Embedding graphs in Lorentzian spacetime

James R. Clough, Tim S. Evans
Centre for Complexity Science, Imperial College London,
South Kensington campus, London, SW7 2AZ, United Kingdom

February 9, 2016

Abstract

Geometric approaches to network analysis combine simply defined models with great descriptive power. In this work we provide a method for embedding directed acyclic graphs into Minkowski spacetime using Multidimensional scaling (MDS). First we generalise the classical MDS algorithm, defined only for metrics with a Euclidean signature, to manifolds of any metric signature. We then use this general method to develop an algorithm to be used on networks which have causal structure allowing them to be embedded in Lorentzian manifolds. The method is demonstrated by calculating embeddings for both causal sets and citation networks in Minkowski spacetime. We finally suggest a number of applications in citation analysis such as paper recommendation, identifying missing citations and fitting citation models to data using this geometric approach.

1 Introduction

One success of network science has been to identify that some complex systems can be simplified by considering just the topology of the pairwise interactions between their parts. Abstracting a complex system as a graph can bring physical insights and predictive power. Yet these graphs can still be very complicated. Network geometry is an approach which further abstracts the system by modelling the nodes of the network as points in a geometric space.

Most existing approaches use Riemannian spaces, the simplest example of which is Euclidean space. Random Geometric Graphs (RGG) are graphs embedded in Euclidean space [1, 2, 3].

Recently there has been much interest in geometric approaches to the study of networks in non-Euclidean spaces [4, 5, 6, 7]. Embedding in hyperbolic spaces can yield scale free, clustered networks with community structure illustrating the remarkable power that geometric approaches have to recover complex network properties.

A well established geometric approach to data analysis is *Multidimensional Scaling* (MDS), a technique to give data expressed as distances or similarities a spatial representation [8]. In most MDS analysis, the space used for that spatial representation has been Euclidean, and the technique, as usually described, requires a Riemannian manifold, where the triangle inequality is maintained.

MDS has been used in network science, to fit models of RGGs to networks from real data, for example, from protein interactions [9]. Normally, the MDS algorithm takes as an argument pairwise distances between objects, so when applying it to simple networks, where only binary pairwise relations exist, these distances have to be inferred from the network structure. In the simplest Euclidean case (as in [9]), the shortest path on the network is used as an estimate for the distances between vertices, from which MDS is used to calculate coordinates. Once these coordinates have been calculated, a new RGG can be built from them, and if it is similar to the original graph, the initial geometric assumption was a good one.

In this paper we will consider networks where each node is associated with a particular time and directed edges between nodes represent causal relations. Such a network forms a **directed acyclic graph** (DAG)[10]. Instead of embedding a network in geometric space alone, the causal ordering of the nodes in a DAG suggests that an embedding in space and time is needed. The causal structure of such a network has the same constraints as the causal structure of spacetime as used in special and general relativity [11, 12].
This suggests that the geometries used in relativity, which are pseudo-Riemannian are the appropriate ones to use because of the special properties of a time dimension. In particular, we will consider *Lorentzian spacetimes*, a special case of pseudo-Riemannian manifolds in which there is one time dimension with some number of spatial dimensions.

Euclidean space, being flat and isotropic, is the simplest Riemannian manifold. Analogously, flat isotropic spacetime is Minkowski spacetime, the simplest Lorentzian manifold. In this paper, we first generalise classical MDS to allow converting distances into coordinates for pseudo-Riemannian manifolds. We then show how this allows embedding of DAGs in Minkowski spacetime[1], and that the coordinates of geometric graphs in Minkowski spacetime can be successfully recovered. We then illustrate this technique by finding coordinates for some citation networks, which naturally form DAGs and suggest applications such as paper recommendation.

2 Lorentzian Multidimensional Scaling

2.1 Review of Standard MDS

We will begin by briefly summarising the details of standard MDS in Euclidean space. Suppose we have N objects, and are given the squared Euclidean distance, S_{ij} between each pair. We wish to find the co-ordinates of the objects, which will be D dimensional vectors, $\mathbf{x}_i$ for each object i, such that they fit the constraint that $|\mathbf{x}_i - \mathbf{x}_j|^2 = S_{ij}$.

The classical MDS algorithm solves this problem by using this $N \times N$ matrix of square distances, S, and then constructing the double centred matrix $\mathsf{B} = -\frac{1}{2}\mathsf{JSJ}$ where $\mathsf{J} = \mathbb{1} - \frac{1}{N}\mathbf{1}.\mathbf{1}^{\mathrm{T}}$.

[1]Somewhat confusingly, in some of the literature on MDS (particularly in psychology - see for instance Shepard [13]), a Minkowski metric refers to a distance measure of the form $d_{xy} = \left[\sum_i (x_i - y_i)^\lambda\right]^{1/\lambda}$. This is *not* what we mean by a Minkowski metric in this paper, rather we mean the Minkowski spacetime of special relativity[14].

It can then be shown (details are available in [8]) that

$$\mathsf{B} = \mathsf{X}\mathsf{X}^{\mathrm{T}} \tag{1}$$

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

$$\mathsf{B} = \mathsf{U}\mathbf{\Sigma}\mathsf{U}^{\mathrm{T}} \tag{2}$$

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

$$\mathsf{X} = \sqrt{\mathbf{\Sigma}}\mathsf{U} \tag{3}$$

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

2.2 Lorentzian MDS

Minkowski spacetime is a combination of a d-dimensional Euclidean space, and one time dimension forming a $(d+1)$-dimensional manifold. A point i in this space, has coordinates $\mathbf{x}_i$ consisting of a time coordinate, x_i^0, and spatial coordinates x_i^k, with $k = 1, 2, ..., d$. The Minkowski separation between two such spacetime points i and j is given by

$$M_{ij} = M(\mathbf{x}_i, \mathbf{x}_j) = -(x_i^0 - x_j^0)^2 + \sum_{k=1}^{d}(x_i^k - x_j^k)^2 \tag{4}$$

Pairs of points can then be classified into three types: for a positive separation the pair is spacelike separated, for a negative separation the pair is timelike separated, while exactly zero separation means the pair is lightlike separated. We can now ask the same question that classical Euclidean MDS poses: given pairwise separations M_{ij}, for points in this space, can we recover coordinates which respect these separations?

Proceeding with the classical Euclidean algorithm we can construct the double centred matrix B as before using $\mathsf{B} = -\frac{1}{2}\mathsf{J}\mathsf{M}\mathsf{J}$. However we now encounter a problem when decomposing B. Previously the eigenvalues were guaranteed to be non-negative, but now we find one negative eigenvalue corresponding to the time dimension's negative sign in equation 4. Since we need to take the square root of these eigenvalues, and we want real coordinates this is a problem.

It turns out that the changes required to the classical MDS algorithm are remarkably simple (details are given in appendix A). Instead of looking for a matrix of coordinates X such that $\mathsf{B} = \mathsf{X}\mathsf{X}^{\mathrm{T}}$, we now search solutions to

$$\mathsf{B} = \mathsf{X}\mathsf{G}\mathsf{X}^{\mathrm{T}} \tag{5}$$

where G is matrix representing the metric of the embedding space. For traditional MDS with its Euclidean space G is just the identity matrix so this factor drops out from the analysis. However for DAGs where nodes are also associated with a time coordinate, we choose G to represent the Minkowski metric. In our conventions, this is a diagonal matrix with -1 in the first column and $+1$ in the others. We again decompose B and now need solutions to

$$\mathsf{X}\mathsf{G}\mathsf{X}^{\mathrm{T}} = \mathsf{U}\mathbf{\Sigma}\mathsf{U}^{\mathrm{T}}\,. \tag{6}$$

The difference to traditional MDS is that the -1 present in G changes the sign of the one negative eigenvalue in Σ. This allows us to take the square root of $\mathbf{\Sigma}$ as we do in classical Euclidean MDS.

2.3 Causal Set DAGs and Minkowski spacetime

To use Lorentzian MDS on networks, we require a method of estimating the separations between every pair of nodes from the network structure. We will do this using ideas from causal set theory. A causal set is a locally finite partially ordered set.

In the causal set approach to quantum gravity, the underlying structure of spacetime is postulated to be a causal set. Spacetime is seen as discrete at the Planck scale, and the continuous spacetime we perceive emerges at larger length scales.[15, 16, 17]. The only structure present are elements of the set and the relation, $\prec$ between pairs of elements. In the correspondence between the discrete causal set, and the continuous spacetime that emerges, the relation $\prec$ corresponds to timelike separation where $x \prec y$ corresponds to x being in the past of y, and spacelike separated pairs are not related. Causal sets give us a natural way of discretising spacetime with related elements, and we can use them in our approach because they have the same structure as a DAG. In physics a timelike separated pair are allowed to be causally related, as information could pass from one to the other, from past to future. However a spacelike separated pair can never be causally related as light could not reach one event from the other. This is why we will associate timelike separation in the continuous spacetime with a link representing a causal relation in the network.

We can now construct the equivalent of an RGG in Minkowski space. We begin by assigning coordinates in $\mathbb{M}^D$ to N points, by sampling uniformly at random[2] from $[0,1]^D$. We will denote the coordinates of point i by x_i^μ, $\mu = 0, 1, \ldots, d$, where $\mu = 0$ is the time coordinate. We then construct the causal set graph, $\mathcal{G}_D$ by placing an edge from i to j if $M_{ij} < 0$ and directing that edge from the past to the future. The fact that edges in the graph now represent causal relations illustrates why the graph is necessarily a DAG, as closed causal loops are forbidden.

Figure 1 shows the distribution of the eigenvalues $\mathbf{\Sigma}$ for causal sets and random DAGs[3].

[2]The range of coordinates is chosen for simplicity but this does not need to be the case in general.

[3]By a random DAG we mean an Erdős-Rényi random graph, with the nodes placed in a random order and edges directed along that order, and then the graph transitively completed. The example in the figure uses an Erdős-Rényi graph with an average degree of 25 but the result is general.

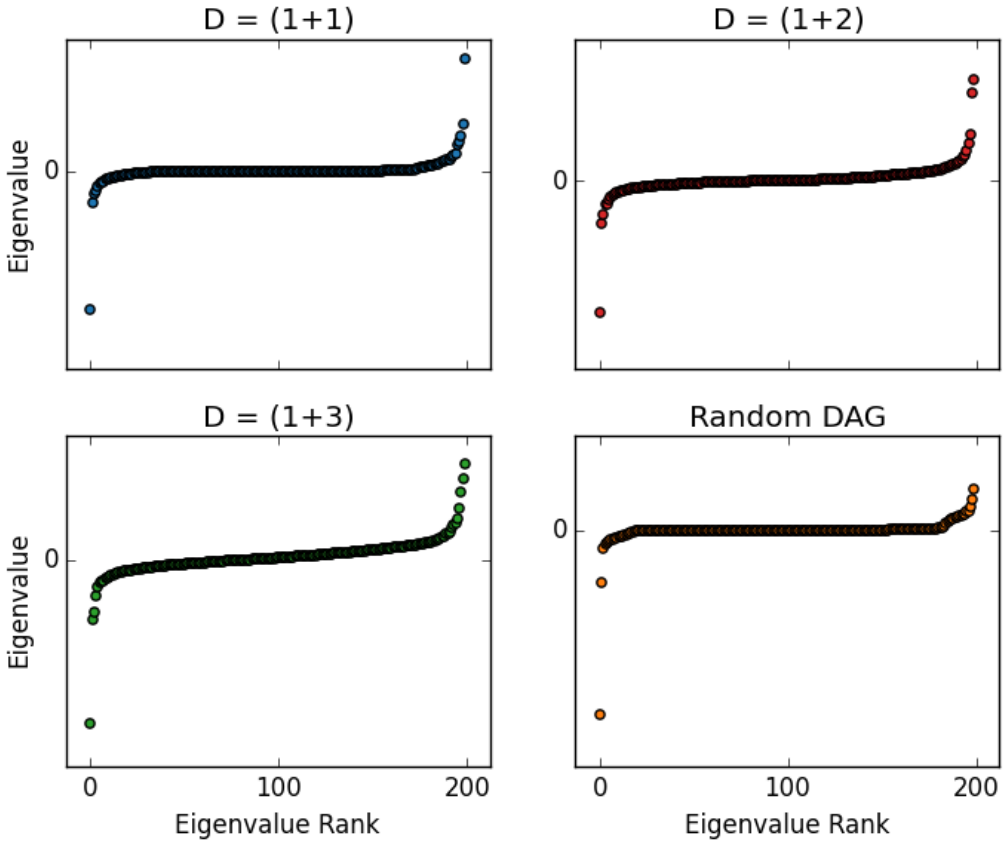


Figure 1: The eigenvalue distribution for MDS of 200 point causal sets in $1+1$, $1+2$, and $1+3$ dimensions, and for a random DAG. In all cases, one large negative eigenvalue is seen, corresponding to the one timelike dimension. For the causal sets, this timelike dimension is the one time dimension in the Minkowski spacetime they are embedded in. For the random DAG, this corresponds to the time ordering which could be created as a consequence of the acyclic property of this graph. We observe d large positive eigenvalues corresponding to the spatial dimensions in each causal set, illustrating that the coarse graining of the causal set does not prevent the MDS algorithm from successfully identifying the principle components of the space. For a given number of points, higher dimensionality will mean fewer relations between the causal set's elements and since these relations are the information used by the MDS algorithm its ability to cleanly pick out d dimensions diminishes as d increases. However for the random DAG there is no clear separation of large eigenvalues suggesting there is no natural dimension to an embedding Minkowski spacetime.

2.4 Embedding Graphs using Lorentzian MDS

To use MDS on a network we must estimate the separation matrix S using the network's structure. For Euclidean MDS, the separation is always a non-negative number and the shortest path between nodes in the graph is a natural and effective estimator for the distance. However in Minkowski spacetime, the separation of points is not always positive, so the the number of steps along some path is not going to be measure of all Lorenztian separations. The solution is to estimate spacelike and timelike separations separately when studying a DAG.

Suppose we have two connected nodes in the graph, meaning they are timelike separated. It was conjectured in [18] and later shown in [19] that for timelike separated points i and j in $\mathcal{G}_D$ the length of *longest path* L_{ij} between two connected nodes, say i and j, is proportional their timelike separation (in the limit of long separations, and where the longest path must respect the edge's direction). So in this case we set $S_{ij} = -1 \times L_{ij}^2$.

Finding the distance between spacelike pairs is more challenging and to our knowledge there is no solution as easily calculated as the longest path is for timelike pairs. Good approximations are known however, and we will use a very simple one, described in [20, 21] as 'naive spatial distance'. Suppose we have two disconnected points i and j in the $\mathcal{G}_D$ meaning they are spacelike separated. We then look for a pair of nodes, k and l, where k is in the future of both nodes i and j while l is in the past of both i and j[4]. We then chose nodes k and l so as to minimise the length of the longest path between k and l (which are necessarily connected, via paths through x and y). The timelike separation between k and l is then used as an estimate for the spacelike separation between i and j. Figure 2 shows an example of how we estimate timelike and spacelike separations.

This estimate is simple and at first appealing, but fails in more than two dimensions for large graphs (hence the 'naive' in the name). Nonetheless we find it is sufficient for our purposes[5]. This is partly because it is inaccurate only for large causal sets but also because in the MDS algorithm each point's coordinates is fixed by many distances, both timelike and spacelike which limits the effect of noise some poor estimations of spacelike separations. We can then set $S_{ij} = L_{kl}^2$ for the chosen k and l.

These timelike and spacelike distances define our separation matrix S (where timelike separation has the $-$ sign in our conventions). Finally we use the algorithm described in the previous section to assign coordinates in some $\hat{D}$-dimensional Minkowski spacetime $\mathbb{M}^{\hat{D}}$ to each vertex in the DAG.

3 Testing Lorentzian MDS on Causal Sets

Given a causal set graph, $\mathcal{G}_D$ it is possible, in principle and for large enough $N = |\mathcal{G}_D|$, to recover all properties of the spacetime (up to a factor of the density of the sprinkling)[11, 22].

[4] If no such pair exists, we set the spacelike distance of i and j equal to some maximal distance which is a parameter of the algorithm. In the examples shown here, we used the length of the longest path in the graph as this parameter.

[5] We tried using the two-link method described in [21] but found that for small graphs with fewer than 500 nodes too many cases had no two-links and so the spacelike separation couldn't be calculated

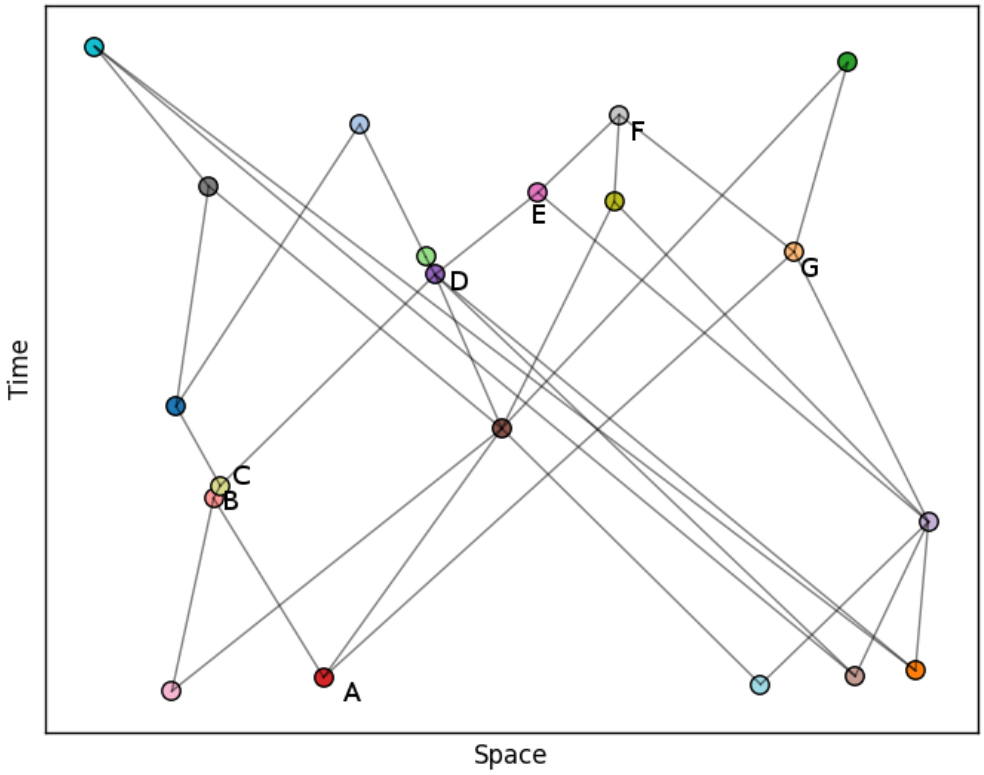


Figure 2: The timelike separation between nodes A and F is approximated as 5 units - as this is the number of edges in the longest direction-respecting path between them. Nodes B and G are spacelike separated. To estimate this separation we find a pair of points in their mutual past and future. In this case, the only such pair is (A, F). The naive spatial separation between (B, G) is then given by the timelike separation between (A, F) so is also in this case 5 units. Note, only the edges not implied by transitivity have been drawn.

In Minkowski spacetime there is only one parameter to recover, D, and this can be estimated for the process described above and for DAGs in general [18, 23, 6]. Our task here is to recover not properties of the manifold in which the nodes are embedded, but to find the full details of that embedding.

If the graph was originally made sprinkling points into $\mathbb{M}^D$ then we know that an exact embedding is possible (since the original sprinkled coordinates must be a solution), and so the embedding algorithm should approximately recover the original coordinates (up to distance preserving factors). If the graph was not, then we may only be able to find an approximate embedding.

As is the case for classical MDS, our Lorentzian MDS is guaranteed to recover the coordinates of points when the exact distances are used as the algorithms input. However,

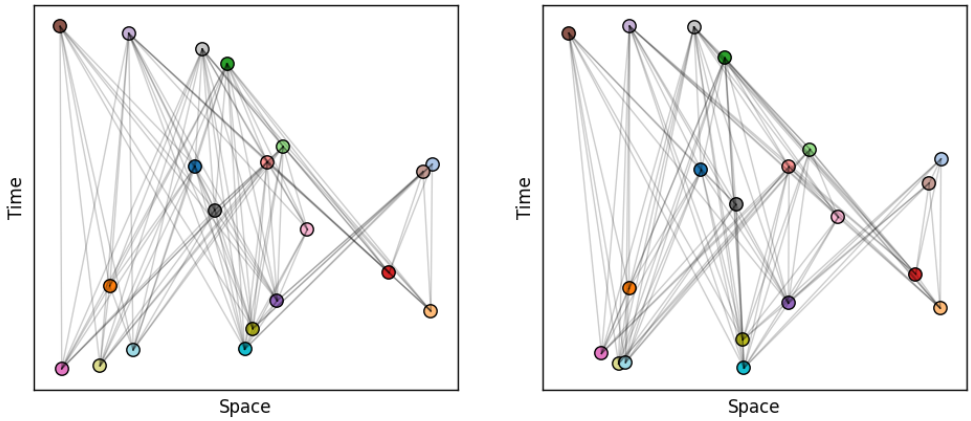


Figure 3: 20 points of a 200 point causal set, (left) and its embedding after the MDS algorithm. The similarity between the two plots shows the success of the embedding algorithm in finding the point's coordinates using only the edges of the graph.

when distances are estimated using graph topology, the pairwise separations will be noisy as they are coarse grained by the discrete graph (see figure 3). To assess the reliability of this algorithm we will test it first on the causal set model described above. We will take the coordinates produced by the Lorentzian MDS algorithm use them to rebuild the graph by again placing edges only between timelike separated pairs. If the overlap between the edges between nodes in the recreated graph, and in the original graph is high, the embedding is an accurate one, and similarly if the overlap is low the embedding is poor. As in [9] we will measure this using the sensitivity (the fraction of the correct edges which were predicted) and specificity (the fraction of correct non-edges which were predicted).

4 Citation Networks in Minkowski spacetime

Our MDS algorithm is able to find accurate embeddings for randomly sprinkled causal sets. We can now attempt to embed networks formed from real social systems, and here we will use citation networks from the arXiv (2003 KDD cup datasets) and the US Supreme Court[24], as well as random DAGs for comparison. Recall that the dimensionality of the embedding is something we can choose, by selecting the $\hat{D}$ largest eigenvalues in the MDS algorithm. To measure the effectiveness of the embedding we will compare the original network, with a new network generated from the coordinates determined by the MDS algorithm. We want to measure the effectiveness of a classifier which predicts edges in the network from the MDS coordinates. To do this we will use the established method of the area under the receiver-operator curve, AUC. Varying a continuous parameter, the sensitivity and specificity of the embedding is measured, and plotted, as in figure 5 and the area under this curve describes the quality of the classifier. The continuous parameter we will vary is the speed of light (or the speed information can be transferred) in the

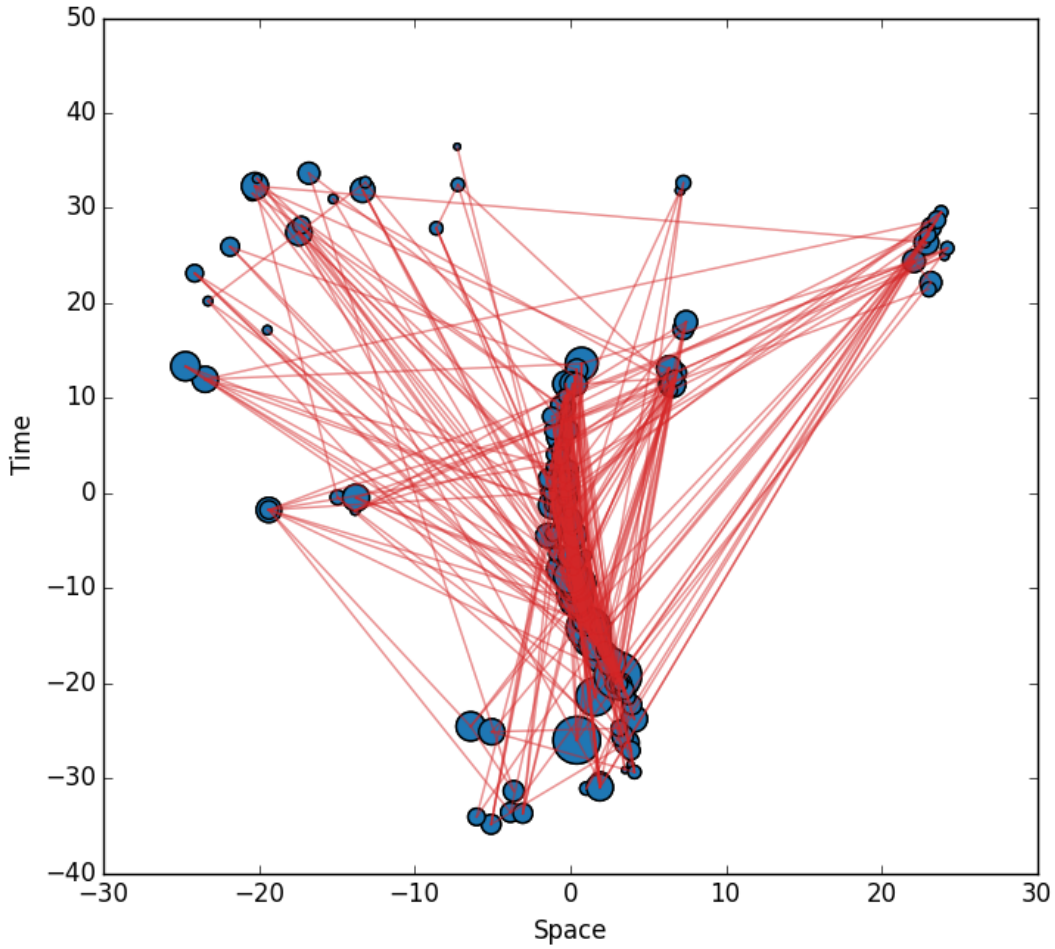


Figure 4: An $D = 1+1$ embedding of the top 200 most cited papers in the `hep-th` citation network. Node size is proportional to the number of citations, and lines correspond to citations amongst these papers. A large group of papers is visible in the middle of the plot, forming a long chain of citations, as well as some more isolated papers on either side. A small number of spacelike citations are visible (those edges more than 45 deg from vertical) because this two-dimensional embedding is not perfect, but only the optimal set of coordinates found by the MDS algorithm.

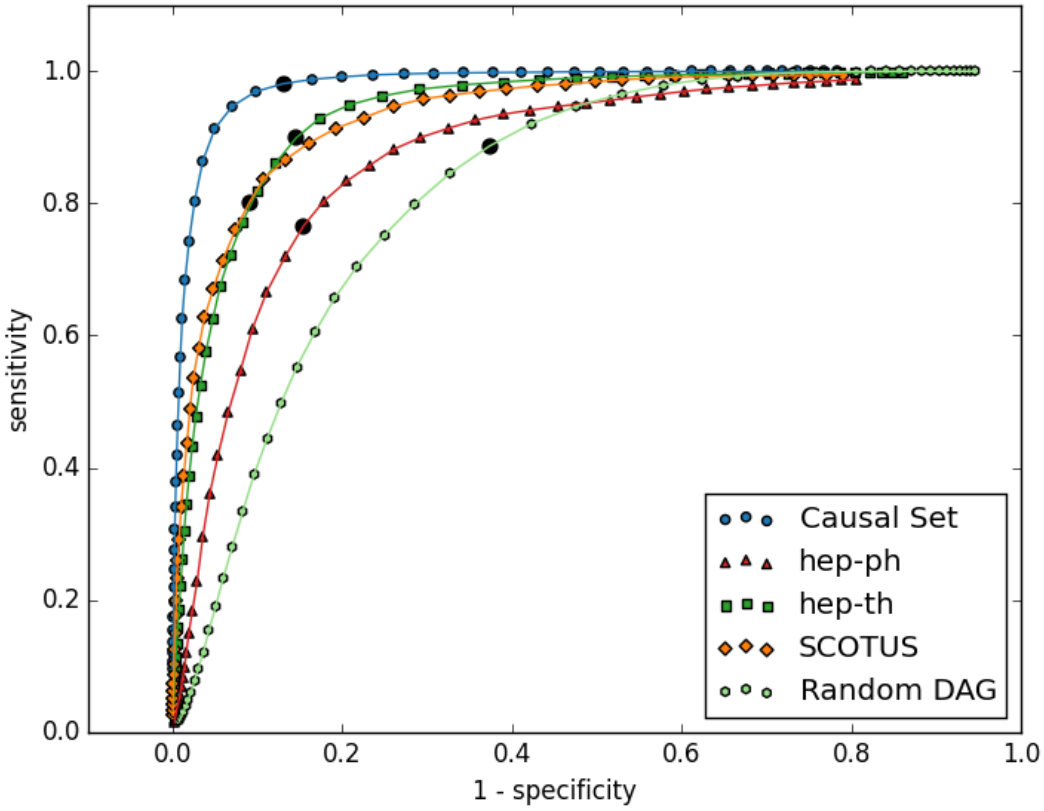


Figure 5: Sensitivity and specificity are measured for various c values for $D = (2+1)$ dimensional embeddings of 5 networks. Where the curve reaches the top-left of the plot we are in the $c \approx 1$ regime ($c = 1$ denoted by the large black point on each curve) where the trade-off between false positives and false negatives is balanced. The shape of these ROC curves measures the effectiveness of the embedding and it is clear that the the causal set performs best, the random DAG worst and the citation networks in between.

Minkowski space, c. Previously, we have set $c = 1$, but varying this speed will change which nodes are connected in new network generated from the MDS coordinates. Now, nodes i and j are connected if their coordinates satisfy

$$-c(x_i^0 - x_j^0)^2 + \sum_{k=1}^{d}(x_i^k - x_j^k)^2 < 0 \tag{7}$$

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

We will measure the ease of embedding a network by taking the mean of the area under this curve for networks of size 100-500. In the case of the citation networks this was done by randomly sampling intervals of this size from the citation network. In the cases of the causal set graph, and the random DAGs, many instances of the required sizes

are stochastically generated.

5 Discussion and Applications

Finding a good geometric embedding of a network provides a powerful tool for the analysis of that network as it allows standard geometric techniques and intuition to be used. Calculations of network properties can be made more efficient, for example, when finding optimal routes from one node to another, the node coordinates provide local information which can improve routing algorithms[25]. Coordinates resulting from geometric embedding also provide a natural visualisation for a network. Such visualisations are used in bibliometrics to help identify distinct fields or assist literature reviews[26].

In the cases of citation analysis, where we conjecture that the spatial dimensions that result from a geometric approach correspond to similarity in the topic of a paper, our approach yields spatial similarities between papers while accounting for the time difference in their publication. Once embedding coordinates are known, the idea that nodes may be 'similar' can be expressed as nodes being close by using an appropriate metric, which need not be the Lorentzian one we used in the construction of the coordinates. Two nodes might be spacelike separated, so have no direct links, yet be close in a Euclidean sense because MDS calculates coordinates globally using information from all vertices and edges. Papers could be recommended if spatially close to others a reader has interest in even if there are no local connections between them, potentially bringing work or authors to the attention of readers who are not aware of them. Spatial similarity can also be used to define clusters. The idea of 'centrality' or importance of a node has a natural representation in terms of the density of points in the geometric neighbourhood of the point linked with the chosen node.

Another use of this approach is where edges in a network are placed primarily according to some geometric rule but their connections are also governed by some smaller second order effect. It may only be possible to measure the smaller effect once we have accounted for the primary geometric one by assigning coordinates using Lorentzian MDS or a similar method. We can see this effect clearly when the geometric embedding is one in real geographic space, such as in [27] where accounting for geographic distance in phone-call data allows more accurate prediction of the second order effect of shared language.

Other approaches to geometric embedding exist in the literature. MDS is characterised by maintaining global separations between pairs. Others, such as Isomap[28, 29] maintain shortest paths between pairs linked by local interactions. That approach may not be as appropriate in pseudo-Riemannian geometries. Firstly, the shortest path locally may not correspond to the geodesic distance like it does in Euclidean space; as discussed above in graphs embedded in Minkowski space it is the longest path which corresponds to the geodesic. Secondly, the idea of local neighbours is less clearly defined if there are different types of separation, or if, as is the case for Minkowski space, the number of nearest neighbours diverges. Another class of embedding approaches are probabalistic,

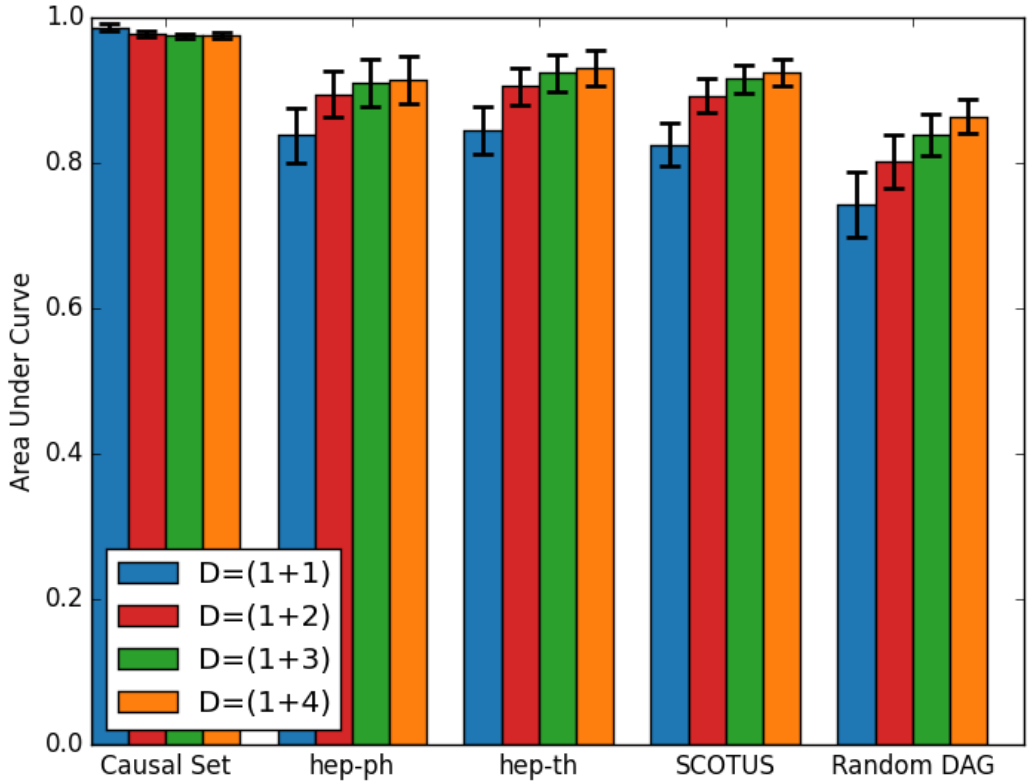


Figure 6: The AUC for 5 types of DAG, embedding in Minkowski spacetime of various dimensions. In this plot, the causal set graphs are created from sprinkling uniformly into a $(d+1)$-dimensional spacetime, then embedded back into that same dimensional spacetime. Their AUC values therefore represent the ideal case, where we know an embedding is possible. The random DAGs show the worst embeddings. Their AUC values represent the success of an embedding for a graph which has no structure. They are noticeably more embeddable in higher dimensions which is because there are more degrees of freedom in which to assign coordinates while maintaining the randomly placed links. The bars show means and standard deviations of the AUC for random DAGs of size $N = 100$ to $N = 500$. Citation networks from the arXiv, and the US Supreme Court fall between these extremes, illustrating the presence of some structure which makes them easier to embed in spacetime. We sampled 250 intervals with between 100 and 500 nodes randomly from each citation network, and the bars show means and standard deviations for the AUC in each case.

such as Stochastic Neighbour Embedding [30]. Although it is beyond the scope of this work, we do not see why such approaches could not be adapted to pseudo-Riemannian manifolds, and the ability to use a mixture of separated images for the same object may prove very useful.

Finally we note that inserting a metric signature into the equations for classical MDS allows it to be used on any metric signature, even though we have focused only on the Lorentzian signature here. To our knowledge this pseudo-Riemannian output is a new development, although some kind of manifold learning techniques exist which can take pseudo-Riemannian manifolds as their input [31, 32]. We note that when performing the Lorentzian network MDS algorithm we often find multiple negative eigenvalues, suggesting that embeddings in spaces with more than one timelike dimension is also possible, as are potential embeddings into Lorentzian manifolds other than Minkowski space, incorporating curvature or preferred directions.

References

[1] J. Dall and M. Christensen. *Random Geometric Graphs*. Phys. Rev. E, 66(1):016121, July 2002.

[2] M. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.

[3] Z. Xie and T. Rogers. *Scale-invariant random geometric graphs*. arXiv:1505.01332, 2015.

[4] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguna. *Hyperbolic geometry of complex networks*. Phys. Rev. E, 82(3):036106, 2010.

[5] D. Krioukov, M. Kitsak, R. Sinkovits, D. Rideout, D. Meyer, and M. Boguna. *Network Cosmology*. Scientific reports, 2:793, 2012.

[6] J. R. Clough and T. S. Evans. *What is the dimension of citation space?* Physica A, 448:235–247, 2016.

[7] Zhihao Wu, Giulia Menichetti, Christoph Rahmede, and Ginestra Bianconi. *Emergent Complex Network Geometry*. Scientific Reports, 5:10073, 2015.

[8] Trevor F. Cox and M.A.A. Cox. *Multidimensional Scaling, Second Edition*. CRC Press, 2000.

[9] D. J. Higham, M. Rasajski, and N. Przulj. *Fitting a geometric graph to a protein-protein interaction network*. Bioinformatics, 24(8):1093–1099, 2008.

[10] J. R. Clough, J. Gollings, T. V. Loach, and T. S. Evans. *Transitive Reduction of Citation Networks*. Complex Networks, 3(2):189:203, 2015.

[11] S. W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-Time*. Cambridge University Press, 1973.

[12] G. Brightwell and M. Luczak. *The mathematics of causal sets*. arXiv:1510.05612, 2015.

[13] Roger N. Shepard. *Attention and the metric structure of the stimulus space*. Journal of Mathematical Psychology, 1(1):54–87, 1964.

[14] H. Minkowski. *Die Grundgleichungen fur die elektromagnetischen Vorgange in bewegten Korpern*, 1910.

[15] L. Bombelli, J. Lee, D. Meyer, and R. D Sorkin. *Space-time as a causal set*. Phys. Rev. Lett. 59(5):521–524, 1987.

[16] L. Bombelli and D. Meyer. *The origin of Lorentzian geometry*. Physics Letters A, 141(5-6):226–228, 1989.

[17] F. Dowker. *Causal sets as discrete spacetime*. Contemporary Physics, 47(1):1–9, 2006.

[18] J. Myrheim. *Statistical Geometry*. CERN–2538, 1978.

[19] B. Bollobás and G. Brightwell. *Box-spaces and random partial orders*. Transactions of the AMS, 324(I):59–72, 1991.

[20] G. Brightwell and R. Gregory. *Structure of random discrete spacetime*. Phys. Rev. Lett. 66(3), 1991.

[21] D. Rideout and P. Wallden. *Spacelike distance from discrete causal order*. Classical and Quantum Gravity, 155013:32, 2009.

[22] F. Dowker, J. Henson, and R. D. Sorkin. *Quantum gravity phenomenology, Lorentz invariance and discreteness*. Modern Physics Letters A, 19(24):1829–1840, 2004.

[23] D. D. Reid. *Manifold dimension of a causal set: Tests in conformally flat spacetimes*. Phys. Rev. D, (October 2002):1–8, 2003.

[24] J. H. Fowler and S. Jeon. *The authority of Supreme Court precedent*. Social Networks, 30(1):16–30, 2008.

[25] D. Krioukov, F. Papadopoulos, A. Vahdat, and M. Boguna. *Curvature and temperature of complex networks*. Phys. Rev. E, 80(3):035101, 2009.

[26] N. J. van Eck and L. Waltman. *CitNetExplorer: A new software tool for analyzing and visualizing citation networks*. Journal of Informetrics, 8(4):802-823, 2014.

[27] P. Expert, T. S. Evans, V. D. Blondel, and R. Lambiotte. *Uncovering space-independent communities in spatial networks.* PNAS, 108(19):7663–8, 2011.

[28] J. B. Tenenbaum, V. de Silva, and J. C. Langford. *A global geometric framework for nonlinear dimensionality reduction.* Science, 290(5500):2319–23, 2000.

[29] S. T. Roweis and L. K. Saul. *Nonlinear dimensionality reduction by locally linear embedding.* Science, 290(5500):2323–6, 2000.

[30] G. E. Hinton and S. T. Roweis. *Stochastic neighbor embedding*. Advances in neural information processing systems, 833–840, 2002.

[31] D. Liu, S. Trajanovski, and P. Van Mieghem. *Reverse Line Graph Construction: The Matrix Relabeling Algorithm MARINLINGA Versus Roussopoulos's Algorithm.* arXiv:1005.0943, 2010.

[32] R. Liu, Z. Lin, Z. Su, and K. Tang. *Feature extraction by learning Lorentzian metric tensor and its extensions*. Pattern Recognition, 43(10):3298–3306, 2010.

A Derivation of Minkowski MDS

Following closely the derivation for Euclidean MDS in [8] we begin with the timelike (negative) and spacelike (positive) square distances between each pair of points, and wish to derive the inner product matrix B, where $B_{ij} = M(\mathbf{x}_i, \mathbf{x}_j)$. We will denote the Minkowski separation between i and j as $S_{ij} = M(\mathbf{x}_i - \mathbf{x}_j, \mathbf{x}_i - \mathbf{x}_j)$.

As usual we first fix the coordinates centre of mass, placing it at the origin, such that $\sum_{i=1}^{n} x_{i\mu} = 0$ for each $\mu = 0, 1, 2, ...d$ in a $d+1$-dimensional space. Remembering that $M(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^{\mathrm{T}} \mathsf{G} \mathbf{x}_j$ with G a diagonal $n \times n$ matrix with -1 in the first row and 1 in every other row (the Minkowski metric) we have:

$$
\begin{aligned}
S_{ij} &= \mathbf{x}_i^{\mathrm{T}} \mathsf{G} \mathbf{x}_i + \mathbf{x}_j^{\mathrm{T}} \mathsf{G} \mathbf{x}_j - 2\mathbf{x}_i^{\mathrm{T}} \mathsf{G} \mathbf{x}_j \\
\frac{1}{n} \sum_{i=1}^{n} S_{ij} &= \mathbf{x}_j^{\mathrm{T}} \mathsf{G} \mathbf{x}_j + \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{\mathrm{T}} \mathsf{G} \mathbf{x}_i - \frac{2}{n} \sum_{i=1}^{n} \mathbf{x}_i^{\mathrm{T}} \mathsf{G} \mathbf{x}_j \\
\frac{1}{n} \sum_{i=1}^{n} S_{ij} &= \mathbf{x}_j^{\mathrm{T}} \mathsf{G} \mathbf{x}_j + \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{\mathrm{T}} \mathsf{G} \mathbf{x}_i - \frac{2}{n} \sum_{i=1}^{n} \sum_{\mu=0}^{d} x_{i\mu} G_{\mu\mu} x_{j\mu}
\end{aligned}
$$

The last term on the right vanishing if the order of the sums is reversed due to the condition we stipulated above. Doing the same summing over s as well gives:

$$
\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} S_{ij} &= \mathbf{x}_j^{\mathrm{T}} \mathsf{G} \mathbf{x}_j + \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{\mathrm{T}} \mathsf{G} \mathbf{x}_i \\
\frac{1}{n} \sum_{j=1}^{n} S_{ij} &= \mathbf{x}_i^{\mathrm{T}} \mathsf{G} \mathbf{x}_i + \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j^{\mathrm{T}} \mathsf{G} \mathbf{x}_j \\
\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} S_{ij} &= \frac{2}{n} \sum_{i=1}^{n} \mathbf{x}_i^{\mathrm{T}} \mathsf{G} \mathbf{x}_i
\end{aligned}
$$

We can now write down each element in the matrix B we are trying to calculate.

$$
\begin{aligned}
B_{ij} &= \mathbf{x}_i^{\mathrm{T}} \mathsf{G} \mathbf{x}_j \\
B_{ij} &= -\frac{1}{2} \left[S_{ij} - \mathbf{x}_i^{\mathrm{T}} \mathsf{G} \mathbf{x}_i - \mathbf{x}_j^{\mathrm{T}} \mathsf{G} \mathbf{x}_j \right] \\
B_{ij} &= -\frac{1}{2} \left[S_{ij} - \frac{1}{n} \sum_{i=1}^{n} S_{ij} + \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{\mathrm{T}} \mathsf{G} \mathbf{x}_i - \frac{1}{n} \sum_{j=1}^{n} S_{ij} + \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j^{\mathrm{T}} \mathsf{G} \mathbf{x}_j \right] \\
B_{ij} &= -\frac{1}{2} \left[S_{ij} - \frac{1}{n} \sum_{i=1}^{n} S_{ij} - \frac{1}{n} \sum_{j=1}^{n} S_{ij} + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} S_{ij} \right]
\end{aligned}
$$

Which gives us the the matrix B from the distances.

In the Euclidean case all of the eigenvalues of B are positive (or zero). In general, the signs of eigenvalues of the matrix B will follow the signature of the metric G. For any

metric, B is symmetric and so can can be decomposed into orthogonal eigenvectors, U and eigenvalues $\mathbf{\Sigma}$.

$$\mathsf{B} = \mathsf{U}\mathbf{\Sigma}\mathsf{U}^{\mathrm{T}}$$

We aim to find a solution to the equation $\mathsf{B} = \mathsf{X}\mathsf{G}\mathsf{X}^{\mathrm{T}}$. Trying $\mathsf{X} = \mathsf{U}\mathsf{D}$ where D is some real diagonal matrix gives

$$\begin{aligned}
\mathsf{U}\mathbf{\Sigma}\mathsf{U}^{\mathrm{T}} &= \mathsf{X}\mathsf{G}\mathsf{X}^{\mathrm{T}} \\
\mathsf{U}\mathbf{\Sigma}\mathsf{U}^{\mathrm{T}} &= \mathsf{U}\mathsf{D}\mathsf{G}\mathsf{D}\mathsf{U}^{\mathrm{T}} \\
\mathbf{\Sigma} &= \mathsf{D}\mathsf{G}\mathsf{D}
\end{aligned}$$

Assuming that the metric G is diagonal we then have that $\Sigma_i = D_i^2 G_i$ and since D is real, the signs of the elements of $\mathbf{\Sigma}$ must equal those of G.