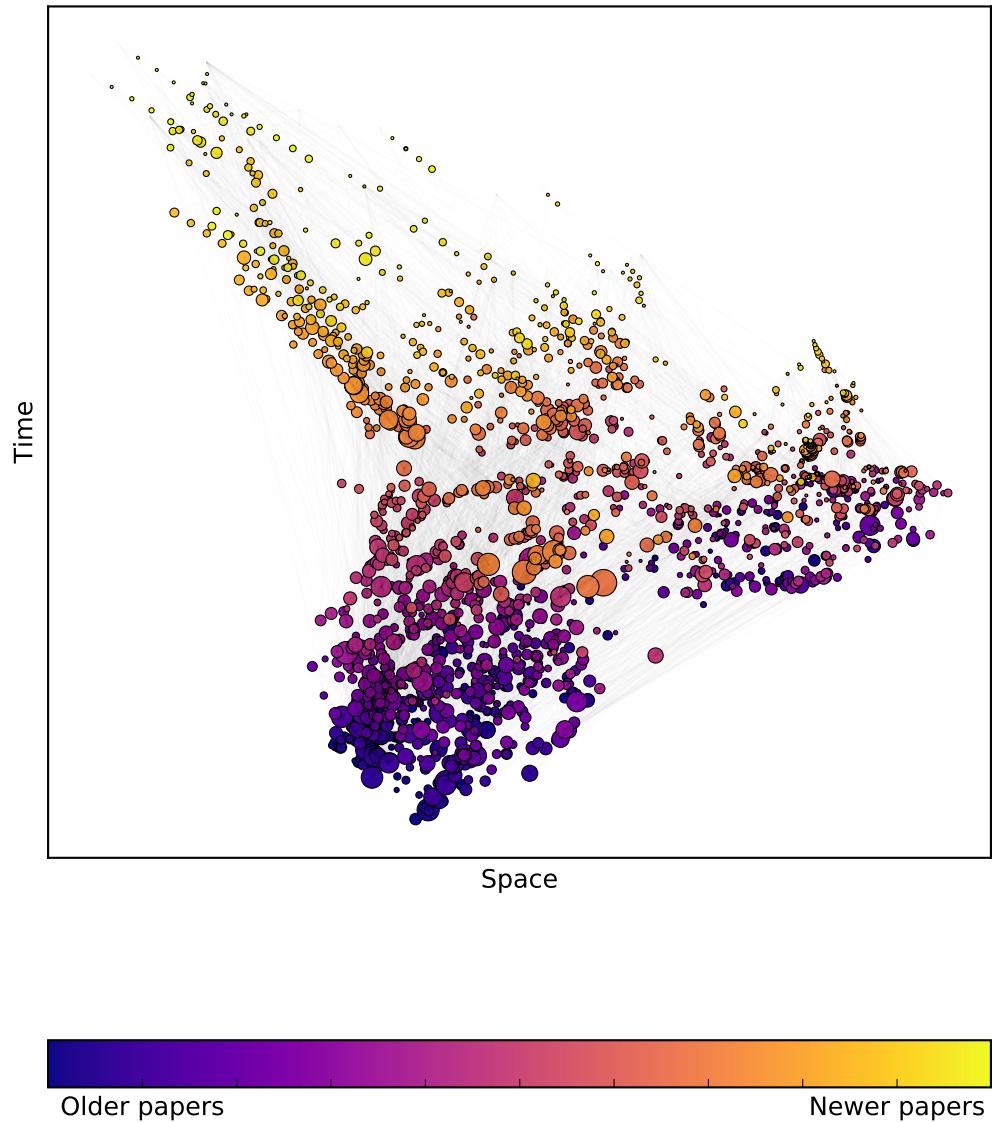
The continuous parameter we will vary is the speed of light (or the speed information can be transferred) in the embedding Minkowski space,  $c$ . Previously, we have set  $c = 1$ , but varying this speed will change which nodes are connected in new network generated from the MDS coordinates. Now, nodes  $i$  and  $j$  are connected if their coordinates  $u_i, u_j$  satisfy

$$-c(u_{i,0} - u_{j,0})^2 + \sum_{k=1}^d (u_{i,k} - u_{j,k})^2 < 0 \quad (7.15)$$

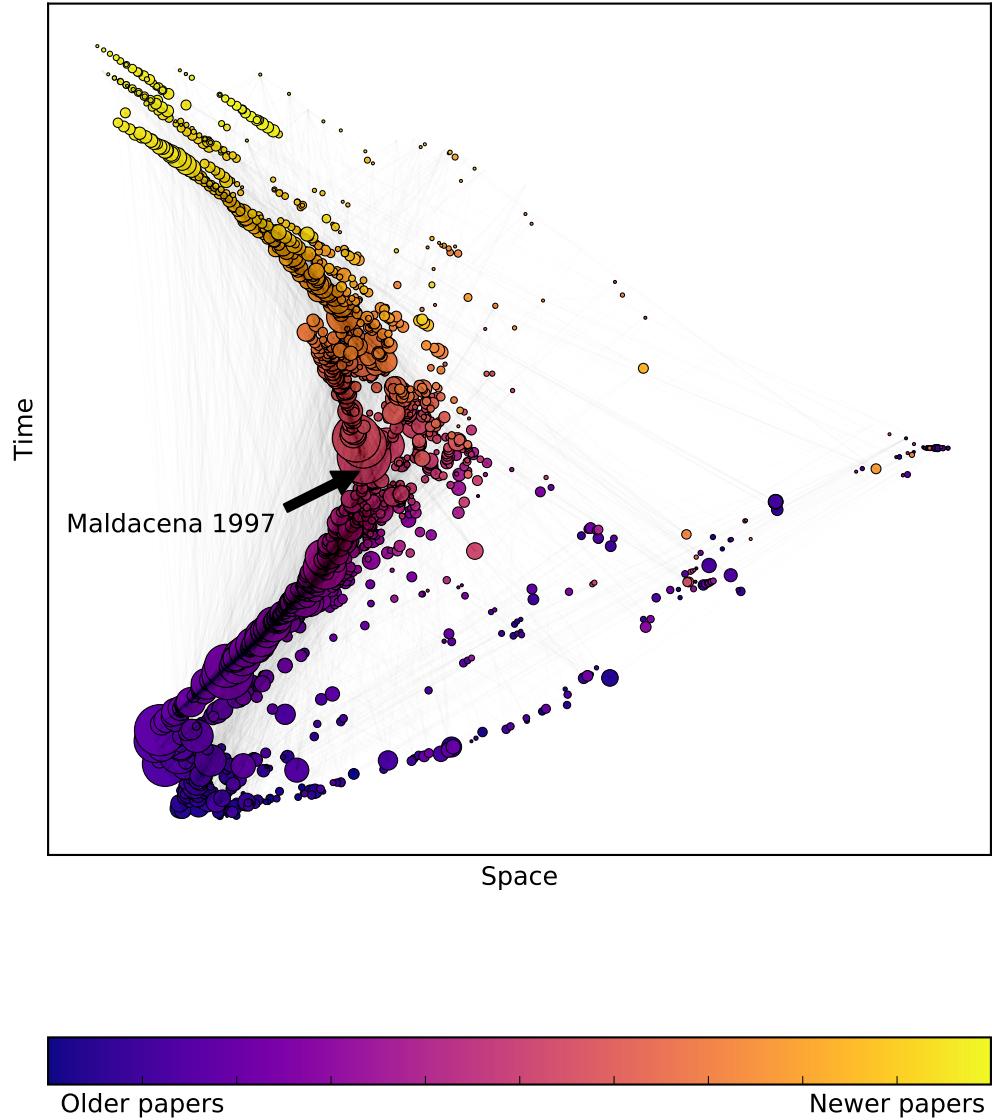
For small values of  $c$ , very few nodes are connected and so the specificity is high (few false positives) but the sensitivity is low (many false negatives). For large values of  $c$ , many nodes are connected and so the reverse is true.

We will compare various networks of size  $N = 1000$ , and results are shown in Figure 7.10. The random DAGs are Erdos-Renyi graphs with the edges then direc-

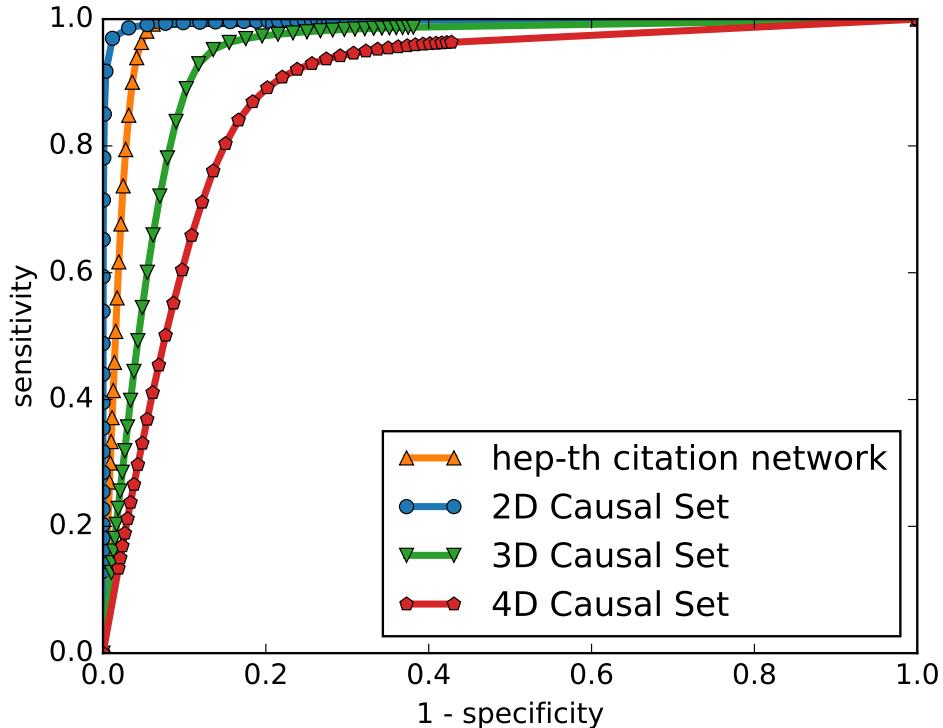
ted according to a random ordering on the nodes, and then transitively completed. The number of edges in the original ER graph is chosen to roughly match the citation networks but the comparatively poor AUC scores of the random DAGs is robust to changes in this parameter.



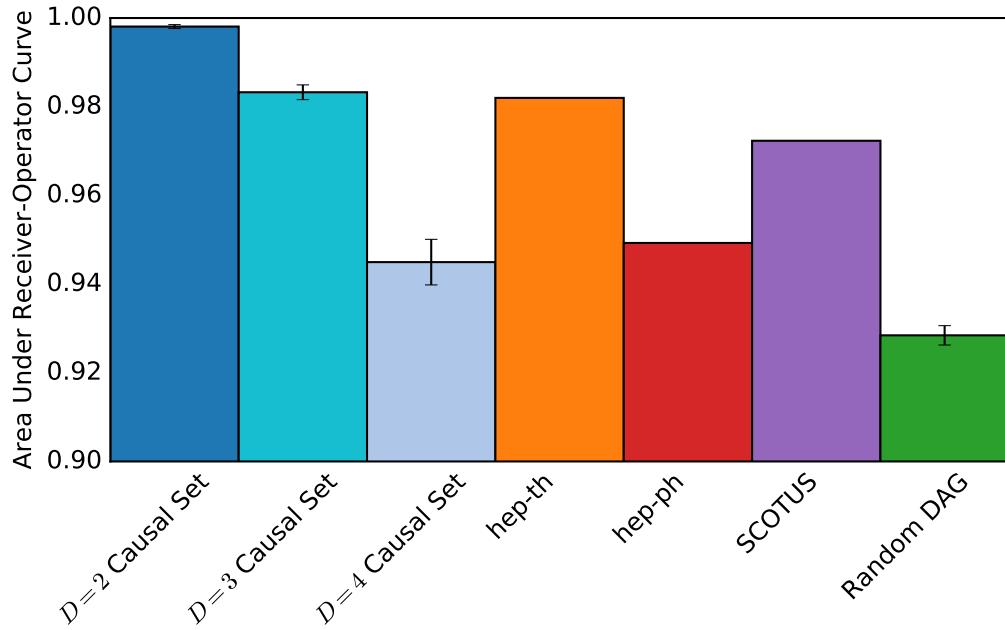
**Figure 7.7:** A visualisation of a  $D = 1 + 1$  embedding of the top 2000 most cited papers in the hep-ph citation network, where node size is proportional to the number of citations. Node colour corresponds to publication date, and correlates strongly with the time coordinate obtained from the embedding algorithm. The hep-ph citation network appears more broad in space indicating more pairs of papers which are spacelike separated from each other.



**Figure 7.8:** A visualisation of a  $D = 1 + 1$  embedding of the top 2000 most cited papers in the hep-th citation network, where node size is proportional to the number of citations. Node colour corresponds to publication date, and correlates strongly with the time coordinate obtained from the embedding algorithm. In contrast to the hep-ph network, the hep-th citation network has most of its papers in a long chain indicating more timelike separated pairs. We highlight the central placement of the most cited paper in that citation network Maldacena's paper [133] which first proposed AdS/CFT correspondence. The visually 'narrow' citation network of hep-th and 'broad' hep-ph agrees with previous findings in [61].



**Figure 7.9:** Curves showing the sensitivity and specificity of embeddings into  $D = 1 + 1$  Minkowski space of causal set graphs, and a citation network from the hep-th section of arXiv, all with  $N = 1000$  vertices. One the network's nodes have assigned coordinates from the embedding algorithm, a new network is built by placing edges between timelike separated pairs. This is done for many values of  $c$ , generating many new networks which are each compared with the original, making a point in this plot. Specificity measures the fraction of edges in the new network which were also in the original, and sensitivity the fraction of non-edges in the new network which were also in the original. By varying  $c$ , these fractions change, forming the continuous curves seen on this plot. The area under these curves gives an overall measurement of how effective the embedding is.



**Figure 7.10:** Area under the curve (AUC) values represent the quality of an embedding. Here we show the AUC values for embedding graphs with 2000 nodes into  $D = 1 + 1$  Minkowski spacetime. A value of 1 represents a perfect embedding, and a value of 0.5 is random chance. The two-dimensional causal set graph has, as expected, the highest value, since there must be coordinates allowing a perfect embedding (the original coordinates used when building that graph). Higher dimensional causal sets can be embedded less well, but still better than a random DAG (far right). Error bars show the standard deviations of this measurement over 20 randomly generated examples. Notably, the three citation networks we use as examples have significantly higher values than the random DAG illustrating that they have structure which allows a better fit to Minkowski spacetime. The comparatively better fit of the hep-th network over the hep-ph network into 2-dimensional spacetime agrees with our result in [61]

## 7.6 Discussion and Applications

There are a number of possible applications to this approach. Firstly, there are applications where the network data of interest is in the form of a DAG and so as is the case for causal sets the natural geometric approach to take is a Lorentzian one. Finding an effective geometric embedding of a network provides a powerful tool for the analysis of that network as it allows standard geometric techniques and intuition to be used. Calculations of network properties can be made more efficient, for example, when finding optimal routes from one node to another, the node coordinates provide local information which can improve routing algorithms [125].

A second application is in dimensionality reduction. Where there is some domain specific reason to think that the appropriate target space should be Lorentzian, because datapoints are associated with points in time and are related in some causal way, this Lorentzian MDS approach can be used as MDS usually is in computer science for finding the projections in the data with the most variance. Generalising the equations for classical MDS allows it to be used to embed into spaces of any metric signature, even though we have focused only on the Lorentzian signature here. To my knowledge this pseudo-Riemannian output for dimensionality reduction is a new development, although some kind of manifold learning techniques exist which can take pseudo-Riemannian manifolds as their input [128, 204]. When performing the embedding one can find multiple negative eigenvalues, suggesting that embeddings in spaces with more than one timelike dimension is also possible, as are potential embeddings into Lorentzian manifolds other than Minkowski space, incorporating curvature or preferred directions. Furthermore dimensionality reduction algorithms which begin by building a graph of nearest neighbours (such as Isomap [194]) could be adapted to have a Lorentzian spacetime as the target space using the kind of longest-path approach we describe here.

The visualisation of networks is a problem in its own right, and two or three dimensional embeddings from this method can be combined with standard plotting software to give network visualisations for DAGs in which the causal ordering is explicit. Such

visualisations are used in bibliometrics to help identify distinct fields or assist literature reviews [196].

When standard dimensionality reduction techniques are used on high-dimensional datasets it is common to see complicated, abstract features of the data represented by directions in the reduced coordinates which may represent the underlying degrees of freedom in the mechanism generating the data. See for example in [194] where this effect is apparent on images and handwritten digits. In the cases of citation analysis we can conjecture that the spatial dimensions that result from a geometric approach correspond to similarity in the topic of a paper, and so this approach yields spatial similarities between papers while accounting for the time difference in their publication. Once estimated coordinates are known, the idea that nodes may be ‘similar’ can be expressed as nodes being close in their spatial coordinates. Two papers may not cite each other, or share authors or citations yet be close in the embedded coordinates since these are calculated globally using information from all vertices and edges. Closeness in the embedded coordinates is then a similarity measure which can be used for applications such as clustering, paper recommendation, and centrality measures, the effectiveness of which is an avenue for future work.

Another use of this approach is where edges in a network are placed primarily according to some geometric rule but their connections are also governed by some smaller second order effect. It may only be possible to measure the smaller effect once we have accounted for the primary geometric one by assigning coordinates. We can see this phenomenon clearly when the geometric embedding is in real geographic space, such as in [78] where accounting for geographic distance in phone-call data allows more accurate prediction of the second order effect of shared language.

## 7.7 Summary

In this chapter an algorithm based on causal set tools and MDS was used to embed the nodes of a DAG in Minkowski spacetime, such that the edges of the graph corresponded as closely as possible to the causal structure of the embedding spacetime.

By approximating the timelike separation between related pairs of nodes as the length of the longest path between them, and using the naive spacelike estimator for unrelated pairs we found the separations between every pair of vertices in the graph. Then we saw how classical Riemannian MDS could be generalised to allow embedding into spaces with any metric signature, and using the Lorentzian signature we could then embed the vertices into Minkowski spacetime.

This method was successful at approximately recovering the coordinates of causal set graphs. By counting the fractions of correctly recovered timelike and spacelike separated pairs the accuracy of the embeddings of complex networks was assessed in Figure 7.10 and it was found that these networks appeared to be better fits to Minkowski spacetime than equivalent random graphs, although they were worse than causal set graph models. As has been the case for Riemannian network geometry, the suggestion of using these embedding coordinates as a similarity space is raised as a hypothesis that is open for investigation using domain specific metadata for various kinds of complex network.

# Chapter 8

## Conclusions and Outlook

The worst thing you can do is to  
completely solve a problem.

---

Dan Kleitman

Complexity science and network science are each still in their infancy. We do not have an overarching framework or theory to describe the wide array of systems and phenomena that show complex emergent behaviour. Perhaps one day we will, or perhaps it will never come. Because of that, it is difficult to be completely precise or rigorous about how the work presented here fits into the wider theory of complex systems. What is possible though is to highlight the progress that has been made in observing and characterising complex networks, and the analogies that can be made between fields of study that previously seemed disparate and unrelated. That is what I hope to have done throughout this thesis, and in this final chapter I will try to tell the whole story again concisely enough that the reader can see everything at once.

We began in chapter 1 by reviewing the established work that has demonstrated the effectiveness of graphs in describing the structure or topology of complex systems of many interacting parts. Graphs provide the mathematical framework for network science, describing the structure, growth and dynamics of the interactions between small parts of complex systems. These mathematical objects are defined simply enough for our analysis, but can contain enough information and complexity to capture exciting emergent phenomena. At its very least, network science provides us with some sur-

prisingly useful tools from which to build numerical analyses of complicated datasets. At its highest, it lets us glimpse at some underlying universalities between complex systems from different areas of science.

Chapter 2 introduces the subfield of relevance to the work presented here, network geometry. These geometric approaches recognise that many real systems of interest are described by a special class of graphs which emerge naturally from simple geometric rules. Thinking about networks in this way allows us to firstly visualise the abstract spaces their nodes live in and describe their structure in more simple and precise ways. Some of the universal properties we *observed* in chapter 1 can be *explained* by appealing to underlying geometries, and important properties which otherwise appeared unrelated (eg. clustering, scale-free degree distributions) emerge naturally from the same geometric rules and so their description is, in some sense, unified.

However, previous geometric approaches have been restricted to Riemannian geometries (speaking in mathematical terms) or to space, rather than spacetime (in the language of physicists). This is all well and good for networks representing systems where edges represent similarity relations which are well modelled by spatial interactions. But other networks exist, with other kinds of relationships. In some systems, edges represent causal relations, not similarity. Crucially this introduces a partial order on the vertices. Often this order represents time, and edges exist between two vertices if they are ‘close enough’ for the earlier one to causally interact with, or pass information to the later one. In either the viewpoint of causality or information flow, there is a natural geometry to use which is Lorentzian geometry. It is the natural choice because it respects the constraints we expect of causal relationships, and because it allows us to draw analogies between the causality and information flow in the abstract spaces of our networks, to the same process in real spacetime as studied in physics. In chapter 3 we introduced causal sets, the mathematical framework used to describe discrete Lorentzian geometries in theoretical physics and quantum gravity. We built an analogy between causal sets and networks with the same causal structure, directed acyclic graphs. Drawing this analogy allows us to bring all of the tools, theorems,

algorithms and techniques of causal set theory into network science via the network geometry framework. This chapter described those tools used in later chapters and gave the background behind causal sets as a physical theory.

Now the groundwork is done and the foundations laid, real data can be brought into play. In chapter 4 we looked at the simplest analysis of causal structure, which is to simply ask which edges in the graph are a necessary part of that structure and which are not. Such a simple question has surprisingly interesting answers, and we saw that in some cases, almost every edge remained after transitive reduction and so was necessary for the network’s causal structure, while in other cases most edges were removed. In citation networks of academic papers we find agreement with previous works on copied citations, seeing around 80% of the citations removed, and so suggest that the remaining network structure may hold significant information which is otherwise obscured. Individual papers showed very different properties under transitive reduction indicating competing types of citation behaviour in the same citation network.

We then began to associate networks and their causal structure with particular Lorentzian geometries. Starting with the simplest such geometry, Minkowski spacetime  $\mathbb{M}^D$ , in chapter 5 we used dimension estimators from causal set theory to give an estimated dimension of a network’s embedding spacetime. We found agreement between independent measures for given networks, whose dimensions appeared to converge to a consistent result in each case. Moreover, these results showed remarkable similarities to those for causal set graphs suggesting that these networks may be, in some regards at least, well described by embedding in Minkowski spacetime. In chapter 6 this idea is taken further by using the geometry of curved spacetime, de Sitter spacetime, to characterise network structure.

The abstract space we embed into was interpreted in terms of ‘narrow’ and ‘diverse’ network growth (or in the case of citation networks, citation behaviour). The techniques developed in these chapters allow for metadata based measures of diversity to be compared to the structural measures presented here, and much future work lies in this area. Interpreting the results of these structural measures is likely to be a domain

dependent task, as networks representing different systems are likely to show different properties, and the meaning of narrow or diverse growth will vary from one application to another. What is significant though is that once these results can be calibrated against well understood measures of diversity in a particular application, they can then be applied to networks where the data required for current standard methods (eg. text analysis of papers, or author similarities in citation networks) is not present, and only the network structure is. We saw that apparently very similar citation networks can be clearly distinguished by these dimension and curvature measures without the information of paper titles or authors being needed.

In chapter 7 we actually mapped directed acyclic graphs to Lorentzian spacetime by embedding the nodes of the network. The algorithm developed in this chapter provides network visualisations which reflect the causal ordering of the vertices in these graphs. More importantly though it allows measurement of the extent to which a given network's structure can be described by a given Lorentzian spacetime. Embedding highly connected subsets of the networks previously considered, demonstrated what had been suggested by the results of previous chapters; namely that many real networks with causal structure can be well described by coordinates in Minkowski spacetime. Counting how many of the network's edges are predicted by the embedding and how many are missed lets us quantify the quality of an embedding. The real networks studied here were, understandably less well fit than the perfect case of a causal set graph, but significantly better fit than equivalent random graphs. There is a vast expanse of possible future work in this direction. Firstly, using improved algorithms as described in this chapter, and optimising the computational implementations used, larger networks can be embedded. Secondly, although flat Minkowski spacetime is the simplest option for an embedding space, it is by no means the only option available. The underlying MDS technique that is generalised in this chapter has, in its classical Riemannian version, been applied to curved spaces [203, 202, 66] and so I hope similar generalisations to curved Lorentzian manifolds can be achieved. Big leaps in network geometry have been made by utilising curved spaces and so embedding

into curved spacetimes is a natural next step which could potentially allow simple, but accurate and predictive descriptions of networks forming DAGs. By further developing methods to fit DAGs to curved spacetime geometries we open up the possibility to embed networks into these spacetimes and, since this approach has worked already in Riemannian space, hopefully achieve even more accurate and descriptive embeddings for DAGs and the complex systems they model.

## 8.1 Future Work

Much of the research detailed in this thesis has been exploratory, drawing analogies between other mathematical and physical approaches and using their techniques to make new observations about complex networks. In some cases, these observations could be compared to existing work on specific complex systems, such as work on redundancy citation networks in chapter 4, and their geometric properties in chapter 5 and chapter 6. However in some cases, it was difficult to make direct comparisons between the observations made here and more domain specific work on the complex systems in question.

There is significant future work to be done in order to verify hypotheses made in earlier chapters. For example, the network geometry approach can let us talk about the ‘diversity’ of citation behaviour, but how does this compare to other measures of diversity in terms of the semantic content of those papers? A proper text analysis was outside the scope of this project but I hope work can be continued in this direction so as to verify or disprove this idea.

A second open path for future work comes in the direction of embedding networks in more complicated spacetimes than flat Minkowski spacetime. Curvature, or more complicated topologies may allow for more accurate and descriptive embeddings. These can both assist in describing a network’s structure but also allow more applications to be derived from the embedding information by making that information more accurate.

Many of the ideas presented here were related to those in causal set theory as

used in theoretical physics as an approach towards a theory of quantum gravity. An idea which collaborators and I have presented elsewhere (but is currently too under-developed to warrant a chapter here) is to develop new measures of network centrality using this causal set approach. Measures of centrality (such as those reviewed in [159, 126, 182]) try to pick out important nodes in a network using the network's structure. In some methods this is done by counting the distance to other nodes (closeness centrality) [28, 174], or the numbers of shortest paths passing through a given node (betweenness centrality) [83, 45, 87]. However we saw that for Lorentzian spacetimes it is not the shortest, but the longest path which corresponds to the geodesic, the path of least action, in the embedding spacetime. Therefore analogous centrality measures to those already commonly used which replace shortest paths with longest paths may turn out to be more appropriate for some DAGs because the implicit geometry used by these centrality measures should in these cases be Lorentzian and not Riemannian.

Other centrality measures which highlight the nodes a random walker is most likely to find itself at [148, 169] also have a natural analogy for DAGs. In causal set theory, where nodes correspond to points in a discrete spacetime, a particle moves stochastically from one discrete spacetime point to another, and summing across all possible paths describes the particles movement in full. Similarly, one could imagine a centrality measure which mimics this calculation where a random particle moves through the ordered vertices of a DAG like a particle moves through the elements of a causal set, with some probability assigned to every possible path but with more central nodes have higher probabilities than others.

These avenues for future work will require further integration of the network geometric and causal set approaches mathematically but I hope they can in time yield further insight into how these networks ought to be analysed.

## 8.2 Available software

It is my hope that much of the work described in previous chapters can be further developed and built upon to reach the point of practical application in solving real problems, and not just those problems we construct in academia for the purpose of solving. When studying complex systems there is always a temptation to create generic methods which do not require detailed knowledge of the system at hand. Their universality and wide potential for application is of course appealing, but is sometimes too much to ask for. However the simplicity and generality of the network approach has thus far proved resilient.

For truly wide use though we need to see people other than those directly involved in network science research to see the value of network methods for their own particular domains, and importantly, see that they can use network methods without first having to become a world-leading expert in the field. It is with this aim in mind that I have endeavoured to make the computer code I (and collaborators) have written as part of the papers and projects undertaken during my PhD available, open source, and hopefully useful, and intelligible to others who are interested. Based on the the popular and easy to use `networkx` package for the Python programming language, our code is available at [59]. The simplest and most widely used algorithms (namely transitive reduction and transitive closure, measuring the length of longest paths etc.) are included in `networkx` itself. Other more specialised code will remain under the sole care of myself and future students of Dr. Tim Evans for now, but I will do my best to maintain our work, and I of course welcome additions from any willing contributors. Since the precise details of the available methods and their implementations is liable to change over time I will not include them here, but it is worth saying that the examples used in this thesis are largely included as examples in the provided code and so should be able to be replicated.

# Bibliography

- [1] Freeciv, available at [www.freeciv.org/](http://www.freeciv.org/), accessed 21/3/2017.
- [2] Kaggle dataset of Hillary Clinton emails, available at [www.kaggle.com/kaggle/hillary-clinton-emails](http://www.kaggle.com/kaggle/hillary-clinton-emails), accessed 21/3/2017.
- [3] Make Hyperbolic Tilings of Images, available at [www.malinc.se](http://www.malinc.se), accessed 21/3/2017.
- [4] OpenFlights, available at [openflights.org](http://openflights.org), accessed 21/3/2017.
- [5] The On-Line Encyclopedia of Integer Sequences, sequence number A000088, available at [oeis.org/A000088](http://oeis.org/A000088), accessed 21/3/2017.
- [6] J. A. and T. M. *Patents, Citations and Innovations*. MIT Press, 2002.
- [7] O. Aharony, S. S. Gubser, J. Maldacena, H. Ooguri, and Y. Oz. Large N Field Theories, String Theory and Gravity. *Physics Reports*, 323(3-4):261, 1999.
- [8] M. Ahmed and D. Rideout. Indications of de Sitter spacetime from classical sequential growth dynamics of causal sets. *Physical Review D*, 81(8):1–11, 2010.
- [9] A. Aho, M. Garey, and J. Ullman. The transitive reduction of a directed graph. *SIAM Journal on Computing*, 1(2):131–137, 1972.
- [10] M. H. Albert and A. M. Frieze. Random graph orders. *Order*, 6(1):19–30, 1989.
- [11] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, Jan 2002.

- [12] R. Aldecoa, C. Orsini, and D. Krioukov. Hyperbolic graph generator. *Computer Physics Communications*, 196:492–496, 2015.
- [13] E. Anderson. Problem of time in quantum gravity. *Annalen der Physik*, 524(12):757–786, 2012.
- [14] C. Andris, D. Lee, M. J. Hamilton, M. Martino, C. E. Gunning, and J. A. Selden. The rise of partisanship and super-cooperators in the U.S. House of Representatives. *PLoS ONE*, 10(4), 2015.
- [15] A. V. Anisimov and D. E. Knuth. Inhomogeneous sorting. *International Journal of Computer & Information Sciences*, 8(4):255–260, Aug. 1979.
- [16] K. Appel, W. Haken, and J. Koch. Every Planar Map is Four-Colorable. *Illinois Journal of Mathematics*, 21(3):429–567, 1977.
- [17] T. Aste, T. Di Matteo, and S. Hyde. Complex networks on hyperbolic surfaces. *Physica A: Statistical Mechanics and its Applications*, 346(1):20–26, 2005.
- [18] L. Babai. Graph isomorphism in quasipolynomial time [extended abstract]. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 684–697. ACM, 2016.
- [19] E. Bachmat. Discrete spacetime and its applications. *Contemporary Mathematics*, 458:347, 2008.
- [20] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality. *Physical review A*, 59(4):381–384, 1988.
- [21] A. T. Balaban. Applications of graph theory in chemistry. *Journal of Chemical Information and Modeling*, 25(3):334–343, Aug. 1985.
- [22] A.-L. Barabási. *Network science*. Cambridge University Press, 2016.
- [23] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(Oct):509–512, 1999.

- [24] A.-L. Barabási, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, (Feb. 2008):1–19, 1999.
- [25] G. A. Barnett and R. E. Rice. Longitudinal non-euclidean networks: Applying Galileo. *Social Networks*, 7(4):287–322, 1985.
- [26] A. Bart and B. Clair. The Math and Art of M. C. Escher, available at [mathstat.slu.edu/escher/index.php/Hyperbolic\\_Geometry](http://mathstat.slu.edu/escher/index.php/Hyperbolic_Geometry), accessed 21/3/2017.
- [27] M. Barthelemy. Spatial networks. *Physics Reports*, 499(1-3):1–101, Feb. 2011.
- [28] M. A. Beauchamp. An improved index of centrality. *Behavioral Science*, 10(2):161–163, 1965.
- [29] D. Benincasa and F. Dowker. Scalar Curvature of a Causal Set. *Physical Review Letters*, 104(18):181301(May):1–4, 2010.
- [30] D. Benincasa, F. Dowker, and B. Schmitzer. The random discrete action for two-dimensional spacetime. *Classical and Quantum Gravity*, 28(10):105018, 2011.
- [31] G. Bianconi. Interdisciplinary and physics challenges of network theory. *EPL (Europhysics Letters)*, 111(5):56001, 2015.
- [32] G. Bianconi. Supersymmetric multiplex networks described by coupled Bose and Fermi statistics. *Physical Review E*, 91(1):1–10, 2015.
- [33] G. Bianconi and A. Barabasi. Bose-Einstein Condensation in Complex Networks. *Physical Review Letters*, 86(24):5632–5635, June 2001.
- [34] G. Bianconi, C. Rahmede, and Z. Wu. Complex quantum network geometries: Evolution and phase transitions. *Physical Review E*, 92(2):022815, 2015.
- [35] L. Biron. Our Flat Universe, Symmetry Magazine, available at <http://www.symmetrymagazine.org/article/april-2015/our-flat-universe>, accessed 21/3/2017.

- [36] T. Bläsius, T. Friedrich, and A. Krohmer. Efficient Embedding of Scale-Free Graphs in the Hyperbolic Plane. *European Symposium on Algorithms*, (16):1–16, 2016.
- [37] M. Boguna, F. Papadopoulos, and D. Krioukov. Sustaining the Internet with Hyperbolic Mapping. *Nature Communications*, 1(6):1–8, 2010.
- [38] B. Bollobás. Random Graphs. pages 215–252. Springer New York, 1998.
- [39] B. Bollobás. *Modern Graph Theory*. Graduate Texts in Mathematics. Springer New York, 2013.
- [40] B. Bollobás and G. Brightwell. Box-spaces and random partial orders. *Transactions of the American Mathematical Society*, 324(1):59–72, 1991.
- [41] T. Bolognesi and A. Lamb. Simple indicators for lorentzian causets. *Classical and Quantum Gravity*, 33(18):185004, 2016.
- [42] L. Bombelli, J. Lee, D. Meyer, and R. Sorkin. Space-time as a causal set. *Physical Review Letters*, 59(5):521–524, 1987.
- [43] L. Bombelli and D. Meyer. The origin of Lorentzian geometry. *Physics Letters A*, 141(5-6):226–228, 1989.
- [44] P. Bonacich. Some unique properties of eigenvector centrality. *Social Networks*, 29(4):555–564, Oct. 2007.
- [45] U. Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(1994):163–177, 2001.
- [46] U. Brandes and C. Pich. Eigensolver methods for progressive multidimensional scaling of large data. In *International Symposium on Graph Drawing*, pages 42–53. Springer, 2006.
- [47] T. J. Brazill and B. Grofman. Factor analysis versus multi-dimensional scaling: Binary choice roll-call voting and the US Supreme Court. *Social Networks*, 24(3):201–229, 2002.

- [48] G. Brightwell, F. Dowker, R. García, J. Henson, and R. D. Sorkin. Observables in causal set cosmology. *Physical Review D*, 67(8):084031, Apr. 2003.
- [49] G. Brightwell and R. Gregory. Structure of random discrete spacetime. *Physical Review Letters*, 66(3), 1991.
- [50] G. Brightwell, J. Henson, and S. Surya. A result in 2d causal set theory: the emergence of spacetime. *Journal of Physics: Conference Series*, 174:012049, June 2009.
- [51] G. Brightwell and M. Luczak. The mathematics of causal sets. In *Recent Trends in Combinatorics*, pages 339–365. Springer, 2016.
- [52] G. Brightwell and P. Winkler. Sphere orders. *Order*, 6(3):235–240, 1989.
- [53] T. Brooks. Evidence of complex citer motivations. *Journal of the American Society for Information Science*, 37(1):34–36, 1986.
- [54] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.
- [55] S. Chandler. The network structure of supreme court jurisprudence. *Public Law and Legal Theory Series*, 2005.
- [56] K. Christensen and N. R. Moloney. *Complexity and criticality*, volume 1. Imperial College Press, 2005.
- [57] V. Ciotti, M. Bonaventura, V. Nicosia, P. Panzarasa, and V. Latora. Homophily and missing links in citation networks. *EPJ Data Science*, 5(1):1–14, 2016.
- [58] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661–703, Nov. 2009.
- [59] J. Clough. *dagology* Python package, available at <https://github.com/JamesClough/dagology>.

- [60] J. R. Clough and T. S. Evans. Embedding graphs in lorentzian spacetime. *arXiv preprint arXiv:1602.03103*, 2016.
- [61] J. R. Clough and T. S. Evans. What is the dimension of citation space? *Physica A: Statistical Mechanics and its Applications*, 448:235–247, 2016.
- [62] J. R. Clough, J. Gollings, T. V. Loach, and T. S. Evans. Transitive reduction of citation networks. *Journal of Complex Networks*, 3(2):189–203, 2015.
- [63] J. Coon, C. P. Dettmann, and O. Georgiou. Full Connectivity: Corners, Edges and Faces. *Journal of Statistical Physics*, 147(4):758–778, 2012.
- [64] J. Coon, C. P. Dettmann, and O. Georgiou. Impact of boundaries on fully connected random geometric networks. *Physical Review E*, 85(1):1–5, 2012.
- [65] T. F. Cox and M. Cox. *Multidimensional Scaling, Second Edition*. CRC Press, 2000.
- [66] A. Cvetkovski and M. Crovella. Multidimensional scaling in the poincaré disk. *Applied Mathematics and Information Sciences*, 10(1):125–133, 2016.
- [67] J. Dall and M. Christensen. Random geometric graphs. *Physical Review E*, 66(1):016121, July 2002.
- [68] V. de Silva and J. B. Tenenbaum. Global Versus Local Methods in Nonlinear Dimensionality Reduction. *Advances in Neural Information Processing Systems 15*, pages 705–712, 2003.
- [69] K. R. Dienes. String theory and the path to unification: A review of recent developments. *Physics Reports*, 287(6):447–525, 1997.
- [70] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Ising model on networks with an arbitrary distribution of connections. *Physical Review E*, 66(1):1–5, 2002.
- [71] F. Dowker. Causal sets as discrete spacetime. *Contemporary Physics*, 47(1):1–9, 2006.

- [72] F. Dowker, J. Henson, and R. Sorkin. Quantum gravity phenomenology, Lorentz invariance and discreteness. *Modern Physics Letters A*, 19(24):1829–1840, 2004.
- [73] B. Dushnik and E. Miller. Partially Ordered Sets. *American Journal of Mathematics*, 63(3):600–610, 1941.
- [74] A. Eichhorn and S. Mizera. Spectral dimension in causal set quantum gravity. *Classical and Quantum Gravity*, 31(12):125007, 2014.
- [75] Y.-H. Eom and S. Fortunato. Characterizing and modeling citation dynamics. *PloS one*, 6(9):e24926, 2011.
- [76] T. Evans and R. Lambiotte. Line graphs, link partitions, and overlapping communities. *Physical Review E*, 80(1):016105, July 2009.
- [77] T. Evans and J. Saramäki. Scale-free networks from self-organization. *Physical Review E*, 72(2):026138, 2005.
- [78] P. Expert, T. Evans, V. D. Blondel, and R. Lambiotte. Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences of the United States of America*, 108(19):7663–8, 2011.
- [79] S. Felsner, P. C. Fishburn, and W. T. Trotter. Finite three dimensional partial orders which are not sphere orders. *Discrete Mathematics*, 201(1–3):101–132, 1999.
- [80] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3–5):75–174, 2010.
- [81] J. H. Fowler and S. Jeon. The authority of Supreme Court precedent. *Social Networks*, 30(1):16–30, 2008.
- [82] J. H. Fowler, T. R. Johnson, J. F. Spriggs, S. Jeon, and P. J. Wahlbeck. Network analysis and the law: Measuring the legal importance of precedents at the us supreme court. *Political Analysis*, pages 324–346, 2007.
- [83] L. C. Freeman. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1):35–41, 1977.

- [84] M. T. Gastner and M. E. J. Newman. The spatial structure of networks. *European Physical Journal B*, 49(2):247–252, 2006.
- [85] E. T. Gawlinski and H. E. Stanley. Continuum percolation in two dimensions: Monte Carlo tests of scaling and universality for non-interacting discs. *J. Phys. A: Math. Gen.*, 14(8):L291–L299, 1981.
- [86] D. George and R. Knegjens. Paperscape, available at, [paperscape.org](http://paperscape.org), accessed 21/3/2017.
- [87] A. P. Giles, O. Georgiou, and C. P. Dettmann. Betweenness centrality in dense random geometric networks. In *Communications (ICC), 2015 IEEE International Conference on*, pages 6450–6455. IEEE, 2015.
- [88] M. Girvan and M. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821–6, June 2002.
- [89] L. Glaser. Causal set actions in various dimensions. In *Journal of Physics: Conference Series*, volume 306, page 012041. IOP Publishing, 2011.
- [90] L. Glaser and S. Surya. Towards a definition of locality in a manifoldlike causal set. *Physical Review D*, 88(12):1–23, 2013.
- [91] S. R. Goldberg, H. Anthony, and T. S. Evans. Modelling citation networks. *Scientometrics*, 105(3):1577–1604, 2015.
- [92] M. L. Goldstein, S. A. Morris, and G. G. Yen. Problems with fitting to the power-law distribution. *European Physical Journal B*, 41(2):255–258, Sept. 2004.
- [93] M. Golosovsky and S. Solomon. Stochastic Dynamical Model of a Growing Citation Network Based on a Self-Exciting Point Process. *Physical Review Letters*, 109(9):098701, Aug. 2012.

- [94] A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes. K-core (bootstrap) percolation on complex networks: Critical phenomena and nonlocal effects. *Physical Review E*, 73(5):1–10, 2006.
- [95] M. Golumbic and E. Scheinerman. Containment Graphs, Posets, and Related Classes of Graphs. *Annals of the New York Academy of Sciences*, 555(1):192–204, May 1989.
- [96] G. Gonthier. Formal Proof – The Four-Color Theorem. *Notices of the AMS*, 2008.
- [97] S. S. Gubser, I. R. Klebanov, and A. M. Polyakov. Gauge Theory Correlators from Non-Critical String Theory. *Physical Letters B*, 428(May):15, 1998.
- [98] O. A. Guerrero and U. Matter. Revealing the Anatomy of Vote Trading. *SSRN*: <https://ssrn.com/abstract=2864445>, (Nov.), 2016.
- [99] A. Hackett, D. Cellai, S. Gómez, A. Arenas, and J. P. Gleeson. Bond Percolation on Multiplex Networks. *Physical Review X*, 6(2):021002, 2016.
- [100] B. Hall, A. Jaffe, and M. Trajtenberg. The NBER Patent Citations Data File. *NBER Working Paper Series*, 2001.
- [101] D. Hanneke, S. Fogwell, and G. Gabrielse. New measurement of the electron magnetic moment and the fine structure constant. *Physical Review Letters*, 100(12):2–5, 2008.
- [102] R. Hartshorne. *Geometry: Euclid and Beyond*, volume 51. Springer Science & Business Media, 2013.
- [103] S. W. Hawking and G. F. R. Ellis. *The large scale structure of space-time*. Cambridge University Press, 2008.
- [104] S. W. Hawking, A. R. King, and P. J. McCarthy. A new topology for curved space-time which incorporates the causal, differential, and conformal structures. *Journal of Mathematical Physics*, 17(2):174–181, 1975.

- [105] J. Henson, D. P. Rideout, R. D. Sorkin, and S. Surya. Onset of the asymptotic regime for finite orders. *arXiv preprint arXiv:1504.05902*, 2015.
- [106] D. J. Higham, M. Rasajski, and N. Przulj. Fitting a geometric graph to a protein-protein interaction network. *Bioinformatics*, 24(8):1093–1099, 2008.
- [107] G. E. Hinton and R. R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786):504–507, 2006.
- [108] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.*, 97(460):1090–1098, 2002.
- [109] P. Holme and J. Saramäki. Temporal networks. *Physics reports*, pages 1–28, 2012.
- [110] H.-K. Hwang. Asymptotic Expansions for the Stirling Numbers of the First Kind. *Journal of Combinatorial Theory, Series A*, 7141:1–7, 1994.
- [111] C. J. Isham. Canonical Quantum Gravity and the Problem of Time. In *Integrable Systems, Quantum Groups, and Quantum Field Theories*, pages 157–287. Springer Netherlands, Dordrecht, 1993.
- [112] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius. The complex network of global cargo ship movements. *Journal of the Royal Society*, 7(48):1093–1103, 2010.
- [113] B. Karrer and M. Newman. Random acyclic networks. *Physical Review Letters*, 128701(Mar.):1–4, 2009.
- [114] B. Karrer and M. Newman. Random graph models for directed acyclic networks. *Physical Review E*, 80(4):046110, Oct. 2009.
- [115] B. Karrer, M. E. Newman, and L. Zdeborová. Percolation on sparse networks. *Physical Review Letters*, 113(20):208702, 2014.
- [116] J. Kim, S. Lee, and H. Shin. Effective Task Scheduling for Embedded Systems Using Iterative Cluster Slack Optimization. 2013(Dec.):479–488, 2013.
- [117] H. J. Kimble. The quantum internet. *Nature*, 453(7198):1023–1030, 2008.

- [118] K.-K. Kleinberg, M. Boguñá, M. Á. Serrano, and F. Papadopoulos. Hidden geometric correlations in real multiplex networks. *Nature Physics*, 2016.
- [119] D. J. Kleitman and B. L. Rothschild. Asymptotic enumeration of partial orders on a finite set. *Transactions of the American Mathematical Society*, 205:205–220, 1975.
- [120] D. E. Knuth. Linear Probing and Graphs. *Algorithmica*, 22(4):561–568, Dec. 1998.
- [121] T. Konopka, F. Markopoulou, and S. Severini. Quantum graphity: A model of emergent locality. *Physical Review D*, 77(10):1–15, 2008.
- [122] D. Krioukov. Clustering means geometry in networks. *Bulletin of the American Physical Society*, 61, 2016.
- [123] D. Krioukov, M. Kitsak, R. S. Sinkovits, D. Rideout, D. Meyer, and M. Boguna. Network cosmology. *Scientific Reports*, 2:793, Jan. 2012.
- [124] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguna. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, Sept. 2010.
- [125] D. Krioukov, F. Papadopoulos, A. Vahdat, and M. Boguna. Curvature and temperature of complex networks. *Physical Review E*, 80(3):035101, Sept. 2009.
- [126] A. Landherr, D. B. Friedl, J. Heidemann, B. Friedl, and J. Heidemann. A Critical Review of Centrality Measures in Social Networks. *Business & Information Systems Engineering*, 2(6):371–385, 2010.
- [127] S. Lehmann, B. Lautrup, and A. Jackson. Citation networks in high energy physics. *Physical Review E*, 68(2):026113, 2003.
- [128] R. Liu, Z. Lin, Z. Su, and K. Tang. Feature extraction by learning Lorentzian metric tensor and its extensions. *Pattern Recognition*, 43(10):3298–3306, 2010.
- [129] C. W. Loe and H. J. Jensen. Comparison of communities detection algorithms for multiplex. *Physica A: Statistical Mechanics and its Applications*, 431:29–45, 2015.

- [130] P. Louridas, D. Spinellis, and V. Vlachos. Power laws in software. *ACM Transactions on Software Engineering and Methodology*, 18(1):1–26, Sept. 2008.
- [131] M. H. MacRoberts and B. R. MacRoberts. Problems of Citation Analysis. *Scientometrics*, 36(3):435–444, 1996.
- [132] D. B. Malament. The class of continuous timelike curves determines the topology of spacetime. *Journal of Mathematical Physics*, 18(7):1399, 1977.
- [133] J. Maldacena. The large N Limit of superconformal field theories and supergravity. *Advances in Theoretical and Mathematical Physics*, 2(2):231–252, 1998.
- [134] N. Masuda and R. Lambiotte. *A guide to temporal networks*. World Scientific, 2016.
- [135] A. S. Mata, M. Boguñá, C. Castellano, and R. Pastor-Satorras. Lifespan method as a tool to study criticality in absorbing-state phase transitions. *Physical Review E*, 91(5):1–8, 2015.
- [136] D. Mather. A framework for building spreadsheet based decision models. *Journal of the Operational Research Society*, 50(1):70–74, 1999.
- [137] B. D. McKay, F. E. Oggier, G. F. Royle, N. J. a. Sloane, I. M. Wanless, and H. S. Wilf. Acyclic digraphs and eigenvalues of (0, 1)-matrices. *Journal of Integer Sequences*, 7(3):1–5, 2004.
- [138] M. A. Méndez, P. Blasiak, and K. A. Penson. Combinatorial approach to generalized Bell and Stirling numbers and boson normal ordering problem. *Journal of Mathematical Physics*, 46(8):083511, 2005.
- [139] S. Mertens and C. Moore. Continuum percolation thresholds in two dimensions. *Physical Review E*, 86(6):061109, 2012.
- [140] D. Meyer. *The dimension of causal sets*. Phd thesis, MIT, 1988.
- [141] D. A. Meyer. Spherical containment and the minkowski dimension of partial orders. *Order*, 10(3):227–237, 1993.

- [142] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. Macmillan, 1973.
- [143] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6:161, 1995.
- [144] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela. Community Structure in Time-Dependent, Multiscale, and Multiplex Networks. *Science*, 328(5980):876–878, 2010.
- [145] J. Myrheim. Statistical Geometry. *Cds.Cern.Ch*, 2538, 1978.
- [146] M. Newman. *Networks: An Introduction*. Oxford University Press.
- [147] M. Newman. The structure and function of complex networks. *SIAM review*, 15(3):247–62, Sept. 2003.
- [148] M. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54, Jan. 2005.
- [149] M. Newman. Power laws, Pareto distributions and Zipf’s law. *Contemporary Physics*, 46(5):323–351, Sept. 2005.
- [150] M. Newman, S. Forrest, and J. Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, 66(3):035101, Sept. 2002.
- [151] M. E. J. Newman and D. J. Watts. Scaling and percolation in the small-world network model. *Physical Review E*, 60(6 Pt B):7332–7342, 1999.
- [152] V. Nicosia, G. Bianconi, V. Latora, and M. Barthelemy. Growing Multiplex Networks. *Physical Review Letters*, 111(5):058701, July 2013.
- [153] J. D. Noh and H. Rieger. Random Walks on Complex Networks. *Physical Review Letters*, 92(11):118701–1, 2004.
- [154] D. Oriti. *Approaches to quantum gravity: Toward a new understanding of space, time and matter*. Cambridge University Press, 2009.

- [155] F. Papadopoulos, R. Aldecoa, and D. Krioukov. Network geometry inference using common neighbors. *Physical Review E*, 92(2):67, 2015.
- [156] F. Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguna, and D. Krioukov. Popularity versus similarity in growing networks. *Nature*, 489(7417):537–40, 2012.
- [157] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63:066117, 2001.
- [158] M. Penrose. *Random geometric graphs*. Oxford University Press, 2003.
- [159] N. Perra and S. Fortunato. Spectral centrality measures in complex networks. *Physical Review E*, 78(3):036107, 2008.
- [160] D. Price. Networks of scientific papers. *Science*, 149(3683):510–515, 1965.
- [161] F. Radicchi. Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America*, 2008.
- [162] S. Redner. Citation statistics from 110 years of physical review. *Physics today*, 58(6):49–54, 2005.
- [163] J. Reichardt and S. Bornholdt. Detecting fuzzy community structures in complex networks with a Potts model. *Physical Review Letters*, 93(21):19–22, 2004.
- [164] D. D. Reid. Manifold dimension of a causal set: Tests in conformally flat space-times. *Physical Review D*, (Oct. 2002):1–8, 2003.
- [165] F. Ren, H. Shen, and X. Cheng. Modeling the clustering in citation networks. *Physica A: Statistical Mechanics and its Applications*, 2012.
- [166] D. Rideout and R. D. Sorkin. Classical sequential growth dynamics for causal sets. *Physical Review D*, 61(2):024002, 1999.
- [167] D. Rideout and P. Wallden. Spacelike distance from discrete causal order. *Classical and Quantum Gravity*, 155013:32, 2009.

- [168] G. Robins, P. Pattison, Y. Kalish, and D. Lusher. An introduction to exponential random graph ( $p^*$ ) models for social networks. *Social Networks*, 29(2):173–191, 2007.
- [169] L. E. C. Rocha and N. Masuda. Random walk centrality for temporal networks. *New Journal of Physics*, 16:1–27, 2014.
- [170] M. Roček and R. M. Williams. Quantum Regge calculus. *Physics Letters B*, 104(1):31–37, 1981.
- [171] C. Rovelli. Time in quantum gravity: An hypothesis. *Physical Review D*, 43(2):442–456, 1991.
- [172] C. Rovelli. Loop quantum gravity. *Living Reviews in Relativity*, 1(1):1, 1998.
- [173] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010.
- [174] G. Sabidussi. The centrality index of a graph. *Psychometrika*, 31(4):581–603, 1966.
- [175] J. Saramäki and P. Holme. Exploring temporal networks with greedy walks. *European Physical Journal B*, 88(12):1–8, 2015.
- [176] J. Saramäki and K. Kaski. Scale-free networks generated by random walkers. *Physica A: Statistical Mechanics and its Applications*, 2004.
- [177] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.
- [178] M. T. Schaub, J.-C. Delvenne, S. N. Yaliraki, and M. Barahona. Markov dynamics as a zooming lens for multiscale community detection: non clique-like communities and the field-of-view limit. *PloS one*, 7(2):e32210, 2012.
- [179] M. T. Schaub, R. Lambiotte, and M. Barahona. Encoding dynamics for multiscale community detection: Markov time sweeping for the map equation. *Physical Review E*, 86(2):026112, Aug. 2012.

- [180] E. R. Scheinerman and J. C. Wierman. On circle containment orders. *Order*, 4(4):315–318, 1988.
- [181] B. Schmitzer. Curvature And Topology On Causal Sets. 2010.
- [182] D. Schoch and U. Brandes. Re-conceptualizing centrality in social networks. *European Journal of Applied Mathematics*, 19(1):1–15, 2016.
- [183] S. Shortreed, M. S. Handcock, and P. Hoff. Positional estimation within a latent space model for networks. *Methodology*, 2(1):24–33, 2006.
- [184] P. Sibani and H. J. Jensen. *Stochastic Dynamics of Complex Systems: From Glasses to Evolution*. Imperial College Press, 2013.
- [185] M. Simkin and V. P. Roychowdhury. Read before you cite! *arXiv preprint cond-mat/0212043*, 2002.
- [186] M. Simkin and V. P. Roychowdhury. Do Copied Citations Create Renowned Papers? *Annals of Improbable Research*, 11(1):24–27, Jan. 2005.
- [187] M. Simkin and V. P. Roychowdhury. Stochastic modeling of citation slips. *Scientometrics*, 62(3):367–384, Mar. 2005.
- [188] L. Speidel, T. Takaguchi, and N. Masuda. Community detection in directed acyclic graphs. *arXiv preprint arXiv:1503.05641*, page 8, Mar. 2015.
- [189] O. Sporns. Contributions and challenges for network models in cognitive neuroscience. *Nature neuroscience*, 17(5):652–60, 2014.
- [190] E. Strano, V. Nicosia, V. Latora, S. Porta, and M. Barthelemy. Elementary processes governing the evolution of road networks. *Scientific Reports*, 2:296, 2012.
- [191] S. Surya. Directions in causal set quantum gravity. *arXiv preprint arXiv:1103.6272*, 2011.
- [192] R. Sverdlov and L. Bombelli. Gravity and matter in causal set theory. *Classical and Quantum Gravity*, 2009.

- [193] T. Tanizawa, G. Paul, R. Cohen, S. Havlin, and H. E. Stanley. Optimization of network robustness to waves of targeted and random attacks. *Physical Review E*, 71(4):047101, 2005.
- [194] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–23, 2000.
- [195] J. R. Ullmann. An Algorithm for Subgraph Isomorphism. *Journal of the ACM*, 23(1):31–42, 1976.
- [196] N. J. van Eck and L. Waltman. CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of Informetrics*, 8(4):802–823, 2014.
- [197] A. Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 8(1):32–39, 2012.
- [198] M. P. Viana, E. Strano, P. Bordin, and M. Barthelemy. The simplicity of planar networks. *Scientific Reports*, 3:3495, 2013.
- [199] M. von Looz, H. Meyerhenke, and R. Prutkin. Generating random hyperbolic graphs in subquadratic time. pages 467–478, 2015.
- [200] M. L. Wallace, V. Larivière, and Y. Gingras. Modeling a century of citation distributions. *Journal of Informetrics*, 3(4):296–303, 2009.
- [201] P. Wallden. Causal Sets Dynamics: Review & Outlook. *Journal of Physics: Conference Series*, 453:012023, 2013.
- [202] J. A. Walter. H-MDS: A new approach for interactive visualization with multi-dimensional scaling in the hyperbolic space. *Information Systems*, 29(4):273–292, 2004.
- [203] J. A. Walter and H. Ritter. On interactive visualization of high-dimensional data using the hyperbolic plane. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 123–132. ACM, 2002.

- [204] L. Wang, R. Liu, and Z. Su. Robust Visual Tracking via Incremental Lorentzian Discriminant Projection. *Journal of Information & Computational Sceince*, 9(7):1749–1759, 2012.
- [205] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(June):440–442, 1998.
- [206] S. Weinberg. *The quantum theory of fields*. Cambridge university press, 1995.
- [207] M. Wiedermann, J. F. Donges, J. Kurths, and R. V. Donner. Spatial network surrogates for disentangling complex system structure from spatial embedding of nodes. *Physical Review E*, 93(4):1–11, 2016.
- [208] H. S. Wilf. Backtrack: An O(1) expected time algorithm for the graph coloring problem. *Information Processing Letters*, 18(3):119–121, 1984.
- [209] H. S. Wilf. Asymptotic Expansions for the Stirling Numbers of the First Kind. *Journal of Combinatorial Theory, Series A*, 64:344–349, 1993.
- [210] P. Winkler. Connectedness and diameter for random orders of fixed dimension. *Order*, 2(2):165–171, 1985.
- [211] P. Winkler. Random orders. *Order*, 1:317–331, 1985.
- [212] E. Witten. Ground ring of two-dimensional string theory. *Nuclear Physics B*, 373(1):187–213, 1992.
- [213] M. Yannakakis. The complexity of the partial order dimension problem. *SIAM Journal on Algebraic Discrete Methods*, 3(3):351–358, 1982.
- [214] K. Zuev, O. Eisenberg, and D. Krioukov. Exponential Random Simplicial Complexes. *Journal of Physics A: Mathematical and Theoretical*, 48(46):21, 2015.

## Appendix A

# Degree after Transitive Reduction

How does the distribution of degrees after transitive reduction we see in real citation networks compare to causal set graphs? Here we will derive the distribution of degree after TR for  $D = 1 + 1$  dimensional causal sets, and also give the mean degree after TR for higher dimensions as derived in [211]. We will exploit the correspondence between a causal set in two-dimensional Minkowski space, and a two-dimensional cube space. Recall that in a cube space of dimension  $D$  we have points  $i = 1, 2, \dots, N$  with coordinates  $z_i^\alpha$  where  $\alpha = 1, 2, \dots, D$  which are uniformly distributed at random in the unit  $D$ -cube  $[0, 1]^D$ . Points  $i$  and  $j$  are related by  $i \prec j$  iff  $z_i^\alpha < z_j^\alpha \forall \alpha$ , meaning that the second point has larger coordinates in all dimensions. In  $D = 2$  we can call the two coordinates  $x_i$  and  $y_i$ . Suppose that the probability of a point in the  $[0, 0]$  corner of an interval containing  $N$  others, having a degree after TR  $k_r$  is  $p(k_r, N)$ . We will first give an argument for the following recursion relation:

$$p(k_r, N) = \frac{N-1}{N} p(k_r, N-1) + \frac{1}{N} p(k_r - 1, N-1) \quad (\text{A.1})$$

We may consider each point in turn ordering them with the largest  $x$  coordinates first for that  $x_i < x_j$  whenever  $i < j$ . Suppose we have already considered the first  $(N-1)$  points and we now look at the point with the  $N$ th largest  $x$  coordinate. This point can only have an edge after TR to the origin if it is minimal in the  $y$  coordinate. Since the coordinates are random and independent, point  $i$  has exactly a  $1/i$  chance of having

the smallest  $y$  coordinate so far, so with this probability a new edge appears, and with probability  $(i - 1)/i$  it does not, and these are the two terms in equation A.1. This view is equivalent to a standard record statistics process[184]. In fact, as long as the  $x$  and  $y$  coordinates are independent they do not even have to be uniformly distributed for this argument to hold. To find the probability distribution  $p(k_r, N)$  we recognise that this recursion relation is similar to that for the unsigned Stirling numbers of the first kind,

denoted by:  $\begin{bmatrix} N \\ k_r \end{bmatrix}$

$$\begin{bmatrix} N + 1 \\ k_r \end{bmatrix} = N \begin{bmatrix} N \\ k_r \end{bmatrix} + \begin{bmatrix} N \\ k_r - 1 \end{bmatrix} \quad (\text{A.2})$$

$$\text{where } \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 1 \quad (\text{A.3})$$

$$\text{and } \begin{bmatrix} N \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ N \end{bmatrix} = 0 \quad (\text{A.4})$$

We can then say that

$$p(k_r, N) = \frac{1}{N!} \begin{bmatrix} N \\ k_r \end{bmatrix} \quad (\text{A.5})$$

To check our answer, note that  $\begin{bmatrix} N \\ 1 \end{bmatrix} = (N - 1)!$  giving  $p(k_r = 1, N) = \frac{1}{N}$  as expected<sup>1</sup>. As noted by Wilf in [209], ‘the Stirling numbers of the first kind are notoriously difficult to compute’, and so we are unlikely to find a nice solution here.

It is useful to find the generating function  $G(z, N)$  where

$$G(z, N) = \sum_{k=0}^{\infty} z^k p(k, N) \quad (\text{A.6})$$

---

<sup>1</sup>The probability that  $k_r = 1$  is simply the probability that the point with the smallest  $x$ -coordinate also has the smallest  $y$ -coordinate

with  $p(k_r, N) = 0$  if  $k_r > N$ . Note that  $G(z=1, N) = 1$  and the first term in this polynomial is  $\frac{z}{N}$  because  $p(k=0, N) = 0$ . From the recursion relation Equation A.1 we now find that

$$G(z, N) = \frac{N-1}{N} G(z, N-1) + \frac{z}{N} G(z, N-1) \quad (\text{A.7})$$

$$G(z, N) = \frac{\Gamma(N+z)}{\Gamma(z)\Gamma(N+1)} = (z+N-1).(z+N-2)...(z) \times \frac{1}{N!}. \quad (\text{A.8})$$

Note that the  $\Gamma(z)$  normalisation factor on the denominator can be seen from the explicit expansion where we know the term  $O(z)$  is  $\frac{z}{N}$ .

The asymptotic limit [209, 110] can be studied from the generating function  $G(z, N)$  in A.8 as

$$\lim_{N \rightarrow \infty} G(z, N) = \frac{N^{z-1}}{\Gamma(z)} \quad (\text{A.9})$$

$$= \frac{1}{N} \frac{\sum_{k=0}^{\infty} (\ln(N))^k z^k / k!}{z^{-1} + \psi(0) + O(z)} \quad (\text{A.10})$$

The first term in the series, the part coming from the  $\Gamma(N+z)/\Gamma(z)$  is just the generating function for the Poisson distribution  $p_{\text{Poisson}}(k) = e^{-\lambda} \lambda^k / k!$  with mean  $\lambda = \ln(N)$ , divided by  $\Gamma(z)$ . However, the non-leading terms coming from the expansion of the denominator,  $\Gamma(z)$ , prevent a simple match so the Poisson-like behaviour as seen in [41] may only be useful for small ranges of  $k_r$ , typically  $|\Delta k_r| \ll \ln(N)$ .

From the generating function in A.8 we can find various moments of  $k_r$  for fixed  $N$ . In two dimensions we then have:

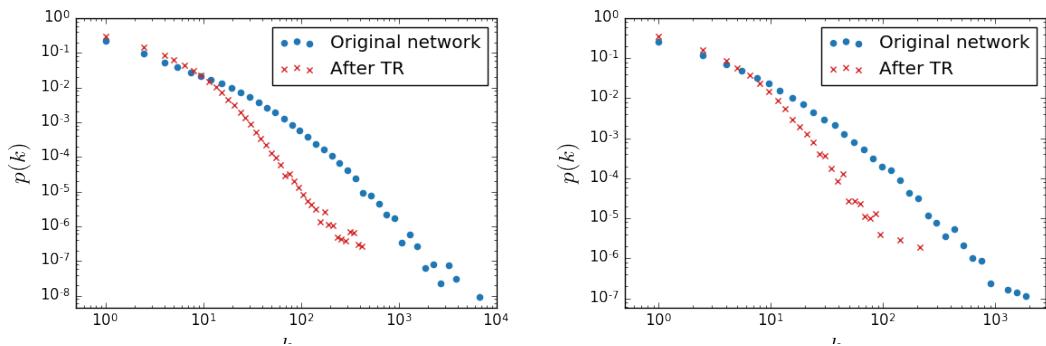
$$\langle k_r \rangle = \sum_{k_r=0}^{\infty} k_r p(k_r, N) = \left. \frac{\partial G(z, N)}{\partial z} \right|_{z=1} = \sum_{i=1}^N \frac{1}{i} = H_n \approx \gamma + \ln(N), \quad (\text{A.11})$$

where  $\gamma \approx 0.577$  is the Euler-Mascheroni constant. This Harmonic number result is the two-dimensional case of  $D$ -dimensional result in [211], which in the large  $N$  limit tends to logarithmic growth, as suggested in [74].

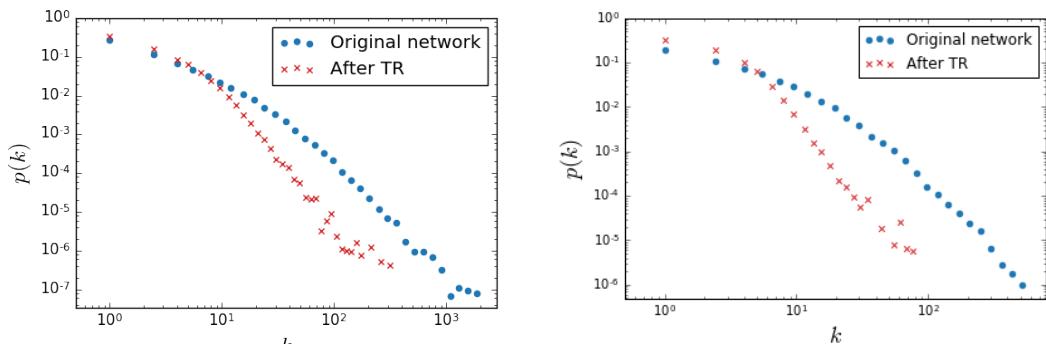
## Appendix B

# Transitive Reduction plots

In this appendix are plots of the degree distributions before and after transitive reduction for those datasets not shown in chapter 4.



**Figure B.1:** Degree distribution of the astro-ph (left) quant-ph (right) citation network before and after transitive reduction.



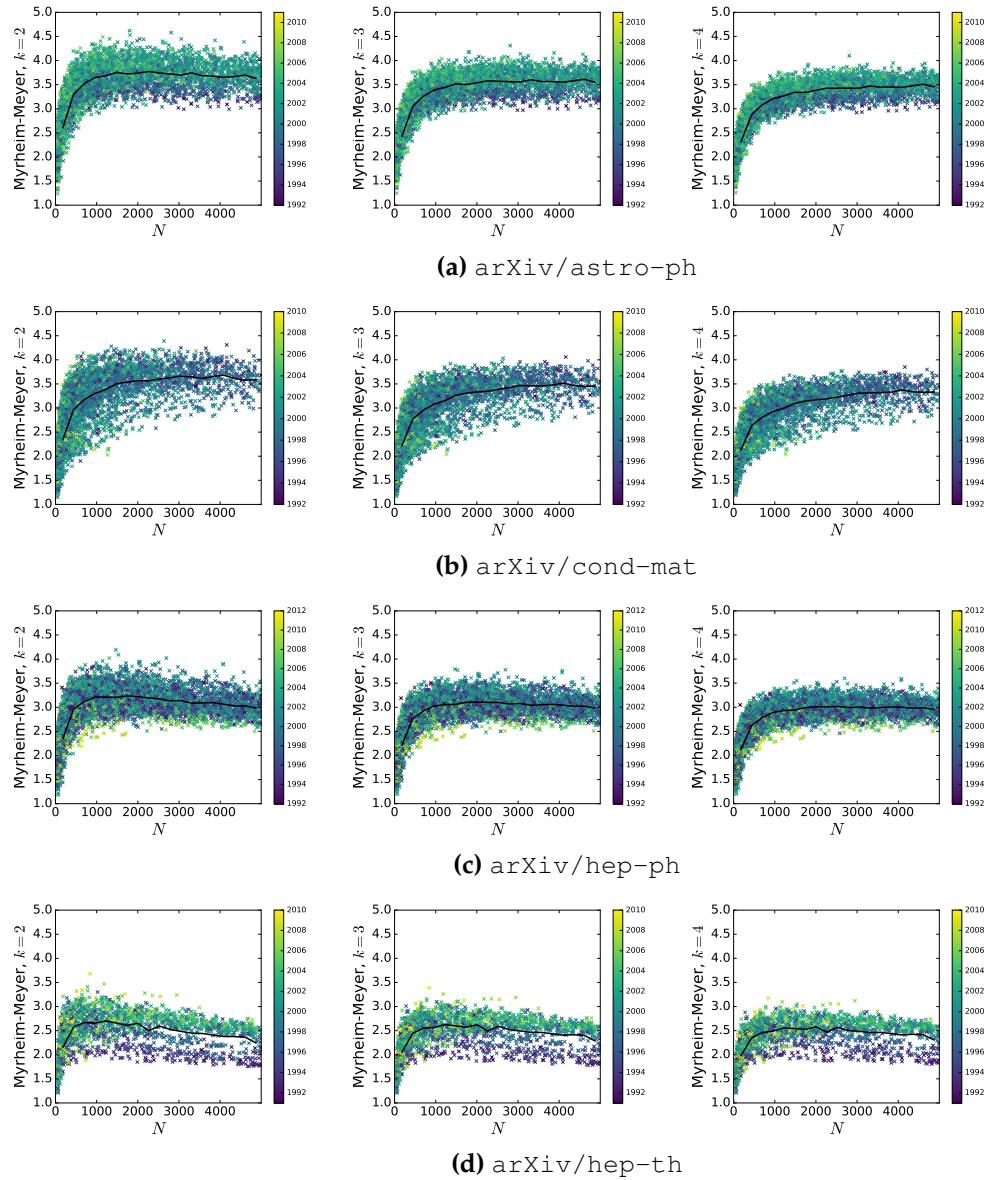
**Figure B.2:** Degree distribution of the cond-mat (left) gr-qc (right) citation network before and after transitive reduction.

## Appendix C

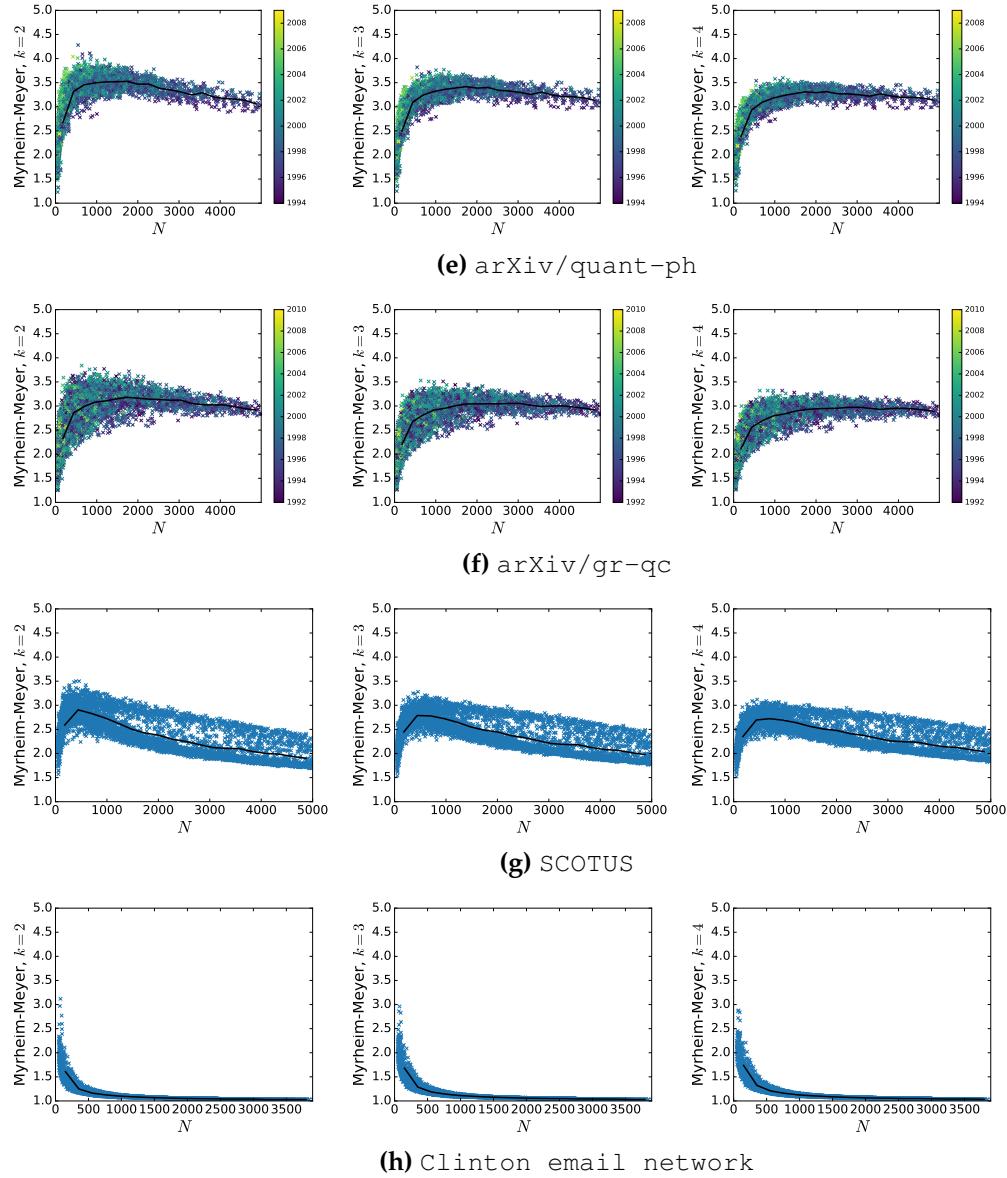
### Myrheim-Meyer Dimension for

$$k > 2$$

In chapter 5 when considering the Myrheim-Meyer dimension estimate, only the  $k = 2$  case was used, and it was claimed that estimating dimension using chains of more than 2 nodes gave very similar results. Those plots are given here. In each case note the similarity across a row of plots for the same network's Myrheim-Meyer dimension with  $k = 2, 3, 4$ . The similarity of these plots for each  $k$  explains why just using the simplest  $k = 2$  was sufficient in previous chapters.



**Figure C.1:** Myrheim-Meyer dimension estimates for real networks, for  $k = 2, 3, 4$ .



**Figure C.1:** Myrheim-Meyer dimension estimates for real networks, for  $k = 2, 3, 4$ .