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, U.K.
30th July 2014

Abstract

We adapt and use methods from the causal set approach to quantum gravity to analyse the structure of citation networks from academic papers on the arXiv, supreme court judgements from the US, and patents. We exploit the causal structure of of citation networks to measure the dimension of the Minkowski space in which these directed acyclic graphs can most easily be embedded explicitly taking time into account as one of the dimensions we are measuring. We show that seemingly similar networks have measurably different dimensions. Our interpretation is that a high dimension corresponds to diverse citation behaviour while a low dimension indicates a narrow range of citations in a field.

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, which 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 also understand the structure of citation networks and the kinds of generative mechanisms which can create them.

These networks have a complex structure which is not easily described by any simple model and so are not easily characterised. The underlying processes which generate the network's structure cannot be directly seen but must be inferred from the structure itself by comparing networks to each other, or to models whose generating mechanisms we know. In order to compare two networks we need to be able to characterise their structure in ways relevant to the dynamics we are interested in. To begin to tackle this problem there has been much recent interest in trying to identify the statistical distribution of citation counts [7–17], and various other aspects of the topology of the citation network [18–20].

When networks exist under some constraints, it is often possible to create new methods of characterising their structure which better take those constraints into account, as is well known for networks embedded in space [21–23]. Citation networks are constrained in time, because authors can only cite something that has already been written. This causal constraint prevents closed loops of directed edges in the graph, since all edges must point the same direction in time, and is the same constraint placed on causally connected events in physics [1]. They can therefore be represented as Directed Acyclic Graphs (DAG) where a directed edge goes from node A to node B if the document represented by node A has cited the document represented by node B.

In this paper we investigate citation networks by considering the temporal constraints they are under, and characterise their structure by making comparisons to simple models with the same constraints. We will look at the structure of these networks from the point of view of causal connections, using tools originally created for use in the causal set approach to quantum gravity. In that context, events in spacetime are connected by causal relations like nodes in a network and it is possible to estimate the dimension of a spacetime by considering the causal relationships of points within it. We will apply these methods to citation networks to give them an analogous dimension.

The rest of this paper is structured as follows. We will first introduce the causal set perspective of DAGs, how they can be embedded in space and time, and the methods of estimating their dimension. In the second section we will adapt these methods for use on citation networks and test them on citation networks from academic papers, patents and court judgements. We will conclude by interpreting our results in terms of using dimension as a measure of 'interdisciplinarity', or diversity.

Dimension estimates for spacetime networks

This paper is intended to be widely accessible so we will first cover the necessary details of causality in spacetime. In the causal set approach to quantum gravity [24–26], spacetime is seen as a set of discrete points and not as a continuous space. These spacetime points are the nodes of a graph and each node has an associated time t and spatial co-ordinates x_i. We will consider only the simplest space times, D-dimensional Minkowski spacetimes of one time dimension and $(D-1)$ spatial dimensions. In this case two nodes have an edge between them if and only if the differences in their co-ordinates satisfy:

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

which is to say their separation in time is larger than their separation in space, using the speed of light to convert between the units of space and time (equivalently we choose units where the speed of light is equal to 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. To form a spacetime network, we add edges between points which are causally connected, i.e. satisfying (1), with a direction reflecting the flow of time. We will use the convention that all edges point backwards in time. Such a network is an example of a DAG.

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

[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

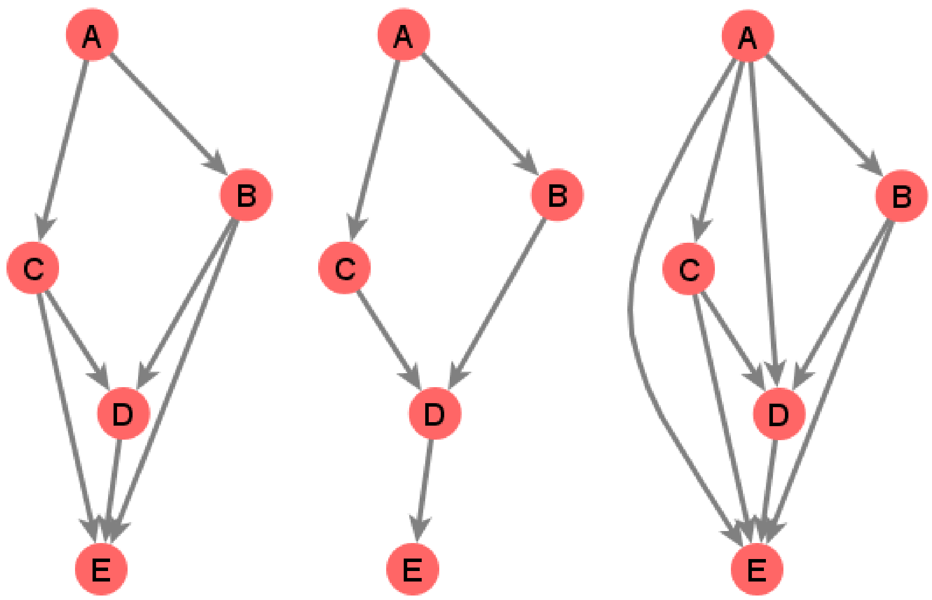

Figure 1: **Left:** A DAG consisting of 5 points. A causal connection exists between any two nodes that can reach one another following edges along their direction. A and B, are causally connected by the edge (A, B); A and D are causally connected by the path {A, B, D}; but B and C are not causally connected. The interval [A, E] contains all of the nodes in this graph since B, C and D are all causally connected to A in one direction, and E in the other direction. Like most citation networks, this graph is neither transitively complete, or transitively reduced.
Centre: The transitive reduction of this graph. All edges except those required to keep all of the causal relations of the DAG have been removed. For example, the edge (B, E) is not required since B and E are already causally related by the path {B, D, E}. No causal connections have been created or destroyed. We will call the edges that remain after TR the 'essential links' though in the literature they are also called 'covering relations', 'nearest neighbour edges' or simply 'links'. In the rest of this paper we will draw networks after TR as it removes many edges making the structure of the network easier to see, but does not break any causal connections.
Right: The transitive completion of this graph. All pairs of causally related nodes now have an edge drawn between them, or alternatively, all edges implied by transitivity are added. Spacetime networks are, by construction, always transitively complete.

causal/temporal ordering as given by the ordering of the time coordinates, and provides a uniquely defined causal relationship. The graph generated by this process must by acyclic since the time coordinates of nodes always decrease when following a path which respects the direction of the edges.

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) [27] as in figure 1.

The simple model we will be using involves randomly scattering points in an interval in Minkowski space. We first create two extremal points with time co-ordinates of 0, and

1, and spatial co-ordinates of 0. We then add the remaining points and by assigning a random time co-ordinate between 0 and 1, and random spatial co-ordinates between -0.5 and 0.5 and allowing them to be in the network if they have edges to the two extremal nodes and so lie within the interval. We will refer to these networks as **spacetime networks** though they are also known as cone spaces in the mathematics literature [28].

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 (1) and make it less likely that two points are connected. So if we were to forget about the space and time coordinates of each point, it would be possible to estimate the number of spatial dimensions by looking at the network's structure, i.e. how its nodes connect to one another. We will use two such methods: the Midpoint-Scaling dimension estimate, and the Myrheim-Meyer dimension estimate.[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. In a Minkowski space, the height, L, of a region is proportional to the length of the longest path through it [31, 32]. We then expect, in a D-dimensional Minkowski space that $V \sim L^D$. Knowing how the size of an interval scales with its height allows the dimension to be inferred.

The Midpoint-Scaling dimension [25] 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 the difference between N_1 and N_2 is minimised.

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 figure 2.

Myrheim-Meyer Dimension

An **n-chain** in a DAG is a sequence of n causally connected nodes. In the spacetime context each point on a chain has all later/earlier points on the sequence in its future/past) light cone. On a general DAG this means there is always a path (respecting the direction of the edges but not necessarily of length one) from each point on the chains to all later points on the chain.

When points are placed at random with uniform probability density in spacetime in some interval the expected number of n-chains S_n, is known to be [31, 33, 34]

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

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

[2]It is possible to define other similar networks, such as a cube-space [29], or a spacetime network using a more complicated spacetime [30]. In our case there was no obvious way to decide which of these models should be preferable, and the interpretations of our results would be the same in either case and so Minkowski space was chosen for its simplicity.

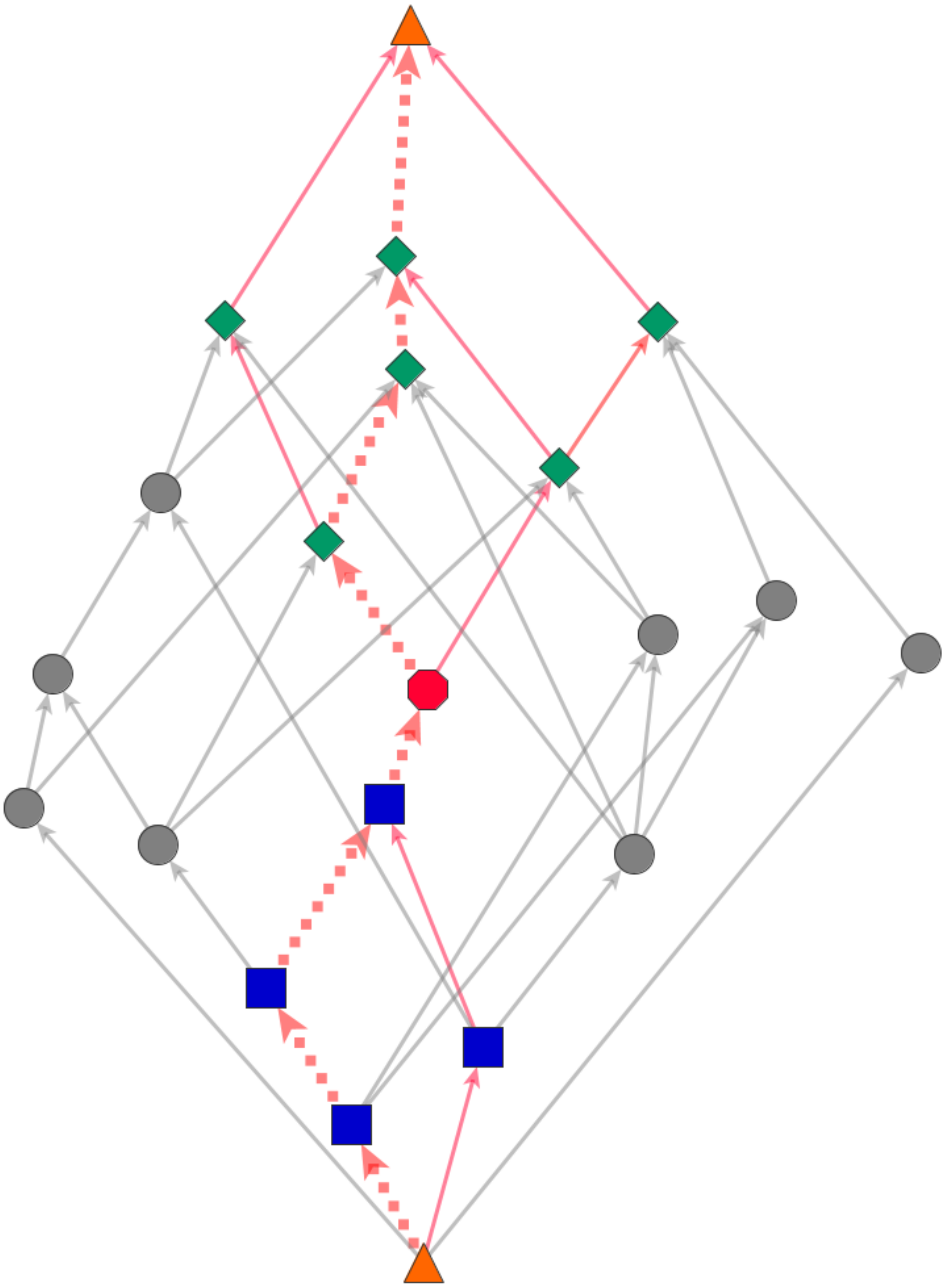

Figure 2: The Midpoint-Scaling dimension in a 2-D spacetime network. The longest path through the interval (defined by the triangular nodes) is shown, and its midpoint is the octagonal node. The diamond nodes are those that lie within a subinterval, from the midpoint to the upper extremal point of the 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. For simplicity, only the the essential links, those remaining after TR are drawn here. It is only the essential links that matter for these dimension estimates, since it is only the causal structure that determines them.

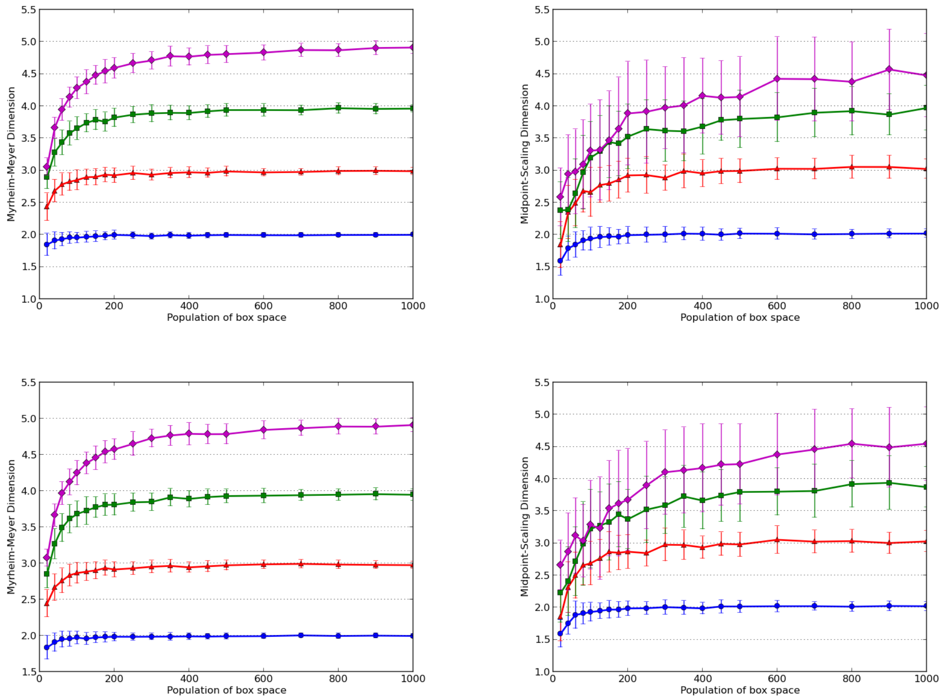

Figure 3: The Myrheim-Meyer dimension (left column) and Midpoint-scaling dimension (right column), against the number of points in the graph, averaged over 100 random spacetime networks for dimensions 2, 3, 4 and 5. The top row shows spacetime networks with all edges present, and the bottom row with only 2% of the allowed edges present. The estimates' convergence from below to the correct dimension is also seen in [34]. Error bars show the standard deviation of the estimated dimension. Errors are larger for higher dimension, but smaller as the size of the space grows.
On the bottom row the dimension estimates are both still very effective, despite only a small fraction of edges being present illustrating the robustness of these dimension estimates.

Given a DAG we can simply count the number of chains and numerically find an estimate for the dimension using this formula. Estimates for dimension can be made using chains of any length n, but there are significantly larger number of 2-chains (just a pair of causally connected points) and so these produce the most accurate estimation of dimension. The expect number of 2-chains is simply

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

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.

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 (see figure 1). We will call the degree after TR the **reduced degree**, k_r. Taking figure 2 as an example, consider the triangular node at the bottom of the diagram. It's degree before TR is $N = 20$ and its degree after is $k_r = 4$ (as shown by the 4 remaining links).

We show in appendix A that for a two-dimensional spacetime network the distribution of reduced degrees k_r is proportional to the Stirling numbers of the first kind. For large N and over a small range of k_r ($\Delta k_r \ll \ln(N)$), the degree distribution is roughly Poissonian with a mean of $\ln(k_r)$. For other dimensions the expected reduced degree is roughly $k_r^{\frac{D}{2}-1}$ [24, 35]. 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 appendix A. In [36] 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.

Estimating the Manifold Dimension of citation networks

Adapting the methods to citation networks

The methods described above are designed to estimate the dimension of the space in which the nodes of a DAG are randomly scattered. DAGs 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.

Despite this, these algorithms to estimate dimension can be applied to any DAG, and a result can be obtained. However, our interpretation of this result does have to change. We are now no longer investigating the properties of a space the points are embedded in, but instead just characterising the DAG's structure in a way that is *analogous* to embedding it in some Minkowski spacetime.

Some work is needed to adapt these dimension estimators to use on citation networks because citation networks do not necessarily share some of the particular properties of the spacetime networks used in these estimators.

Firstly, examples of spacetime networks are usually constructed to be an interval, that is there is only one 'start', or 'source' node (with zero in-degree) and one 'end' or 'sink' node (with zero out-degree), both of which are reachable from any node in the network. This is almost always not the case for citation networks. So instead of estimating the dimension of the whole citation network, we look at many small 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 we ignore this pair of nodes. We then plot a histogram of the population of the interval against its estimated dimension.

Secondly, the spacetime networks are always **transitively complete**. That is 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 network the edges (A, B) and (B, C) imply (A, C). In citation networks this constraint is not present 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 [37] (adding edges between any two nodes if there is a path between them) before counting the 2-chains which are now just the edges.

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 [38]. 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 [39] 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 [40], in total around 25,000 nodes. Further discussion of these particular datasets is available in our previous paper [36] and our datasets will be made available on figshare [53].

Discussion and Interpretation

Figures 4, 5, 6 and 7 show 2D histograms for each arXiv section, plotting the population of an interval against its estimated dimension. The trend line is created by binning data and showing the mean and standard deviation for each bin.

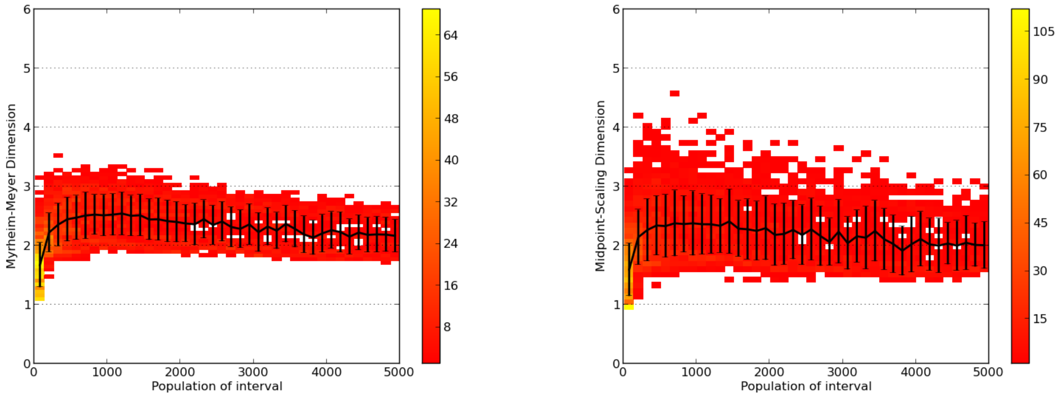

Figure 4: The Myrheim-Meyer (left) and Midpoint-scaling (right) dimensions for the `hep-th` citation network appears to settle at a value around 2 for large intervals

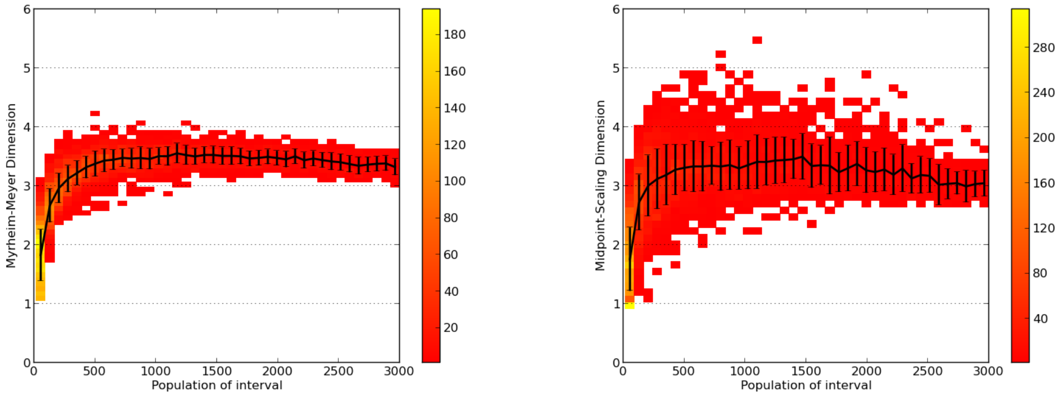

Figure 5: The Myrheim-Meyer (left) and Midpoint-scaling (right) dimensions for the `quant-ph` citation network appears to settle at a value around 3-3.5 for large intervals

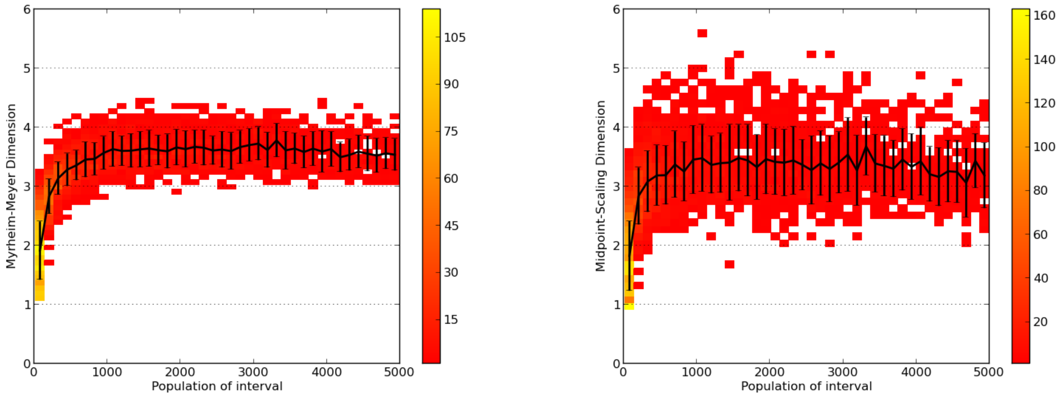

Figure 6: The Myrheim-Meyer (left) and Midpoint-scaling (right) dimensions for the `astro-ph` citation network appears to settle at a value around 3-3.5 for large intervals

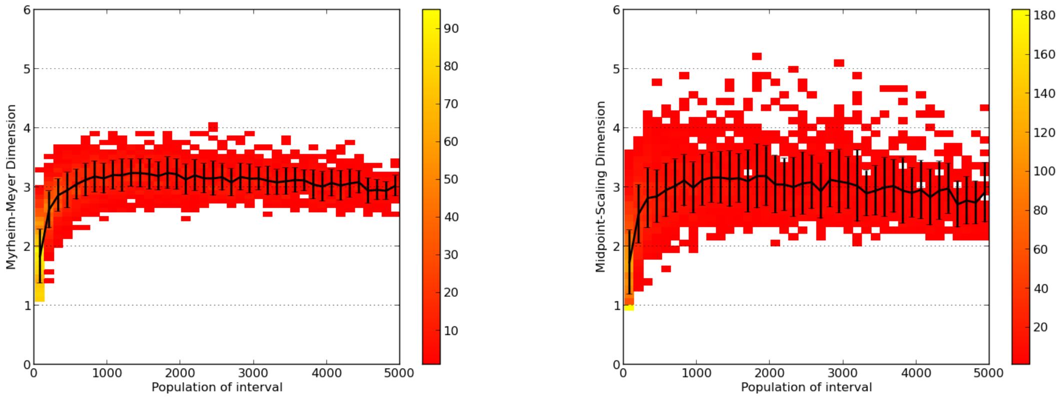

Figure 7: The Myrheim-Meyer (left) and Midpoint-scaling (right) dimensions for the `hep-ph` citation network appears to settle at a value around 3 for large intervals

It is immediately clear from the differing shapes of the histograms in figures 4 to 11 that there are structural differences in the citation networks analysed here which have been revealed by these dimension estimates. Unsurprisingly, they do not have the same structure as a DAG generated by scattering points in a Minkowski space (see the difference in the spread of estimates in figure 3 and the real citation networks). There is a much greater spread of values in the real citation networks, suggesting structural heterogeneity unlike the homogeneous Minkowski space. This is again 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 [18].

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. Although there is significant spread of estimates for each interval size most of the weight of the histogram is near the trend line. For networks generated from scattering points in a Minkowski space there is an underlying dimension being estimated and so it is reasonable to expect independent methods 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. They have significantly different estimated dimensions strongly suggesting that the arXiv citation network is structurally heterogeneous, with its different communities having measurably different citation behaviours. We suggest that these estimates could provide a novel method of measuring these differences in other large, heterogeneous citation networks, or DAGs representing other systems.

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 a DAG. Figure 8 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 show a similar shape to the original network, but with a different final estimate. This suggests that this shape is due only these constraints since that is all that remains after rewiring. The estimated dimension for large intervals though is different to the original `hep-th` network, which suggests that this number is a characteristic of the particular structure of that network and not an expected feature of a random or rewired network under the same constraints.

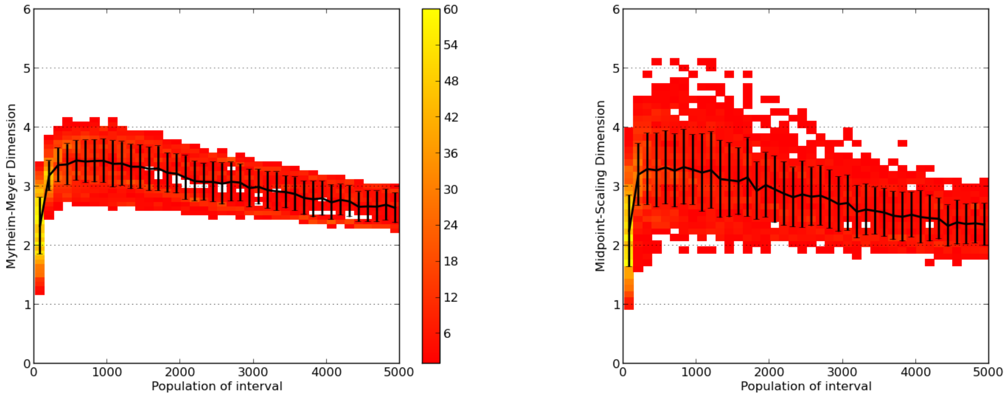

Figure 8: The Myrheim-Meyer (left) and Midpoint-scaling (right) dimensions for the rewired `hep-th` citation network. The 10^5 edges have been rewired randomly 10^7 times, so all structure other than the degree distribution and causality constraints has been removed.

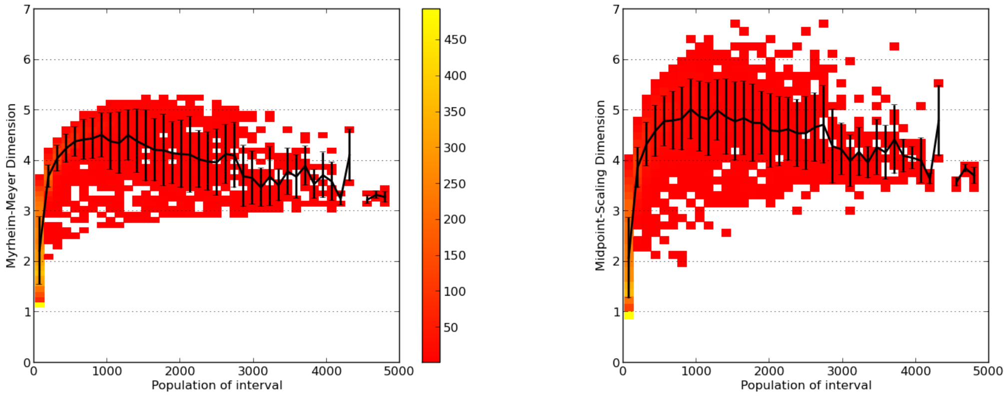

Figure 9: The Myheim-Meyer (left) and Midpoint-scaling (right) dimensions for the Price model network, with the same size and average degree as the `quant-ph` citation network. The spread of values is much larger than in the real citation network and the dimension estimate is higher, illustrating how these dimension estimates can show differences in structure.

Null models

It may seem as if any network with the same degree distribution and subject to the same causal constraints will have a similar looking dimension estimate. Here we further investigate the extent to which causal constraints and degree distribution determine estimated dimension with a simple null model.

We generate a network with the same degree distribution as the `quant-ph` network

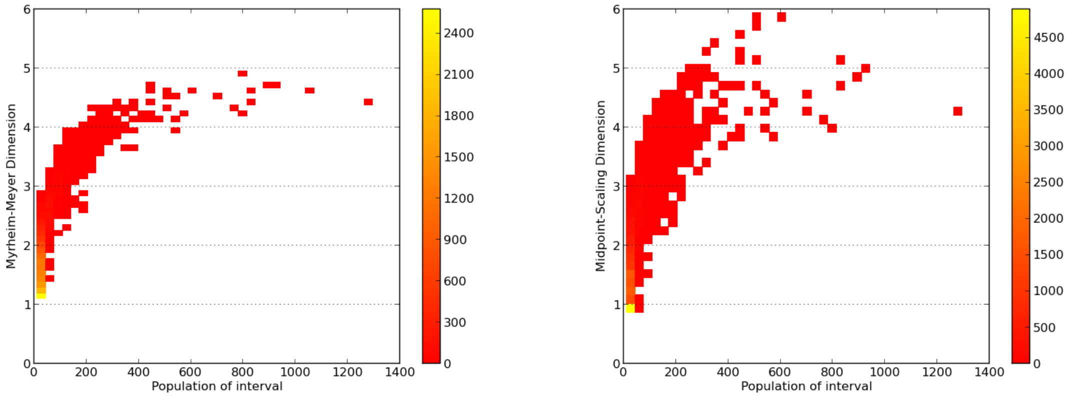

Figure 10: The Myheim-Meyer (left) and Midpoint-scaling (right) dimensions for the patent citation network. In this citation network larger intervals are much rarer than in the others, as a large interval usually contains many different paths from that starting node to the ending node, which is rare in patent citation networks.

using the simple cumulative attachment model for citations[3] due to Price [**?**] and measure its dimension.

Figure 9 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.

Differing citation behaviour

Citation networks from outside academia illustrate different behaviour. The US patent network's histogram is much sparser for larger intervals, and almost all measured intervals were very small. This network, as discussed in [36], has a much lower clustering than the others. Few of its edges are implied by transitivity, which is common in large intervals, explaining the rarity of these large intervals. For large intervals the measured dimension is around 5, much larger than the arXiv citation networks.

The histogram for the US Supreme Court citation network has a different shape to all the others. In the arXiv networks, and the patent network we see a slow growth in estimated dimension as interval size increases. These plots have a similar shape to our earlier tests on DAGs generated from Minkowski spaces although the convergence on the final value appears slower, in terms of how large an interval must be before the estimated dimension plateaus.

The US Supreme Court network, figure 11 seems to show the opposite effect. Small intervals have a higher dimension estimate, and dimension falls as interval size increases.

[3] 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, $p = 0.6$.

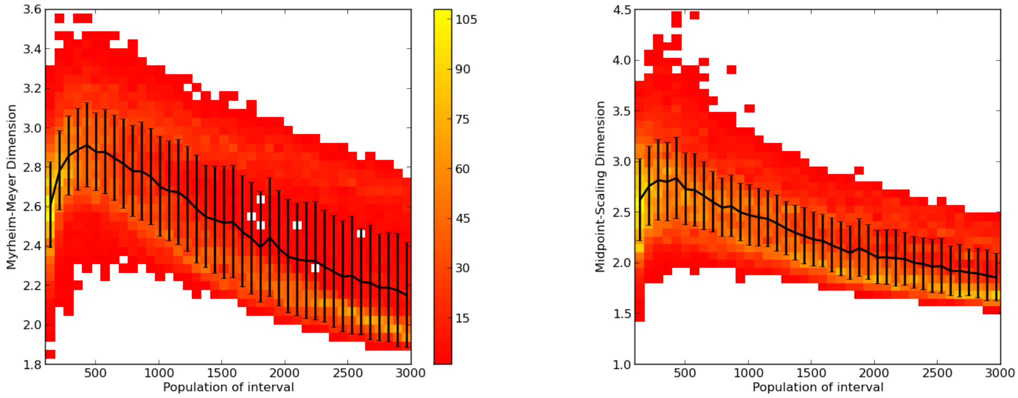

Figure 11: The Myrheim-Meyer (left) and Midpoint-scaling (right) dimensions for the US Supreme Court citation network.

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 appears to have a different estimated dimension on different time scales. For intervals smaller than this characteristic time scale, the measured dimension is around 3, meaning that the size of the interval grows cubically with the length of the longest chain through it (roughly proportional to the temporal length of the interval). For longer intervals it is 2, meaning the interval population only grows quadratically. This temporal heterogeneity of the network is easily revealed by these dimension estimates.

It is not immediately clear how to interpret these results, or what having a low or high estimated dimension means about a network without thinking back to Minkowski spaces. A network with dimension 1 (no spatial dimensions, just a time dimension) can be thought of a single line, where nodes form an unbroken chain of edges, each linking the previous node. 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.

We can therefore interpret a small dimension as a more narrow field, where most of the papers are citing most others (or are at least causally connected to most others) corresponding to a small number of different areas of study. A large dimension then corresponds to a more diverse field, where many independent authors can cite the same paper without citing each other. 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.

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. We have also shown that these dimension measures highlight important differences in the causal structure.

For instance the two particle physics sections of arXiv, `hep-th` and `hep-ph`, are similar in many ways but clearly differ in the dimension measures which quantify how 'broad' or 'narrow' the citation behaviour of authors in these fields is. Furthermore, these methods can be used as a way of differentiating citation networks. 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 that section of the network has come from, using only the topology of the citation network.

The message here is that the networks' structures can differ measurably in seemingly very similar fields of study, and that this is potentially useful information we can extract about different citation behaviours, something of interest to any scientists who want to improve their use of bibliometric measures as an aid to research.

The dimension of a general network is a concept which has been considered before, for example [41]. 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 [36]. This is why it was important to develop methods which includes time as one of its dimensions. For this reason we defined the 'dimension' of a DAG as the dimension of a Minkowski space into which our citation networks can be embedded. The dimension measures we use explicitly involve time and so take the the causal constraints of citation networks into account, recognising that in-edges (being cited) and out-edges (citing someone else) must point in different directions of time.

One view of our choice of Minkowski spacetime for the embedding space is that it is one of the simplest choice to make, defined by a single parameter D. It is also justified by our results as we have obtained consistent measurements which clearly distinguish different fields. However there are studies where networks are embedded in other types of space, be they for quantum gravity [26,34,42] or standard growing network models [30,43]. So an interesting question is to ask why Minkowski space works so well?

The models of growing networks studied so far [30] are equivalent to the simple Price model [7] of citations. The Price model captures the fat-tailed nature of the total degree distribution but it has been shown that the degree-distribution not a good indicator of the nature of any embedding space [42]. In any case such simple models are known to miss most other features of real citation networks. Indeed, using other causally aware measures on citation data reveals important new features in real citation network data [36, 44]. So we find little encouragement to use other spacetimes. One particular feature noted in [36] 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. These are the essential links, i.e. the citations left after transitive reduction. This suggests that the essential structure surrounding a document, the 'nearest neighbours', may be relatively similar, statistically, for all points and hence a Minkowski spacetime may make sense. We can find no motivation to use a space time where there are significant differences at different points such as the exponential growth in the de Sitter space of [30]. Put another way, by looking at the essential links, which are necessary to the causal structure, we get a picture where a constant propagation speed for information makes sense. It would only be in the other links, those not essential for the causal structure but which form the vast majority (80%) of real citation networks, where we see vast differences between the connections between papers well known in citation analyses.

We note that other dimension estimators exist for graphs. We have tried using the reduced degree, but as explained in the 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's spectral dimension [35]. 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 [45,46]. This method requires a DAG to be perfectly embedded in a manifold as it gives integer dimension estimates. They are 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. 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 and small deviations from DAGs which are faithfully embedded in a Minkowski manifold only lead to small deviations in the estimated dimension. Furthermore we note that 'dimension' has been used to describe many aspects of a network's topology, such as the number of parameters in a model [47], the time a random walker takes to return to its starting position [48], or the number spatial (but not temporal) dimensions measured using influence regions [20].

Acknowledgments

We would like to thank Jamie Gollings, Amy Hughes, Tamar Loach, Kim Christensen and Mauricio Barahona for useful discussions.

References

[1] A.W. Harzing, *Publish or Perish.* Tarma Software Research, 2010.

[2] A. Abbas, L. Zhang, and S. U. Khan, "A literature review on the state-of-the-art in patent analysis", *World Patent Information*, vol. 37, pp. 3–13, Jun 2014.

[3] P. W. Martin, *Introduction to basic legal citation.* Legal Information Institute, 2007.

[4] T. A. Brooks, "Evidence of complex citer motivations", *Journal of the American Society for Information Science*, vol. 37, no. 1, pp. 34–36, 1986.

[5] M. Simkin and V. Roychowdhury, "Read before you cite!", *Complex Systems*, vol. 14, pp. 269–274, 2003.

[6] L. Bornmann and H.D. Daniel, “What do citation counts measure? a review of studies on citing behavior”, *Journal of Documentation*, vol. 64, no. 1, p. 4580, 2008.

[7] D. S. Price, “Networks of scientific papers”, *Science*, vol. 149, pp. 510–515, 1965.

[8] D. S. Price, “A general theory of bibliometric and other cumulative advantage processes”, *J.Amer.Soc.Inform.Sci.*, vol. 27, pp. 292–306, 1976.

[9] S. Redner, “How popular is your paper? An empirical study of the citation distribution”, *The European Physical Journal B*, vol. 134, pp. 131–134, 1998.

[10] C. Tsallis and M. de Albuquerque, “Are citations of scientific papers a case of nonextensivity?”, *Physical Journal B*, 2000.

[11] S. Lehmann, B. Lautrup, and A. D. Jackson, “Citation networks in high energy physics”, *Physical Review E*, vol. 68, p. 026113, Aug. 2003.

[12] M. V. Simkin and V. P. Roychowdhury, “Stochastic modeling of citation slips”, *Scientometrics*, vol. 62, no. 3, pp. 367–384, 2005.

[13] F. Radicchi, S. Fortunato, and C. Castellano, “Universality of citation distributions: Toward an objective measure of scientific impact”, *Proceedings of the National Academy of Sciences*, vol. 105, pp. 17268–17272, 2008.

[14] L. Bornmann and H.-D. Daniel, “Universality of citation distributions — a validation of radicchi et al.’s relative indicator $c_f = c/c_0$ at the micro level using data from chemistry”, *J. Am. Soc. Inf. Sci.*, vol. 60, pp. 1664–1670, Aug 2009.

[15] M. J. Stringer, M. Sales-Pardo, and L. A. N. Amaral, “Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal”, *J. Am. Soc. Inf. Sci.*, vol. 61, no. 7, pp. 1377–1385, 2010.

[16] E. Vieira and J. Gomes, “Citations to scientific articles: Its distribution and dependence on the article features”, *Journal of Informetrics*, vol. 4, pp. 1–13, Jan. 2010.

[17] Y.-H. Eom and S. Fortunato, “Characterizing and modeling citation dynamics.”, *PloS one*, vol. 6, p. e24926, Jan. 2011.

[18] M. Newman and J. Park, “Why social networks are different from other types of networks”, *Physical Review E*, vol. 68, p. 036122, Sept. 2003.

[19] O. Shanker, “Complex network dimension and path counts”, *Theoretical Computer Science*, vol. 411, pp. 2454–2458, June 2010.

[20] A. Bonato, D. Gleich, M. Kim, and D. Mitsche, “Dimensionality of social networks using motifs and eigenvalues”, *arXiv:1405.0157*, pp. 1–26, 2014.

[21] M. Barthelemy, “Spatial networks”, *Physics Reports*, vol. 499, pp. 1–101, Feb. 2011.

[22] 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*, vol. 108, pp. 7663–8, May 2011.

[23] L. Daqing, K. Kosmidis, A. Bunde, and S. Havlin, "Dimension of spatially embedded networks", *Nature Physics*, vol. 7, pp. 481–484, Feb. 2011.

[24] L. Bombelli, J. Lee, D. Meyer, and R. Sorkin, "Space-time as a causal set", *Physical review letters*, vol. 59, no. 5, pp. 521–524, 1987.

[25] G. Brightwell and R. Gregory, "Structure of random discrete spacetime", *Physical review letters*, vol. 66, no. 3, 1991.

[26] F. Dowker, "Causal sets as discrete spacetime", *Contemporary Physics*, no. October 2012, pp. 37–41, 2006.

[27] J. Henson, "Discovering the Discrete Universe", *arXiv preprint arXiv:1003.5890*, pp. 1–24, 2010.

[28] B. Bollobás and G. Brightwell, "Box-spaces and random partial orders", *Transactions of the American Mathematical Society*, vol. 324, no. I, pp. 59–72, 1991.

[29] B. Bollobás and P. Winkler, "The longest chain among random points in Euclidean space", *Proceedings of the American Mathematical Society*, pp. 347–353, 1988.

[30] D. Krioukov, M. Kitsak, R. S. Sinkovits, D. Rideout, D. Meyer, and M. Boguñá, "Network cosmology.", *Scientific reports*, vol. 2, p. 793, Jan. 2012.

[31] J. Myrheim, "Statistical Geometry", *CERN*, vol. 2538, 1978.

[32] G. Thompson and D. D. Reid, "Paths in Discrete Spactime: A Numerical Investigation of the Myrheim Length Conjecture", 2003.

[33] D. Meyer, *The dimension of causal sets*. PhD thesis, 1988.

[34] D. D. Reid, "Manifold dimension of a causal set: Tests in conformally flat spacetimes", *Physical Review D*, no. October 2002, pp. 1–8, 2003.

[35] A. Eichhorn and S. Mizera, "Spectral dimension in causal set quantum gravity", *Classical and Quantum Gravity*, vol. 31, p. 125007, June 2014.

[36] J. Clough, J. Gollings, T. Loach, and T. Evans, "Transitive reduction of citation networks", *arXiv:1310.8224v2*, pp. 1–17, 2014.

[37] A. Aho, M. Garey, and J. Ullman, "The transitive reduction of a directed graph", *SIAM Journal on Computing*, vol. 1, no. 2, pp. 131–137, 1972.

[38] `http://paperscape.org/`

[39] B. Hall, A. Jaffe, and M. Trajtenberg, "The NBER Patent Citations Data File", *NBER Working Paper Series*, 2001.

[40] J. H. Fowler and S. Jeon, “The authority of Supreme Court precedent”, *Social Networks*, vol. 30, pp. 16–30, Jan. 2008.

[41] C. Song, L. K. Gallos, S. Havlin, and H. a. Makse, “How to calculate the fractal dimension of a complex network: the box covering algorithm”, *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2007, pp. P03006–P03006, Mar. 2007.

[42] T. Bolognesi and A. Lamb, “Simple indicators for Lorentzian causets”, *arXiv preprint arXiv:1407.1649*, 2014.

[43] L. Ferretti, M. Cortelezzi, and M. Mamino, “Duality between preferential attachment and static networks on hyperbolic spaces”, *EPL (Europhysics Letters)*, pp. 1–8, 2014.

[44] S.R. Goldberg, H. Anthony and T.S. Evans “Modelling Citation Networks”, in preparation.

[45] L. Bombelli, *Space-time as a causal set.* PhD thesis, 1987.

[46] G. Brightwell and P. Winkler, “Counting linear extensions”, *Order*, pp. 225–242, 1991.

[47] L. Freeman, “Spheres, cubes and boxes: graph dimensionality and network structure”, *Social Networks*, vol. 5, pp. 139–156, 1983.

[48] S. Bilke and C. Peterson, “Topological properties of citation and metabolic networks”, *Physical Review E*, vol. 64, p. 036106, Aug. 2001.

[49] P. Winkler, “Random orders”, *Order*, vol. 1, pp. 317–331, 1985.

[50] P. Sibani and H. J. Jensen, *Stochastic Dynamics of Complex Systems: From Glasses to Evolution*, vol. 2 of *Series on Complexity Science.* Imperial College Press, 2013.

[51] H. Wilf, “The asymptotic behavior of the Stirling numbers of the first kind”, *Journal of Combinatorial Theory, Series A*, vol. 349, pp. 344–349, 1993.

[52] H. Hwang, “Asymptotic expansions for the Stirling numbers of the first kind”, *Journal of Combinatorial Theory, Series A*, vol. 7141, no. 7141, pp. 1–7, 1995.

[53] `http://dx.doi.org/10.6084/m9.figshare.1122521`

A 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 $i = 1, 2...N$ with coordinates z_i^α where $\alpha = 1, 2...D$. 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 [49] 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.

$$p(k_r, N) = \frac{N-1}{N} p(k_r, N-1) + \frac{1}{N} p(k_r - 1, N-1) \tag{4}$$

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 largest 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 largest x coordinate, $i = N$. This point $i = N$ can only be a new link to the origin if it is minimal in the y coordinate. That is, since we already know that every existing point has a larger x coordinate by our ordering, we have $x_i > x_j$ but $y_i < y_j$ for $N = i > j$. Because the coordinates are just random numbers, the probability that y_N is the smallest 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 equation 4. This view is equivalent to a standard record statistics process [50]. Indeed the points don't even have to be uniformly distributed here, the only requirement is that the 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 $\left[\begin{array}{c} N \\ k_r \end{array} \right]$, namely

$$\begin{aligned} \left[\begin{array}{c} N+1 \\ k_r \end{array} \right] &= N \left[\begin{array}{c} N \\ k_r \end{array} \right] + \left[\begin{array}{c} N \\ k_r - 1 \end{array} \right] && (5) \\ \text{where} \left[\begin{array}{c} 0 \\ 0 \end{array} \right] &= 1 && (6) \\ \text{and} \left[\begin{array}{c} N \\ 0 \end{array} \right] &= \left[\begin{array}{c} 0 \\ N \end{array} \right] = 0 && (7) \end{aligned}$$

We can then say that

$$p(k_r, N) = \frac{1}{N!} \left[\begin{array}{c} N \\ k_r \end{array} \right] \tag{8}$$

To check our answer, note that $\left[\begin{array}{c} N \\ 1 \end{array} \right] = (N-1)!$ giving $p(k_r = 1, N) = \frac{1}{N}$ as expected [4]. As noted by Wilf in [51], 'the Stirling numbers of the first kind are notoriously difficult to compute', and so we are unlikely to find a nice solution here.

[4] The probability that $k_r = 1$ is simply the probability that the point with the smallest x-coordinate also has the smallest y-coordinate

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

$$G(z,N)=\sum_{k=0}^{\infty} z^k p(k,N) \tag{9}$$

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 4 we now find that

$$\begin{aligned} G(z,N) &= \frac{N-1}{N}G(z,N-1)+\frac{z}{N}G(z,N-1) && (10)\\ G(z,N) &= \frac{\Gamma(N+z)}{\Gamma(z)\Gamma(N+1)}=(z+N-1).(z+N-2)...(z)\ \times\frac{1}{N!}\,. && (11)\end{aligned}$$

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 [51,52] can be studied from the generating function $G(z,N)$ in 11 as

$$\begin{aligned}\lim_{N\to\infty} G(z,N) &= \frac{N^{z-1}}{\Gamma(z)} && (12)\\ &= \frac{1}{N}\frac{\sum_{k=0}(\ln(N))^k z^k/k!}{z^{-1}+\psi(0)+O(z)} && (13)\end{aligned}$$

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 $p_{\mathrm{Poisson}}(k)=e^{-\lambda}\lambda^k/k!$ 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 [42] may only be useful for small ranges of k_r, typically $|\Delta k_r|\ll\ln(N)$.

From the generating function in 11 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.

$$\langle k_r\rangle=\sum_{k_r=0}^{\infty} k_r\, p(k_r,N) \;=\; \left.\frac{\partial G(z,N)}{\partial z}\right|_{z=1}=\sum_{i=1}^{N}\frac{1}{i}=H_n\approx\gamma+\ln(N)\,, \tag{14}$$

where $\gamma\approx 0.577$ is the Euler-Mascheroni constant. This Harmonic number result is the two-dimensional case of D-dimensional result in [49], which in the large N limit tends to logarithmic growth, as suggested in [35].

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

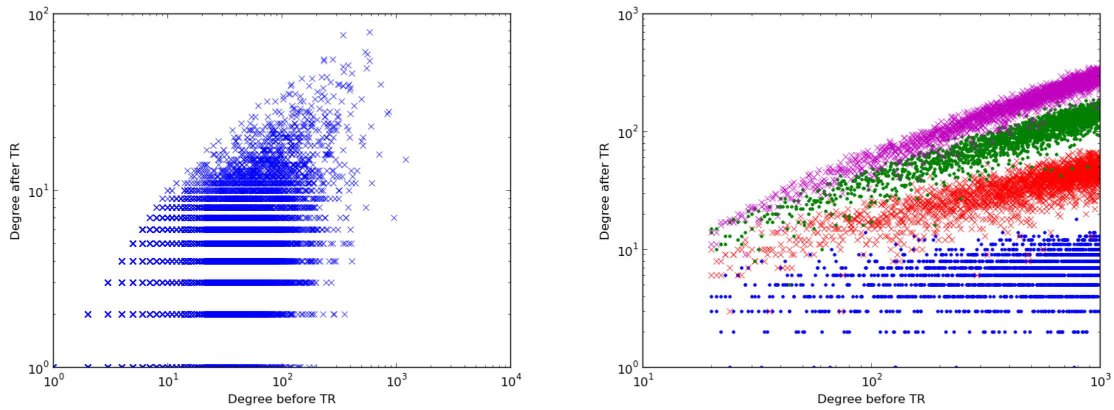

Figure 12: **Left:** the degree before, and after transitive reduction for the `hep-ph` citation network. The spread of k_r is very wide for a given k_r, indicating a heterogeneity in the papers.
Right: the degree before, and after transitive reduction for spacetime networks of dimension 2-5. Lower dimension appear 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.