Physica A 448 (2016) 235–247

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

What is the dimension of citation space?

James R. Clough *, Tim S. Evans

Centre for Complexity Science, Imperial College London, South Kensington campus, London, SW7 2AZ, UK

HIGHLIGHTS

- Causal constraints in network structure require the development of novel methods.
- Tools from causal set theory can be applied to study citation networks.
- This can reveal differences in citation behaviour between subfields.
- We measure the dimension of the spacetime each citation network is closest to.
- Other directed acyclic graphs can be studied with these methods.

ARTICLE INFO

Article history:
Received 24 April 2015
Received in revised form 24 September 2015
Available online 23 December 2015

Keywords:
Citation network
Directed acyclic graph
Minkowski space
Dimension
Causal set
Network geometry

ABSTRACT

Citation networks represent the flow of information between agents. They are constrained in time and so form directed acyclic graphs which have a causal structure. Here we provide novel quantitative methods to characterise that structure by adapting methods used in the causal set approach to quantum gravity by considering the networks to be embedded in a Minkowski spacetime and measuring its dimension using Myrheim–Meyer and Midpoint-scaling estimates. We illustrate these methods on citation networks from the arXiv, supreme court judgements from the USA, and patents and find that otherwise similar citation networks have measurably different dimensions. We suggest that these differences can be interpreted in terms of the level of diversity or narrowness in citation behaviour.

© 2016 Published by Elsevier B.V.

1. Introduction

Citation analysis has great potential to help researchers find useful academic papers [1], for inventors to find interesting patents [2], or for judges to discover relevant past judgements [3]. It is not, however, enough to simply count citations, because they can be made for a variety of reasons beyond an author genuinely finding a document useful [4–6]. To interpret the information encoded in a citation network we must be able to identify and describe the important features, such as the fat-tailed citation distributions [7–12], clustering [13], motifs [14], distance measures [15] and others [16,17]. Beginning with Price's cumulative advantage principle [18,19] there have been numerous attempts to construct models which replicate some of these features [9,20–26] and in doing so highlight potentially important mechanisms for the flow of information which drives the growth of real citation networks.

Citation networks are constrained in time, because authors can only cite something that has already been written.[1] This causal constraint prevents closed loops of directed edges in the graph, since all edges must point the same direction in time,

* Corresponding author.
E-mail address: jrc309@ic.ac.uk (J.R. Clough).

[1] Occasionally this is not the case for real citation networks. For instance, two authors may share and cite each others work before either is published, leading to two papers which both cite each other, clearly forming a cycle. Such 'acausal' edges are rare, making up less than 1% of edges in all citation networks considered here, and so were removed from the network since many techniques used here assume that the network forms a DAG.

http://dx.doi.org/10.1016/j.physa.2015.12.053
0378-4371/© 2016 Published by Elsevier B.V.

and is the same constraint placed on causally connected events in physics. Therefore they naturally form Directed Acyclic Graphs (DAG) where a directed edge going from node A to node B represents document A having cited the document B.

By taking constraints into account it is often possible to create new methods of characterising network structure as is well known for networks embedded in space [27–29]. This kind of geometric approach can also be used by postulating the existence of a hidden geometric space which describes some aspects of a network's structure [30–32]. It is our view that the same is true of citation networks (as well as other networks which naturally form directed acyclic graphs because their edges represent causal connections). The role of time in citation networks means that instead of an underlying Riemannian manifold (such as usual Euclidean space of spatial networks, or a Hyperbolic space of Ref. [30]) we will consider a Lorentzian manifold.

In a Lorentzian manifold one dimension, usually representing time is treated differently to the other, spatial dimensions. Relations between points can be classified as timelike, null, or spacelike. The simplest such manifold, and the one we will consider here is Minkowski space which is the geometry of Einstein's special relativity, in which timelike relations correspond to causal connections. We will consider applying this geometry to DAGs by equating the directed edges in the graph with timelike separation between the nodes. With this approach we will characterise network structure using tools from the causal set approach to quantum gravity, in which spacetime (which is a Lorentzian manifold) is discretised and has the structure of a DAG. In particular, we will use methods which estimate the dimension of a Minkowski space from its causal structure and apply them to citation networks.

The rest of this paper is structured as follows. In Section 2 we introduce in detail the Lorentzian perspective of DAGs, seeing how they can be embedded in space and time, and the methods of estimating their dimension. In Section 3 we will adapt these methods for use on citation networks and test them on examples from academic papers, patents and court judgements. In Section 4 we interpret these results in terms of using dimension as a measure of citation diversity, and finally in Section 5 we discuss applications and the similarities and differences of this approach with others in the literature.

2. Dimension estimates for spacetime networks

In the causal set approach to quantum gravity, spacetime is seen as a set of discrete points with a partial order relation, called a causal set whose structure approximates the continuous space we perceive. Whether causal sets are a useful approach to understanding the physical universe is not the focus of this paper and so we direct the reader to Refs. [33–35] for more details.

We will consider only the simplest spacetimes, D dimensional Minkowski spacetimes of one time dimension and $D-1$ spatial dimensions.[2] To create a causal set which approximates the structure of Minkowski space, we begin by randomly and uniformly scattering points in Minkowski space by randomly assigning each point an associated time t and spatial co-ordinates x_i. Two points are causally connected if and only if they are timelike separated, meaning the differences in their co-ordinates satisfy:

$$(\Delta t)^2 > \sum_i (\Delta x_i)^2. \tag{1}$$

If this relationship is satisfied we then say that the point with the larger/smaller t coordinate is in the future/past lightcone of the other. In special relativity it is this relationship that defines whether two events in spacetime can causally affect one another. The direction of the edges is determined by the causal/temporal ordering as given by the ordering of the time coordinates, and provides a uniquely defined causal relationship. To translate this structure into the language of networks, we say each point is a node, and we add edges between nodes which are causally connected, i.e. their coordinates satisfy Eq. (1). We will use the convention that all edges point backwards in time. This process necessarily generates a DAG since all edges point the same direction in time.

An **interval** [A, B] in a DAG is the set of nodes which can be reached from A (are in its causal past) in one direction, and from B in the other direction (in its causal future) [36] as in Fig. 1. The dimension estimates used here are defined on an interval in a DAG.

To illustrate these estimates we will use a simple network model in which points are scattered uniformly at random in an interval of Minkowski space. We first create two extremal points with time co-ordinates of 0, and 1 respectively, with all spatial co-ordinates equal to 0. We then add more nodes to the network by assigning a random time co-ordinate between 0 and 1, and random spatial co-ordinates between -0.5 and 0.5 and allowing this node to be in the network if it has edges to the two extremal nodes and so lie within the interval such that $G_D(N)$ is a network created by this process with N nodes, which are described by D coordinates. We will refer to these networks as **spacetime networks** though they are also known as cone spaces in the mathematics literature [37].

The number of spatial dimensions will determine the structure of the graph this process creates. Extra spatial dimensions add further terms to the summation on the right hand side of Eq. (1) and make it less likely that two points are connected.

---

2 It is possible to define other similar networks, such as a cube-space [41], a spacetime network using a curved spacetime [31] or geometric space with other rules [26]. Minkowski spacetime is the simplest, being defined by just one parameter D, the measurement of which we will use to characterise the network's structure.

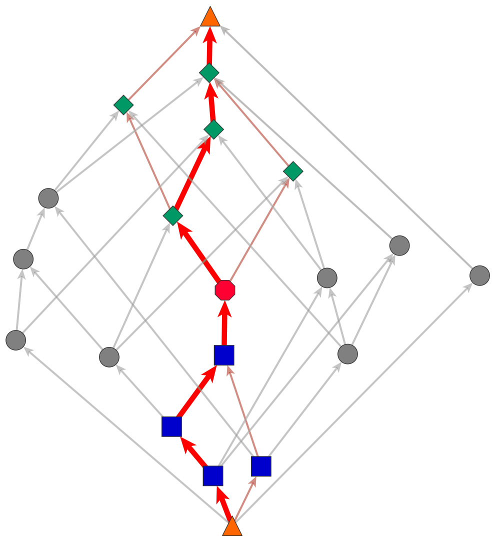

Fig. 1. The Midpoint-Scaling dimension in a 2-D spacetime network. The longest path through the interval (defined by the triangular nodes) is shown, and its midpoint is the octagonal node. The diamond nodes are those that lie within a subinterval, from the midpoint to the upper extremal point of the network, and the square nodes lie in the lower subinterval. In a 2D network we expect the number of nodes lying in these subintervals (the diamonds and squares) to be approximately half of the total population of the whole network.

So if we were to forget about the space and time coordinates of each point it should still be possible to estimate the number of spatial dimensions from the network's structure alone. We will use two such methods: the Myrheim–Meyer dimension estimate, and the Midpoint-Scaling dimension estimate.

2.1. Myrheim–Meyer dimension

An **n-chain** in a DAG is a sequence of n nodes which are all causally connected to each other. When points are placed at random with uniform probability density in a spacetime interval the expected number of n-chains S_n, is given by [38,39]

$$\langle S_n \rangle = \frac{N^n \Gamma(D/2)\Gamma(D)\Gamma(D+1)^{n-1}}{2^{n-1} n \Gamma(nD/2)\Gamma((n+1)D/2)} \tag{2}$$

where D is the dimension of the Minkowski spacetime and $\Gamma(z)$ is the Gamma function.

For a given DAG we can simply count the number of chains and numerically find an estimate for the dimension using this formula. Estimates for the dimension can be made using chains of any length, but there are significantly larger number of 2-chains (just a pair of causally connected points). Because of this we found the 2-chain case to be more accurate than others as well as being computationally easier. The expected number of 2-chains is simply

$$\frac{\langle S_2 \rangle}{N^2} \equiv f(D) = \frac{\Gamma(D+1)\Gamma(D/2)}{4\Gamma\left(\frac{3}{2}D\right)}. \tag{3}$$

For a given interval, the left hand side of this equation can be measured, and the right hand side, $f(D)$ is a monotonically decreasing function, so we can estimate D by inverting it numerically [40].

2.2. Midpoint-scaling dimension

When nodes are uniformly and randomly scattered in a space, the number of points in a region is proportional to the volume, V, of that region [39]. In a Minkowski space, the longest chain through an interval corresponds, in the large-n limit, to the geodesic through embedding spacetime (the Myrheim length conjecture[3]) [38,41,42]. This means that in an interval the length of the longest path, L is proportional to the time difference between the starting and ending nodes.[4] We then expect that in a D-dimensional Minkowski space $V(L) \sim L^D$. Knowing how the size of an interval scales with its height allows the dimension to be inferred.

3 It is proven in the case of flat spacetime.

4 The causal structure of Minkowski space and therefore the structure of the spacetime network is invariant under Lorentz transformations, so any interval can be transformed such that the starting and ending nodes have the same spatial coordinates, so the length of the geodesic in the continuum limit is simply the difference in the node's time coordinates.

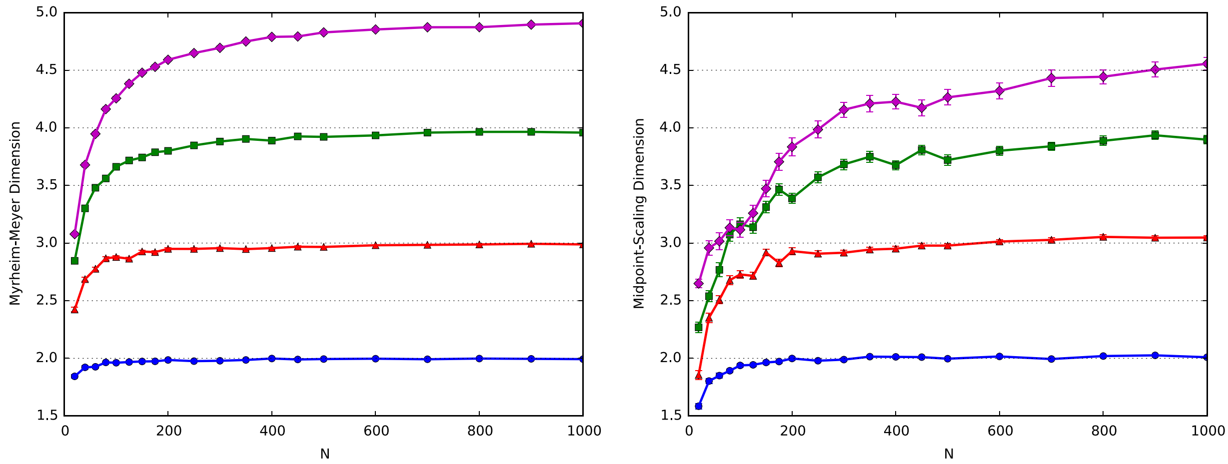

Fig. 2. The Myrheim–Meyer dimension (left column) and Midpoint-scaling dimension (right column), against the number of points in the graph, averaged over 100 random spacetime networks $G_D(N)$ for D equal to 2, 3, 4 and 5. The estimates' convergence from below to the correct dimension is also seen in Ref. [40]. Error bars (smaller than the markers on the left figure) show the standard error of the estimated dimension. Errors are larger for higher dimension, but smaller as the size of the space grows.

The Midpoint-Scaling dimension [34] measures how the size of two subintervals scale with the size of a larger interval between two nodes. The two subintervals of interval [A, B] are [A, C] and [C, B], which have populations N_1 and N_2. The midpoint, C, is the node on the longest path such that smaller population of N_1 and N_2 is maximised.

Since [A, C] and [C, B] each have around half the height of [A, B] we can estimate the manifold dimension of this interval using $N_1 \simeq N_2 \simeq \frac{N}{2^D}$. This is illustrated in Fig. 1 (see Fig. 2).

2.3. Reduced degree

It is also possible to estimate the dimension of a spacetime network by comparing the average in/out degree of a node before and after **Transitive Reduction** (TR). TR is an operation on directed graphs which removes all edges implied by transitivity, the result of which is uniquely defined if the graph is acyclic [43]. We will call the degree after TR the **reduced degree**, k_r. Taking Fig. 1 as an example, consider the triangular node at the bottom of the diagram. Its degree before TR is $N = 20$ and its degree after is $k_r = 4$ (as shown by the 4 remaining links).

We show in the Appendix that for a two-dimensional spacetime network the distribution of reduced degrees k_r is proportional to the unsigned Stirling numbers of the first kind. For large N the degree distribution is roughly Poissonian with a mean of $\ln(N)$. For other dimensions the expected reduced degree is roughly[5] $k_r = k^{\frac{D-2}{D}}$ [33]. However, we found that the dimension estimate given by this method does not display the consistency of the other two methods described here when used on citation networks. This is primarily because in a given citation network nodes which have the same degree can have reduced degrees which differ by more than an order of magnitude, as shown in the Appendix. In Ref. [44] we suggest that the reduced degree of a node reveals particular properties of the paper it represents, and given such variation this method is too noisy to use as a way of characterising the network as a whole.

3. Estimating the Minkowski dimension of citation networks

The methods described above are designed to estimate the dimension of the spacetime in which the nodes of a DAG are randomly scattered. Graphs which represent citation networks do not originate from points scattered in a Minkowski space, and so there is no original 'dimension' for us to estimate.

This means that the interpretation of the dimension estimate has to change. We are now no longer investigating the properties of a space in which we know the nodes are embedded, but instead just characterising the graph's structure in a way that is *analogous* to embedding it in some Minkowski spacetime. We do not claim that citation networks are like the spacetime networks described above in all regards, only that these tools are useful characterisations of different citation networks.

Some work is needed to adapt these dimension estimators for use on citation networks because citation networks do not share some of the particular properties of causal sets or the spacetime networks described above.

Firstly, the spacetime networks are constructed to be an interval, that is there is only one source node (with zero in-degree) and one sink node (with zero out-degree), both of which must be reachable from any other node in the network. This

5 We note that some versions of this formula in the literature contain typographical errors and are incorrect. Our analysis, numerical and analytical, gives the form we quote here.

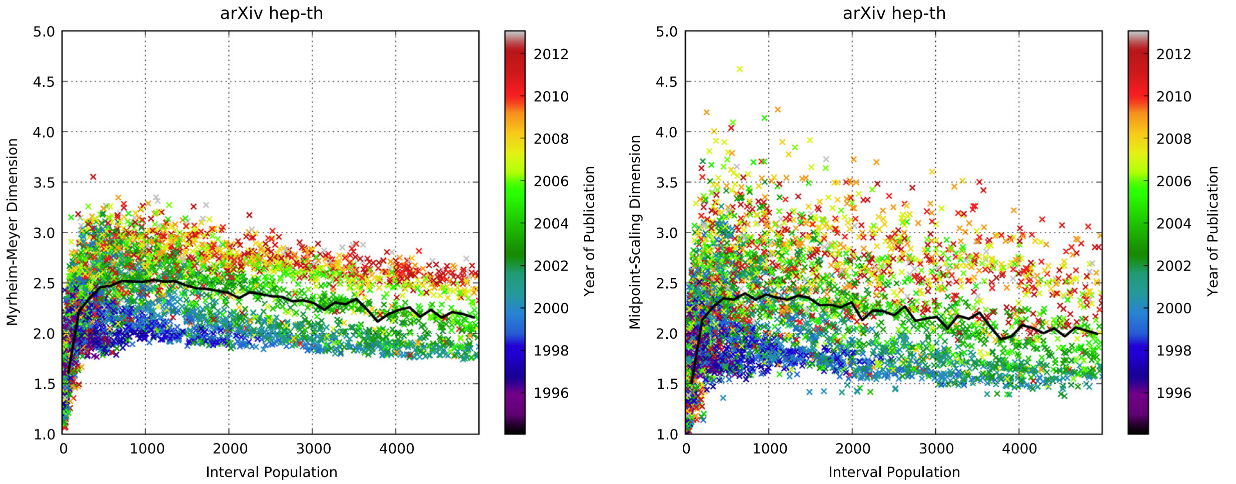

Fig. 3. The Myrheim–Meyer (left) and Midpoint-scaling (right) dimensions for the hep-th citation network appears to settle at a value around 2 for large intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

is almost always not the case for citation networks. So instead of estimating the dimension of the whole citation network, we look at many intervals within the citation network and apply the estimators to these intervals. To find intervals we choose two nodes uniformly at random from the network, and if an interval exists between them we estimate its dimension, otherwise ignoring this pair of nodes. To visualise the results we plot the population of the interval against its estimated dimension.

Secondly, the spacetime networks are always transitively complete meaning that if node A is in the future lightcone of B, and B is in the future lightcone of C then A is necessarily in the future lightcone of C. In the language of networks the edges (A, B) and (B, C) imply the edge (A, C) also exists. In citation networks this is not necessarily true since if an author cites a paper, they do not also have to cite its entire bibliography. A consequence of this is that there is no distinction in spacetime networks between edges and causal connections, but in citation networks they are different. So in our implementation of the Myrheim–Meyer dimension estimator we seek to count chains of causally connected nodes and not just edges. To do this we first transitively complete the network [43] (adding edges between any two nodes if there is a path between them) before counting the 2-chains which are now just the edges.

3.1. Data

To test these dimension estimates we used citation networks from academic papers, patents and court judgements. The academic citation networks are from subsections of the arXiv online research paper repository, from the citation network visualiser Paperscape. The citation network is separated out into the subsections of the arXiv, and each consists of the citations from one paper in that subsection to another also in that subsection. Here we will look at the 'high energy theory', 'high energy phenomenology', 'astrophysics', and 'quantum physics' sections, labelled by their tags on the arXiv, `hep-th`, `hep-ph`, `astro-ph`, `quant-ph` respectively. Their sizes range from around 20,000 to around 120,000 nodes and stretch in time from 1991 to 2013.

Since patents must cite other patents that contain 'prior art' they also form a citation network. We use data derived from patents registered in the USA between 1975 and 1999 [45] and in total there are around 4,000,000 patents.

Court decisions also cite previous decisions as precedent so form a citation network. We will analyse the network formed by all decisions and citations made by the US Supreme Court from its inception in 1754 to 2002 [46], in total around 25,000 nodes.

Further discussion of these particular datasets is available in our previous paper [44] and our datasets will be made available on figshare [47].

4. Results and interpretation

Figs. 3–6 show scatter plots for each arXiv section, plotting the population of 5000 randomly chosen intervals against their estimated dimension. Each point is coloured by the publication date of the last node in the interval.

It is immediately clear from the differing shapes of the histograms in Figs. 3–10 that there are structural differences in the citation networks analysed here which have been revealed by these dimension estimates. In general there is a large spread of measured dimensions suggesting structural heterogeneity unlike the homogeneous Minkowski space embedded networks. This is unsurprising given that citation networks show high levels of clustering and usually contain many different communities with strong intra-community links but weak inter-community links [48].

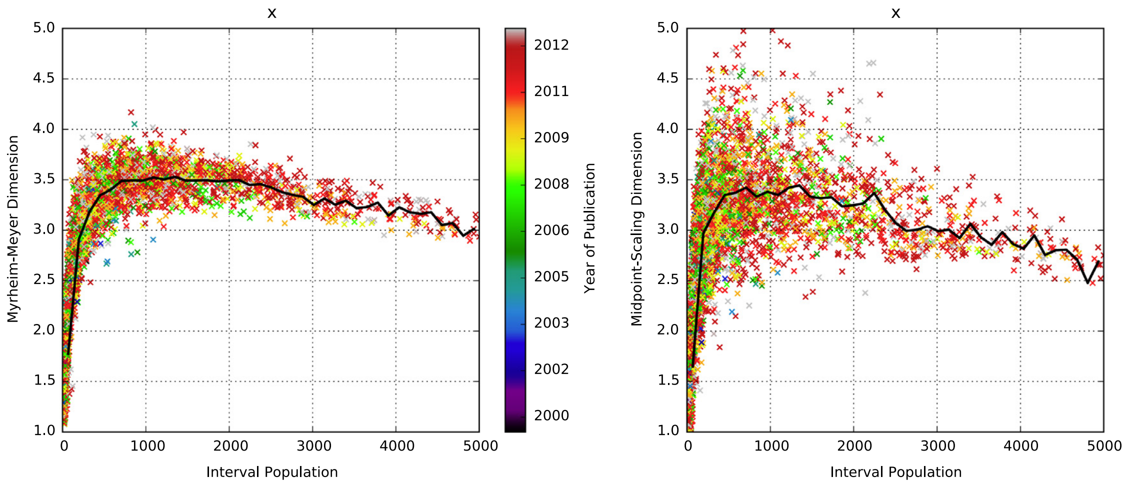

Fig. 4. The Myrheim–Meyer (left) and Midpoint-scaling (right) dimensions for the `quant-ph` citation network appears to settle at a value around 3 for large intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

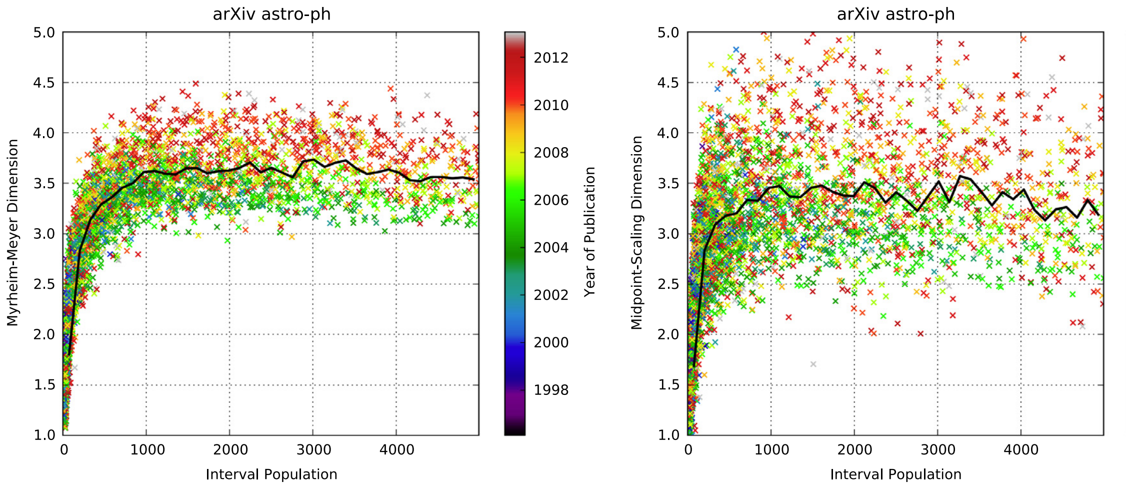

Fig. 5. The Myrheim–Meyer (left) and Midpoint-scaling (right) dimensions for the `astro-ph` citation network appears to settle at a value around 3.5 for large intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

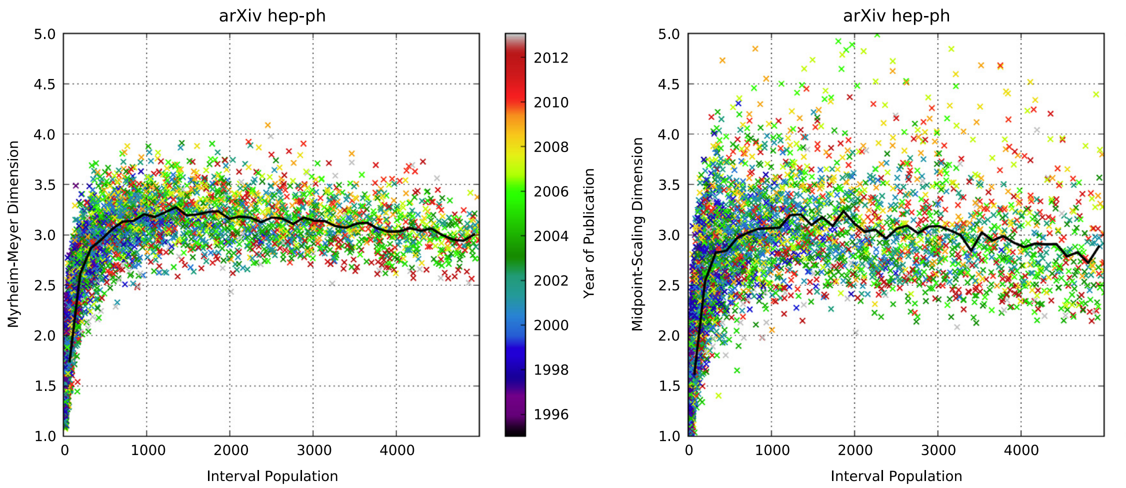

Fig. 6. The Myrheim–Meyer (left) and Midpoint-scaling (right) dimensions for the `hep-ph` citation network appears to settle at a value around 3 for large intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

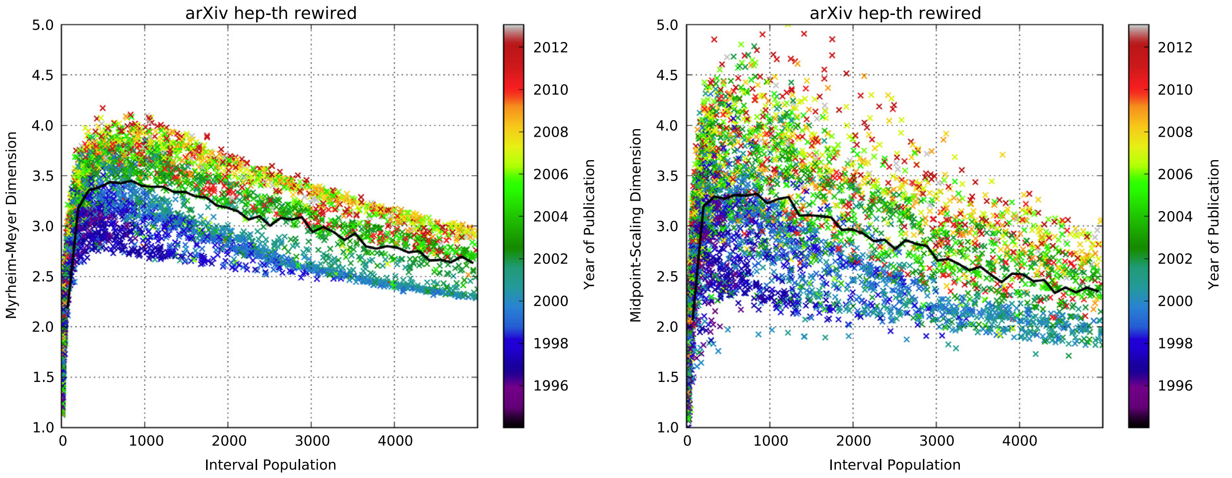

Fig. 7. The Myrheim–Meyer (left) and Midpoint-scaling (right) dimensions for the rewired `hep-th` citation network. The 10^5 edges have been rewired randomly 10^7 times, so all structure other than the degree distribution and causality constraints has been removed.

In all four arXiv citation networks the two plots, for Myrheim–Meyer dimension and Midpoint-Scaling dimension, show similar shapes, and converge on a consistent dimension value for large interval sizes. For the Minkowski spacetime networks there actually is an underlying dimension being estimated and so it is reasonable to expect independent estimations of it to agree. This is not obviously the case in other networks so it is encouraging to see consistency between the two methods in real social networks.

Crudely, the 'dimension' of the `hep-th` network appears to be around 2, and the `hep-ph` network around 3, `astro-ph` around 3.5, and `quant-ph` also around 3. We note that each of the individual arXiv citation networks, containing only intra-section links are themselves sub-networks of the larger arXiv citation network. Our estimated dimensions for these four sections show significant differences strongly suggesting that the arXiv citation network is structurally heterogeneous, with its different communities having measurably different citation behaviours. This suggests that an application of this technique would be to provide a novel method of measuring these differences in other large, heterogeneous citation networks, or DAGs representing other systems, and identifying relevant subgroups even without externally applied labelling of the nodes.

4.1. Similar causal constraints give similar structure

Citation networks are under causal constraints which impose some structure. We can see the effect of this structure by rewiring the edges of the network but maintaining the causal constraints. This is done by taking two edges, [A, B] and [C, D] and rewiring them to [A, D] and [C, B] if both of the new edges respect causality, thereby retaining the original in, and out degrees of each node and ensuring the network remains acyclic. Fig. 7 shows that after all structure other than the causal constraints, and in and out degree of each node in a network is removed, the dimension estimate plots give a significantly different result to the original `hep-th` network showing that other structure is involved in determining the estimated dimension and importantly that it is not an expected feature of a random network with the same degree distribution and causal constraints.

To further investigate the extent to which causal constraints and degree distribution we will determine estimated dimension with a simple null model.

We generate a network with the same degree distribution as the `quant-ph` network using the simple cumulative attachment model for citations[6] due to Price [19] and measure its dimension.

Fig. 8 shows that we can easily create a scale-free network with the same degree distribution as a citation network does not look the same according to these dimension estimates.

4.2. Differing citation behaviour

Citation networks from outside academia illustrate different behaviour. The US patent network's plot is much sparser for larger intervals, and almost all measured intervals were very small. For large intervals the measured dimension is around 5, much larger than the arXiv citation networks.

[6] We begin with a small number of nodes connected in a line. We add nodes one by one, and when a node is added it attaches $\langle k_{in} \rangle$ edges to existing nodes, where $\langle k_{in} \rangle$ is the mean in degree in the network whose degree distribution we are replicating. With probability p, edges attach preferentially, that is, proportionally to nodes according to their current in-degree, and with probability $1 - p$ they attach randomly. By manually tuning p, we can create a network with a very similar degree distribution to a real citation network. In this instance, $p = 0.6$.

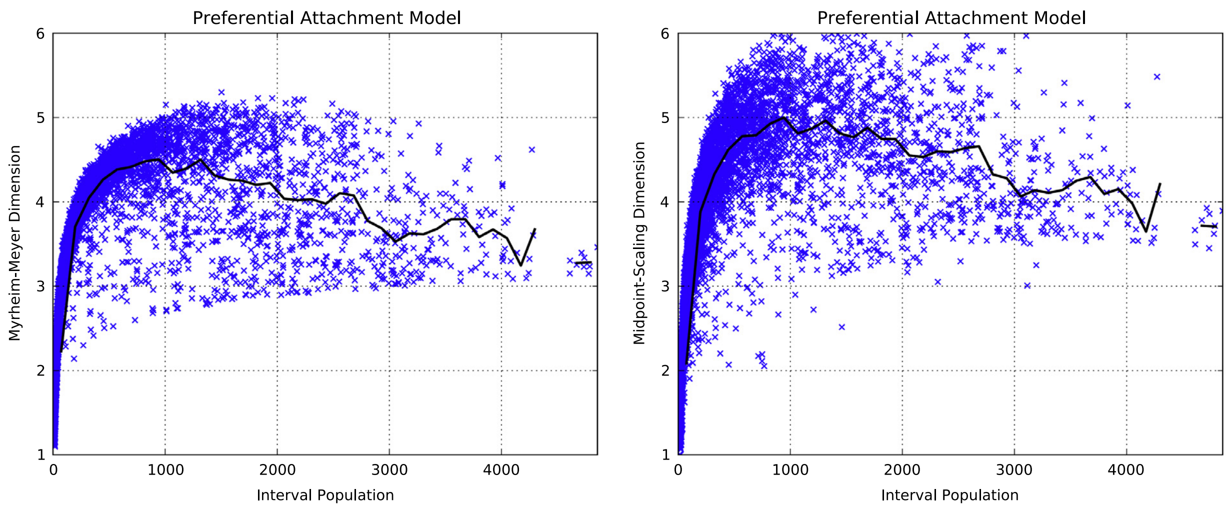

Fig. 8. The Myrheim–Meyer (left) and Midpoint-scaling (right) dimensions for the Price model network, with the same size and average degree as the `quant-ph` citation network. The spread of values is much larger than in the real citation network and the dimension estimate is higher, illustrating how these dimension estimates can show differences in structure.

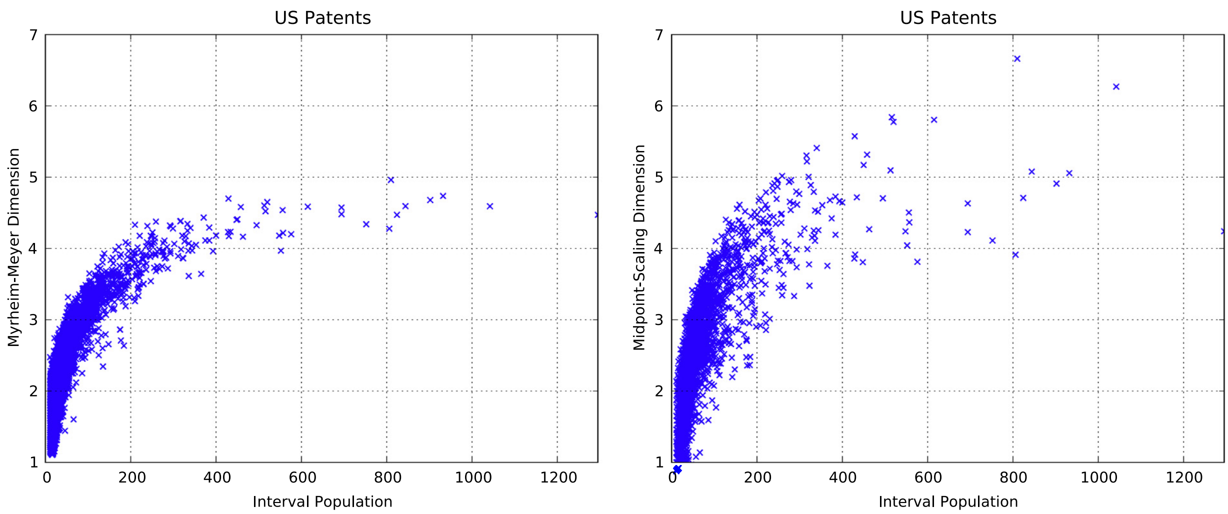

Fig. 9. The Myrheim–Meyer (left) and Midpoint-scaling (right) dimensions for the patent citation network. In this citation network larger intervals are much rarer than in the others, as a large interval usually contains many different paths from that starting node to the ending node, which is rare in patent citation networks.

Fig. 10 shows estimates for the US Supreme Court citation network and shows a different shape to all the others we investigated. In the arXiv and patent citation networks we see a slow growth in estimated dimension as interval size increases. The US Supreme Court network seems to show the opposite effect. Small intervals have a higher dimension estimate, and dimension falls as interval size increases. Our suggestion is that this effect is caused by this network stretching over an unusually long time period (it covers all judgements made in the Supreme Court since 1754). In the same way that a large, thin plane appears three-dimensional on length scales much smaller than the plane's thickness, but two dimensional on length scales much larger than this thickness, this network may also appear to have a different estimated dimension on different time scales. We also note that newer cases (those with a larger case number on the colour bar) have a lower measured dimension than older cases which is again the opposite trend to that in the arXiv citation networks. This temporal heterogeneity is another aspect of citation network structure which can be revealed by measuring dimension.

4.3. *Measuring citation diversity*

To find an interpretation for these results, we can look at simple examples of low, and high dimensional networks. A network with dimension 1 (no spatial dimensions, just a time dimension) would be a single line, where nodes form an unbroken chain of edges, each linking the previous node.

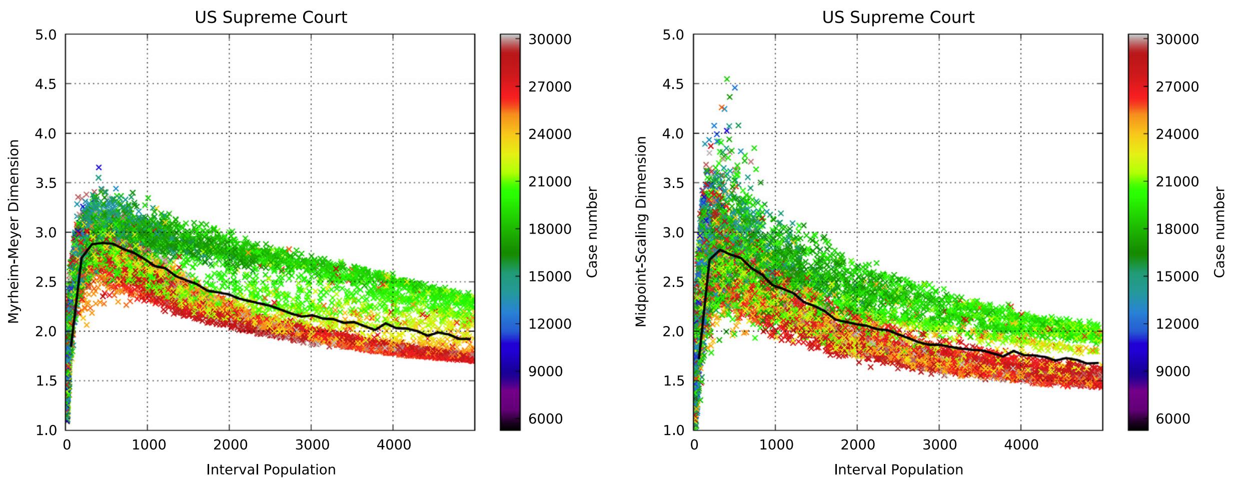

Fig. 10. The Myrheim–Meyer (left) and Midpoint-scaling (right) dimensions for the US Supreme Court citation network. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

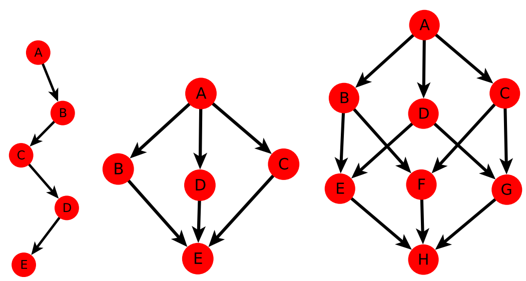

Fig. 11. From left to right, simple DAGs embeddable in 1, 2 and 3 dimensional Minkowski spacetimes.

As the number of spatial dimensions grow it is more likely that two nodes do not have a causal relationship, and so do not link to each other.[7] Furthermore in any spatio-temporal model of citation networks the number of spatial dimensions corresponds to the number of coordinates required to parametrise a paper. It is appealing then to interpret a small dimensional citation network as a more narrow field, where most of the papers are causally related to most others, and any paper can be described by a small number of parameters, corresponding to a small number of different areas of study (see Fig. 11).

A large dimension would then correspond to a more diverse field, where many independent authors can cite the same paper without citing each other and each paper requires a large number of parameters to be described. Our high-dimensional networks, such as the astrophysics section of the arXiv can then be interpreted as being more diverse in terms of citation behaviour than high energy physics section, and patents more diverse than physics papers or court judgements.

5. Discussion

In this paper we have illustrated the effectiveness of manifold dimension estimates as novel ways of characterising networks which form DAGs, and in particular networks constrained by causality. We have shown that in citation networks, the Midpoint-Scaling dimension and Myrheim–Meyer dimension estimators show strong agreement and highlight important differences in the causal structure.

We see a number of possible applications and uses of this technique in network analysis. Firstly, they provide another method of quantitatively characterising structure in a way that allows different networks to be differentiated. For instance the two particle physics sections of arXiv, `hep-th` and `hep-ph`, are similar in many ways. They have similar degree distributions and clustering coefficients [44] but clearly differ in the dimension measures which quantify how ‘broad’ or ‘narrow’ the citation behaviour of authors in these fields is. Given two intervals, one from the `hep-th` network, and the other from `hep-ph` we can estimate their dimensions using these methods and we could deduce which section of the arXiv those papers came from, without knowing anything about their authors or their content, and using only the information in the topology of the citation network. The message here is that citation networks from different areas of study have measurably

7 In the language of causal sets, the *ordering fraction* decreases.

different citation behaviours, and that this is information we can extract which is potentially of interest to any scientists who want to improve their use of bibliometric measures as an aid to research.

Secondly there has been recent interest in embedding networks derived from real data in 'hidden metric spaces', such as in Ref. [49]. A program to find hidden Lorentzian spaces to embed citation networks (and other DAGs) in could yield similarly useful applications for citation analysis such as showing the evolution of communities over time, or providing a distance/similarity measure between individual documents which can be used for paper recommendation. Before the nodes of a network can be assigned coordinates in a space, it is necessary to know how many dimensions that space has. Therefore, analysis like ours is a necessary first step on this program.

Our approach here has been to view the citation network as a static object, which is a natural view if it is embedded in a spacetime where time is just another coordinate each node has. It is possible to view the citation network with a growing model instead, but in many cases including random geometric graphs and causal sets, static equilibrium models and growth approaches are equivalent as shown in Ref. [50]. If viewed from a dynamic growth perspective, our dimensions estimates can be considered as characterising the dimensionality of the available space for nodes to appear in that growth process.

Though a citation network represents information spreading process between authors, not every citation necessarily corresponds to useful information being passed from one writer to another. It is difficult to know from the structure of the citation network alone which edges correspond to genuine insight and which are citations inserted for other reasons. In Ref. [44] it is explained that transitive edges (those implied by some other directed path) are less likely to be necessary for information flow between two authors and the effect of their removal is investigated. The methods for dimension estimation shown here are effectively invariant with respect to the existence of transitive edges since only the existence of a path from one node to another is relevant and so it is therefore reasonable to suggest that they depend primarily on the information carrying edges of the network.

5.1. *Why use Minkowski space?*

The dimension of a network is a concept which has been considered before, with differing meanings; for example the number of parameters in a model [51], the time a random walker takes to return to its starting position [52], fractal box counting dimension in Euclidean space [53], the number spatial (but not temporal) dimensions measured using influence regions [54] and more [55].

However causal constraints are a key feature of citation networks and DAGs and it is essential that such a constraint is taken into account when analysing these networks [44]. This is why it was important to develop methods which include time as dimension which is not the same as the spatial dimensions. The dimension measures we use explicitly involve time and take the causal constraints of citation networks into account, recognising that in-edges (being cited) and out-edges (citing someone else) are fundamentally different things. While in a purely spatial embedding of a network, one might reasonably suggest that the probability of an edge existing between two nodes will always increase if the distance between them decreases, this is not necessarily the case in time. The information in one paper takes time to reach other researchers, and this delay is larger if researchers work in separate field.

In the arXiv citation network, we found that the time difference of a citation between two papers in the same arXiv subsection was 1.6 years, and for papers in different Section 2.1 years, suggesting that citations to papers which are less similar (or further away in some abstract spatial representation) are more likely to span larger time intervals, reflecting the finite speed of information propagation.

The primary reason for our choice of Minkowski spacetime for these dimension estimates over some other Lorentzian spacetime is that it is one of the simplest choices to make, defined by a single parameter D. The scope of this work was to characterise different network structures by measuring spacetime dimension and not to find an actual embedding of the network into that spacetime by assigning coordinates to each point.

This would be much more difficult to do for spacetimes which are described by more than just one parameter (e.g. for curved spacetimes). Regarding studies of networks embedded in other types of space, for causal sets we refer the reader to Refs. [35,40,56] and for growing network models in hyperbolic spaces to Refs. [31,57].

There are suggestions in the literature, that it may be possible to find better correspondences between particular networks and spaces. The de Sitter spacetime studied in Ref. [31] can give the same fat-tailed citation distribution as the simple Price model [18] but the development of other network measures is needed before we can truly say that there is a correspondence between citation networks and these kinds of spaces. Indeed, using other causally aware measures on citation data reveals important new features in real citation network data [44,25] which are not present in generic preferential attachment models.

One particular feature noted in Ref. [44] is that only small number of papers published shortly before the referencing paper are needed to define the causal structure of a real citation network, the edges left after transitive reduction.

We note that other dimension estimators exist for networks embedded in Minkowski spaces. We have tried using the reduced degree, but as explained in Appendix it is not a feasible method. A recently published method not implemented here, but potentially appropriate for citation networks uses a random walker on a causal set to estimate it is spectral dimension [58]. However many other methods are inappropriate for analysis of the causal structure of a citation network. One such method is to find the smallest dimension in which any subgraph can be faithfully embedded [33,59]. This method requires a DAG to be perfectly embedded in a manifold as just one subgraph which cannot be embeddable in D dimensions means that the entire network is given a dimension higher than D.

Furthermore, there are some finite DAGs which cannot be perfectly embedded in a Minkowski space of any dimension [60]. This method is less appropriate for analysis of our citation networks, since the integer result such dimension estimates give can be increased by one by the rewiring of only one edge, which is an unhelpful property when dealing with noisy real-world data, or it may not even be defined at all.

Conversely the two estimates we use here are robust to noise, a useful property when analysing data from social interactions. They produce a real number value and small deviations from DAGs which are faithfully embedded in a Minkowski manifold only lead to small deviations in the estimated dimension.

Acknowledgements

We would like to thank Jamie Gollings, Amy Hughes, Tamar Loach, Kim Christensen, Mauricio Barahona and Nick Jones for useful discussions. We also like to thank Rob Knegjens for providing us with a copy of the citation network data from Paperscape.

Appendix. Reduced degree

Derivation of distribution of reduced degree in 2D spacetime networks

In 2D Minkowski space the exact distribution of degrees after TR can be calculated because in 2 dimensions the spacetime network is equivalent to another structure called a cube space. In a cube space of dimension D we have points $i = 1, 2 \dots N$ with coordinates z_i^α where $\alpha = 1, 2 \dots D$. Point i is connected to point j iff $z_i^\alpha < z_j^\alpha \ \forall \ \alpha$, which is to say that j has larger co-ordinates in all dimensions. The nodes of the network are randomly and uniformly scattered in this space and connected using this rule.

Though this result will turn out to be of little help for analysis of citation networks, for reasons discussed later, it is to the best of our knowledge a novel extension of the result of Ref. [61] which derives the expected value of the reduced degree in cube spaces of all dimensions and so we include it here.

Suppose that the probability of a node in the corner of an interval containing N other nodes, having a degree after TR (reduced degree) k_r is $p(k_r, N)$. We will first give an argument for the following recursion relation.

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

Since we are only considering 2 dimensions, let us call the first coordinate x, $z_i^{\alpha=1} = x_i$, and the second coordinate y, $z_i^{\alpha=2} = y_i$. We may consider each point in turn, ordering them with smallest x coordinates first so that $x_i < x_j$ if $i < j$. Suppose we have already considered the first $(N-1)$ points and now look at the point with the N-th smallest x coordinate, $i = N$. All previously considered points have smaller x coordinates, so this point $i = N$ can only be a new link to the origin if it is minimal in the y coordinate. Because the coordinates are just random numbers, the probability that y_N is the smallest so far is simply $\frac{1}{N}$. So with this probability, a new TR-surviving-edge will appear, and with probability $\frac{N}{N-1}$ it will not, explaining both terms in Eq. (A.1). This view is equivalent to a standard record statistics process [62]. Indeed the points do not even have to be uniformly distributed here, the only requirement is that the D coordinates are independent random variables.

To solve this, we then recognise the recursion relation for the unsigned Stirling numbers of the first kind $\begin{bmatrix} N \\ k_r \end{bmatrix}$,namely

$$\begin{bmatrix} N+1 \\ k_r \end{bmatrix} = N \begin{bmatrix} N \\ k_r \end{bmatrix} + \begin{bmatrix} N \\ k_r - 1 \end{bmatrix} \tag{A.2}$$

$$\text{where } \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 1 \tag{A.3}$$

$$\text{and } \begin{bmatrix} N \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ N \end{bmatrix} = 0. \tag{A.4}$$

We can then say that

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

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.[8] As noted by Wilf in Ref. [63], 'the Stirling numbers of the first kind are notoriously difficult to compute', and so we are unlikely to find a nice solution here.

8 The probability that $k_r = 1$ is simply the probability that the point with the smallest x-coordinate also has the smallest y-coordinate.

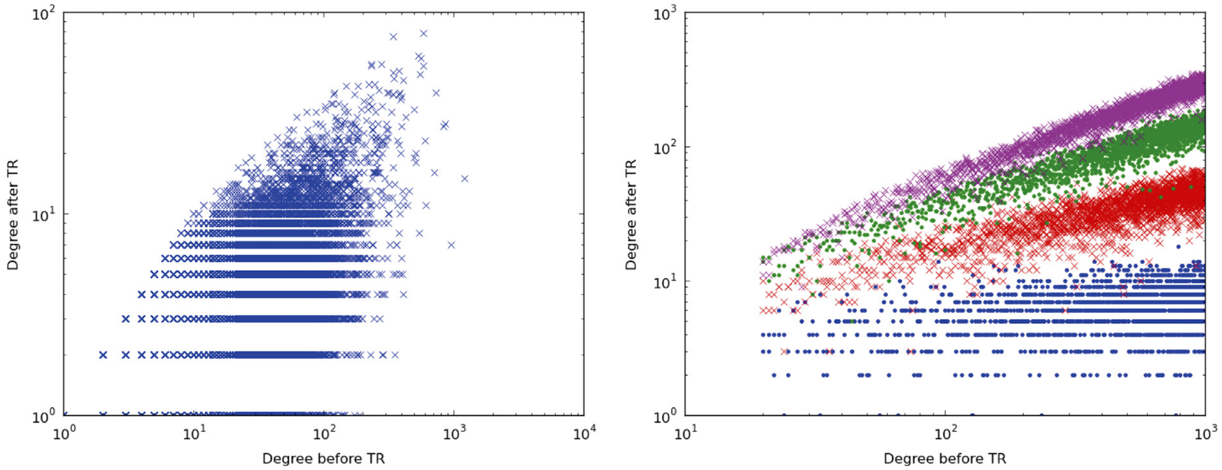

Fig. A.12. Left: the degree before, and after transitive reduction for the `hep-ph` citation network. The spread of k_r is very wide for a given k_r, indicating a heterogeneity in the papers. Right: the degree before, and after transitive reduction for spacetime networks of dimension 2–5. Lower dimension appears lower on the plot. To try and use the reduced degree method to estimate dimension is essentially to ask which of the scatter plots on the right figure best fits the left figure, given the large spread of values on the left, the estimated dimension for individual subgraphs has very large variation, unlike the other dimension estimates which have similar answers throughout the network and so better achieve the goal of characterising the whole network's structure. The reduced degree method is more useful as a characterisation of individual nodes within the network.

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

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

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 (A.1) we now find that

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

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

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 [63,64] can be studied from the generating function $G(z, N)$ in (A.8) as

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

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

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 Ref. [56] 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. Here we will derive the expected k_r for a given N in two-dimensions.

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

where $\gamma \approx 0.577$ is the Euler–Mascheroni constant. This Harmonic number result is the two-dimensional case of D-dimensional result in Ref. [61], which in the large N limit tends to logarithmic growth, as suggested in Ref. [58].

Comparison of measured reduced degree in citation networks, and spacetime networks

See Fig. A.12.

References

[1] A.W. Harzing, Publish or Perish, Tarma Software Research, 2010.
[2] A. Abbas, L. Zhang, S.U. Khan, A literature review on the state-of-the-art in patent analysis, World Pat. Inf. 37 (2014) 3–13.
[3] P.W. Martin, Introduction to Basic Legal Citation, Legal Information Institute, 2007.
[4] T.A. Brooks, Evidence of complex citer motivations, J. Am. Soc. Inf. Sci. 37 (1986) 34–36.
[5] M. Simkin, V. Roychowdhury, Read before you cite!, Complex Syst. 14 (2003) 269–274.
[6] L. Bornmann, H.D. Daniel, What do citation counts measure? A review of studies on citing behavior, J. Doc. 64 (2008) 45–80.
[7] S. Redner, How popular is your paper? An empirical study of the citation distribution, Eur. Phys. J. B 134 (1998) 131–134.
[8] C. Tsallis, M. de Albuquerque, Are citations of scientific papers a case of nonextensivity ? Phys. J. B (2000).
[9] M.V. Simkin, V.P. Roychowdhury, Stochastic modeling of citation slips, Scientometrics 62 (3) (2005) 367–384.
[10] F. Radicchi, S. Fortunato, C. Castellano, Universality of citation distributions: Toward an objective measure of scientific impact, Proc. Natl. Acad. Sci. 105 (2008) 17268–17272.
[11] Y.-H. Eom, S. Fortunato, Characterizing and modeling citation dynamics, PLoS One 6 (2011) e24926.
[12] T.S. Evans, N. Hopkins, B.S. Kaube, Universality of performance indicators based on citation and reference counts, Scientometrics 93 (2012) 473–495. http://dx.doi.org/10.1007/s11192-012-0694-9.
[13] F.-X. Ren, H.-W. Shen, X.-Q. Cheng, Modeling the clustering in citation networks, Physica A 391 (2012) 3533–3539. http://dx.doi.org/10.1016/j.physa.2012.02.001.
[14] C.J. Carstens, A uniform random graph model for directed acyclic networks and its effect on motif-finding, J. Complex Netw. 2 (2014) 419–430. http://dx.doi.org/10.1093/comnet/cnu028.
[15] M.J. Bommarito, D.M. Katz, J.L. Zelner, J.H. Fowler, Distance measures for dynamic citation networks, Physica A 389 (2010) 4201–4208. http://dx.doi.org/10.1016/j.physa.2010.06.003.
[16] D. Lucio-Arias, L. Leydesdorff, Main-path analysis and path-dependent transitions in HistCite-based historiograms, J. Am. Soc. Inf. Sci. 59 (2008) 1948–1962. http://dx.doi.org/10.1002/asi.20903.
[17] L. Šubelj, D. Fiala, M. Bajec, Network-based statistical comparison of citation topology of bibliographic databases, Sci. Rep. 4 (2014) 6496. http://dx.doi.org/10.1038/srep06496.
[18] D.S. Price, Networks of scientific papers, Science 149 (1965) 510–515.
[19] D.S. Price, A general theory of bibliometric and other cumulative advantage processes, J. Am. Soc. Inf. Sci. 27 (1976) 292–306.
[20] K. Börner, J.T. Maru, R.L. Goldstone, The simultaneous evolution of author and paper networks, PNAS 101 (2004) 5266–5527. http://dx.doi.org/10.1073/pnas.0307625100.
[21] M.V. Simkin, V.P. Roychowdhury, A mathematical theory of citing, J. Am. Soc. Inf. Sci. 58 (2007) 1661–1673. http://dx.doi.org/10.1002/asi.20653.
[22] Z.-X. Wu, P. Holme, Modeling scientific-citation patterns and other triangle-rich acyclic networks, Phys. Rev. E 80 (2009) 80. http://dx.doi.org/10.1103/physreve.80.037101.
[23] D.R. Amancio, O.N. Oliveira, L. da Fontoura Costa, Three-feature model to reproduce the topology of citation networks and the effects from authors' visibility on their h-index, J. Informetr. 6 (2013) 427–434. http://dx.doi.org/10.1016/j.joi.2012.02.005.
[24] Y. Wu, T.Z.J. Fu, D.M. Chiu, Generalized preferential attachment considering ageing, J. Informetr. 8 (2014) 650–658. http://dx.doi.org/10.1016/j.joi.2014.06.002.
[25] S.R. Goldberg, H. Anthony, T.S. Evans, Modelling citation networks, arXiv:1408.2970, 2014, http://dx.doi.org/10.1007/s11192-015-1737-9.
[26] Z. Xie, Z. Ouyang, P. Zhang, D. Yi, D. Kong, Modeling the Citation Network by Network Cosmology, PLoS One 10 (2015) e0120687. http://dx.doi.org/10.1371/journal.pone.0120687.
[27] M. Barthelemy, Spatial networks, Phys. Rep. 499 (2011) 1–101.
[28] P. Expert, T. Evans, V.D. Blondel, R. Lambiotte, Uncovering space-independent communities in spatial networks, Proc. Natl. Acad. Sci. USA 108 (2011) 7663–7668.
[29] L. Daqing, K. Kosmidis, A. Bunde, S. Havlin, Dimension of spatially embedded networks, Nat. Phys. 7 (2011) 481–484.
[30] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, M. Boguna, Hyperbolic geometry of complex networks, Phys. Rev. E 82 (2010) 036106.
[31] D. Krioukov, M. Kitsak, R.S. Sinkovits, D. Rideout, D. Meyer, M. Boguñá, Network cosmology, Sci. Rep. 2 (2012) 793.
[32] Z. Xie, J. Zhu, D. Kong, J. Li, A random geometric graph built on a time-varying Riemannian manifold, Physica A 436 (2015) 492–498. http://dx.doi.org/10.1016/j.physa.2015.05.076.
[33] L. Bombelli, J. Lee, D. Meyer, R. Sorkin, Space–time as a causal set, Phys. Rev. Lett. 59 (5) (1987) 521–524.
[34] G. Brightwell, R. Gregory, Structure of random discrete spacetime, Phys. Rev. Lett. 66 (3) (1991).
[35] F. Dowker, Causal sets as discrete spacetime, Contemp. Phys. 47 (1) (2006) 37–41.
[36] J. Henson, Discovering the discrete universe, 2010, arXiv:1003.5890.
[37] B. Bollobás, G. Brightwell, Box-spaces and random partial orders, Trans. Amer. Math. Soc. 324 (I) (1991) 59–72.
[38] J. Myrheim, Statistical Geometry, CERN TH-2538, 1978.
[39] D. Meyer, The dimension of causal sets (Ph.D. thesis) 1988.
[40] D.D. Reid, Manifold dimension of a causal set: Tests in conformally flat spacetimes, Phys. Rev. D 67 (2) (2003) 1–8.
[41] B. Bollobás, P. Winkler, The longest chain among random points in Euclidean space, Proc. Amer. Math. Soc. (1988) 347–353.
[42] R. Ilie, G. Thompson, D.D. Reid, A numerical study of the correspondence between paths in a causal set and geodesics in the continuum, Classical Quantum Gravity 23 (2006) 3275–3285. http://dx.doi.org/10.1088/0264-9381/23/10/002.
[43] A. Aho, M. Garey, J. Ullman, The transitive reduction of a directed graph, SIAM J. Comput. 1 (2) (1972) 131–137.
[44] J. Clough, J. Gollings, T. Loach, T. Evans, Transitive reduction of citation networks, J. Complex Netw. 3 (2) (2015) 189–203. http://dx.doi.org/10.1093/comnet/cnu039.
[45] B. Hall, A. Jaffe, M. Trajtenberg, The NBER patent citations data file, NBER Working Paper Series, 2001.
[46] J.H. Fowler, S. Jeon, The authority of Supreme Court precedent, Social Networks 30 (2008) 16–30.
[47] What is the dimension of citation space? http://dx.doi.org/10.6084/m9.figshare.1122521.
[48] M. Newman, J. Park, Why social networks are different from other types of networks, Phys. Rev. E 68 (2003) 036122.
[49] M. Serrano, D. Krioukov, M. Boguna, Self-similarity of complex networks and hidden metric spaces, Phys. Rev. Lett. 100 (7) (2008) 078701.
[50] D. Krioukov, Massimo Ostilli, Duality between equilibrium and growing networks, Phys. Rev. E 88 (2) (2013) 022808.
[51] L. Freeman, Spheres, cubes and boxes: graph dimensionality and network structure, Social Networks 5 (1983) 139–156.
[52] S. Bilke, C. Peterson, Topological properties of citation and metabolic networks, Phys. Rev. E 64 (2001) 036106.
[53] C. Song, L.K. Gallos, S. Havlin, H.A. Makse, How to calculate the fractal dimension of a complex network: the box covering algorithm, J. Stat. Mech. (2007) P03006. http://dx.doi.org/10.1088/1742-5468/2007/03/p03006.
[54] A. Bonato, D. Gleich, M. Kim, D. Mitsche, Dimensionality of social networks using motifs and eigenvalues, PLoS One 9 (2014) e106052. http://dx.doi.org/10.1371/journal.pone.0106052.
[55] O. Shanker, Complex network dimension and path counts, Theoret. Comput. Sci. 411 (2010) 2454–2458.
[56] T. Bolognesi, A. Lamb, Simple indicators for Lorentzian causets, 2014. arXiv Preprint arXiv:1407.1649.
[57] L. Ferretti, M. Cortelezzi, M. Mamino, Duality between preferential attachment and static networks on hyperbolic spaces, Europhys. Lett. EPL (2014) 1–8.
[58] A. Eichhorn, S. Mizera, Spectral dimension in causal set quantum gravity, Classical Quantum Gravity 31 (2014) 125007.
[59] G. Brightwell, P. Winkler, Counting linear extensions, Order (1991) 225–242.
[60] S. Felsner, P.C. Fishbur, W.T. Trotter, Finite three dimensional partial orders which are not sphere orders, Discrete Math. 201 (1999) 101–132.
[61] P. Winkler, Random orders, Order 1 (1985) 317–331.
[62] P. Sibani, H.J. Jensen, Stochastic Dynamics of Complex Systems: From Glasses to Evolution, in: Series on Complexity Science, vol. 2, Imperial College Press, 2013.
[63] H. Wilf, The asymptotic behavior of the Stirling numbers of the first kind, J. Combin. Theory Ser. A 349 (1993) 344–349.
[64] H. Hwang, Asymptotic expansions for the Stirling numbers of the first kind, J. Combin. Theory Ser. A 7141 (7141) (1995) 1–7.