Imperial College London

An Exploration of Possibly Natural Geometries on Directed Acyclic Graphs

Author: 01346808
Supervisor: Dr. Tim Evans
Assessor: Professor Kim Christensen
THEO-Evans-1
Word Count: 9740

Acknowledgements

Thank-you to my supervisor for invaluable guidance and discussion, and thank-you to my friends and family for their unconditional wisdom and support.

Abstract

Network geometry is a young field mainly focused on undirected, Riemannian geometries. Recently, however, work has been undertaken to understand directed acyclic graphs from a (mainly Lorentzian) geometric point of view. In this writing, by building off of already understood Lorentzian network geometry, we will explore natural extensions of the undirected communicability, conservative diffusion, and general transition metrics. We will also compare the properties of various path algorithms to actual geodesics on a Lorentzian embedding space, then use analogies between the embedding space and the graph to formulate a semi-local, embedding-space-independent path algorithm which mimics the motion of a particle on a geodesic. We will find that this mimicry is ruined in the dense limit, and that to truly mimic this motion requires the introduction of a secondary length scale. We propose that this is a feature of all purely graph-dependent, non-global algorithms, such that any natural dynamics of information on a directed acyclic graph will inform a length scale around which to centre procedures for choosing an MHKM conformal factor. We will also study how the various path algorithms behave with regard to their deviations away from the true geodesic, and their length scalings.

1 Introduction

Graph theory has been an incredibly important tool in the field of complexity for understanding highly interconnected systems. With a simple graph, one is able to model social circles, telecommunications, the internet, the spread of disease, the spread of rumours, economic trade, and much more [13]. Normally, these networks do not start embedded in any background space; where they are, the embedding is typically Euclidean. Some networks, however, are embedded in (or otherwised coupled to) time. As such, they have an element of causality, disposing them to directed acyclic graph models. Examples include food webs, citations, software dependencies, and logical arguments. In spite of their ubiquity, directed acyclic graphs have mainly been explored from the point of view of causal set gravity, where they are used to model spacetime, rather than from a complexity point of view [21]. This writing will thus be concerned with futher exploring directed acyclic graphs from a complexity point of view.

One of the most interesting aspects of any complex network is its geometry: how can this graph be embedded into a background, or modelled as a metric space? While the network geometry of undirected graphs is relatively well undertsood [2], the network geometry of directed acyclic graphs has only been studied from the perspective of spacetimes, with a corresponding spacetime longest path geodesic structure. The analogue of this type of geodesic structure in the undirected case is simply Riemannian shortest path stucture. Of course, the dynamics of information on complex networks is not always conducive to the most obvious geometries. As such, the purpose of this investigation will be to explore the different geometries on directed acyclic graphs which might be useful for modeling information dynamics on complex systems.

We will take two approaches to achieve this. In the first approach, we will start from dynamical metrics already used in the study of undirected graphs, and try to find natural generalizations to the directed acyclic case. In the second approach, we will analyze various path algorithms to determine what path properties are conducive to spacetime geodesic structure. We will then cross reference this with the basic criteria for a dynamics to be physical, and attempt to formulate a path algorithm which gives both the well studied spacetime structure, and which might be found in nature. From this we will try to extract some general restrictions on natural dynamical spacetime geodesic path algorithms. Our second approach will be executed mainly computationally.

Throughout the writing we take $c = 1$ and $(+ - --)$.

2 Graph Theory

2.1 Graphs, DiGraphs, DAGs, and Causets

Basic Definitions A graph G is a set of nodes V connected by a set of edges E, such that $G = (V, E)$. An edge $e \in E$ can be written as either a set $e = \{a, b\}$ for $a, b \in V$, or an ordered pair $e = (a, b)$, where we say a precedes b and b succeeds a. In the former case we have an undirected edge, and in the latter case we have a directed edge. Typically graphs only have one type of edge, and are called either undirected or directed graphs. We will refer to the latter case as digraphs.

The degree of a node is the number of nodes it is directly connected to, and is usually denoted by k. In the undirected case, we distinguish between the nodes indegree and outdegree, which count the number of predecessors and successors respectively.

Paths Of particular interest to us will be walks and paths. A walk is a sequence of nodes with edges connecting neighbouring nodes in the sequence, while a path is a walk with no repeating nodes. In the case of digraphs, the direction of the edges must follow the order of the sequence (note that some authors use a slightly different definition). A cycle is a path which starts and ends at the same node.

Directed Acyclic Graphs and Causets Using these ideas we can identify a type of digraph which has no cycles, typically refered to as a Directed Acyclic Graph (DAG). Directed acyclic graphs are conducive to our notion of causality, and can be used to model citation networks, food webs, data compression trees, and even the conformal structure of spacetime [4, 14, 6, 15]. In this sense, they are closely related to causal sets, or causets.

A causet is a set equipped with a partial order $\prec$, where $a \prec b$ means a is in the past of b. Partial orders obey [19]

- Reflexivity – $a \preceq a$
- Antisymmetry – $a \preceq b$ and $b \preceq a \implies a = b$
- Transitivity – $a \preceq b$ and $b \preceq c \implies a \preceq c$

and are distinct from total orders in that they may contain elements that are not related, which we denote by $a \not\prec b$. We say the interval, or sometimes causal diamond, between two points a and b is all points c such that $a \prec c \prec b$.

Clearly all DAGs have a natural partial order, where $\forall a, b$ in the DAG, we say $a \prec b$ if there exists a path from a to b. If $a \prec b$ implies there is an edge directly from a to b, then we say the DAG is transitively complete. Conversely, if $a \prec b$ implies there is an edge from a to b iff there is no node between a and b, ie. an edge if $\nexists c$ such that $a \prec c \prec b$, then we say the graph is transitively reduced [4]. Edges in transitively reduced graphs tell us about the immediate future of nodes, while edges in transitively complete graphs tell us about the entire future of nodes.

2.2 Matrix Notation

Basic Notation It is useful to encode graph structure using matrix notation. Let us label each node in a graph of size N by the set of integers up to N. Then define x_i as an N dimensional vector which is zero everywhere except for the ith position, where it is 1. We can then define a scalar field on the graph as the vector ϕ such that $\phi_i = \phi \cdot x$ is the value of the field at the ith node.

Adjacency We can also define the adjacency matrix A, where $A_{ij} = 1$ if nodes i and j are connected by an edge, and zero otherwise. For digraphs, we instead define a successor matrix F, and a predecessor matrix $P = F^T$. For these, $F_{ij} = 1$ if there is an edge pointing from i to j, and similarly for P. For transitively complete DAGs, F is interpreted as a future adjacency matrix, while P is a past adjacency matrix. These matrices have the property that $(A^n)_{ij}$ is the number of walks of length n connecting i and j [17]. The acyclicity of DAGs means all walks are necessarilly paths, such that F^n and P^n encode the number of length n paths between nodes.

Symmetry With this notation we also introduce a gauge symmetry in the labeling of the nodes. This symmetry can be represented using a permutation matrix S, where each row and each column of the matrix contains a single 1, and is zero everywhere else. This can be represented by $S = x_{i_1} \oplus x_{i_2} \oplus \cdots \oplus x_{i_N}$, where i_n is the old label of the nth node. Applying this to an N dimensional vector just rearranges all the components of the vector, hence, a relabeling.

3 Manifolds

3.1 Distance Metrics

A distance, or metric, is a map $d : X \times X \to [0, \infty)$

- Indiscernibility: $d(x, y) = 0 \iff x = y$
- Symmetry: $d(x, y) = d(y, x)$
- Triangle Identity: $d(x, z) \leq d(x, y) + d(y, z)$

where X is called the metric space. In this thesis we will be working with separations in spacetime, where the triangle identity is reversed, indiscernibility doesn't apply, and $d(x, y)$ maps into all $\mathbb{R}$; and DAGs, where the symmetry condition is perhaps not desirable. It is known that

- $d(x, y) = d(y, x)$
- $d(x, z) \geq d(x, y) + d(y, z)$

implies $d(x,y) \leq 0$. This of course makes the reversed triangle identity moot. Therefore we must have $d(x,y) \neq d(y,x)$. Let us try to extend these ideas. From the reversed triangle identity we have $d(x,x) = 0$, and so

$$0 \geq d(x,y) + d(y,x) \tag{1}$$

We generally don't want much asymmetry between the past and the future, and so the only natural choice is to saturate this inequality $d(x,y) = -d(y,x)$. But then we can manipulate the reversed triangle identity into the form

$$d(x,z) \leq d(x,y) + d(y,z) \tag{2}$$

meaning

$$d(x,z) = d(x,y) + d(y,z) \tag{3}$$

and we are forced to work in one dimension. It is not difficult to use such a function to impose, or nearly impose, a total order. Simply define $x < y$ if $d(0,x) < d(0,y)$ for some origin 0 of choice. This hints that if we wish to use the reversed triangle identity in multiple dimensions, we need to order the elements in X. In the case of spacetime separations, the symmetry condition $d(x,y) = d(y,x)$ is actually maintained, but the reversed triangle identity only applies if $x \prec y \prec z$ or $x \succ y \succ z$.

For most spaces X with a natural metric, the triangle identity implies that the straightest path between two points is also the shortest path. In the case of spacetime separation, the opposite holds: the longest path is the straightest (see Fig. 1). These ideas hint that a longest path dynamics is associated with a partial ordering, ie. a causal structure. We will see that longest path structures will act as a starting point for many of our investigations.

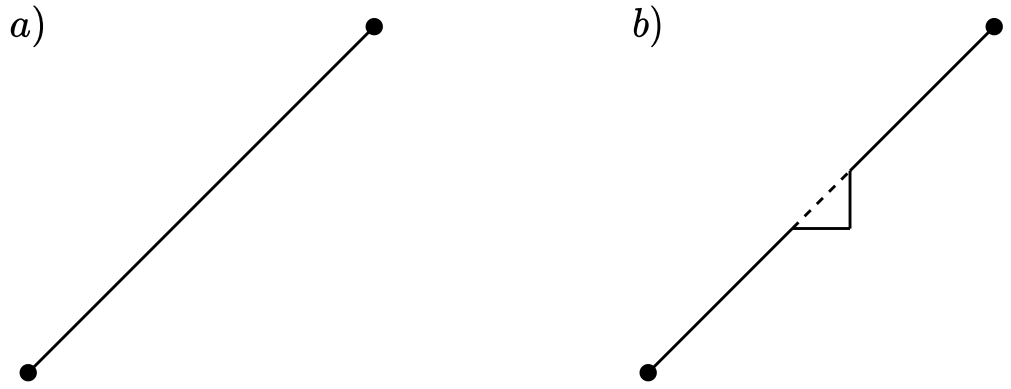


Figure 1: The curve a) depicts the straightest path in a metric space X equipped with a distance metric obeying the triangle identity. b) depicts the same curve perturbed by a small deviation. The triangle identity ensures curve a) is shorter than curve b). Because all curves fixed between the two endpoints can be formed through a series of such perturbations, curve a) is the shortest path in the space. This is trivially generalized to separation functions obeying the reversed triangle identity and longest paths in the space.

3.2 Riemannian and Lorentzian Manifolds

Basic Definition A differential manifold is a geometric space $\mathcal{M}$ equipped with a bundle of tangent spaces $T\mathcal{M}$. The tangent space $T_x\mathcal{M}$ at the point $x \in \mathcal{M}$ is a vector space of the same dimension as the manifold. The manifold can be equipped with a metric $g(x)$ at each of its points $g(x) : T_x\mathcal{M} \times T_x\mathcal{M} \to [0,\infty)$, where $T_x\mathcal{M}$ is the metric space. In this case, the manifold is Riemannian. If we relax the metric conditions as in the previous section, then our manifold is pseudo-Riemannian. If we use the spacetime separation in particular, then our manifold is Lorentzian [23].

Geodesics A path is a continuous map $\gamma : [0,1] \to \mathcal{M}$, and has tangent vectors $X(\tau) = \frac{d\gamma(\tau)}{d\tau}$, where $\tau \in [0,1]$ parameterizes the path. The metric is related to infinitesimal separations along paths in the sense that

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = g(X(\tau), X(\tau)) d\tau^2 \tag{4}$$

where we have used the Einstein summation convention, and can see the metric is represented by a matrix. In the case of d dimensional Riemannian manifolds, the metric will have d positive eigenvalues, whereas in the Lorentzian case it will have 1 positive eigenvalue and $d-1$ negative eigenvalues. The sign of these eigenvalues is called the metric signature, and in the Lorentzian case is related to the fact that points can have negative separation. We classify the separations as timelike ($ds^2 > 0$), spacelike ($ds^2 < 0$), or lightlike $ds^2 = 0$. We say a path is causal if $g(X(\tau), X(\tau)) \geq 0 \ \forall \tau$, ie. if every tangent on the curve is either timelike or lightlike. Two points x and y are causally related if there exists a causal path between them. The causal diamond between x and y is the set of points which are causally related to both x and y. In terms of partial orders, if $x \prec y$, then $\{z | x \prec z \prec y\}$ is the set of points forming the causal diamond [23].

We relate vectors between tangent spaces by a connection/covariant derivative ∇. We say the covariant derivative $\nabla_X T$ tells us how the tensor T changes as we move along the tangent vector X. A path is a geodesic if motion along the path induces the least change in that motion, ie. if $\nabla_{X(s)} X(s) = 0$. In this sense, geodesics are the straightest paths on the manifold, and are governed by a second order differential equation[24].

Causal Structure MHKM showed that given just the causal structure of a $d > 2$ Lorentzian manifold, one can determine its metric up to a conformal factor [15]. That is, if we know whether $x \prec y$, $x \succ y$, or $x \not\sim y$ for all $x, y \in \mathcal{M}$, then we know $\Omega(x)^2 g(x)$ up to the (real) local scaling $\Omega(x)$. The simple intuition for this is that a positive scaling factor doesn't change the sign of $g(X, X)$, so timelike-forward, timelike-backward, and spacelike vectors remain as they are. Thus, up to a volume element (or some other parameter), a causet can represent a spacetime.

Symmetry Finally, we note that, just like graphs, manifolds cannot be physically altered by the coordinate system we use to describe them. This gauge symmetry is true of all manifolds, and is called diffeomorphism invariance.

3.3 1+1d Minkowski Space

We will frequently be using 1+1d Minkowski space, where the metric in Cartesian coordinates is

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu = dt^2 - dx^2 \tag{5}$$

and $\nabla = \partial$, so the geodesics are just straight lines. The geodesic separation between points is then just $s^2 = t^2 - x^2$. We can use the separation to write down hyperbolic coordinates, $t = z \cosh \eta$ and $x = z \sinh \eta$, where z is the separation from the origin, and η is the rapidity. Hyperbolic coordinates reveal that curves of constant separation are hyperbolas. The metric in these coordinates is

$$ds^2 = dz^2 - z^2 d\eta^2 \tag{6}$$

We will also make use of lightcone coordinates, where $u = t + x$ and $v = t - x$. The metric is

$$ds^2 = du dv \tag{7}$$

The causal diamond between the points $(0, 0)$ and $(x, t) = (s \cosh \eta, s \sinh \eta) = (u, v)$ is the set of points in the intervals $u \in [0, 1]$ and $v \in [0, 1]$. The causal diamond has an area $A = uv = t^2 - u^2 = s^2$. We call this a separation s causal diamond.

3.4 Graphs from Manifolds

The Poisson Point Process For our purposes we want to generate a graph $G = (V, E)$ which can be embedded into some manifold $\mathcal{M}$. We start by choosing points from $\mathcal{M}$ uniformly at random and putting them into V. This will lead to a set of points distributed according to

$$p_n(B) = \frac{(\rho V_B)^n}{n!} e^{-\rho V_B} \tag{8}$$

where $p_n(B)$ is the probability of finding n points from V in a region $B \subset \mathcal{M}$ of volume V_B. Here, the distribution of points has an average density ρ, and hence a characteristic length scale $\xi = 1/\rho^{\frac{1}{d}}$. This is called a Poisson Point Process, or PPP [11].

Hard Connection We now need a way of connecting nodes that reflects the manifold structure. A typical process is the hard connection model, where two points are connected by an edge if they fall within some distance r of one another (a graph formed using geometrical rules like this is called a Random Geometric Graph, or RGG[25])

Let us consider what happens when we saturate this process on a Riemannian manifold $\mathcal{M}$, such that the size of the graph $N = |\mathcal{M}|$. Naively we might expect the graph to imitate the manifold exactly, however this is not the case. This becomes clear when considering the shortest path between two points on the graph, which appears as a discontinuous set of jumps on the embedding manifold. This is in contrast to the shortest path on the embedding manifold itself, which is a continuous geodesic (see Fig. 2). The obvious solution to this problem is to adjust the connection rule to respond to the density of points; in this case we should set $r = \xi$. From this thought experiment, we see that for the generated graph to reproduce the appropriate behaviour, we need the length scale of the PPP and the length scale of the connection process to communicate. If they don't, there will be disparities in the graph and manifold dynamics.

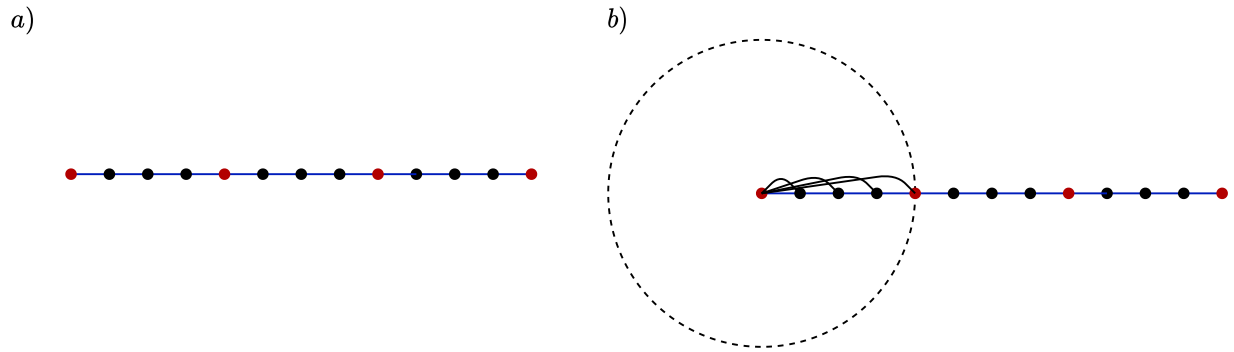


Figure 2: a) depicts nodes in a graph (black and red) chosen from an embedding manifold $\mathcal{M}$ lying on a geodesic (blue) of $\mathcal{M}$. b) depicts the hard RGG process through which the nodes become connected by edges (black) (the process is only illustrated for the leftmost node). Even though the shortest path on $\mathcal{M}$ is a continuous line, because the radius of connection is finite, the shortest path on the graph is a discontinuous set of jumps between nodes (red).

Directed Acyclic Graph Generation If we wish to form a DAG from a manifold, we will have to specify a directionality on the edges while also ensuring the resulting graph contains no cycles. The easiest way to do this is to start with a Lorentzian manifold and simply draw an edge from a to b if $a \prec b$ on the manifold. This is actually equivalent to a hard connection process with a separation radius of zero.

Again, we can consider what happens as $N \to |\mathcal{M}|$. From the MHKM theorem, we know that this will only tell us about the manifold up to a conformal factor. Thus, if we want to recreate the manifold, we must specify the volume element or similar. A natural way to do this might be to coarse grain subgraphs of the manifold into single nodes with volumes proportional to the sizes of the corresponding subgraphs. This is analogous to the thought experiment from the previous problem, although we do not have a natural way to constrain our ambiguous parameter.

4 Network Geometry and Dynamics

Network geometry is a very broad field concerned with applying geometric techniques to the study complex systems. Some of the main topics include network self-similarity, hyperbolic latent spaces, topological analysis, dynamical metrics, and, as above, manifold embedding [20, 12, 5, 18, 2] .

In this section we will focus on generalizing dynamical metrics on undirected graphs into the directed acyclic case. In particular, we will generalize metrics typically used to characterize connectivity and conservative diffusion (although we will omit infectious diffusion metrics). We will also take inspiration from path integral theory to formulate a possible route between general transition dynamics on graphs and geometry.

4.1 Communicability Distance

A First Generalization The first natural metric is based upon the likelihood of information exchange between nodes i and j. Using the properties of the adjacency matrix, we can find the number of walks of length

n connecting node pairs. If we weigh shorter walks as more likely we get a communicability matrix [8]

$$G = \sum_{n=0} a_n A^n \tag{9}$$

where $a_n < a_m$ for $n > m$. It is common to use $a_n = 1/n!$, such that

$$G = e^A \tag{10}$$

from which we define the communicability metric

$$d_{ij} = G_{ii} + G_{jj} - G_{ij} - G_{ji} \tag{11}$$

For generic, undirected graphs, it is necessary for shorter walks to have higher weights, otherwise the graph would have diverging communicability between all its nodes. The directed acyclicity of DAGs, however, prevents infinite walks between two finitely separated points, and so this series truncates. This is important as we expect DAGs to have a natural longest path structure, so we want $a_n > a_m$ for $n > m$, which would not converge if the series were infinite. In the context of geodesic paths, the path length is the action, thus motivating $a_n = e^{\chi n}$ [24]. This gives

$$G^{(F)} = \sum_{n=0} e^{\chi n} F^n \tag{12}$$

from which the communicability metric is defined similarly to (11). F, $G^{(F)}$, and the metric will in general not be symmetric, but under the appropriate node labeling, will be triangular. If we still want the metric to be symmetric, then we can do $d^+_{ij} = d^{(F)}_{ij} + d^{(F)}_{ji}$, which is equivalent to $d^+_{ij} = d^{(F)}_{ij} + d^{(P)}_{ij}$. For consistency we might also wish to do $a_n = e^{-\chi n}$ in the undirected case. This calculation is insignificant regarding DAG geometry, and is left in appendix A for interest.

Generalization Through Entropy The acyclicity of DAGs also ensures all walks on the graph are paths. Thus, F^n counts the number of length n paths between the node pairs. The total number of connecting paths is thus

$$\Omega = \sum_{n=1} F^n \tag{13}$$

As previously mentioned, DAGs are closely related to Lorentzian manifolds. Let us then consider D, the number of causal paths between two points, A and B, which we choose to be separated by s in a continuous spacetime. If Σ is the locus of points a separation $s - \delta s$ from B, and which are inside the causal diamond between A and B, then clearly all paths connecting A and B must pass through Σ. If we choose a spacetime with sufficient symmetry that the number of paths between two points is only dependent on their separation, then

$$D(s) = \frac{V(\Sigma)}{\delta s^{d-1}} D(s - \delta s) \tag{14}$$

where $V(\Sigma)$ is the $d-1$ dimensional volume of Σ. Taylor expanding around s

$$\begin{aligned} \Big(1 - \frac{\delta s^{d-1}}{V(\Sigma)}\Big) D(s) &= \delta s \frac{dD(s)}{ds} \\ \implies D(s) &= D(0) e^{s/\xi} \end{aligned} \tag{15}$$

where in a continuous spacetime, ξ is some infinitesimal separation, and $D(0)$ is vanishing. In such a case, this expression is meaningless. However, by noting the similarity between the interpretation of Ω and $D(s)$, we write down the metric

$$d = \xi \log \Omega \tag{16}$$

where $D(0) := 1$, and ξ is now a characteristic length scale for the graph. Although the derivation of (15) is only valid when D is solely dependent on s, it is clear this metric is an entropy of random walks, and so has a natural interpretation for all DAGs (not just those mimicking the spacetime we used). Thus, we have a natural generalization of the information communicability in a DAG which is literally an entropy term. Note that this does not obey $d_{ij} + d_{jk} = d_{ik}$, as (15) is only "valid" because the continuity of spacetime makes it divergent.

Expanding to higher order in δs we can get a more general expression for (15), although it is not sufficiently tractable to lead to any particular metric. Its calculation is left in Appendix B for interest.

4.2 Diffusion Distance

Generalization Through Conservation Another natural metric of interest is based upon diffusive processes. For an undirected graph, diffusion is governed by the graph Laplacian, $L = K - A$, where K is a diagonal matrix of the degrees of each node ie. such that K_{ii} is the degree of the ith node. Graph diffusion is posed in the following way: let $\phi(t)$ be a time dependent scalar field on the graph. If the rate of diffusion of the field along edges between nodes is proportional to the difference between the field at those nodes, then the overall scalar field will obey

$$\frac{d\phi}{dt} = -kL\phi \tag{17}$$

where k is the proportionality constant controlling the diffusion [17]. In graph theory this leads to the identification of L as an analogue for the differential Laplacian ∇^2. This is obviously very useful for understanding the dynamics of information on graphs, and can be used for topological characterization [7].

The analogue of the Laplacian on DAGs would be the d'Alembertian. Unfortunately, the d'Alembertian describes wavelike behaviour, and therefore cannot be retrieved by simple diffusion arguments. Current efforts for understanding the graph d'Alembertian come mainly from the causet gravity program, and take a more brute force principled deduction approach [1]. For now, we will not concern ourselves with differential operators, but instead focus on the metrics retrieved from diffusion dynamics. In this case, the typical metric is

$$d_{ij}(t) = k|\phi_i(t) - \phi_j(t)| \tag{18}$$

Let us now try to formulate something similar for DAGs. In this case, time is embedded directly into our system, and so it is sufficient to enforce conservation directly between the nodes. The flow should be isotropic, so if node i has k_i^{out} successors, then each successor should receive $\phi \cdot x_i / k_i^{out}$. The set of nodes in the past of i is $P^T x_i$. If K^{out} is a diagonal matrix of outdegrees, then the predecessors of i contribute $(K^{out})^{-1} P^T x_i \cdot \phi$ to the field at ϕ_i. Thus

$$\phi \cdot x_i = (K^{out})^{-1} P^T x_i \cdot \phi \tag{19}$$

If we tensor sum over the x_i, we get $x_1 \oplus x_2 \oplus \cdots \oplus x_N = \mathbb{1}$. Thus, summing over (19) like this

$$\phi = P(K^{out})^{-1}\phi \tag{20}$$

therefore reducing this to an eignvector problem. If we wish to solve this we will have to enforce some kind of boundary conditions.

Example Let us consider a simple example to see how this might work. The DAG depicted in Fig. 3 is described by

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \tag{21}$$

and $K^{out} = \mathrm{diag}(2, 1, 1, 0)$. If we blindly apply (20) and solve for ϕ, we get $\phi_1 = 0$, $\phi_2 = \phi_1/2$, $\phi_3 = \phi_1/2$, $\phi_4 = \phi_2 + \phi_3 = \phi_1$. These equations would be perfectly sensical if ϕ weren't unsourced, ie. if $\phi_1 \neq 0$. One simple solution to this would be to connect the last node back up to the first node to close the system. Then we get

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \tag{22}$$

and $K^{out} = \mathrm{diag}(2, 1, 1, 1)$. The unit eigenvalue eigenvector of the matrix $P(K^{out})^{-1}$ is $\phi \propto (1, 1/2, 1/2, 1)^T$, which is exactly what we would expect. The metric is again of the form

$$d_{ij} = |\phi_i - \phi_j| \tag{23}$$

which in this example is

$$d \propto \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix} \tag{24}$$

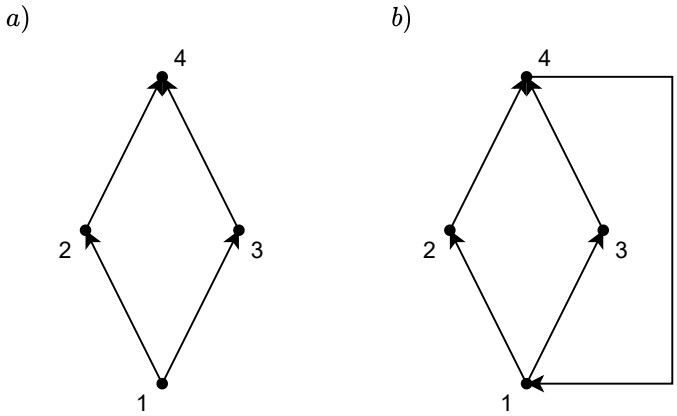


Figure 3: The DAG a) is described by the predecessor matrix given in (21). The digraph b) is just a) with circular boundary conditions. b) is described by (22), and thus has a nontrivial diffusion eigenvector.

4.2.1 A General Geometry from Transition Dyanmics

A Loose Derivation The above two dynamics might be categorized under a more general "transition dynamics", where motion on the graph is determined by the flow of information from node to node. We can formulate a general transition dynamics using the transition matrix T, where T_{ij} is the probability of movement from node i to node j. On graphs themselves, this is not much more spectacular than the adjacency matrix. On a manifold, however, the transition matrix is more interesting.

On a manifold, the transition matrix becomes a two argument distribution $T(x, y)$. We can represent the probability density of moving along some path $x(\tau)$ as

$$\begin{aligned}\rho[x(\tau)] &= \prod_{\tau} T(x(\tau), x(\tau + \delta\tau)) \\ &= \prod_{x(\tau)} T(x, x + \delta x) \\ &= \exp\left(\sum_{x(\tau)} \log T(x, x + \delta x)\right)\end{aligned} \tag{25}$$

As with the adjacency matrix, we assume $T(x, x) = 0$. Thus, Taylor expanding the logarithm would be nonsense. To fix this we can simply allow information to stay still, such that $T(x, x) \neq 0$. We will see, however, that even keeping the singularity does not affect the interpretation of the result. Thu, Taylor expanding

$$\rho[x(\tau)] = \exp\left(\sum_{x(\tau)} \log T(x, x) + \delta x^i \frac{\partial \log T}{\partial x^i}_{(x,x)} + \frac{1}{2}\delta x^i \delta x^j \frac{\partial^2 \log T}{\partial x^i \partial x^j}_{(x,x)} + \cdots\right) \tag{26}$$

where the derivatives act on the *second argument* of the function. We will now assume that $T(x, y) \geq 0$ for $x \neq y$, and so $\log T$ has an extremum at $x = y$. As for the first term, we absorb it into the normalization of ρ, removing that part of the singularity (when $T(x, x) \neq 0$, we must assume two related, but stronger, conditions: 1. T(x,y) has an extremum at $x = y$, and 2. $T(x, x)$ takes the same value for all x). This leaves us with

$$\begin{aligned}\rho[x(\tau)] &= \exp\left(\sum_{x(\tau)} \frac{1}{2}\delta x^i \delta x^j \frac{\partial^2 \log T}{\partial x^i \partial x^j}_{(x,x)} + \cdots\right) \\ &= \exp\left(\frac{1}{2}\int_{x(\tau)} dx^i dx^j \frac{\partial^2 \log T}{\partial x^i \partial x^j}_{(x,x)}\right) \\ &= \exp\left(\frac{\delta\tau}{2}\int_{x(\tau)} d\tau \frac{dx^i}{d\tau}\frac{dx^j}{d\tau}\frac{\partial^2 \log T}{\partial x^i \partial x^j}_{(x,x)}\right)\end{aligned} \tag{27}$$

In path integral theory it is common to have

$$p(x, y) = \int \mathcal{D}x(\tau) e^{S/S_0} \tag{28}$$

where $p(x,y)$ is the probability of following some path from x to y, where S is a free energy or action to be extremized, and S_0 is a scale for that action. Comparing this with (27) we see that the integral in the exponent can be identified with the action. We therefore have

$$S = \int_{x(\tau)} d\tau \frac{dx^i}{d\tau}\frac{dx^j}{d\tau} g_{ij} \propto \int_{x(\tau)} dx^i dx^j g_{ij} \tag{29}$$

with

$$g_{ij} = \frac{1}{2} S_0 \frac{\partial^2 \log T}{\partial x^i \partial x^j}_{(x,x)} \tag{30}$$

where the $\delta\tau$ has been absorbed into S_0. With the appropriate parameterization of the path, the extremization of S is identical to the extremization of a path length on a manifold equipped with the metric g_{ij} [24]. We therefore have a route between the transition function and geometry.

We note, however, that we have not completely removed the logarithmic singularity. Again, the most straightforward way to deal with this would be to choose the transition function more carefully, and avoid the singularity altogether. In some cases, however, the singularity can be absorbed into the action scaling. We will now work a simple example to illustrate this.

Example Let us consider a one dimensional line with a particle transitioning according to a simple random walk procedure. That is, at each instant, the particle has an equal chance of stepping some infinitesimal distance either left or right. The corresponding transition matrix $T(0,y)$ is two delta functions hugging the origin. We can write this as the limit of three gaussians as follows

$$T(0,y) = \frac{T_0}{\sqrt{2\pi}\sigma}\Big(\frac{1}{2}e^{-\frac{(y-\sigma)^2}{2\sigma^2}} + \frac{1}{2}e^{-\frac{(y+\sigma)^2}{2\sigma^2}} - \epsilon e^{-\frac{y^2+\sigma^2}{2\sigma^2}}\Big) \tag{31}$$

This is two gaussians hugging the origin, with a third gaussian in the middle ensuring $T(0,0) \propto 1-\epsilon$. When $\epsilon = 1$, $T(0,0) = 0$, and we get the logarithmic singularity. If we replace $-y/\sigma$ with $Q = \frac{x-y}{\sigma}$, then do some algebra, we can write the full transition function as

$$T(x,y) = \sqrt{\frac{1}{2\pi e}}\frac{T_0}{\sigma}\Big(\cosh Q - \epsilon\Big)e^{-Q^2/2} \tag{32}$$

The transition function then has an extremum at $x = y$, and the same value of $T(x,x)$ for any x, such that we can apply the above derivations. The metric is then

$$g = \frac{1}{2} S_0 \frac{d^2 \log T}{dy^2}_{(x,x)} = \frac{S_0}{2\sigma^2}\frac{\epsilon}{1-\epsilon} \tag{33}$$

Taking either $\epsilon \to 1$ or $\sigma \to 0$ causes this to diverge. The solution is that we can simply absorb these divergences into S_0, giving the general result that the metric associated to a one dimensional homogenous isotropic walk is simply a constant. This is, of course, exactly as we would expect.

One might worry that this is trivially always constant, however if we deform Q into the form $Q = \sqrt{G(x)}(x-y)$, then the same calculation reveals

$$g = \frac{S_0}{2\sigma^2}\frac{\epsilon}{1-\epsilon}G(x) \tag{34}$$

As long as $G(x)$ does not vanish, or approach zero anywhere too quickly, then the deformed transition function will limit, again, to two spikes hugging the origin. Thus, a slightly altered random walk limits to a different metric.

Back To Graph Theory We have now shown that (30) represents a physical metric on a manifold for the appropriate transition function. We must now pull this expression back onto a discrete space; the derivatives obviously make this problematic. While there is a large area of research concerned with finding objects on graphs which somehow limit to differential operators [9], such investigations are out of the scope of this writing. If instead we work directly with the integral (29) over a finite path, we can get rid of the differentials. For a path $\mathcal{P} = (i_1, i_2, ..., i_n)$ we thus have an action

$$S(\mathcal{P}) = \sum_{k=1}^{n-1} \log T_{i_k i_{k+1}} \tag{35}$$

This is not the path length itself, but an action which characterizes it nontheless. Intuitively, it is an entropy term associated with the probability of taking a specific path. While we do not expect this geometry on its own to be very useful, with a good definition for the discrete Hessian, one would be able to retrieve a more exact analogue to (30). These formulations have also been worked with Riemannian geometry in mind. Generalization to Lorentzian geometry is mostly trivial, although one would have to take care regarding the form of the transition matrix, and its interpretion.

5 Paths

A secondary route to identifying geometries on DAGs is through the various possible algorithms for generating paths between points. The most obvious algorithm is simply the longest path algorithm, which has already been studied in the context of causal sets [10] [16].

One of the main results concerning the longest path is that, for DAGs generated from a Minkowski spacetime, at high node density, the longest path between points of sufficient separation converges to the geodesic on the spacetime [3]. But, these are not the only path algorithms which we expect to converge to the geodesic on the embedding space. This can be understood in the context of symmetry breaking. While the embedding spaces have full coordinate invariance, the graph has at most node label invariance, which is a much smaller symmetry. We therefore expect a degeneracy splitting, manifesting itself in the many mathematical objects on the graph which all converge to the same object on the manifold.

Furthermore, longest paths don't arise from a natural dynamics like the metrics from section 4. In section 3.4, we saw that graph dynamics converging to manifold dynamics is non-trivial. The purpose of this section will then be to identify the properties of natural dynamical frameworks on DAGs which converge to Lorentzian longest path dynamics. We will do this computationally, and so should determine good measures that will tell us about the convergence.

5.1 Path Measures

Let us put path measures into context. We will imagine generating a DAG from infinite, 1+1 dimensional Minkowski space, with some arbitrary node density according to the RGG process from section (3.4). We will then choose two nodes with an interval of size N and ignore everything in the DAG except the subgraph in this interval. We will refer to these two nodes as the source and the target, and scale the Minkowski space so that they are separated by s. The node density in this interval will be $\rho = N/s^2$, and so the characteristic length scale will be $\xi = 1/\rho^{\frac{1}{2}} = s/N^{\frac{1}{2}}$.

We will then generate a path between the source and target using some path algorithm, and apply and record our path measure. Where the path algorithm produces more than one path, we will record the path measure over all possible paths. We will then repeat this process for all source-target pairs with a size N DAG interval, and average over all recorded path measures. Doing this across many interval sizes will tell us how our convergence measures behave as the local node density increases. The goal is to choose measures which behave a specific way if, and hopefully only if, the generated path becomes more similar to the geodesic between the soure and target node in the dense limit.

Note that this is not an experimental procedure, but a definition for the process of measuring convergence. Obviously, in reality, we will not be using infinite spaces, nor identifying all possible N interval node pairs. This defining procedure is also easily generalizable to more complicated manifolds, however in this investigation we will only focus on the $1 + 1$ dimensional Minkowski case.

5.1.1 Scaling

The first and most obvious path measure is the scaling. From Bollobás and Brightwell [3], we know that the discrete length of the longest path in d dimensional Minkowski space (on average) scales as

$$\bar{n}_{long} = a(s/\xi) + b(s/\xi)^{-\alpha} \tag{36}$$

where the discrete length n_{long} refers to the number of nodes in the path. We can see that when $\xi \ll s$, we get $\bar{n}_{long} \propto s$. The actual geoedesic will of course scale $\propto s$, and so this works as a necessary, but not sufficient, condition for geodesic convergence.

The insufficiency of this measure becomes apparent when considering paths of linked geodesics. For example, the path in Fig. 4 consists of two linked geodesics, and so scales as $\frac{s}{\cosh \eta} \propto s$. Lorentz invariance of the embedding manifold and uniformity of the PPP ensure (most normal) path algorithm scalings don't alter with a change in rapidity, so we can expect this issue to carry over from the embedding manifold to the graph.

5.1.2 Deviation

To solve this we introduce path deviation. We define the deviation of a single node to be its perpendicular separation from the true geodesic on the embedding manifold. The path deviation is then the average of the deviations of the nodes in the path. If the path in question approaches the geodesic in the limit, then the average deviation across many paths should asymptotically approach zero. A vanishing average path deviation then also works as a necessary condition for geodesic convergence, and covers cases where there is appropriate scaling but still crookedness.

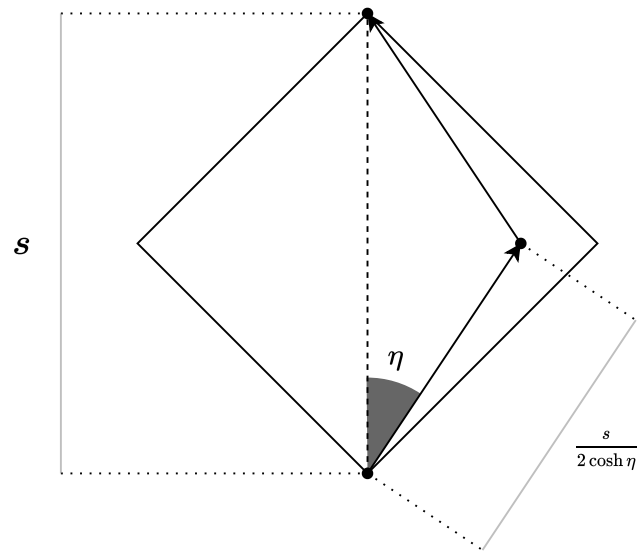


Figure 4: The separation s causal diamond between a source and target node in $1+1$d Minkowski spacetime. The crooked path (solid arrows) has a rapidity η, and therefore a total separation of $\frac{s}{\cosh\eta}$. Meanwhile, the geodesic (dashed line) has a separation s. Though the crooked path is not a geodesic, it still scales as $\propto s$. Therefore scaling is a necessary, but insufficient, geodesic property.

Scale of Deviations Let us get a basic idea of the behaviour of the path deviation. Using lightcone coordinates in $1+1$d Minkowski space, the total average node deviation in the diamond is

$$\bar{\sigma}_N = 2\int_0^s du \int_0^u dv \rho(u-v) = \frac{\rho s^3}{6} \tag{37}$$

where the deviation of a node at (u,v) is $u-v=x$ (because the geodesic is the vertical line at $x=0$). Using $\rho = N/s^2$ the expected deviation of a single node is

$$\bar{\sigma} = \bar{\sigma}_N/N = \frac{s}{6} \tag{38}$$

Note that, naively, this is an upper bound on path deviation. However, the deviation of a given node in a path depends on its predecessor in that path, such that if a path has a large deviation early on, it might very well maintain higher than average deviations until it nears the end node. Nonetheless, this calculation sets the scale for what we should expect from the deviation in practice.

Deviation Lower Bound We can also get an idea of the minimum average deviation for a path of length n in the $1+1$d Minkowski diamond. Consider a region of width $2x_0$ hugging the geodesic, as in Fig 5. The rectangular region has an area of $2x_0(s-2t_0)$, while the top and bottom regions form a small diamond of separation $2t_0$. Therefore the whole region has an area of $2x_0(s-2t_0)+(2t_0)^2$. Noting that $t_0 = x_0$, this becomes $2x_0 s$. For a path to fit inside this region we then require $2x_0 s\rho = n$. If the path is a geodesic, then $n \sim s/\xi$, and so $2x_0 = \xi$. The width of this "region of closest approach" is therefore just the characteristic length scale of the graph.

To get an approximate lower bound we calculate the expected deviation for all the nodes in this region. As before, the ends contribute a diamond of separation $2t_0$. We already calculated the contribution for diamonds in (37), so $\bar{\sigma}_N^{ends} = \frac{\rho(2t_0)^3}{6}$. The contribution from the rectangle is

$$\bar{\sigma}_N^{rect} = 2\rho(s-2t_0)\int_0^{x_0} dx x = \rho x_0^2(s-2x_0) \tag{39}$$

where again we have used $t_0 = x_0$. Using $2x_0 = \xi$,

$$\bar{\sigma}_N = \bar{\sigma}_N^{ends} + \bar{\sigma}_N^{rect} = \frac{s}{4} - \frac{\xi}{12} \tag{40}$$

The geodesic has $n = s/\xi$ nodes and should just fit into this region, so

$$\bar{\sigma} = \bar{\sigma}_N/n = \xi\Big(\frac{1}{4} - \frac{\xi}{12s}\Big) \tag{41}$$

This serves as an approximate lower bound on the scaling of the deviation. For high s/ξ, we get $\bar{\sigma} \propto \xi$.

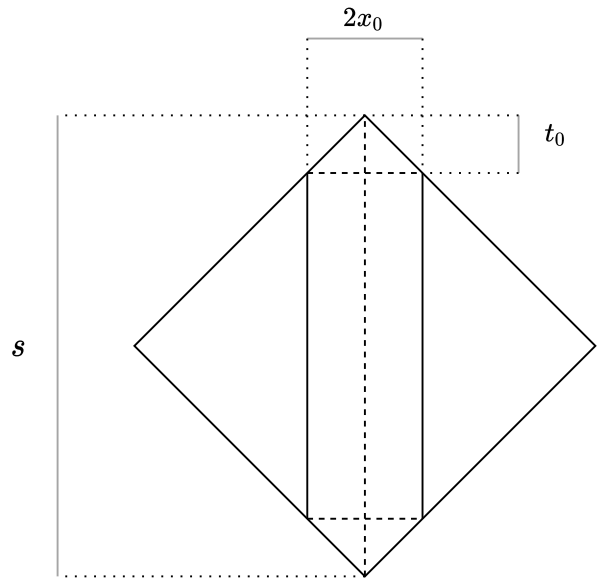


Figure 5: A region of width $2x_0$ hugging the geodesic (vertical dashed line) between a source and target node. The region consists of a rectangle of width $2x_0$ and height $s - 2t_0$, and two small triangles. Together the triangles form a diamond of height (separation) $2t_0$ and width $2x_0$. The smallest this region can be while still, on average, fitting the geodesic is $2x_0 = \xi$. In that the case the average deviation of the geodesic is given by (41)

5.2 Path Algorithms

We will now consider various path algorithms which we know should converge, as well as those which we know shouldn't. We will then discuss the properties of these paths, including their scaling and deviations. Inspired by these discussions, we will develop a new path algorithm which we intend to mimic geodesic behaviour on the embedding manifold, but whose dynamical properties are more physical than the other convergent algorithms. In later sections we will simulate these algorithms and study their behaviour empirically.

5.2.1 Longest Path

The longest path is well studied and, as already mentioned, scales according to our expectations (see (36)). Although the longest path is obviously equivalent to the geodesic in the limit, its algorithm requires global graph information, and is therefore generally unsuitable for describing a natural, nontrivial dynamics of information. The shortest path on undirected graphs has similar problems, and so is typically forgone in favour of more interesting metrics, such as those from section 4.

5.2.2 Greedy Path

We can try to formulate a more natural dynamics by optimizing locally. Consider starting at the source node, then proceeding to the successor (of the source) which is furthest (not closest) from the target node. Then repeat this procedure from the new node until you make it to the target node. This algorithm defines the greedy path, and mimics the "straightest path" dynamics of geodesics. We thus expect it to scale and deviate as if converging to the geodesic. Note that this algorithm is trivial for transitively complete DAGs, so here we will consider it for the transitively reduced case.

While this path is locally optimization, it still has two major problems. Firstly, it depends on separations, and thus the embedding manifold. Most complex systems have no intrinsic embedding manifold, making this algorithm more difficult to apply in practice. Furthermore, it takes a source and target node as arguments, meaning it has to have some nonlocal information about the graph to make those optimizations. Typically, a particle moving on a manifold optimizes locally by maintaining its velocity vector as it moves between tangent spaces (intertia and the geodesic equation). This indicates that if we want a natural local dynamics we should use an algorithm with memory.

5.2.3 Diamond Path

We wish to come up with an algorithm that we know makes geodesics in $1 + 1$d flat space and which optimizes locally, has memory, and only uses tools available on the graph. Insofar as nodes on the graph are analogous to

points on the manifold, edges are analagous to moevement between those points, ie. velocities. The algorithm should mimic the geodesic equation, in that it compares its previous velocity with its current velocity. Thus, the algorithm should compare its last choice of edge to its next choice of edge.

Let us consider a particle in $1+1$d Minkowski space which moves in jumps of separation Δs. Working in lightcone coordinates, we imagine it has jumped from (u_A, v_A) to $(u_B, v_B) = (0, 0)$ and must now choose a third point (u_C, v_C) to jump to. The point C will be Δs into the future, and so must lie on the curve depicted in Fig. 6. A is similarly Δs in the past, and so both points will form a causal diamond of area $u_A v_A = u_C v_C = \Delta s^2$ with B. The area of the causal diamond between A and C is

$$\begin{aligned} s_{AC}^2 &= (u_A + u_C)(v_A + v_C) \\ &= (u_A + u_C)(v_A + \frac{\Delta s^2}{u_C}) \\ &= 2\Delta s^2 + \frac{u_A \Delta s^2}{u_C} + u_C v_A \end{aligned} \tag{42}$$

The points A and B have already been chosen, so the only remaining free parameter here is u_C. If we vary u_C and maximize s_{AC} we get the condition

$$\frac{v_C}{u_C} = \frac{v_A}{u_A} \tag{43}$$

But this can only happen if ABC forms a straight line, and thus the particle will follow a geodesic.

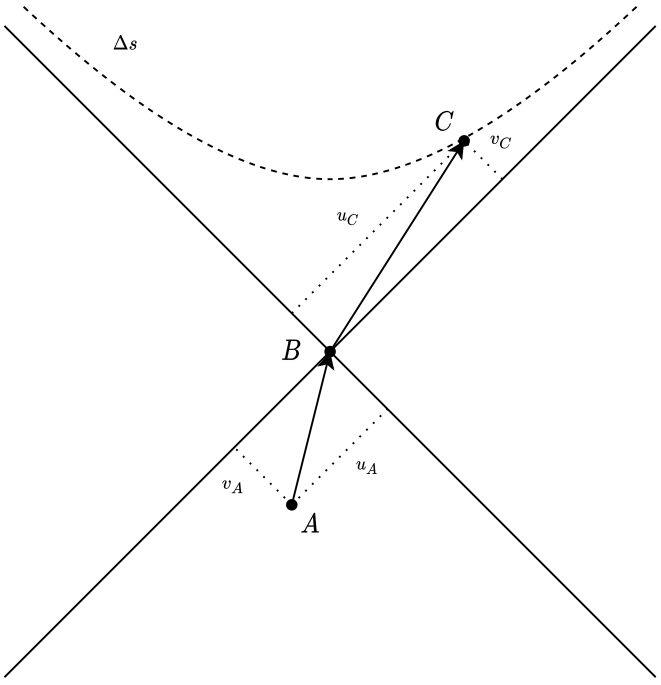


Figure 6: A particle jumps a separation Δs from point A to point B on a $1+1$d Minkowski spacetime. It again jumps Δs to a point C lying somewhere on the hyperbola depicted (dashed curve). If the area of the causal diamond between A and C is maximized, then $\frac{v_C}{u_C} = \frac{v_A}{u_A}$, and ABC will form a straight line.

The area of diamonds is easily translated onto graphs, so we can recreate this algorithm for the discrete case as follows: start at a source node A and move to a node in A's immediate future. Call this new node B. Then, calculate the sizes of the intervals between A, and all the nodes in the immediate future of B. The next node in the path will be the node in the future of B that shares the biggest interval with A. Then, repeat this process using B in place of A, and the new node in place of B (see Fig. 7*a*). We thus have an algorithm which is embedding independent, locally optimizing, and that mimics the geodesic equation in its memory of its past motion. There are a couple subtleties that should be addressed. Firstly, the algorithm might select multiple nodes in the immediate future of B, as it is unlikely there will always be a unique maximum interval . To fix this we simply specify a process for selecting from that set of maxima (this is depicted in Fig. 7*b*). Here, we will select from that set in two ways: randomly, and using the greedy algorithm. This means we actually have two algorithms, one for each supplemental selection process. We call these two algorithms the random diamond and greedy diamond algorithms.

Although the greedy selection process ruins the algorithms embedding independence, the greedy diamond path still gives us valuable information about the diamond algorithm itself. Because we expect the greedy

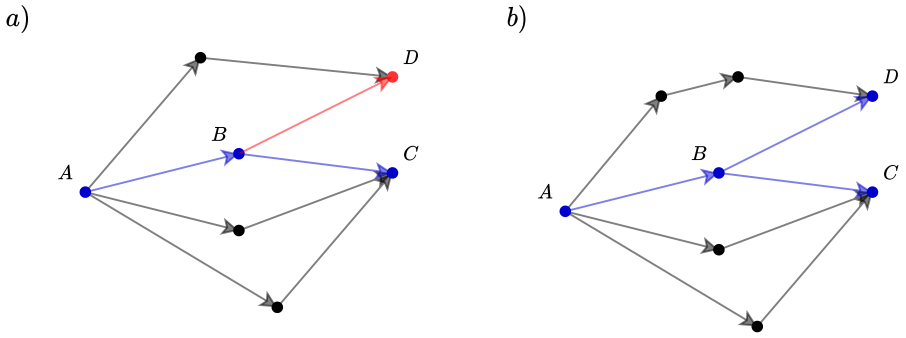


Figure 7: The DAG a) depicts a particle moving according to the diamond algorithm. The particle starts by jumping from A to B. The particle then jumps to C, rather than D, because the interval AC is bigger than the interval AD. The DAG b) depicts a similar situation, except the intervals AC and AD are the same size, and so the particle must apply some supplementary selection process to decide between C and D.

algorithm to converge, if the diamond greedy algorithm doesn't converge, then we know this must be because of a weakness in the diamond algorithms ability to reproduce geodesic behaviour. Similarly, the random diamond algorithm converging suggests the diamond algorithm is reproducing this behaviour.

That being said, simply combining two algorithms with geodesic behaviour does not necessarily mean the resulting algorithm will have vanishing deviation or s/ξ scaling. The key here is that the path deviation measure is a combination of local properties, while the paths themselves are a series of local decisions. Consider an algorithm which, at each step, chooses randomly from the 5 greediest nodes of the current nodes successors. Though this will have a higher deviation than the greedy path, it still on average chooses nodes with lower deviations than the random path. Thus the deviation responds predictably to combining paths like this. The scaling, on the other hand, is a more global property, and is thus more easily ruined by interruptions.

The second subtlety is that neither diamond algorithm takes a source and target node as arguments. Instead, it takes an initial node and a secondary node, ie. a source node, and an edge. This is actually in line with the behaviour of geodesics. The motion of a particle along a geodesic in a Lorentzian spacetime can be specified by its initial position and velocity. However, this means that we won't necessarily be able to predict whether a diamond path we initialize will ever hit a given target. In our case, this is not a problem. Because we are working with a DAG inside a source-target interval, the generated path is guaranteed to hit the target node. Our experiments will mimic exactly this scenario, so we will have to be wary of a bias towards convergence regarding the path deviations when doing simulations.

5.2.4 Random Path

The random path is the only possible maximally local path, as it uses no external information to make decisions about the next node in the sequence. The paths we will be using, however, are not perfectly random. By confining ourselves to the causal diamond between the source and target nodes, we guarantee that the random path starts at the source and ends at the target. Thus, we are not choosing from the full set of paths, but from the set of paths satisfying this condition. This is normally called a Brownian chain; here we will use the term random path.

Unlike the longest path, which is transitively invariant, or the greedy path, which is trivial upon transitive completion, the random path behaves uniquely under both transitive reduction and transitive completion of the DAG. We will call these cases TR-random and TC-random.

Transitively Complete Random Path We can use a simple argument to get an estimation of the length of the TC-random path. This argument is adapted from discussions with the projects supervisor. At each jump, all the nodes to the future of the current node are equally likely. Thus, every position in the causal future is equally likely. We can calculate how many nodes there will be on average between our new node and the target node by doing

$$N_1 = \rho \int_0^{s_0} dudv \frac{uv}{s_0^2} = \frac{\rho s_0^2}{4} = \frac{N_0}{4} \tag{44}$$

where s_0 and N_0 are the separation and interval size (respectively) between the current and target nodes. Therefore after m jumps there are $N_m = \frac{N_0}{4^m}$ nodes remaining. When $N_m = 1$ we are one node away from the target, so $n = \log_4 N + 2$, where, since the starting node is the source, $N = N_0$. The addition of 2 is to account for the source and target nodes.

This is not an exact calculation, as the position of a node is dependent on all the nodes before it, and thus subject to fluctuations away from the average throughout the whole path. To fix this we include a proportionality on the logarithm, giving

$$\bar{n}_{rand} = \log_{4a} N + 2 \tag{45}$$

where $a > 0$.

The path deviation is trickier to estimate accurately, so we will analyze it qualitatively. Naively, the deviations will be equivalent to (38). But, the path is linked to the source and target, so we expect those numbers to be pulled down. It is clear that the average deviation across all paths in the continuous case will be nonzero, so the deviations should plateau. Inspired by Bollobás and Brightwell we expect

$$\bar{\sigma}_{rand} = \sigma^{(0)}_{rand} + \frac{b}{N^{\alpha}} + \cdots \tag{46}$$

where b and α are different from those in (36), and the ellipsis indicates lower order terms.

This TC-random algorithm creates many paths per source-target pair, so we expect large fluctuations in any path measure.

Transitively Reduced Random Path Because the diamond algorithm acts on a transitively reduced graph, it is the TC-random path we should compare the random diamond path to to understand the convergence properties of the diamond algorithm. In the continuous limit, the TR-random path is equivalent to a random walk through spacetime. Such a model is not worth solving for our purposes, so we will analyze this path qualitatively. A model for a spacetime random walk describing the TR-random path in the continuous limit can be found in Appendix C.

Because this algorithm limits to a diffusion process, it will not scale logarithmically. In the case of non-relativistic diffusion, the RMS distance goes as $\sqrt{t}$ [22]. In our case we are working with the limit of a relativistic diffusion model confined to a causal diamond. Thus, while we do not expect an exact $1/2$ power law, there should still be some power law. This also means the TR-random path will in general be longer than the TC-random path. This makes sense, as the TC-random paths contain all the TR-random paths, plus those with sudden jumps into the future.

For deviations, the arguments for the TC-random path applies exactly. Thus we expect the TR-random path to obey (46), but with different constants. The TR-Random path also has a smaller set of possible paths, so we expect lower fluctuations than in the TC-Random case.

6 Method

In this section we will go over the techniques used to generate the DAGs and paths. Various packages and languages were considered for these purposes, including Cython, Numba, and the C++ Boost Graph Library. Cython and Numba lacked the necessary support for the appropriate graph and path generation; the C++ Boost Graph Library was able to produce graphs of size $N \sim 1.6 \times 10^4$ natively in a couple minutes, although is designed for a much broader range of application, and thus demanded longer production times; using C++ alone was not sufficient, as it lacked a necessary path generation algorithm. Ultimately, the code was written in python using numpy and the networkx graph library, and was run on the Imperial high performance computing cluster.

6.1 DAG Generation

We will now be looking to translate the process described at the beginning of section 5.1 into a tractable form. The process is as follows:

1. Generate two ordered sets of $N-2$ random numbers on the interval $(0,1)$. Label these sets u and v. These will be the lightcone coordinates of the nodes in the interval.
2. Add two nodes, one at $(0,0)$ and one at $(1,1)$, to u and v. These will be the source and target node.
3. Generate a set of ordered pairs containing all (A,B) where $u_A < u_B$ and $v_A < v_B$. This is the directed edge list.

This process essentially chooses a random interval-N-source-target pair from a DAG generated on an infinite 1+1d Minkowski space, and implicitly scales the space so that $s = 1$. Note, this DAG is transitively complete. A networkx method was used for transitively reducing the DAG when needed.

6.2 Path Algorithm Implementation

Longest Path Algorithm Implementation The longest path was implemented using the source-target Bellman-Ford shortest path algorithm with uniform negative edge weights. Minimizing the paths negative edge weights maximizes the number of edges in the path, and therefore gives the longest path between the source and the target. The Bellman-Ford source-target shortest path algorithm was pulled from the networkx graph package.

Greedy Path Algorithm The greedy path is implemented recursively. The recursive function takes a starting node, an end node, a node list, and the (transitively reduced) graph structure. It finds the separations between the target node and each of the successors of the starting node, then finds the successor of highest separation. Denote this successor the max node. The function then adds the max node to the node list, and calls itself with the max node, the target node, the updated node list, and the graph structure. If the starting node has no successors, it breaks the stack and returns the node list.

If we run this function with the starting node as the source node, the end node as the target node, and with an initial node list containing just the source node, then the node list it returns will be the greedy path between the source and the target. This process is illustrated in Fig. 8, and mimics the procedure described in section 5.2.2.

Diamond Algorithm Implementation The diamond path is also implemented recursively. The recursive function takes a starting node, a secondary node, a node list, and the (transitively reduced) graph structure. It finds the size of the interval between the starting node and each of the successors of the secondary node, then finds the successors with the largest interval. Let us call this set of successors the max nodes. The function then applies some procedure (see below) for selecting a single node from the set of max nodes. Denote this single node the selected max node. The function then adds the selected max node to the node list, and calls itself with the secondary node, the selected max node, the updated node list, and the graph structure. If the secondary node has no successors, it breaks the stack and returns the node list.

The procedure applied for selecting from the max nodes depends on whether we wish to use the greedy diamond path or the random diamond path. In the former case, the function finds the separations between the target node and each of the max nodes, then selects the max node with the highest such separation. In the latter case, the function simply selects from the max nodes randomly.

If we run this function with the starting node as the source node; the secondary node as either a greedy or random successor of the starting node; and with an initial node list containing the source node and this secondary node; then the node list it returns will either be the greedy diamond path or the random diamond path, depending on the selection procedure used. This process is illustrated in Fig. 9, and mimics the algorithm described in section 5.2.3.

Random Path Algorithm Implementation The random path algorithm works essentially identically to the previous two algorithms, although it only takes the source node, and can take either a transitively complete, or transitively reduced graph structure. For the sake of avoiding unnecessary redundancy, we will omit its description.

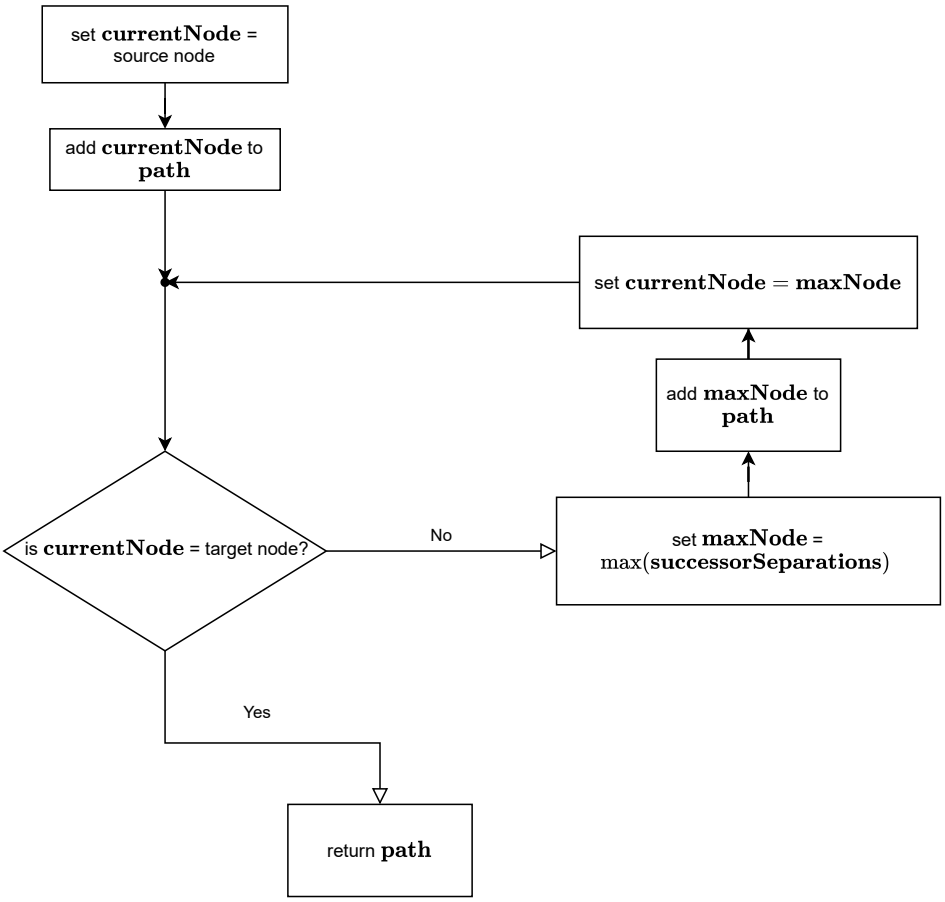


Figure 8: The greedy path algorithm. Here, **successorSeparations** are the (embedding manifold) separations between the successors of **currentNode** and the target node. Similarly, max(**successorSeparations**) is the successor with the highest such separation.

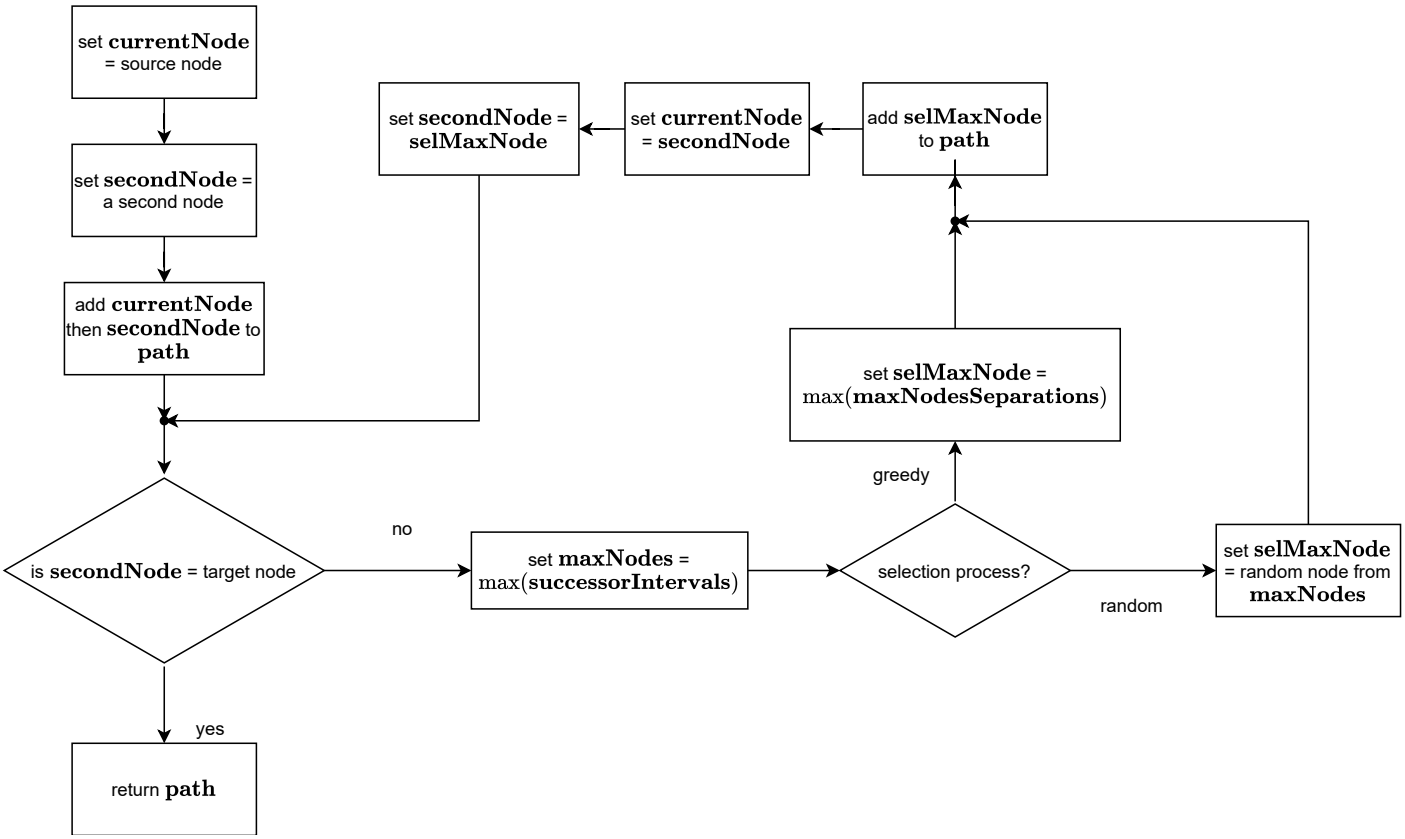


Figure 9: The diamond algorithm. Here, **successorIntervals** are the sizes of the (graph) intervals between the successors of the **secondNode** and the **currentNode**. Similarly, max(**successorIntervals**) is the set of successors with the highest such interval (note that there are often more than one). This algorithm also applies a secondary selection process. The greedy selection process is similar to the greedy algorithm depicted in Fig. 8. The random selection process is trivial. Both processes act on the set of maximum interval successors.

6.3 Data Analysis

The average discrete path length and path deviation was taken for each path for a range of interval sizes, typically from $N = 10^3$ to $N \sim 10^4$. The average of a path measure was ascertained from a sample of possible intervals and (where applicable) path instances.

Let us take the deviation measurements of the random path at $N = 10^3$ for example. First an interval $N = 10^3$ causal diamond was generated. Then 20 different random paths were generated. The deviation of each of these paths was measured and recorded. This was then repeated on 50 different interval $N = 10^3$ causal diamonds. The average of the recorded random path deviations was taken to approximate the true average random path deviation at $N = 10^3$, with a standard error calculated from this distribution as well. This process was repeated over a range of N.

The number of graph and path iterations for each algorithm is listed in table 1.

Path Algorithm	Path Iterations	Graph Iterations
Longest	1	50
Greedy	N/A	50
TC-Random	20	50
TR-Random	20	50
Greedy Diamond	N/A	50
Random Diamond	20	50

Table 1: Sampling of paths at a given N

Path Algorithm	A	$\pm\sigma_A$	α	$\pm\sigma_\alpha$
Longest	1.816	0.015	0.5074	0.001
Greedy	1.650	0.030	0.496	0.002
TC-Random	3.121	0.112	0.086	0.004
TR-Random	3.771	0.120	0.186	0.004
Greedy Diamond	1.211	0.096	0.504	0.010
Random Diamond	3.625	0.248	0.221	0.009

(a) Path lengths n fitted to $n = AN^\alpha$ (with errors)

Path Algorithm	A	$\pm\sigma_A$	α	$\pm\sigma_\alpha$
Longest	0.168	0.023	-0.173	0.016
Greedy	0.128	0.012	-0.226	0.109
TC-Random	0.098	0.007	-0.050	0.009
TR-Random	0.149	0.007	-0.026	0.006
Greedy Diamond	0.246	0.063	-0.189	0.032
Random Diamond	0.125	0.007	0.054	0.006

(b) Deviations σ fitted to $\sigma = AN^\alpha$ (with errors)

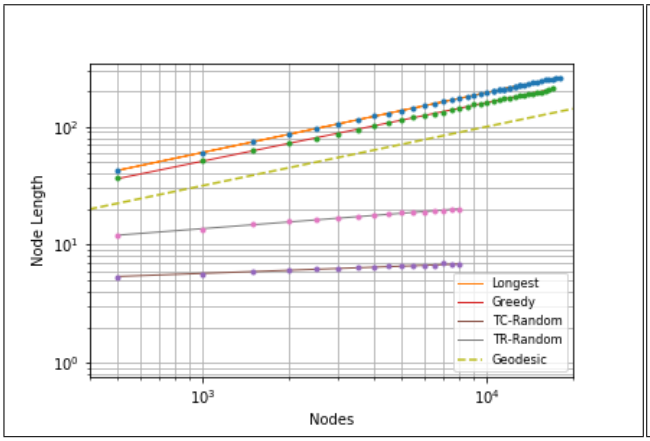
Table 2: Power law fittings

Longest, Greedy, TC-Random, and TR-Random Doing this for the longest, greedy, TC-Random, and TR-Random paths we find that the longest and greedy paths approximately follow geodesic scaling laws, and have vanishing deviations as expected. Only fitting data points for $N > N_{cutoff}$ brought the leading order power law closer to that in (36) as N_{cutoff} was increased. This indicates validity in the low order corrections from (36).

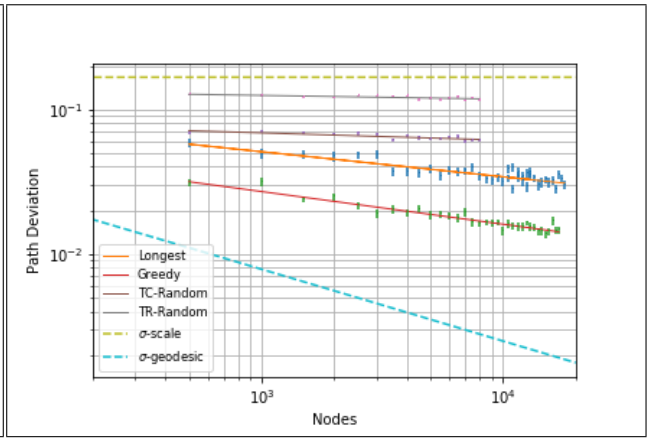
The random paths also have increasing length scaling, but with different power laws, such that the TC-random path is consistently smaller than the TR-random path. The data was not taken over a large enough range to decidedly conclude whether the TC-random path length scales logarithmically. Fitting (45) to the data regardless we find the $\log_4$ proportionality correction comes out to $a = 1.6067 \pm 0.0131$. Furthermore, logarithmic scalings with constant addition terms other than 2 did not fit the data effectively in the low N range, giving more weight to the prediction (45).

The short range of the data also meant it was impossible to get an accurate estimate for $\bar{\sigma}^{(0)}_{rand}$ in either the TC or TR cases. Fitting a pure power law to the random path deviation data instead showed that the deviations were approximately constant, with a slightly negative scaling exponent. Deviations of all four paths were also bounded between (38) and (41). All this can be seen in Figs. 10a and 10b, and the full set of fittings can be found in tables 2a and 2b.

Greedy and Greedy Diamond In the case of the greedy diamond path the length scale power law was equivalent to the greedy paths to within ∼one standard deviation, while the path deviation was equivalent to within ∼two. In both cases the diamond path behaved as if converging to a geodesic. Because the greedy path itself converges to a geodesic, this does not confirm that the diamond algorithm is geodesic-like, although it does not rule it out (see section 5.2.3). This can be seen in Figs. 11a and 11b.

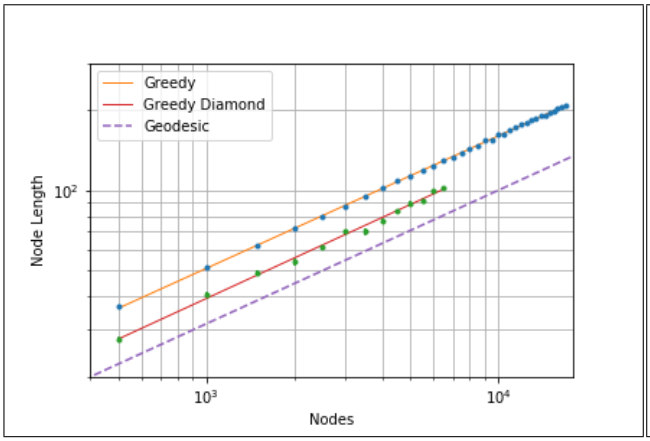


(a) The discrete path length n of the longest, greedy, TR-, and TC-Random paths between the source and target nodes of a separation $s = 1$ causal diamond in $1 + 1$d Minkowski space, log-log plotted against the number of nodes N in the DAG, then fitted to power laws. The $n = N^{1/2}$ scaling of an ideal geodesic is depicted as a dashed line. The longest and greedy paths scale like the geodesic for sufficiently high N, while the random paths do not. The error bars are too small to be visible on this plot, and smaller than the markers depicted.

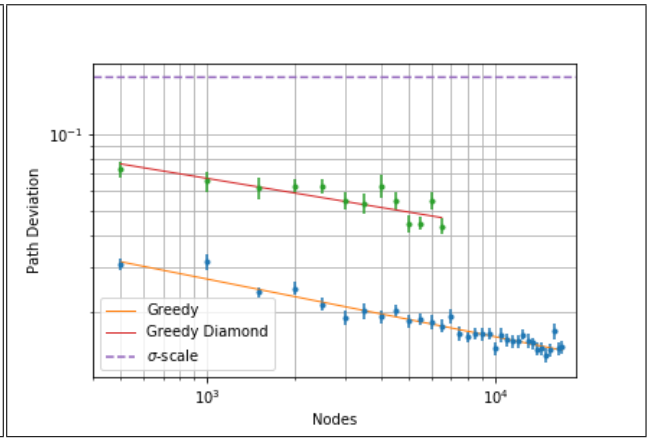


(b) The path deviation of the longest, greedy, TR-, and TC-Random paths generated between the source and target nodes of a separation $s = 1$ causal diamond in $1 + 1$d Minkowski space, log-log plotted against the number of nodes N in the DAG, then fitted to power laws. The yellow dashed line is the average node deviation, and the blue dashed line is the approximate lower bound. The longest and greedy paths have steeply falling deviations, while the random path deviations are nearly constant.

Figure 10

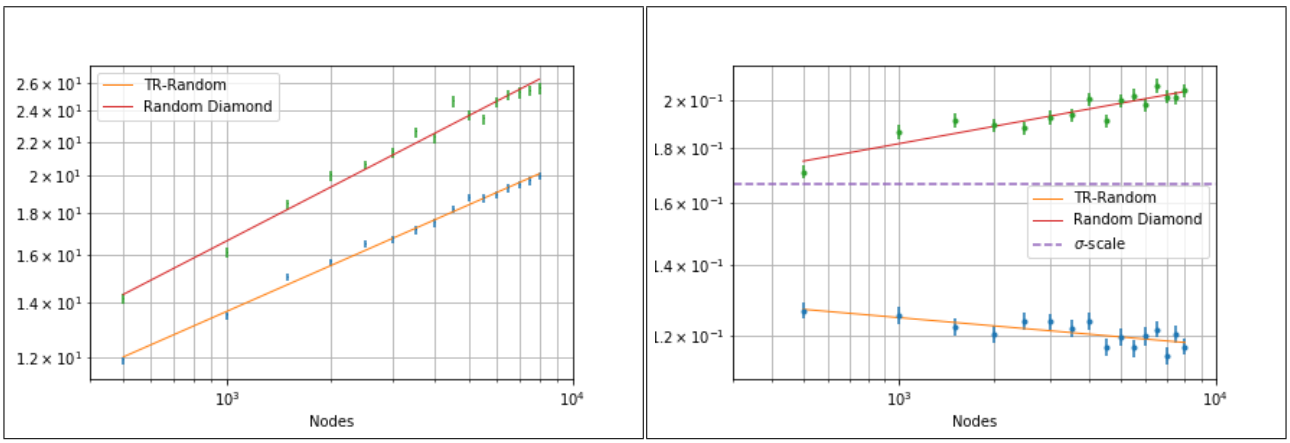


(a) The discrete path length n of the greedy and greedy diamond paths between the source and target nodes of a separation $s = 1$ causal diamond in $1 + 1$d Minkowski space, log-log plotted against the number of nodes N in the DAG, then fitted to power laws. The $n = N^{1/2}$ scaling of an ideal geodesic is depicted as a dashed line. Both paths scale like the geodesic for sufficiently high N. The error bars are too small to be visible on this plot, and smaller than the markers depicted.



(b) The path deviation of the greedy and greedy diamond paths between the source and target nodes of a separation $s = 1$ causal diamond in $1 + 1$d Minkowski space, log-log plotted against the number of nodes N in the DAG, then fitted to power laws. The purple dashed line is the average node deviation, while the approximate lower bound is not depicted. The greedy path has a slightly steeper decline than the greedy diamond path, hinting that the diamond path increases the overall deviation.

Figure 11



(a) The discrete path length n of the TR-random and random diamond paths between the source and target nodes of a separation $s = 1$ causal diamond in $1 + 1$d Minkowski space, log-log plotted against the number of nodes N in the DAG, then fitted to power laws. The ideal geodesic scaling is omitted. The TR-random path has no geodesic scaling, so the random diamond path not having geodesic scaling does not mean the diamond algorithm is non-geodesic.

(b) The path deviation of the TR-random and random diamond paths between the source and target nodes of a separation $s = 1$ causal diamond in $1 + 1$d Minkowski space, log-log plotted against the number of nodes N in the DAG, then fitted to power laws. The purple dashed line is the average node deviation, while the approximate lower bound is not depicted. While the TR-Random path has a slight downward scaling, the diamond random path has upward scaling. This implies that the diamond algorithm cannot be geodesic-like.

Figure 12

TR-Random and Random Diamond Once again, the random diamond path has a similar length scaling power law to the pure random case. The deviations, however, are more interesting. Fig. 12b shows that the random diamond algorithm not only has a higher power law, but a positive power law. The TR-Random power law, on the other hand, is negative in this range, and presumably constant for higher N. Beyond that, the random diamond path deviation is consistently higher than the average node deviation (38). This suggests that the diamond algorithm does not behave convergently, but divergently. This is at odds with the discussion from section 5.2.3.

Diamond Algorithm Discrepancy This discrepancy can be understood by the simple fact that the diamond algorithm does not see the dense limit. This is because the dense limit depends on having an infinite number of nodes, while the diamond algorithm always makes decisions based upon its finite sight into the past and future. In the thought experiment which motivated the diamond algorithm (see section 3.4) the test particle was moving a fixed separation Δs at each iteration. In the diamond algorithm, our test particle always moves $\Delta s = \xi \times (1 \text{ node})$. But then, as $\xi \to 0$, $\Delta s \to 0$. This does not mimic the thought experiment, since there we maintained a finite Δs. The natural solution is to modify the diamond algorithm so that the test particle moves $\sim \xi_\Delta / \xi$ nodes each time, such that $\Delta s = \xi_\Delta$ is a constant value as $\xi \to 0$. This then defines a secondary length scale ξ_Δ on the graph. This is exactly in line with our discussion from section 3.4, wherein the recreation of the manifolds dynamic structure from purely local considerations required a secondary length scale. This length scale can then help us coarse grain the graph, giving a natural procedure for choosing an MHKM conformal factor (noting that the extended diamond algorithm is easily generalized to any DAG). This also tells us that our analogy wherein nodes were positions and edges were velocities was not quite right. Instead, we should draw analogy between these coarse grained regions and local neighbourhoods on the manifold.

7 Conclusion and Discussion

The purpose of these investigations was to explore possible geometrical descriptions of complex systems with causal structure. In our first approach we did this by extending known geometrical descriptions from undirected graph theory into the theory of directed acyclic graphs. More specifically, we did this for the communicability and diffusion distances. In the former case, the finite connectivity between points meant we could not only give higher weights to longer paths, but also formulate the communicability as an entropy of random walks. In the latter case, we found we could reduce diffusion down to an eigenvector problem. We also showed that certain transition dynamics on continuous manifolds had associated metrics, which went as the logarithm of the transition function. Translating this directly back to graphs would have required the definition of a discrete

Hessian, and so was forgone for a simpler, finite path action metric. This metric also took the form of an entropy of the probability of motion, and was about equally applicable to DAGs and undirected graphs.

In our second approach we designed a path algorithm which corrected the unphysicalness of the longest and greedy paths, and which mimicked a simple algorithm in continuous Minkowski space for tracing out geodesics. Simulating this algorithm we found that it did not converge to the geodesic in the limit due to its high locality. This then motivated the design of a new algorithm which sees the length scale of the causal diamond and reacts accordingly. It was noted that this new algorithm mimics the geodesic tracing procedure better than the previous algorithm, leading to the conclusion that any natural path dynamics on a directed acyclic graph will have to set some length scale separate from the length scale associated with its embedding. This is the same as in the directed case, and a link between the MHKM theorem and motion on the graph.

Along the way we also analyzed the scaling and deviations of the longest, greedy, and random paths on 1+1d Minkowski embeddable DAGs. These paths behaved qualitatively as expected, and obeyed the predicted scaling and deviation at leading order where applicable. Insufficient data was taken to appropriately predict next to leading order coefficients and power laws.

This writing was largely exploratory, and presents alternative routes for studying the geometry of directed acyclic graphs. Next steps include studying the relationship between the derived geometries and known measures of centrality, connectivity (etc.), generalizing the transition metric to directed acyclic graphs more carefully, simulating the adjusted form of the extended diamond algorithm, and taking better measurements to estimate more of the coefficients from section 5.2. In section 4, we also neglected generalizing non-conservative diffusion processes, which might be useful for understanding something like the spread of information through a citation network.

References

[1] Dionigi M. T. Benincasa and Fay Dowker. "Scalar Curvature of a Causal Set". In: *Physical Review Letters* 104.18 (May 2010). ISSN: 1079-7114. DOI: 10.1103/physrevlett.104.181301. URL: http://dx.doi.org/10.1103/PhysRevLett.104.181301.

[2] Marián Boguñá et al. "Network geometry". In: *Nature Reviews Physics* 3.2 (Jan. 2021), pp. 114–135. ISSN: 2522-5820. DOI: 10.1038/s42254-020-00264-4. URL: http://dx.doi.org/10.1038/s42254-020-00264-4.

[3] Béla Bollobás and Graham Brightwell. "Box-Spaces and Random Partial Orders". In: *Transactions of the American Mathematical Society* 324.1 (1991), pp. 59–72. ISSN: 00029947.

[4] J. R. Clough et al. "Transitive reduction of citation networks". In: *Journal of Complex Networks* 3.2 (Sept. 2014), pp. 189–203. ISSN: 2051-1329. DOI: 10.1093/comnet/cnu039. URL: http://dx.doi.org/10.1093/comnet/cnu039.

[5] Ronald Coifman et al. "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion Maps". In: *Proceedings of the National Academy of Sciences of the United States of America* 102 (June 2005), pp. 7426–31. DOI: 10.1073/pnas.0500334102.

[6] Maxime Crochemore and Renaud Vérin. "Direct construction of compact directed acyclic word graphs". In: *Combinatorial Pattern Matching*. Ed. by Alberto Apostolico and Jotun Hein. Berlin, Heidelberg: Springer Berlin Heidelberg, 1997, pp. 116–129. ISBN: 978-3-540-69214-0.

[7] K.Ch. Das. "The Laplacian spectrum of a graph". In: *Computers Mathematics with Applications* 48.5 (2004), pp. 715–724. ISSN: 0898-1221. DOI: https://doi.org/10.1016/j.camwa.2004.05.005. URL: https://www.sciencedirect.com/science/article/pii/S0898122104003074.

[8] Ernesto Estrada, Naomichi Hatano, and Michele Benzi. "The physics of communicability in complex networks". In: *Physics Reports* 514.3 (May 2012), pp. 89–119. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2012.01.006. URL: http://dx.doi.org/10.1016/j.physrep.2012.01.006.

[9] Matthias Hein, Jean-Yves Audibert, and Ulrike von Luxburg. *Graph Laplacians and their convergence on random neighborhood graphs*. arXiv: math/0608522v2.

[10] Raluca Ilie, Gregory B Thompson, and David D Reid. "A numerical study of the correspondence between paths in a causal set and geodesics in the continuum". In: *Classical and Quantum Gravity* 23.10 (Apr. 2006), pp. 3275–3285. ISSN: 1361-6382. DOI: 10.1088/0264-9381/23/10/002. URL: http://dx.doi.org/10.1088/0264-9381/23/10/002.

[11] Achim Klenke. "The Poisson Point Process". In: *Probability Theory: A Comprehensive Course*. London: Springer London, 2014, pp. 543–561. ISBN: 978-1-4471-5361-0. DOI: 10.1007/978-1-4471-5361-0_24. URL: https://doi.org/10.1007/978-1-4471-5361-0_24.

[12] Dmitri Krioukov et al. "Hyperbolic geometry of complex networks". In: *Physical Review E* 82.3 (Sept. 2010). ISSN: 1550-2376. DOI: 10.1103/physreve.82.036106. URL: http://dx.doi.org/10.1103/PhysRevE.82.036106.

[13] Vito Latora, Vincenzo Nicosia, and Giovanni Russo. *Complex Networks: Principles, Methods and Applications*. Cambridge University Press, 2017. DOI: 10.1017/9781316216002.

[14] Tomasz Luczak and Joel E. Cohen. "Stratigraphy of a Random Acyclic Directed Graph: The Size of Trophic Levels in the Cascade Model". In: *The Annals of Applied Probability* 3.2 (1993), pp. 403–420. DOI: 10.1214/aoap/1177005431. URL: https://doi.org/10.1214/aoap/1177005431.

[15] David B. Malament. "The class of continuous timelike curves determines the topology of spacetime". In: *Journal of Mathematical Physics* 18.7 (1977), pp. 1399–1404. DOI: 10.1063/1.523436. eprint: https://doi.org/10.1063/1.523436. URL: https://doi.org/10.1063/1.523436.

[16] Jan Myrheim. "Statistical geometry". In: (Aug. 1978).

[17] M. E. J. Newman. *Networks: an introduction*. Oxford; New York: Oxford University Press, 2010. ISBN: 9780199206650 0199206651. URL: http://www.amazon.com/Networks-An-Introduction-Mark-Newman/dp/0199206651/ref=sr_1_5?ie=UTF8&qid=1352896678&sr=8-5&keywords=complex+networks.

[18] Patrick Rubin-Delanchy. *Manifold structure in graph embeddings*. 2021. arXiv: 2006.05168 [stat.ML].

[19] Dan A. Simovici and Chabane Djeraba. *Mathematical Tools for Data Mining: Set Theory, Partial Orders, Combinatorics*. 2nd. Springer Publishing Company, Incorporated, 2014, pp. 67–95. ISBN: 1447164067.

[20] Chaoming Song, Shlomo Havlin, and Hernán A. Makse. "Self-similarity of complex networks". In: *Nature* 433.7024 (Jan. 2005), pp. 392–395. ISSN: 1476-4687. DOI: 10.1038/nature03248. URL: http://dx.doi.org/10.1038/nature03248.

[21] Sumati Surya. "The causal set approach to quantum gravity". In: *Living Reviews in Relativity* 22.1 (Sept. 2019). ISSN: 1433-8351. DOI: 10.1007/s41114-019-0023-1. URL: http://dx.doi.org/10.1007/s41114-019-0023-1.

[22] J Topping. "Investigations on the Theory of the Brownian Movement". In: *Physics Bulletin* 7.10 (Oct. 1956), pp. 281–281. DOI: 10.1088/0031-9112/7/10/012. URL: https://doi.org/10.1088/0031-9112/7/10/012.

[23] Robert M. Wald. *General Relativity*. Chicago, USA: Chicago Univ. Pr., 1984, pp. 11–23. DOI: 10.7208/chicago/9780226870373.001.0001.

[24] Robert M. Wald. *General Relativity*. Chicago, USA: Chicago Univ. Pr., 1984, pp. 41–46. DOI: 10.7208/chicago/9780226870373.001.0001.

[25] Mark Walters. "Random geometric graphs". In: *Surveys in Combinatorics 2011*. Ed. by RobinEditor Chapman. London Mathematical Society Lecture Note Series. Cambridge University Press, 2011, pp. 365–402. DOI: 10.1017/CBO9781139004114.009.

A Adjusted Communicability for Undirected Graphs

In section 4.1 the coefficients for the directed acyclic communicability matrix were taken to be $a_n = e^{\chi n}$. This motivates the use of $a_n = e^{-\chi n}$ in the undirected case.

$$G = \sum_{n=0} e^{-\chi n} A^n \tag{47}$$

If $e^{-\chi}|A| \leq 1$, then this converges to

$$G = (1 - e^{-\chi} A)^{-1} \tag{48}$$

where $|A|$ is the Frobenius norm

$$|A| = \sqrt{\sum_i \sum_j |A_{ij}|^2} \tag{49}$$

For the sum to converge, we therefore require $\chi \geq \frac{1}{2} \log 2|E|$, where $|E|$ is the number of edges in the graph.

B A More General Calculation of D(s)

In section 4.1, the number of paths of paths between two points in a sufficiently symmetric spacetime was shown to take the form of an exponential. This was only the case insofar as the expression itself was fundamentally intractable and divergent. In fact, the form of the expression implies $D(s_1)D(s_2) = D(s_1 + s_2)$, which is clearly not the case. This manifests itself in that $d_{ij} + d_{jk} \neq d_{ik}$. The more accurate form of these expressions in a continuous, d dimensional spacetime is

$$D(x, z) = D(x, y) * D(y, z) := \int_{\Sigma(x,z)} \frac{d^{d-1}y}{V_\Sigma} D(x, y) D(y, z) \tag{50}$$

where $\Sigma(x, z)$ is a $d - 1$ dimensional hypersurface "cutting off" x from z, and which does not extend outside of the causal diamond of x and z. This is basically saying, for every way to get from x to the point y on the hypersurface Σ, there are $D(y, z)$ ways to get from that point y to the endpoint z. Going from x to z one must always cross Σ, therefore giving us the above expression.

This expression is difficult to hold down. In section 4.1 we used a highly symmetric spacetime, and chose a Σ close to x which was a constant separation $s_0 = s - \delta s$ away from z. This was

$$\omega D(s) = \frac{dD(s)}{ds} \tag{51}$$

where

$$\omega = \frac{1}{\delta s}\Big(1 - \frac{\delta s^{d-1}}{V(\Sigma)}\Big) \tag{52}$$

For δs sufficiently small, $V(\Sigma)$ is independent of s. However, for a general δs, then

$$\begin{aligned} V(\Sigma) &= \Delta\eta s_0 \\ &= \Big(\log \frac{s}{s_0} - \log \frac{s_0}{s}\Big) s_0 \\ &= 2 s_0 \log \frac{s}{s_0} \\ &= 2(s - \delta s) \log\left(\frac{1}{1 - \frac{\delta s}{s}}\right) \end{aligned} \tag{53}$$

where $s_0 = s - \delta s$, and $\Delta\eta$ is the range of the s_0 hyperbola which falls inside the causal diamond. Then

$$\omega = \frac{1}{\delta s}\left(1 - \frac{1}{2} \frac{1}{\big(\frac{s}{\delta s} - 1\big) \log\big(\frac{1}{1 - \frac{\delta s}{s}}\big)}\right) \tag{54}$$

This is still difficult to integrate. The Laurent series around $\delta s = 0$

$$\omega \approx \frac{1}{\delta s}\left(1 - \frac{1}{2}\left(1 + \frac{\delta s}{2s} + \frac{5\delta s^2}{12 s^2}\right)\right) \tag{55}$$

Solving for $D(s)$ gives

$$D \approx D(0)\big(s/\xi\big)^{\frac{1}{4}} \exp\left(\frac{s}{2\xi} + \frac{5\xi}{24 s}\right) \tag{56}$$

where we have now substituted $\delta s = \xi$. Higher order corrections would require expanding to higher order in the derivative. Once again, for $\xi = 0$ this is divergent, however for finite ξ is not. In fact, we see similar low order corrections here as we did in section 5.2, implying it might be usable as an approximate expression for Ω (rather than as a dense limit analogue, as in section 4.1). Also note that this does not become equivalent to the expression (15) with vanishing ξ/s because we used the Laurent series expansion of ω.

C Random Walks in 1+1d Minkowski Space

The TR-Random path, in the dense limit, is essentially a random walk in spacetime. Let us consider a particle in $1+1$d Minkowski space taking such a random walk. At each instant, the particle has some probability $\rho(\eta)$ of moving a separation ξ in the direction η. Following Einstein's derivation of the heat equation, the probability density $\phi(x)$ of being at a position x is therefore given by the master equation

$$\phi(x) = \int_{\mathbb{R}} d\eta \rho(\eta) \phi(x - \xi X(\eta)) \tag{57}$$

ie. the probability of being at x is the probability of being at a position $x - \xi X(\eta)$ and moving $\eta X(\eta)$, where

$$X(\eta) = \begin{bmatrix} \cosh \eta \\ \sinh \eta \end{bmatrix} \tag{58}$$

in Cartesian coordinates. Wick rotating with the substitution $u = i\eta$ we can Taylor expand ϕ in the complex plane for small ξ. This turns out to be equivalent to taking the expansion as though X were bounded, although the convergence conditions on ϕ are stronger. For sufficient convergence

$$\begin{aligned} \phi(x) &= \int_{\mathbb{R}} d\eta \rho(\eta) \Big(\phi(x) - \xi X^\mu \partial_\mu \phi(x) + \frac{1}{2} \xi^2 X^\mu X^\nu \partial_\mu \partial_\nu \phi(x) + \cdots \Big) \\ &= \phi(x) - \alpha^\mu \partial_\mu \phi(x) + H^{\mu\nu} \partial_\mu \partial_\nu \phi(x) \end{aligned} \tag{59}$$

where

$$\alpha^\mu = \xi \int_{\mathbb{R}} \rho(\eta) X^\mu(\eta) \tag{60}$$

$$H^{\mu\nu} = \frac{1}{2} \xi^2 \int_{\mathbb{R}} d\eta \rho(\eta) X^\mu(\eta) X^\nu(\eta) \tag{61}$$

such that

$$\alpha^\mu \partial_\mu \phi(x) = H^{\mu\nu} \partial_\mu \partial_\nu \phi(x) \tag{62}$$

This equation only preserves Lorentz invariance for $\alpha^\mu = 0$ and $H^{\mu\nu} \propto \eta^{\mu\nu}$. This gives a purely wave-like solution, which would be odd. In the case of the TR-random path, Lorentz invariance is present insofar as $\rho(\eta) = \rho_0$ is a constant. Let us then assume that the Lorentz invariance has been broken due to the Taylor expansion, and simply use the natural Cartesian coordinate system of the diamond. Constant $\rho(\eta)$ then means

$$\alpha^\mu = \rho_0 \xi \int_{-\infty}^{\infty} d\eta \begin{bmatrix} \cosh \eta \\ \sinh \eta \end{bmatrix} \tag{63}$$

By symmetry, $\alpha^1 = 0$. Similarly,

$$H = \frac{1}{2} \xi^2 \rho_0 \int_{-\infty}^{\infty} d\eta \begin{bmatrix} (\cosh \eta)^2 & \cosh \eta \sinh \eta \\ \cosh \eta \sinh \eta & (\sinh \eta)^2 \end{bmatrix} \tag{64}$$

Again, by symmetry, the off-diagonals vanish. The diagonals, however, follow from

$$\begin{aligned} \int d\eta (\cosh \eta)^2 &= \frac{1}{2} (\cosh \eta \sinh \eta + \eta) \\ \int d\eta (\cosh \eta)^2 &= \frac{1}{2} (\cosh \eta \sinh \eta - \eta) \end{aligned} \tag{65}$$

which means H can be written as

$$H = \begin{bmatrix} A + B & 0 \\ 0 & A - B \end{bmatrix} \tag{66}$$

and therefore the diffusion equation in these coordinates is

$$A \Big(\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2} \Big) \phi + B \Big(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} \Big) \phi = \alpha^0 \frac{\partial \phi}{\partial t} \tag{67}$$

The first term has decaying solutions, while the second term has wave-like solutions. The third term picks out an "arrow of time". This serves as a Lorentz broken first approximation to relativistic diffusion in the form of the TR-random path. Solving this should give the average position of nodes in the TR-random path in the dense limit, from which we can get the path deviation.