Simple indicators for Lorentzian causets

Tommaso Bolognesi - CNR/ISTI, Pisa
Alexander Lamb - Princeton University

v.5 - December 2, 2014

Abstract

Several classes of DAGs (directed acyclic graphs), and associated growth dynamics, have been investigated over the last two decades, mainly in the context of the Causal Set Program, with the purpose of finding satisfactory discrete models of spacetime. We introduce some simple statistical indicators that can be used for comparing these graphs, and for assessing their closeness to the ideal Lorentzian causets -- those obtained by uniformly sprinkling points in a Lorenztian manifold. In particular, we introduce 'longest/shortest path plots' as a way to visually detect the extent to which a DAG matches the reversed triangular inequality of special relativity (and the twin paradox), and we use it for assessing causets both of stochastic and of deterministic, algorithmic origin. We identify a very simple deterministic algorithm that behaves optimally in this respect.

1. Introduction

A most direct way to obtain a *realistic* discrete model of spacetime -- one of physical significance -- consists in choosing a continuous model of it, namely a flat or curved Lorentzian manifold M that satisfies the Einstein field equations, and deriving a *causal set* from it by applying the *sprinkling* technique.

A *causal set,* or *causet*, is a set of unstructured spacetime events provided with a partial order relation, denoted by the symbol '$\prec$', expressing causal dependencies between them. The relation is irreflexive ($x \nprec x$) and must be finitary, that is, for any pair (s, t) of events, the set of points between them must be finite: $|\{x \mid s \prec x \prec t\}| < \infty$. A causet can be represented by a DAG (directed acyclic graph). The *sprinkling* technique consists in uniformly distributing points, by a Poisson distribution, in a Lorentzian manifold M and letting them inherit the causal (lightcone) structure of the latter, thus obtaining a causet; in the sequel, we shall conveniently call these objects *sprinkled causets* (find more details in Section 2, and in [12, 21]).

A challenging goal of causet-based quantum gravity research is to reverse the above logic, and to try and build causets of physical significance *without* resorting to an underlying continuum. Several techniques for doing this have been investigated in the last two decades, and some of them will be described in Section 2. Under this perspective, the manifold is only obtained a posteriori, as an asymptotic approximation.

One way to assess a causet C obtained by some of these techniques is to check whether C is 'faithfully embeddable' in some manifold M of appropriate dimension, that is, whether it could have arisen from sprinkling events in M. In [16] Henson introduces a method for building an actual embedding of a given causet in 2D Minkowski space, but recognizes the need of additional, complementary, possibly more efficient criteria for defining scales of causet 'manifoldlikeness'.

In the approach described here, we are interested in extracting some distinguishing feature of sprinkled causets *other than, or weaker than embeddability*; in particular, we shall try to identify statistical features that could reflect the peculiarities of the Lorentz metrics, most notably the reversed triangular inequality that underlies the twin paradox of Special Relativity. We shall then use these indicators as a benchmark for the graphs obtained by other techniques.

Our special interest for how causets might reflect the features of Lorentzian manifolds can be concisely motivated by the following quote from [18]:

"*Fundamental discreteness is a very old and attractive idea but it remains to be seen whether it can be reconciled with observable physics, and, in particular, with quantum mechanics and Lorentz invariance.*"

In Section 2 we briefly recall four stochastic causet construction techniques (we shall use the terms 'causet' and 'DAG' interchangeably), including sprinkling in flat (Minkowski) and positively curved (de Sitter) manifolds.

In Section 3 we concentrate on statistical properties based on counting graph edges. Keeping in mind that sprinkled causets are *transitively closed* by construction, we shall apply the counts to transitively closed versions of the DAGs obtained by all other considered techniques. This analysis reveals that power-law distributions of node degrees are quite common, and provides information that is relevant for estimating causet dimensionality, that we shall visualize in what we call *ordering fraction spectra*.

In Section 4 we focus on edge counts for the *transitive reduction* of the considered graphs, which appears to be the correct setting for addressing 'Lorentzianity' issues. We first analyze the degree of the root node of an interval, as a function of the number of nodes in the latter. Then we introduce a special type of diagram - *longest/shortest path plots* - meant to expose the peculiar wide range of the path lengths between two generic nodes, a feature of Lorentzian sprinkled graphs sometimes referred to as 'non-locality'. We illustrate, again, how the various causet classes perform with respect to these indicators.

In Section 5 we introduce a *deterministic* causet construction technique based on permutations of tuples of natural numbers and on their manipulation by a stateful control unit; we show that one instance of this model (out of 65,536) yields a pseudo-random causet with excellent non-locality properties comparable to those of sprinkled causets.

In Section 6 we summarize our results.

2. Stochastic causet construction techniques

In this section we introduce the four stochastic causet construction techniques to be analyzed in the rest of the paper. In addition, and for the sake of comparison, we shall also consider causets that take the form of a regular lattice. In the sequel we use the concept of *order interval*, often abbreviated to '*interval*': if s and t are two points of a discrete or continuous set with partial order $\prec$, then the *order interval* between them, denoted $I[s, t]$, is the set $\{x \mid s \prec x \prec t\} \cup \{s, t\}$.

2.1 Minkowski sprinkling

A Minkowski sprinkled causet is obtained by applying the above mentioned sprinkling technique to a portion of D-dimensional Minkowski space. For example, if $M^{(1,2)}$ is 3D Minkowski space with time dimension t and space dimensions x and y, L^2 is the squared Lorentz distance $L^2\big(p(t_p, x_p, y_p),\ q(t_q, x_q, y_q)\big) = +(t_p - t_q)^2 - (x_p - x_q)^2 - (y_p - y_q)^2$, and C is a bounded, connected subset of it, e.g. a unit cube, then a set S of points uniformly distributed in C yields a sprinkled causet graph $G(S, E)$, where the set of graph nodes is S itself and the edges E are the ordered pairs of nodes (p, q) such that q is in p's future lightcone: $E = \{(p, q) \in S^2 \mid L^2(p,\ q) \geq 0 \wedge t_p < t_q\}$. (Once the graph is constructed, node coordinates become irrelevant.) In general a sprinkled causet may have several sources - nodes whose in-degree is zero - and sinks - nodes whose out-degree is zero. However, given two nodes s and t of a (sprinkled) causet $G(S, E)$, the order interval $I[s, t]$ has, by definition, only one source (s) and one sink (t).

2.2 De Sitter sprinkling

De Sitter spacetime is an exact solution of the Einstein's field equations of General Relativity, conjectured to model the universe both at the Plank era (at time $t < 10^{-44}$ sec) and in the far future, at thermodynamic equilibrium. De Sitter *spacetime* has constant positive curvature, and describes the universe as an empty and flat *space* - one with zero curvature - which grows exponentially under the effect of a positive cosmological constant.

We shall restrict to the special case of two-dimensional de Sitter spacetime $DS^{(1,1)}$, which can be conveniently represented (see, e.g., [17]) as a one-sheeted, 2D hyperboloid embedded in flat 3D Minkowski space $M^{(1,2)}$. Let t, x, y be the axes of the latter, where t is the vertical time axis around which a hyperbola rotates to create the hyperboloid, and let τ and θ be the time and space coordinates in $DS^{(1,1)}$: in the embedding, τ flows vertically while angular coordinate θ spans space - a circle of constant coordinate t (and τ). Given the hyperboloid equation $x^2 + y^2 - t^2 = 1$, the differential element of Lorentzian length $ds^2 = dt^2 - dx^2 - dy^2$, and its vertical projection $d\tau^2 = dt^2 - dx^2$, one can readily derive, by integration of the latter along a vertical hyperbolic segment, the functional dependences $\tau(t) = \text{ArcSinh}(t)$ and $t(\tau) = \text{Sinh}(\tau)$. Similarly, the radius of the space circle at time τ is $r(\tau) = \text{Cosh}(\tau)$, hence the size of de Sitter space is $2\pi\text{Cosh}(\tau) = \pi(e^{\tau} + e^{-\tau})$. Thus, de Sitter space grows exponentially with time. (For a visualization of this result, see the interactive demo [10].) It is finally easy to see that the differential element of Lorentzian length in $DS^{(1,1)}$is $ds^2 = d\tau^2 - \text{Cosh}(\tau)\, d\theta^2$.

For implementing de Sitter sprinkling we need to know how to distribute points on the surface of the hyperboloid, and how to build causal edges. If δ is the desired uniform density, the expected number of points on the circular hyperboloid section S between τ and $\tau + d\tau$ is $\delta * Area(S) = \delta * 2\pi\text{Cosh}(\tau)d\tau$. Thus, if we uniformly sprinkle in a section of the hyperboloid with $0 \leq \tau \leq \tau_{\max}$, the time coordinate of these points is a *random variable* $\boldsymbol{\tau}$ whose normalized density can be calculated to be:

$$f_{\tau}(x) = Cosh(x)/Sinh(\tau_{\max}). \tag{1}$$

The Minkowski coordinate t of these points is itself a random variable $\boldsymbol{t}$. Keeping in mind the above functional relations $t(\tau)$ and $\tau(t)$, and using a fundamental theorem on functions of random variables, we find that the density f_t of $\boldsymbol{t}$ is constant:

$$f_t(x) = f_{\tau}(\tau(x))/t'(\tau(x)) = Csch(\tau_{\max}) \tag{2}$$

('*Csch*' is the hyperbolic cosecant). Then, in light of the circular symmetry of the distribution, implementing a Poisson sprinkling in $DS^{(1,1)}$, with τ ranging in $[0, \tau_{\text{Max}}]$, is straightforward. We create points with coordinates (r, θ, t), where polar coordinates (r, θ) replace (x, y), such that:

- t is distributed *uniformly* in $[0, t_{\max}]$, where $t_{\max} = Sinh(\tau_{\text{Max}})$,
- $r = \sqrt{t^2 + 1}$
- θ is distributed *uniformly* in $[0, 2\pi)$.

Once the points are uniformly distributed in $DS^{(1,1)}$as described, causal edges can be established among them by referring directly to their Lorentz distance in $M^{(1,2)}$: although the squared Lorentz distance changes when moving from p to q on a geodesic in $DS^{(1,1)}$or one in $M^{(1,2)}$, the signs of these two distances always agree, thus yielding the same causal structure.

2.3 Percolation dynamics

Transitive percolation dynamics has been widely studied in the context of the Causal Set Programme. An N-node percolation causet is built by progressively numbering nodes 1, 2, ..., N, and by creating an edge $i \to j$, for any i and j such that $i < j$, with fixed probability p (typically $p \ll 1$). Note that a percolation causet may be disconnected and may include several source nodes. A variant called *originary percolation*, which guarantees the existence of a single source node, or *root*, is studied in [1], where it is also observed that any subset S of a standard, transitive percolation causet C, consisting of a node x and all nodes in its future, is an instance of an originary percolation causet. However, for our statistical analyses we shall always take a whole, transitive percolation causet, which can be seen as the union of a number of such smaller, originary percolation causets.

2.4 Popularity/Similarity dynamics

This technique is described in [17], where it is also shown that the growth dynamics is *asymptotically* identical to that of sprinkled causets from de Sitter space. Each node is assigned a progressive natural number n, as in percolation dynamics, and is placed in 2D Euclidean space, with polar coordinates (r, θ), where r is a monotonic, increasing function of n (whose precise nature is not essential for building the

causet), and θ is a random angle uniformly chosen in $[0, 2\pi)$. When new node n is created, a fixed number m of edges reaching n from previously created nodes is introduced: $s_i \to n$ ($i = 1, 2, ..., m$). Typically, $m = 2$. Nodes s_i are those that minimize the product $s_i \Delta\theta_i$, where $\Delta\theta_i$ is the angular distance of node s_i from node n.

The name 'popularity/similarity', abbreviated as 'pop/sim' in the sequel, reflects the tradeoff between popularity and similarity that determines the structure and drives the growth of various complex networks, including the Internet. Node labels (birth dates) represent 'popularity', while angular distances express 'similarity': when the from-nodes s_i are chosen, for being connected to new node n, older nodes - with smaller labels - are preferred, and in the long run they become more and more 'popular', thus increasing their out-degrees; at the same time, the chosen nodes must be as 'similar' as possible (small angular distance) to the target node n.

2.5 Regular grids

For the sake of comparisons we find it useful to consider also directed regular grids, in particular square or cubic grids whose edges are oriented parallel to the cartesian axes of their embedding euclidean 2D or 3D spaces. These DAGs manifest maximum locality, as opposed to the non-locality typical of sprinkled causets from a Lorentzian manifold.

Randomized versions of maximally local DAGs can be obtained by sprinkling points in Euclidean space of some dimension, where one of the dimensions is time, and by creating an edge from node p to node q whenever q has higher time coordinate than p and the Euclidean distance between p and q is smaller than a given threshold.

3. Counting edges in transitively closed causets

In this section we try to characterize causets by collecting statistical information based on edge counts for transitively closed graphs. We focus on two types of statistical indicator: *node degree distributions* and *ordering fraction spectra*. These are first applied to Minkowski sprinkled causets, and then to the other causet classes. Sprinkled causet are transitively closed by construction, thus we shall take transitive closures for all other causet types, before counting edges.

3.1 Node degree distribution for transitively closed causets

Minkowski space $M^{(1,1)}$ represents a spacetime that extends to infinity in both dimensions; a finite causet is obtained by uniformly sprinkling points in a bounded region of it. The statistics of such causets, however, are affected by the shape of the region border. One way to avoid this problem is to analyze causet intervals, as defined above. The three upper *LogLog* plots in Figure 1 show the distributions of the node degrees for three interval causets: an 8000-node sprinkled 2D Minkowski interval, an 8000-node sprinkled 2D de Sitter interval, and a 3025-node regular, square grid, which is also an interval, with source $s(1, 1)$ and sink $t(55, 55)$ (nodes have integer coordinates (n, m) ranging in $\{1, ..., 55\}$, and edges are oriented upward or to the right). Note that we shall always refer to the *out*-degrees.

The histograms for the three interval causets contain 50 points each; they were obtained by partitioning the range of possible degrees into 50 slots of equal width and by counting the nodes whose degree fall in each slot. In the first plot the continuous line represents the theoretical density -Log(z), appropriately scaled. This negative logarithm density can be derived by assimilating the node degree to a random variable (r.v.) $\boldsymbol{r}=\boldsymbol{x}*\boldsymbol{y}$, where r.v.'s $\boldsymbol{x}$ and $\boldsymbol{y}$ have uniform densities in the unit interval: $f_{\boldsymbol{x}}(z) = 1, f_{\boldsymbol{y}}(z) = 1, z \in [0,1]$. The reason is that, by a -45 degree rotation, we can represent the sprinkled Minkowski interval as a unit square $[0, 1]\times[0, 1]$ in Euclidean 2D space, and model the degree of the generic sprinkled point $p(1-\boldsymbol{x}, 1-\boldsymbol{y})$ as the area $\boldsymbol{x}*\boldsymbol{y}$ of rectangle whose sides have lengths $\boldsymbol{x}$ and $\boldsymbol{y}$. The distribution function for $\boldsymbol{r}$ is: $F_{\boldsymbol{r}}(z) \doteq \text{Prob}[\boldsymbol{x}*\boldsymbol{y} \le z] = z(1-\text{Log}(z))$, yielding, by derivation, the density function $f_{\boldsymbol{r}}(z) = -\text{Log}(z)$. Note that this same analysis is valid for the square grid (third plot).

In the central part of Figure 1 we show the node degree densities for transitively closed causets with rotational symmetry. The first plot refers to a causet obtained by uniformly sprinkling points on a rectangular portion of $M^{(1,1)}$ that wraps around a cylinder of unitary height - a 'Minkowski cylinder'. The out-degree of a node is computed as the area of the future lightcone of that node. The distance $\boldsymbol{h}$ of the generic node p from the upper border of the cylinder is a r.v. with uniform density: $f_{\boldsymbol{h}}(x) = 1$, for $x \in [0,1]$. Assuming that the cylinder radius is a large enough, so that no lightcone overlaps with itself, the area of p's future lighcone is r.v. $\boldsymbol{A} = \boldsymbol{h}^2$. By applying well known results on functions of r.v.'s, we obtain $f_{\boldsymbol{A}}(y) = \frac{1}{2} y^{-1/2}$, $y \in [0, 1]$ for the density of r.v. $\boldsymbol{A}$. Thus the density is a power-law, which explains the linearity of the *LogLog* plot in figure.

The second case of rotationally symmetric causet is a square grid arranged around a cylinder, with edges at +45 and -45 from the cylinder axis - a regular, euclidean counterpart of the sprinkled Minkowski cylinder. The analysis of the node degree density is somewhat analogous to that of the Minkowski cylinder. The histogram is obtained by partitioning the degrees into 50 slots, and reveals a power-law distribution.

The next two plots show node degree densities for causets obtained by sprinkling in full sections of 2D de Sitter spacetime, that is, in portions of the hyperboloid discussed in subsection 2.2 that span the whole spatial (angular) dimension and a time interval $[0, \tau_{\max}]$, with $\tau_{\max} = 1$ (l.h.s. plot) and $\tau_{\max} = 5$ (r.h.s. plot). Note that in these cases we directly plot the number of occurrences of individual degree values, without need to segment the spectrum of node degrees into (50) classes. The two *LogLog* plots confirm the power-law character of these distributions, that was first discovered in [17]; the interested reader can find in that reference a discussion and cosmological interpretation of the upward convexity particularly apparent in the second plot, which corresponds to a non constant γ exponent of the involved power law expression.

In the lower part of Figure 1 we show the node degree densities for percolation and pop/sim causets (described in Section 2). Note that the latter, described in terms of polar coordinates (r, θ), could also be regarded as rotationally symmetric; one difference with the other causets of that class, however, is that the r 'coordinate' plays here a somewhat abstract role, being used purely for ordering nodes as they are created, not for defining a particular metric of the manifold where those causets are created by sprinkling.

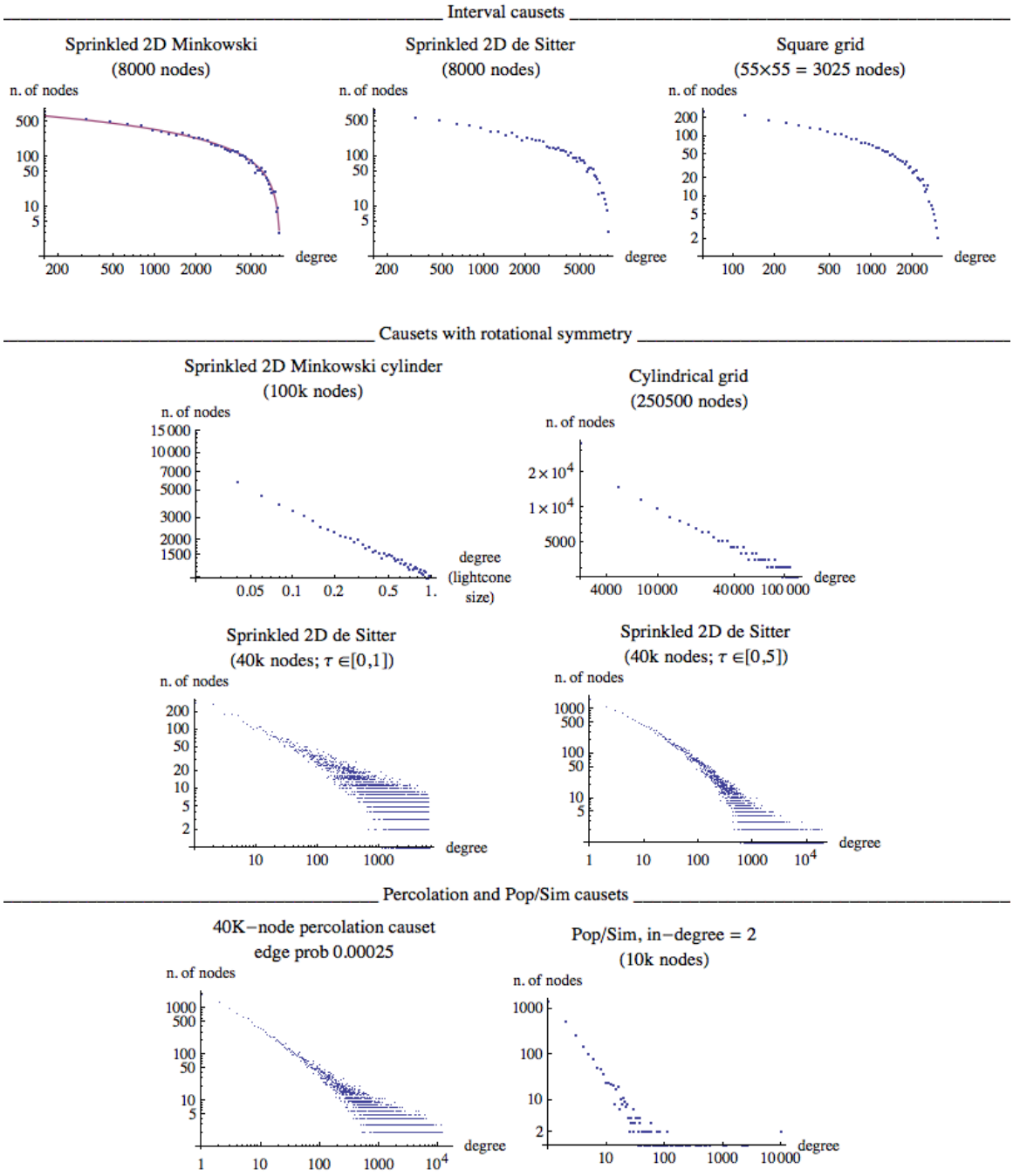

Figure 1 - Node degree distributions for transitively closed causets

The first plot shows the node degree density of a 40k-node percolation causet in which the probability of finding an edge between any two nodes i and j is $p(i, j) = 0.00025$. Node degrees are computed after taking the transitive closure of the graph. The second plot shows the node degree density of the transitive closure of a 10k-node pop/sim causet. In both cases the *LogLog* plots appear roughly linear, suggesting a power law distribution, although the γ exponent in the two cases is clearly different (in particular, the analysis in [17] calculates an asymptotic value $\gamma = 2$ for pop/sim causets). Note that the presence of exactly two nodes with degree 10^4 in the pop/sim plot is not accidental: these are the two initial nodes necessary for starting the procedure, which consists in iteratively adding a new node with two incoming edges from exactly two previously generated nodes. One can easily see that any created node can be reached by the initial two nodes, so that these have maximum degree in the transitively closed graph.

Two conclusions can be drawn by a qualitative analysis of the plots in Figure 1. First, node degree distributions of transitively closed causets fail to characterize Lorentzianity, since they do not distinguish between a sprinkled Minkowski graph -- the Lorentzian causet of reference -- and a regular grid, which is the typical example of a non-Lorenzian causet. Second, while [17] seems to attribute a special status to de Sitter sprinkled causets - in that they share a power law distribution of node degrees with many other complex networks (social, biological, or artificial) - it is clear that this property, in itself, is not so discriminatory, since it can be found also among the other transitively closed causets usually discussed in connection with discrete models of spacetime - sprinkled Minkowski, percolation, pop/sim (and grid).

3.2 Ordering fraction spectra

Counting the edges of transitively closed causets, or of their order intervals, is also at the basis of the Myrheim-Meyer dimension estimator. Let $I_k^D[s, t]$ be a k-node interval with source s and sink t, obtained from a causet sprinkled in D-dimensional Minkowski space. The number $R(D, k)$ of edges in the interval is given by [19, 22]:

$$R(D,\ k)\ =\ f(D)\binom{k}{2}$$

where

$$f(D) = 3/2\binom{3\,D/2}{D}.$$

In general, the *ordering fraction* of a k-node causet (interval) is defined as the ratio $R/\binom{k}{2}$ between the number R of edges in it and the maximum number $\binom{k}{2}$ of directed edges that could connect so many nodes. In light of the above equations, the ordering fraction of sprinkled D-dimensional Minkowski intervals is exactly defined by function $f(D)$. We can then obtain an estimation of the Myrheim-Meyer dimension D of a generic causet by counting its nodes (k) and edges (R) and numerically inverting function $f(D)$:

$$D = f^{-1}\left(R\Big/\binom{k}{2}\right).$$

The upper-left plot of Figure 2 was obtained by creating 100 intervals of random volume (number of points) lower than 1000 in 2D-, 3D- and 4D-Minkowski space, and by plotting, for each interval I, the point (*Volume*(I), *OrderingFraction*(I)). The points nicely align to the expected values - the horizontal gridlines - which mark the values of $f(D)$ for the indicated dimensional values. We call these diagrams *ordering fraction spectra*.

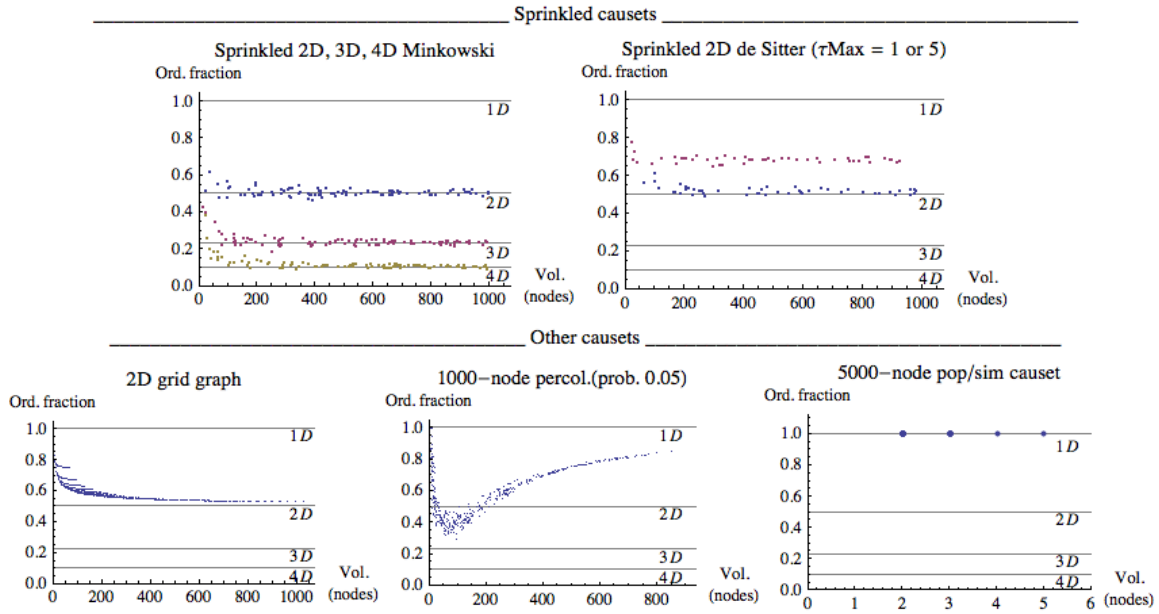

Figure 2 - Ordering fraction spectra for causet classes

The upper-right plot of Figure 2 was obtained by creating 50 intervals of random volume -- again less than 1000 points -- in 2D de Sitter spacetime; for the lower data set, that matches well the 2D ordering fraction value 0.5, the time variable τ ranges in [0, 1], while for the upper data set the range is [0, 5]. Let us stress that *both* data sets are derived from sprinkling in (the hyperboloid model of) *2-dimensional* de Sitter spacetime. This plot thus indicates that, while the Myrheim-Meyer dimension estimator operates correctly when sprinkling in flat (e.g. Minkowski) spacetime, it becomes unreliable when sprinkling in curved manifolds, becoming sensitive to the interval time span. In any case, based on the above definition of ordering fraction, the flat nature of the plots for sprinkled (Minkowski and de Sitter) causets implies that in these causets the growth rate of function $R(k)$, counting the edges as a function of the nodes, is $O(k^2)$.

In the lower part of Figure 2 we show three more ordering fraction spectra. The first diagram plots the (*Volume*(I), *OrderingFraction*(I)) pairs for *all* intervals I that can be found in a 32×32 directed square grid. The ordering fraction value for an n×m directed square grid interval is:

$$\frac{\sum_{x=1}^{m}\sum_{y=1}^{n} x\,y - m\,n}{\binom{m\,n}{2}} \qquad (3)$$

The plot indicates that the Myrheim-Meyer dimension estimation is only asymptotically correct for this simple case.

The second plot shows the typical shape of the ordering fraction spectra for causets obtained from percolation dynamics using constant edge probability. The plot was obtained by randomly sampling 500 intervals from the transitive closure of a 1000-node causet with edge probability 0.05. On the large scale, percolation causets tend to be one-dimensional.

The poor ordering fraction spectrum in the third plot was obtained by sampling 500 intervals from a 5000-node pop/sim causet. The repertoire of possible volume-ordering fraction pairs is very limited here, with volumes below 6 nodes, and the single value 1 for the ordering fraction. The preferential attachment policy implicit in the pop/sim algorithm -- new nodes prefer to connect with the popular nodes with lowest birth times -- yields causets with very short chains, thus small intervals, that are totally ordered (a necessary and sufficient condition for the ordering fraction to be 1) or *almost* totally ordered. An exhaustive search of all the intervals from this 5000-node causet reveals that the lowest possible ordering fraction value is indeed 21/22, achieved by only three intervals of volume 12.

In light of the fact, claimed in [17], that pop/sim and sprinkled deSitter causets share the same asymptotic dynamics, we would be interested in comparing the ordering fraction spectra in the upper-right plot (de Sitter) and lower-right plot (pop/sim) of Figure 2. Unfortunately, the limited interval volumes that we can attain for pop/sim causets, as just discussed, makes the comparison unfeasible.

However, much lower ordering fraction values are achieved for *whole* pop/sim causets, rather than for their intervals. In particular, we experimentally find that the ordering fraction of a k-node such causet is not constant: it *decreases with* k, and is an $O(k^{-1})$ function. Since the ordering fraction is $R(k)/\binom{k}{2}$, where $\binom{k}{2}$ is $O(k^2)$, this result indicates that $R(k)$ -- the number of edges in the transitively closed causet -- is $O(k)$, like the number $RR(k)$ of edges in the original, 'raw' causet. (The linearity of $RR(k)$ is a direct consequence of adding a fixed number of edges with each new node.) Thus, in spite of the asymptotic similarity of full de Sitter causets and pop/sim causets, we find that these causet classes differ in the growth rates of their edges, which are, respectively, $O(k^2)$ and $O(k)$.

Note that the strong effect of the preferential attachment in pop/sim causets is also revealed by comparing the plots for node degree distributions of *full* sprinkled 2D de Sitter causets (in the central part of Figure 1), and pop/sim causets (lower part of the same figure): clearly, the latter exhibits a power law distribution whose exponent ($-\gamma$) has higher absolute value.

The conclusion we can draw from the inspection of the above plots is that ordering fraction spectra, while useful for possibly detecting causet dimensionality, are not relevant for revealing causet Lorentzianity, since, again, they do not separate sprinked (Lorentzian) causets from a regular grid (non-Lorentzian).

4. Counting links in transitively reduced causets

In this section we try to characterize causets by collecting statistical information based on edge counts and paths for *transitively reduced* graphs. 'Links' is the name commonly adopted in the Causal Set programme for indicating the essential edges of a causet -- those left after transitive reduction. (Recall that the transitive reduction of a relation R is the smallest relation that admits the same transitive closure of R. When R is acyclic, its transitive reduction is unique.)

A first simple indicator is suggested by this quote from D. Rideout [20]:

"*The 'usual' discrete structures which we encounter, e.g. as discrete approximations to spatial geometry, have a 'mean valence' of order 1. e.g. each `node' of a Cartesian lattice in three dimensions has six nearest neighbors. Random spatial lattices, such as a Voronoi complex, will similarly have valences of order 1 (or perhaps more properly of order of the spatial dimension). Such discrete structures cannot hope to capture the noncompact Lorentz symmetry of spacetime. Causal sets, however, have a `mean valence' which grows with some finite power of the number of elements in the causet set. It is this 'hyper-connectivity' that allows them to maintain Lorentz invariance in the presence of discreteness.*"

It should perhaps be clarified that we cannot meaningfully apply the concept of Lorentz invariance *directly* to a causet C, simply because there is no coordinate system to which to apply the Lorentz transformation. But when the causet C comes already embedded in some manifold M, with its coordinate system, we may indeed decide to *keep the information on node coordinates*, apply the transformation to M, and see its effect on C, which is now dragged and reshaped by the operation. When C is obtained by a Poisson sprinkling in M, it can be shown that in the equivalence class of all Lorentz-transformed *embedded versions* of C there is no preferred element that we can pick out, while if we embed, for example, a regular grid, this is not the case. In the above quote, Rideout establishes a connection between this invariance property of sprinkled causets (embeddable by definition) and the growing degree (valence) of their nodes.

Note that this degree refers, here and in the rest of the section, to the *links,* as defined above. If we consider the familiar discrete structures mentioned in the quote, such as the Cartesian lattice - a regular grid - or a Voronoi complex, we have the following situation. The node degrees of the *transitive closure* do grow with the number of nodes; for example, the degree of the root of a transitively closed $m \times n$ directed square grid DAG of the type already considered is $m*n$, implying a linear growth w.r.t. the number of nodes. But the node-degree growth rate becomes O(1) in the transitive reduction, which coincides with the grid itself. The peculiarity of sprinkled causets is that they have a 'mean valence' that grows with the number of nodes *both* in the transitive closure and, more significantly, in the transitive reduction.

Our general approach here is to look for statistical properties that are necessarily satisfied by causets whose embeddability is guaranteed, namely sprinkled causets, and then use these properties for checking *directly* the given C, without emphasizing on its embeddability. As a first property, let's look at the growth rate of node degrees in transitively reduced sprinkled causets.

4.1 Growth rate of node degrees

In [13] Bombelli et al. mention that, considering the causet $C[s, t]$ obtained from uniformly sprinkling points in an order interval $I[s, t]$ of height T of d-dimensional Minkowski space (T being the Lorentz distance between s and t), the number of nearest neighbors of root s in the interval -- the number of outgoing links -- grows like $Log(T)$ for $d = 2$, and like T^{d-2} for $d \geq 3$, provided that the sprinkling density is

kept constant. Equivalently, we can consider an interval of fixed, unitary size (height), and increase the number k of nodes sprinkled in it (thus, the density), in which case the root degree $deg(k)$ has growth $O(Log(k^{1/2})) = O(Log(k))$ for $d = 2$, and $O(k^{(d-2)/d})$ for $d \geq 3$. (Note. We detect a possible typo in [15], equation (8), which indicates that the number of links from the source of a $(d+1)$-dimensional interval with N nodes is proportional to $N^{(d-1)/2}$ for $d > 1$. In light of the above results, the formula should read $N^{(d-1)/(d+1)}$).

In this subsection we carry out explicity the analyses of cases $d = 2$ and $d = 3$ (not given in [13]); we obtain the precise expressions for $deg(k)$, not just their growth rate, and verify their agreement with experimental data.

For the analysis of the two-dimensional case we find it convenient, as done before, to represent the sprinkled interval as a unit square $[0, 1]\times[0, 1]$ in Euclidean 2D space, where the k points are uniformly distributed, and where the interval root is $s(0, 0)$. By the definition of 'link', $deg(k)$ corresponds now to the number of points p_i, out of k, that form empty rectangles $R_i(s, p_i)$; the latter are identified by their lower-left and upper-right vertices. Three such rectangles are depicted in Figure 3-left; their purple diagonals are the only links from the root s, for the causet (not shown) associated with the 13-point sprinkling.

Given that the distribution is Poisson with density k, the probability $P^n_{x,y}$ of finding n points in the generic rectangle $R_{x,y}(s, p(x, y))$, identified by point $p(x, y)$, and excluding s from the count, is $(kxy)^n e^{-kxy}/n!$, and the probability for the rectangle to be empty is $P^0_{x,y} = e^{-kxy}$. Thus, the probability of each sprinkled point $p_i(x, y_i)$ to be the endpoint of a link (s, p_i) is $e^{-k x_i y_i}$, and we can write:

$$deg(k) = \sum_{i=1}^{k} e^{-k x_i y_i} \tag{4}$$

Since the k points are uniformly distributed, we can use the summation in (4) as an approximation of the integral of the exponential function over the unit square:

$$\sum_{i=1}^{k} (1/k)\, e^{-k x_i y_i} \approx \int_{x=0}^{1}\int_{y=0}^{1} e^{-kxy}\, dx\, dy \tag{5}$$

where $(1/k)$ represents the fraction of unit square area charged to each 'sample' $e^{-k x_i y_i}$. Since the integral at the l.h.s. evaluates to $\frac{\text{EulerGamma} + \text{Gamma}[0,k] + \text{Log}[k]}{k}$, we immediately obtain

$$deg(k) \approx EulerGamma + Gamma[0, k] + Log[k] \tag{6}$$

where *EulerGamma* is the Euler-Mascheroni constant (0,5772...), and the contribution of the 'incomplete gamma function' $Gamma[0, k]$ is vanishingly small.

The plot in Figure 3-right plots experimental data for the degree of the root of a sprinkled, transitively reduced 2D Minkowski interval causet, for densities from $k = 100$ to $k = 3000$, in steps of 100 (the dots), matched by the theoretical degree $deg(k)$ of equation (6) (solid line). Each data point was obtained by averaging over 20 intervals of fixed density.

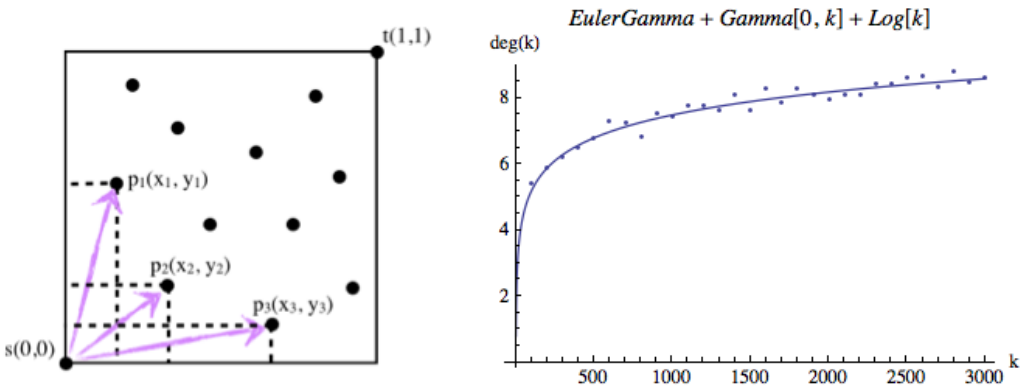

Figure 3 - Logarithmic growth of the degree of the root node of a causet obtained by sprinkling k points in a fixed (and rotated) 2D-Minkowski interval, as a function of k. Match between data and analytical expression.

For the analysis of the 3D case we use a similar approach. We sprinkle a set S of k points in a 3D interval $I[s, t]$ of Minkowski space $M^{(1,2)}$, assuming, without loss of generality, $s = (0, 0, 0)$ and $t = (0, 0, 2)$; this is depicted as the yellow double cone in Figure 4.

Let V denote the volume of the double cone $I[s, t]$, and δ denote the density of the sprinkling: $V = 2\pi/3$ and $\delta(k) = k/V$. Any point u inside interval $I[s, t]$ identifies a sub-interval $X[s, u]$, delimited, in Figure 4, by the lower yellow cone and the blue cone. Analogous to the 2D case, the probability of sub-interval $X[s, u]$ to be empty is:

$$P^0_u = e^{-\delta\, \text{Vol}(X[s,u])} \tag{7}$$

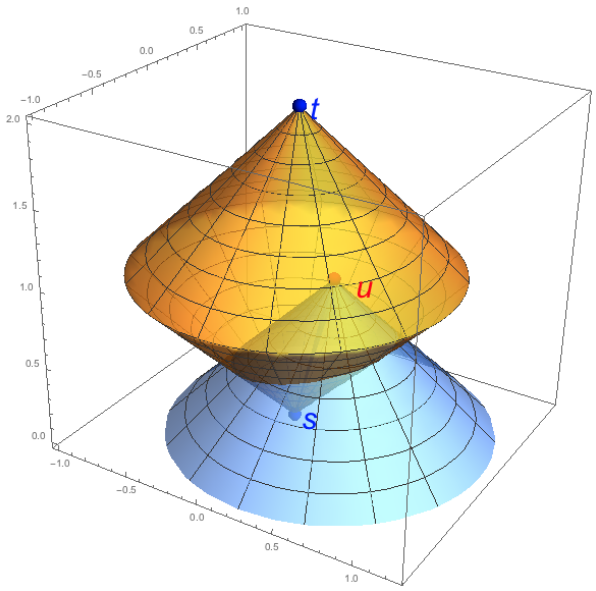

Figure 4 - Interval *I*[*s*, *t*] in Minkowski space $M^{(1,2)}$ (in yellow), and a sub-interval *X*[*s*, *u*], for a point *u* inside *I*[*s*, *t*] (in blue). The analysis of the degree of *s* makes use of the volume of *X*[*s*, *u*] as a function of the position of *u*.

Thus we need to compute the volume $Vol(X[s,u])$ of the sub-interval identified by point u. It is convenient to use polar coordinates (r,θ,z) for u, where z corresponds to the time coordinate of $M^{(1,2)}$. The intersection between the two cones is an ellipsis whose long and short axes have lenghts, respectively, $\sqrt{z^2+r^2}$ and $\sqrt{z^2-r^2}$, yielding an area of $\frac{\pi}{4}\sqrt{z^4-r^4}$. The surface of interval $X[s,u]$ is the combination of two skew cones, each of height $\frac{1}{2}(r^2-z^2)\Big/\sqrt{r^2+z^2}$. We conclude that:

$$Vol(X[s,u]) = \frac{\pi}{12}\left(z^2-r^2\right)^{3/2} \tag{8}$$

(In light of the cylindrical symmetry, θ plays no role in the above calculation.)

Analogous to eq. (4) for the 2D case, we express the degree of s as a summation:

$$deg(k) = \sum_{u\in S} e^{-\delta\,\mathrm{Vol}(X[s,u])} \tag{9}$$

and approximate it by the integral, in cylindical coordinates, in interval $I[s,t]$:

$$\sum_{u\in S}(V/k)\,e^{-\delta\,\mathrm{Vol}(X[s,u])} \approx \int_{z=0}^{1}\int_{\theta=0}^{2\pi}\int_{r=0}^{z} r*P^0_{u(r,\theta,z)}\,dr\,d\theta\,dz \;+\; \int_{z=1}^{2}\int_{\theta=0}^{2\pi}\int_{r=0}^{2-z} r*P^0_{u(r,\theta,z)}\,dr\,d\theta\,dz \tag{10}$$

where (V/k) represents the fraction of the $I[s,t]$ volume charged to each of the k 'samples' in the summation, and integration is done in two steps, corresponding, respectively, to the lower and upper cones of $I[s,t]$. If we analogously partition the sprinkled set S into a lower and an upper part, and express

$$deg(k) = \mathrm{deg}_{lower}(k) + deg_{upper}(k) \tag{11}$$

with the obvious meaning of the symbols, we obtain:

$$deg_{\mathrm{lower}}(k) \approx (k/V)\int_{z=0}^{1}\int_{\theta=0}^{2\pi}\int_{r=0}^{z} r*Exp\left[\left(\delta(k)\,\frac{\pi}{12}\left(z^2-r^2\right)^{3/2}\right]dr\,d\theta\,dz \tag{12}$$

which yields (using *Mathematica* for integration):

$$deg_{lower}(k) = -8+8\,e^{-k/8}-k\,ExpIntegralE\left[\frac{1}{3},\frac{k}{8}\right]+6\,k^{1/3}\,Gamma\left[\frac{5}{3}\right] \tag{13}$$

Similarly:

$$deg_{upper}(k) \approx (k/V)\int_{z=1}^{2}\int_{\theta=0}^{2\pi}\int_{r=0}^{2-z} r*Exp\left[\left(\delta(k)\,\frac{\pi}{12}\left(z^2-r^2\right)^{3/2}\right]dr\,d\theta\,dz \tag{14}$$

yielding:

$$deg_{upper}(k) = k\, ExpIntegralE\left[\frac{1}{3}, \frac{k}{8}\right] + \frac{4}{3}\left(3\, e^{-k} - 6\, e^{-k/8} - 3\, k\, ExpIntegralE\left[\frac{1}{3}, k\right] + \frac{Gamma\left[\frac{1}{3}\right]}{k^{1/3}} - \frac{Gamma\left[\frac{1}{3}, k\right]}{k^{1/3}}\right) \quad (15)$$

All components of $deg_{upper}(k)$ vanish as $k \to \infty$, while the dominant component of $deg_{lower}(k)$ is $6\, k^{1/3}\, Gamma\left[\frac{5}{3}\right]$ (where $6*Gamma\left[\frac{5}{3}\right] = 5.416$), confirming the growth rate $O(k^{(d-2)/d})$ for dimension $d = 3$, indicated in [13].

Figure 5 plots experimental data for $deg(k)$, for k from 300 to 5400, in steps of 300 (the dots), in comparison with the theoretical degree $deg(k)$ of equation (11) (with (13) and (15)). Each data point was obtained by averaging over 50 intervals of fixed density. The match is quite satisfactory.

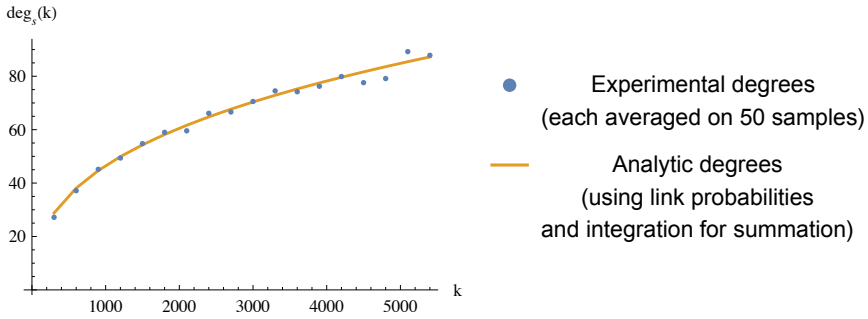

Figure 5 - $O(k^{1/3})$ growth of the degree of the root node of a causet obtained by sprinkling k points in 2D-Minkowski intervals, as a function of k. Match between data and analytical expression.

How about the other causet classes?

For the case of sprinkled de Sitter intervals the theoretical analysis of node degree growth appears harder, and is not carried out (the node degree *distribution* for de Sitter causets is studied in [17]). Problems arise even for the experimental analysis. In the 2D case we can no longer take advantage of the interval rotation and empty rectangle count, thus we must actually build the causet and then face the computational bottleneck of transitive reduction; however, for sprinkling densities up to $d = 2000$ points we could verify, with *Mathematica* code, that node degrees exhibit an apparent $O(Log(d))$ growth, analogous to that of the Minkowski case but with smaller values.

For percolation causets, and looking at the largest scales, the node degree growth turns out to be bounded. This is related to the phenomenon of 'posts', which was first studied in [4]. A 'post' is a node x that creates a bipartition of the *whole* set of causet nodes into the future and the past lightcone of x. In other words, *all* causet nodes are related to x. Reference [4] proves that posts almost surely keep appearing undefinitely, yielding a 'bouncing universe' picture; a direct consequence is that the degree of the causet root must be bounded by the number of nodes in the interval between the root and the first post. Note that the presence of infinite posts in an infinite percolation causet is consistent with the idea of an asymptotic dimension 1; the same conclusion is suggested (at a smaller scale) by the ordering fraction spectrum in the lower-central plot of Figure 2.

Pop/sim causets suffer from the already encountered problem of extremely reduced interval sizes. In Figure 6 we show a 500-node pop/sim causet in its original and transitively reduced forms. Typically the transitively reduced causet is almost a tree, where very few nodes have very high degree, due to the implicit preferential attachment policy. Thus, order intervals are almost always just sequences of links, and we can roughly estimate that, for an n-node causet, the size of the generic interval grows like $Log(n)$: we cannot expect to obtain intervals of the order of magnitude of those used for the experimental analysis of the root degree of sprinked intervals, which consisted of thousands of nodes.

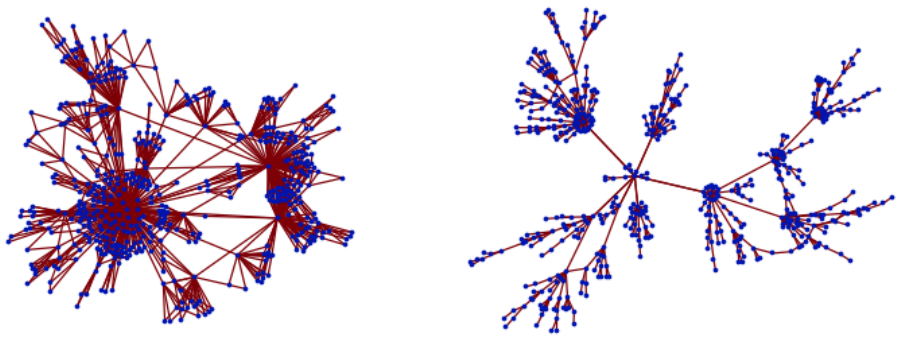

Figure 6 - A 500-node pop/sim causet in its original form (left) and transitively reduced form (right). The average node-degree of the latter is always smaller than 2 (it is here 1.026)

The case of a regular grid is trivial: node degrees do *not* grow with the number of nodes.

In conclusion, it seems natural to propose that a necessary condition for a causet to qualify as a 'good' discrete model of physical spacetime is to offer the hyperconnectivity of the causets obtained by sprinkling in a Lorentzian manifold, that is, to have intervals whose root

degrees grows unboundedly with the interval size. Additionally, in the presence of a clear logarithmic or $O(k^{(d-2)/d})$ growth we would obtain information on dimensionality. This type of analysis allows us to rule out percolation and regular grid causets; however, we found it difficult to apply it to pop/sim causets.

Let us now introduce an ad-hoc statistical indicator related to counting links, which deals with information on longest and shortest paths between nodes.

4.2 Longest/shortest path plot

In this subsection we introduce a statistical indicator meant to reflect Lorentzianity more effectively than those proposed above. More precisely, we try to characterize as directly as possible *Lorentzian non-locality*. Consider two events $p_1(t1, x1, y1, z1)$ and $p_2(t2, x2, y2, z2)$ in flat Minkowski space $M^{(1,3)}$. The (+ - - -) signature of the Lorentz metric is at the root of the twin paradox: the proper time delay $\Delta\tau$ experienced by a clock travelling from p_1 to p_2 is maximum when the trajectory is a straight line connecting the two points (first twin):

$$\Delta\tau = \sqrt{+(t1-t2)^2 - (x1-x2)^2 - (y1-y2)^2 - (z1-z2)^2}\ .$$

All other trajectories register shorter time delays (second twin); the lower limit - zero time - is represented by a trajectory formed by two segments of light rays that connect p_1 and p_2 via one intermediate point p_x, at which the trajectory forms a $\pi/2$ angle.

When a causet C is derived by sprinkling in Minkowski space M, we can approximate the Lorentz distance in M between two points $p1$ and $p2$ in timelike relation by the length (number of edges) of the *longest* chain P between them in C [19], which represents a geodesic. This correspondence is conjectured to hold also for sprinklings in curved manifolds. Note that all edges of chain P must be links, that is, edges of the transitive reduction C_{red} of C (if one were not a link, it could be replaced by two or more links, yielding a longer path and conflicting with P being the longest). Thus C_{red}, which is unique, codes all the necessary information for measuring the Lorentz distance between any two events. In analogy with the existence, in M, of infinite trajectories between $p1$ and $p2$ that are shorter, or even *much* shorter than the straight one, we may expect the (finite) number of alternative paths between *any* pair $p1$ and $p2$ in C_{red} to widely range in length, from very long to very short.

For capturing this peculiar feature of Minkowski sprinklings in a simplified way, we concentrate on the extreme cases of the longest and shortest paths between event pairs in C_{red}, and collect statistical data on them in a two dimensional array A of grey level cells, that we call *longest/shortest path plot*. Array A generally refers to a *transitively reduced* interval causet $C[s, t]$ (dropping the 'red' subscript) with source s and sink t, and is built as follows. For each event x in $C[s, t]$ we compute the pair $(longest(x), shortest(x))$ of longest and shortest path lengths from s to x, and then we let array cell $A(l, s)$ represent, by grey levels, the number of nodes found in C that, like x, correspond to those two lengths: $A(l, s) = |\{x \in Nodes(C_{\text{red}}[s, t]) \mid longest(x) = l \wedge shortest(x) = s\}|$.

Figure 7 shows longest/shortest path plots for four transitively reduced sprinkled causets from intervals in, respectively, 2D, 3D and 4D Minkowski space, and 2D de Sitter spacetime. The remarkable horizontal development of these plots directly reflects the systematic presence of very short shortest paths between nodes that are separated by increasingly long longest paths, which somehow reflects, in the discrete setting, a key feature of Lorentzian manifolds -- the reversed triangular inequality. We shall therefore consider a flat longest/shortest path plot of this type as a *necessary* requirement for any directed graph aiming at 'Lorentzianity'.

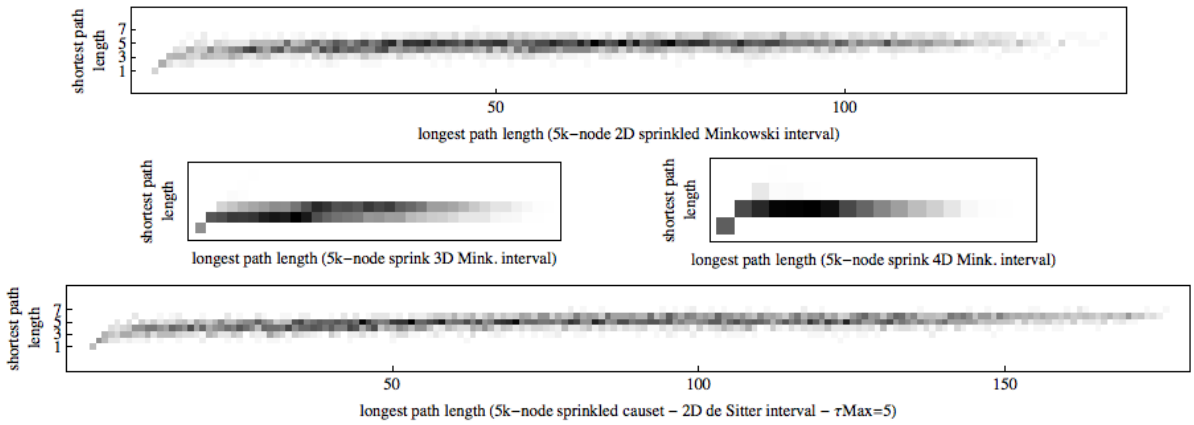

Figure 7 - Longest/shortest path plots for 5k-node causets obtained by sprinkling in 2D (upper), 3D, 4D (central row) Minkowski intervals; similar plot for 5k-node sprinkled causet from 2D de Sitter interval, with time span (0, 5).

It would be interesting to obtain an analytical account for the slow growth of the above longest/shortest path plots, at least in the case of 2D Minkowski space. For doing this, in the sequel we start by considering separately functions $lpl(k)$ and $spl(k)$, which define, respectively, the *average* longest and shortest path length from source to sink of a sprinkled 2D Minkowski interval as a function of the sprinkling density k, where the average is taken over many intervals.

Longest path length

In [14] and [3] it is shown that $lpl(k)*k^{-1/d}$ converges to some constant m_d as the number k of points sprinkled in a d-dimensional Minkowski interval grows to ∞. The value of m_d, however, has been calculated only for $d = 2$, and found to be $m_d = 2$. This immediately yields an asymptotic estimate of $2\sqrt{k}$ for $lpl(k)$. Figure 8 shows experimental data for the longest paths under investigation, a fitting function $(1.57*k^{0.526})$, and the theoretical growth $2\sqrt{k}$.

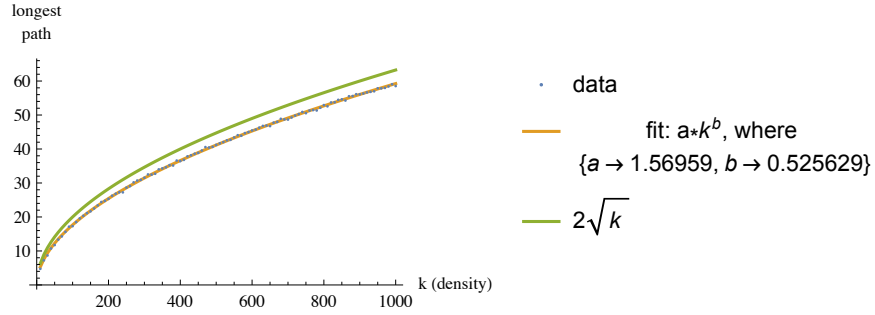

Figure 8 - Longest path length of a transitively reduced sprinkled Minkowski 2D causet as a function of the sprinkling density *k*. The function has a polynomial growth O(k^x), and we experimentally find x = 0.526. Each data point has been obtained by averaging over 50 intervals of fixed density.

The plot suggests that the convergence to the asymptotic regime $2\sqrt{k}$ is slow. (Further experiments with higher values of k indicate an ever decreasing difference between theoretical and experimental values; for $k = 5000$, for example, we find an average longest path length of 138.6, where the theoretical value is $2\sqrt{k} = 141.42$.)

Shortest path length

A detailed analysis of the shortest path length *spl*(k) is carried out in a forthcoming paper [11]. For the purposes of the present paper, it is enough to use an upper bound for function *spl*(k) based on an expectedly short (not *shortest*) path that we call *lower peripheral path*.

Given a transitively reduced causet interval $C_k[s, t]$ with source $s(0, 0)$ and sink $t(1, 1)$, obtained from a set S of k points uniformly distributed in the unit box, the *lower peripheral path* $p^0 \to p^1 \to \ldots \to p^h$ of $C_k[s, t]$, abbreviated *LPP*(k), is a path of length h (edge count) connecting s to t, with $p^0 = s$ and $p^h = t$, simply defined as the sequence of the lowest links in $C_k[s, t]$: all points of S that are not in it, fall above the polygonal line formed by its points. (Clearly, we could equivalently use the symmetric, upper peripheral path.) Note that, being a sequence of *links*, this path has the property that no point of S is found inside any of the rectangles $R[p^{i-1}, p^i]$, $i = 1, \ldots, h$; for this reason, *LPP*(k) is only vaguely similar but not the same as the lower border of the convex hull of S, which is generally shorter and may well violate this property.

The length of the shortest path is bounded above by the length *LPPlength*(k) of *LPP*(k). Since the precise value of *LPPlength*(k) is approximated by $log(k) + 1$ [11], the shortest path length has O($log(k)$) growth.

Figure 9 shows the data collected for the mean lengths *LPPlength*(k), each averaged over 50 samples, for $k = 1000$ up to 40,000 in steps of 1000, in comparison with the analytical prediction.

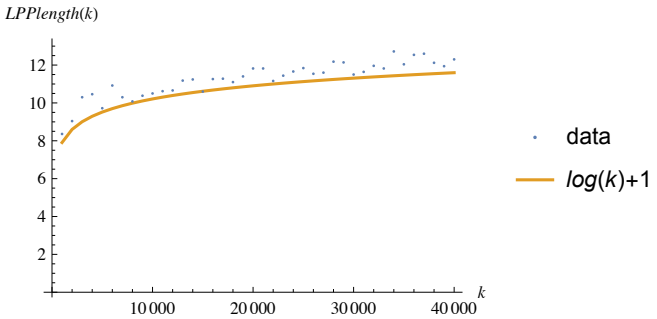

Figure 9 - Mean lengths of the lower peripheral paths *LPP*(k), each averaged over 50 samples, for sprinklings of k points in the unit box, with k ranging from 1000 to 40,000, in steps of 1000. Fit by natural logaritm.

We can now use the above results - the exact form $2\sqrt{k}$ for *lpl*(k) and the upper bound $LPPlength(k) = log(k) + 1$ for *spl*(k) - for trying to reproduce the general character of the experimental longest/shortest path plot of Figure 7 (upper plot). We do it by defining a function f that combines these two functions by eliminating variable k: $f(x) := LPPlength(lpl^{-1}(x)) = log(\frac{x^2}{4}) + 1 = 2log(x/2) + 1$. The plot for f is shown in Figure 10.

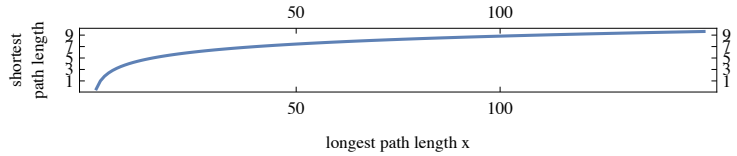

Figure 10 - Function $f(x) = log(\frac{x^2}{4}) + 1$ approximates the longest/shortest path plot for 2D Minkowski sprinkled causets under the assumptions *lpl*(k) = $2\sqrt{k}$ and *spl*(k) < log(k) + 1. Compare with Figure 7.

In spite of having used only an upper bound for the shortest path length, the plot still conveys the general idea of the dramatically growing disproportion between longest and shortest path lengths, as the number of points between the points to be connected increases.

Let us now turn to the other causet classes.

It was not possible to effectively discriminate between a regular grid and a sprinkled causet by looking at transitive closures, and using the node degree densities and ordering fraction spectra of Section 3. However, the examination of transitively reduced graphs yields visible differences. The longest/shortest path plot for a grid graph, in which the lengths of the longest and shortest paths from the root s to any node x trivially coincide, is shown in Figure 11-left.

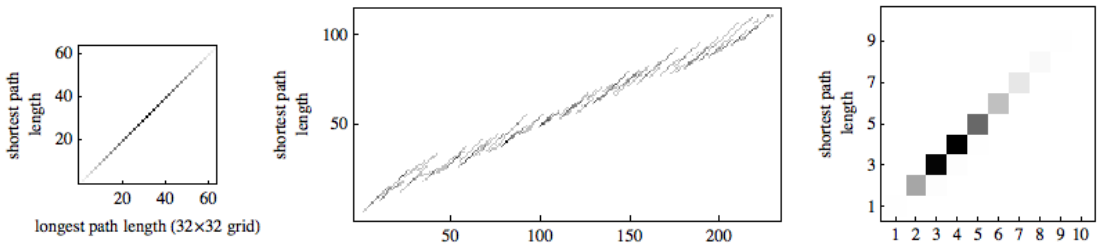

Figure 11 - Longest/shortest path plots for three causets. Left: 32×32-node grid. Center: graph obtained by sprinkling 1000 points in a unit square, where edges connect points at Euclidean distance smaller than 0.1. Right: 10k-node pop/sim causet.

The diagram provides now a particularly effective visual account of the fact that these graphs occupy the opposite extreme of the spectrum (maximum locality), relative to sprinkled causets (maximum non-locality). For the sake of comparison, we also considered a sort of randomized version of the regular grid, a 'proximity' graph obtained by sprinkling points in a 2D unit square and using the Euclidean metric and a threshold for creating edges. The longest/shortest path plot in Figure 11-center refers to a graph obtained in this way, where a directed edge from point $p_1(x_1,\ y_1)$ to $p_2(x_2,\ y_2)$ is created whenever $y_2 > y_1$ and $d(p_1, p_2) < \delta$, where d is the Euclidean distance and $\delta = 0.1$. Finally Figure 11-right shows the longest/shortest path plot for a 10k-node pop/sim causet. While the asymptotic behaviour of this class of causets approximates that of de Sitter sprinkled causets [17], our analysis reveals, again, a remarkable difference between the two causet types, at the finite scale, with pop/sim causets unable to develop short paths as alternatives to long paths to the same node. No clue of non-locality seems to emerge in pop/sim causets.

In the case of graphs from percolation dynamics, high ratios between longest and shortest path lengths can be achieved, as documented by the experimental longest/shortest path plots of Figure 12.

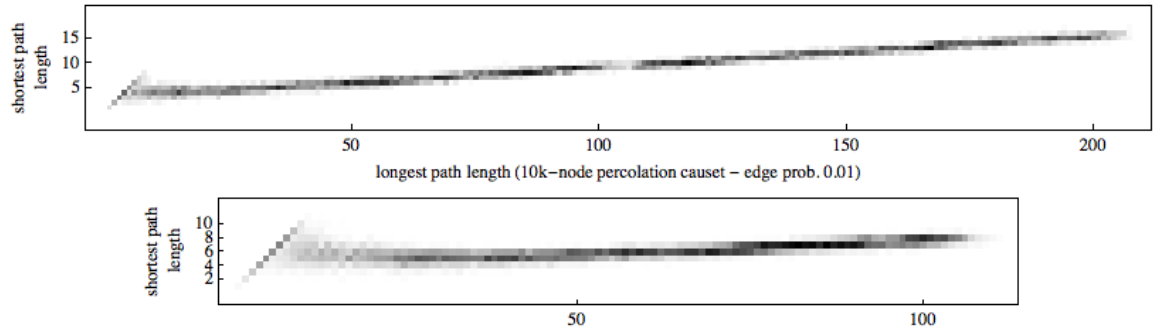

Figure 12 - Longest/shortest path plots for two causets from percolation dynamics. Upper: 10k-node causet with edge probability 0.01. Lower: 20k-causet with edge probability 0.0025.

However, as the number of nodes grows these graphs tend to become 1D structures, as revealed by their ordering fraction spectra, and as implied, on a longer run, by the already discussed phenomenon of posts. As a consequence, both the longest and the shortest paths must exhibit a roughly linear growth with respect to the number k of nodes, although the two multiplicative factors may largely differ. This implies that the longest/shortest path plot eventually assumes a linear character too. This fact is rather apparent in the upper plot of Figure 12.

Having proposed longest/shortest path plots as a useful and adequately discriminative tool for comparing causets, we observe that none of the causet classes covered in Figures 11 and 12 achieves plots comparable to the one obtained for the sprinkled causets of Figure 7. Surprisingly, a very satisfactory longest/shortest path plot is found by switching to a deterministic causet construction technique.

5. New deterministic 2D causet construction techniques

In [6, 7, 8] the first author has described various *deterministic*, *algorithmic* causet construction techniques based on simple models of computation, and conceived as completely independent from the sprinkling technique, which heavily depends on an underlying manifold. (The application to these algorithmic causet classes of the Lorentzianity indicators proposed here is the subject of a forthcoming paper.)

In this section we introduce a new family of deterministic techniques that mediates between the stochastic technique of sprinkling and the purely abstract (i.e. manifold-independent) algorithmic approach, in an attempt to achieve longest/shortest path plots -- our main indicator of Lorentzian non-locality -- comparable to those obtained for sprinkled causets, while retaining the benefits of the deterministic approach.

The main benefit to be expected from deterministic over stochastic techniques, is the emergence of structure, in one form or another, and the mix of structure and (deterministic) chaos, as widely discussed and shown in [6, 7, 8, 23]

We introduce a family of automata that we call *permutation ants* (PA), in which the control unit (the 'ant') moves on a finite array of cells by short steps or perhaps jumps while performing simple operations on them, such as reading, writing, comparing and swapping cells, or adding new ones. At each step the cell array A of length n contains a permutation π:[1.. n]⟶[1.. n] of the first n integers: $A(i) = \pi(i)$, i = *1, ..., n*.

The n-element permutation π is directly transformed into an n-node causet as follows: each node is labeled by the pair $(i, \pi(i))$, which can be understood as a pair of integer coordinates, and a directed edge $(x_1, y_1)\longrightarrow(x_2, y_2)$ is created between two nodes if and only if $x_1 < x_2$ and $y_1 < y_2$. In doing so, we are essentially still reasoning in terms of lightcones, which are now rectangles with edges parallel to the x and y axes; however, an important difference with sprinkling is that not only the ant introduces new nodes sequentially, but it may also go back and modify the structure of the causet, potentially at any location.

In the next two subsections we describe two types of PA automaton. Three important pieces of information for these descriptions are: (i) the *data structure* on which the ant operates; (ii) the set S of *situations* that are recognized by the ant; and (iii) the set R of possible ant *reactions*, that depend on the situation. Following the approach of [23], we are interested in enumerating and exploring exhaustively the complete space of instances of each automaton. If all reactions are applicable to any situation, the size of this space is $|R|^{|S|}$.

5.1 Stateful PA automaton

In the *stateful PA automaton* the ant can be in a finite number of states, as in Turing machines; we shall restrict to the set of states {0, 1}.

Data structure

The ant operates on the cells of array A, which keeps permutation π as described above.

Situation - coded by 2 bits: b_1 and b_2 ⇒ 4 cases.
b_1 - This bit represents the current state of the ant - 0 or 1.
b_2 - Assuming the ant positioned at cell c with content x, b_2 detects whether $x \le c$ or $x > c$.

Reaction - coded by 4 bits: $b_1 ... b_4$ ⇒ 16 cases.
b_1 and b_2 - these 2 bits identify 4 possible reactions, numbered from 0 to 3:

- 0 - Swap contents of cell c and c-1 (fails if c-1 does not exist);
- 1 - Swap contents of cell c and c+1 (fails if c+1 does not exist);
- 2 - Create new cell at the right of cell c, with value max+1, where max is the current number of cells;
- 3 - Add new cell at the left of cell c, with value max+1.

b_3 - This bit defines the new state of the ant.
b_4 - This bit defines the ant's move:

- 0 - Move one step to the left (fails if cell c-1 does not exist);
- 1 - Move one step to the right (fails if cell c+1 does not exist).

For each of the 4 situations there is a choice among 16 reactions, thus there are $16^4 = 65\,536$ different automaton instances, that we number by the decimal representation of the 16 bits that characterize each of them (4 reaction bits per situation). We have simulated and inspected all of them, starting from initial configuration A_{init}= (1, 2) and the ant in state s_{init}= 0, positioned at cell 1. Note that when the reaction fails - the ant attempting to access cells beyond the array limits - the whole computation is aborted. Out of the 12278 automata that survive after 100 steps, we have selected two interesting cases.

Automaton 1925 has an irregular and quite remarkable behavior, documented in Figure 13.

The first diagram plots the positions occupied by the ant on the growing array A during a one-million-step computation. Quite remarkably, the ant keeps sweeping the full range of available cells from one extreme to the other, without ever entering a regular behavior, and yet without ever attempting to cross the boundaries. The final permutation is represented in the second plot, as the set of points $(i, \pi(i))$, i = 1, ..., 2811. The diagonal shadow indicates that the permutation is a sort of compromise between the identity function and a random scattering - a mix of order and randomness. These two diagrams illustrate what we mean when we claim that algorithmic, deterministic causet construction techniques can offer surprising emergent properties that cannot be expected from stochastic techniques.

The third plot of Figure 13 provides the ordering fraction spectrum for 500 intervals of the final causet - a DAG with 2811 nodes and 2,456,824 edges, that decrease to 18,531 after transitive reduction - and suggests a Myrheim-Meyer dimension slightly less than 2D. The lower diagram is the longest/shortest path plot for the transitively reduced causet, which appears quite satisfactory and comparable to the analogous plot for 2D sprinkled Minkowski intervals (Figure 7). Note that we cannot exclude apriori the existence of multiple sources (and sinks) in the causet obtained from a permutation as described. In this case, the longest/shortest path plot is build by considering the lengths of the longest and shortest link-paths from the root node 1 to all the nodes reachable from it, which may therefore be less than the total number of nodes.

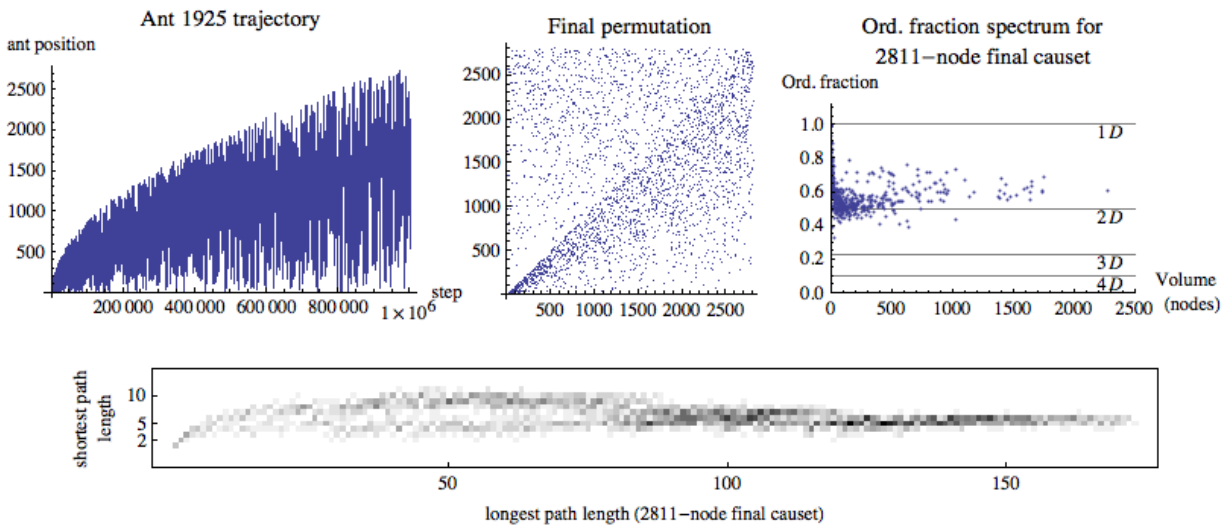

Figure 13 - Stateful PA automaton n. 1925. Ant trajectory, final permutation, ordering fraction spectrum, longest/shortest path length.

The stateful PA automaton 1929 has a definitely more regular behavior, as illustrated in Figure 14.

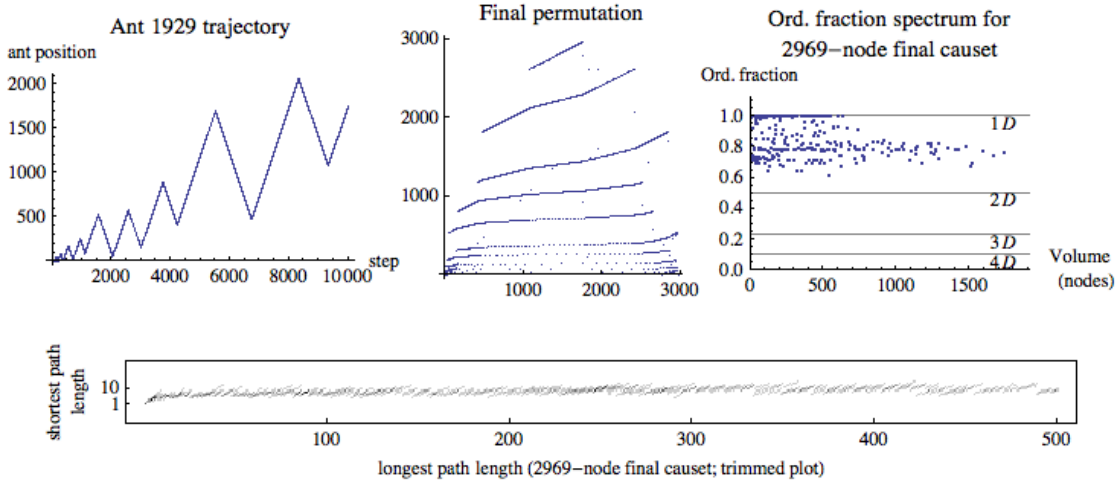

Figure 14 - Stateful PA automaton n. 1929. Ant trajectory, final permutation, ordering fraction spectrum, longest/shortest path length.

The ant trajectory is shown in the first plot for 10,000 steps (it proceeds similarly up to at least one million steps). The final permutation is represented in the second plot. The third plot shows the ordering fraction spectrum for 500 intervals of the final causet - a DAG with 2969 nodes and 2,566,500 edges, that decrease to 5933 after transitive reduction; this plot differs considerably from that of the previous automaton n. 1925, providing a vague indication of a low, non-integer Myrheim-Meyer dimension. Finally, at the bottom of Figure 14 we provide the longest/shortest path plot, which was trimmed to longest path length 500 for clarity (the full plot hits value 1100, with essentially no growth of the shortest path length). In spite of the relative regularity of the ant trajectory and final permutation, this longest/shortest path plot still looks quite satisfactory when compared to the analogous plot for 2D sprinkled Minkowski intervals.

5.2 Stateless PA automaton

The second type of algorithm that we consider is a *stateless PA automaton*. The control head, or ant, is now stateless, but this simplification is compensated by a slightly more complex array structure, and a type of reaction similat to a GOTO statement.

Data structure

Array cell $A(i)$ is now a pair $(bit(i), \pi(\mathrm{i}))$, where $bit(\mathrm{i})$ is a bit and $\pi(\mathrm{i})$ is the permutation element, as before.

Situation - coded by 2 bits: b_1 and $b_2 \Rightarrow$ 4 cases.

$b_1 = bit(c)$. Cell c is where the ant is currently positioned.
$b_2 = bit(c\text{-}1)$. (Fails if cell c-1 does not exist.)

Reaction - coded by 3 bits: b_1, b_2, b_3 ⇒ 8 cases.

$b_1 = 0 \Longrightarrow$ Write b_2 in cell c and swap cells c and c-1 (potential failure is detected when checking the situation, as just described).
$b_1 = 1 \Longrightarrow$ Create a new cell $(b_2, max+1)$, where max is the current number of cells, and insert it at position c.
b_3 - This bit defines the ant's move:
0 - Ant does not move
1 - Ant jumps to cell $\pi(c)$ (GOTO statement; c is the current ant position).

Reasoning as in the previous case, we now obtain $8^4 = 4096$ automaton instances. We have simulated and inspected all of them, starting from an initial two-cell configuration $A_{init} = ((0, 1), (0, 2))$, with the ant positioned at cell 2. As with the stateful ant, the computation aborts when the ant attempts to access locations beyond the array limits. Out of the 2876 automata that survive after 100 steps we have selected two interesting cases.

Stateless PA automaton n. 2560, many copies of which are indeed found in this family, is remarkable for its perfect random-like behavior, illustrated in Figure 15.

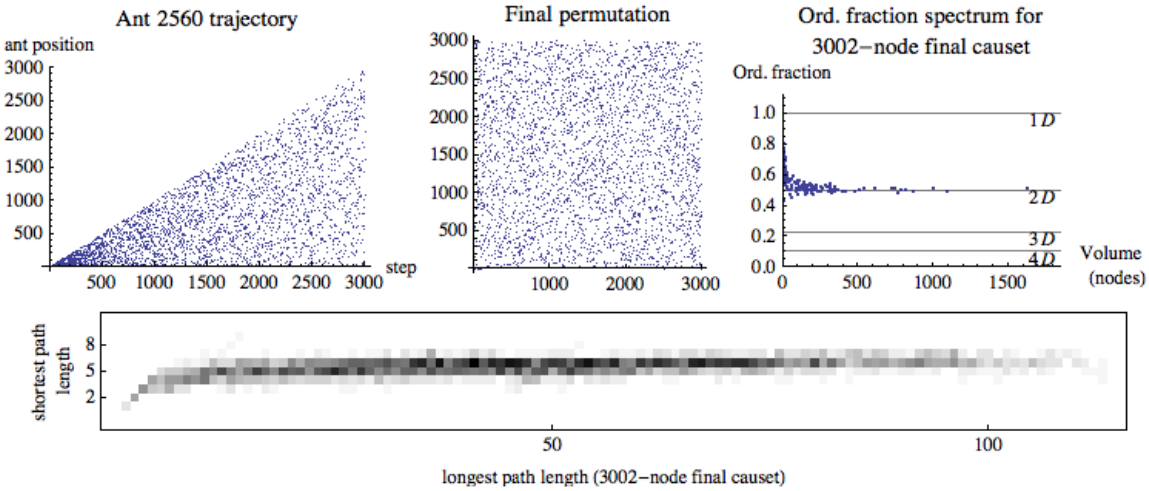

Figure 15 - Stateless PA automaton n. 2560. 3000-step computation. Ant trajectory, final permutation, ordering fraction spectrum, longest/shortest path length.

This automaton can be regarded as a 'perfect sprinkler', since its ordering fraction spectrum and its longest/shortest path plot are indistinguishable from those of sprinkled 2D Minkowski causets. An interesting feature of these causets is that they only have one source node (root node 1), so that all nodes in the graph contribute to the longest/shortest path plot.

Figure 16 shows the growth of the longest path length, starting from the root, as a function of the number k of nodes for the causets built by automaton 2560; the growth rate $k^{0.521}$ is remarkably close to that found for the sprinkled Minkowski case shown in Figure 8 ($k^{0.526}$, where k represents the number of nodes sprinkled in the unit box).

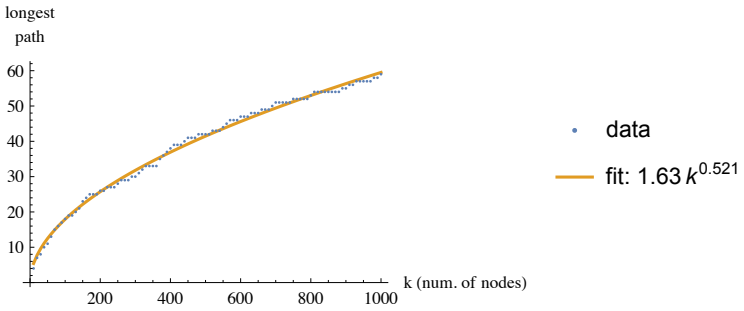

Figure 16 - Longest path length of transitively reduced causet from stateless PA automaton 2560 as a function of the number of nodes k. The function has an $O(n^{0.521})$ growth. Compare with Figure 8.

Note that the final causet, built in 3000 steps, has 3002 nodes. This means that, starting from the two-cell initial array, the ant reaction is always of one type: create a new cell at each step. A closer look at the behavior of the computation reveals that several options in the algorithm are never used by this specific instance n. 2560. In particular, even the bits that decorate array cells can be eliminated! We can then provide a very concise algorithm that performs exactly the same computation. Here is the tiny *Mathematica* code for the simplified ant step:

```
step[{array_, pos_}] := {Insert[array, Length[array] + 1, pos], array[[pos]]}
```

By iterating the function call 3000 times (`Nest[step, {{1, 2}, 2}, 3000]`), with initial array (1, 2) and initial ant position 2, one obtains exactly the same results of Figure 15. We believe that this concise randomization algorithm might be of interest, independent of the application to causets.

Going back to the general, stateless PA automaton, we present in Figure 17 just the ant trajectory for a second instance of the model, namely n. 3593.

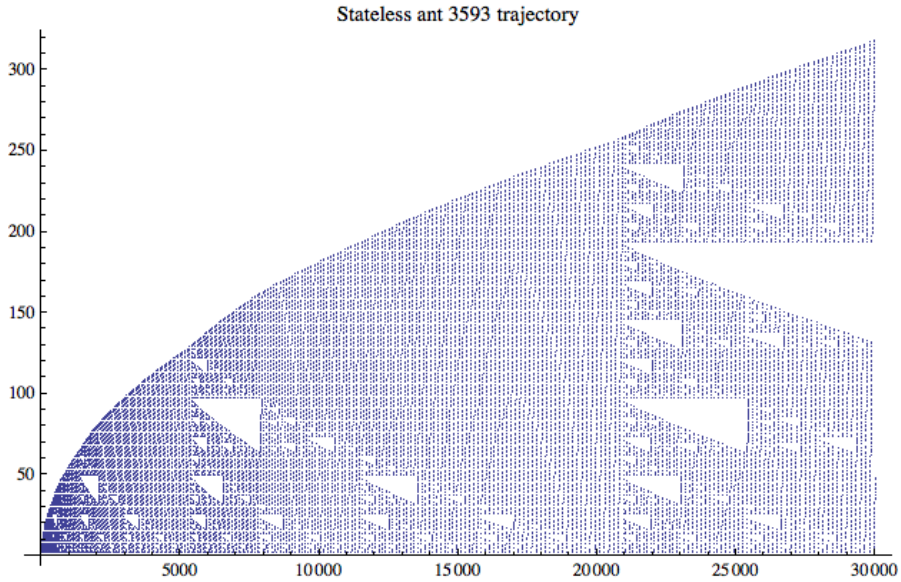

Figure 17 - Ant trajectory for stateless PA automaton n. 3593

In this case the causet obtained by the permutation is essentially one-dimensional, and totally uninteresting in terms of 'Lorentzianity' and non-locality. The reason for presenting this conclusive plot is to show yet another instance of the typical, triangle-based nested patterns that emerge in cellular automata and other models, as widely illustrated in [23]: it is precisely this richness of emergent phenomena, from regular to pseudo-random patterns, that makes the study of algorithmic causets so attractive.

6. Conclusions

In this paper we have introduced some statistical indicators for the analysis and comparison of various classes of causal sets, with the main objective to capture the 'Lorentzianity' of the latter, intended as the manifestation, in the discrete setting, of the inverse triangular inequality and the 'non-local' nature of Minkowski space. Our experimental work, based on extensive computer simulations, has led us to a number of conclusions.

First, we have established a clear distinction between indicators based on transitively closed (Section 3) and transitively reduced (Section 4) causets. In spite of their usefulness for Myrheim-Meyer dimension estimation, indicators of the first type cannot characterize 'Lorentzianity', failing to discriminate between cases as different as the highly non-local sprinkled causets, and highly local directed graphs such as a regular grid. Furthermore we have shown that a power-law distribution of node degrees is not a rare property of transitively closed causets, and does not single out sprinkled de Sitter causets [17] as a special case of discrete spacetime, being present, for example, in percolation causets and in what we called sprinkled Minkowski cylinder (Figure 1). Note that, whenever the estimation of causet node degrees is obtained by measuring lightcone areas in the embedding manifold, as done in [17], the analysis is inevitably bound to address only the transitively closed version of the causet, with the mentioned limitations.

Considering transitively reduced causets proves more useful, although, unfortunately, transitive reduction is computationally costly [2], being O(|V|*|El) for a directed acyclic graph G(V, E) - this has been the main computational bottleneck of our investigation. We have proved, analytically and experimentally, that the node degrees of a transitively reduced sprinkled Minkowski interval causet exhibit $O(log(k))$ growth in 2D and $O(k^{1/3})$ growth in 3D, where k is the number of sprinkled nodes, confirming what concisely stated, in slightly different terms and without proof, in [13].

Still in the context of transitively reduced causets, we have introduced longest/shortest path plots. These simple but remarkably effective visual indicators are designed to directly reflect the interplay between longest and shortest link-paths, whose length differences grow very large in causets derived from sprinkling in Lorentzian manifolds. We have shown that the very slow growth that these plots exhibit for sprinkled Minkowski intervals (of which we provide partial analytical account) and for de Sitter intervals, is not observed in the other considered causet classes.

We have then introduced two new classes of *deterministic*, algorithmic causets, built by what we called stateless and stateful PA (Permutation Ant) automata, and have identified a 'perfect sprinkler' - a deterministic 'ant' able to build, *without* resorting to an underlying Lorentzian manifold, a causet that appears, under the lens of the introduced indicators, basically indistinguishable from a sprinkled causet. For this 'ant' we have also provided an alternative, extremely concise implementation - one line of code.

We believe that concentrating exclusively on stochastic causets, and over-emphasizing the requirement of *direct* embeddability in a Lorentzian manifold, prevents us from discovering causets with rich internal structure and a mix or regularity and pseudo-randomness as observed in algorithmic causets; these causets might still turn out to be embeddable in asymptotic sense, e.g. after coarse-graining.

In this respect, more work is needed for exploring the space of algorithmic causets and their properites, in search for cases where an optimal performance in terms of longest/shortest path plots, i.e. non-locality, combines with rich internal structure. It would also be interesting, moving above the ground level of spacetime, to look for other cases in nature (e.g. in the biosphere) where the feature of 'non-locality', as revealed by our plots, plays some role, and to study how it is implemented. For example, the evolution of plant leaf venation has led to patterns that optimize hydraulics and tolerance to cuts, and involve long paths for nutrient transportation between points at a short distance from each other.

References

1 M. Ahmed, D. Rideout, 'Indications of de Sitter spacetime from classical sequential growth dynamics of causal sets', Physical Review D **81**, 083528 (2010).

2 A. V. Aho, M. R. Garey, J. D. Ullman, 'The Transitive reduction of a Directed Graph', *SIAM Journal on Computing* 1 (2), pp. 131-137, 1972, doi:10.1137/0201008

3 B. Bollobás , G. Brightwell, 'Box-Spaces and Random Partial Orders', *Trans. Amer. Math. Soc.*, 324(1):59-72, March 1991.

4 B. Bollobás , G. Brightwell, 'The Structure of Random Graph Orders', *SIAM J. Discrete Math.*, Vol. 10, No. 2, pp. 318-335 (1997).

5 T. Bolognesi, 'Planar Trinet Dynamics with Two Rewrite Rules'. In: *Complex Systems*, vol. 18 (1) pp. 1 – 41. Complex Systems Publications, Inc, 2008.

6 T. Bolognesi, 'Causal sets from simple models of computation'. In: *International Journal of Unconventional Computing*, vol. 6 (6) pp. 489 – 524. OCP Science, 2010. (See also arxiv.org/abs/1004.3128.)

7 T. Bolognesi, 'Algorithmic causets'. In: *Journal of Physics: Conference Series*, vol. 306 (1) article n. 012042. Special issue: 5th International Workshop DICE 2010 Space-Time-Matter – Current Issues in Quantum Mechanics and Beyond. Lajos Diósi, Hans-Thomas Elze, Leone Fronzoni, Jonathan Halliwell, Enrico Prati, Giuseppe Vitiello, James Yearsley (eds.). IOP Science, 2011.

8 T. Bolognesi, 'Algorithmic Causal Sets for a Computational Spacetime'. In: *A Computable Universe - Understanding and Exploring Nature as Computation*, H. Zenil (editor), World Scientific, 2013.

9 T. Bolognesi, 'Do Particles Evolve?'. In: *Irreducibility and Computational Equivalence - 10 Years After Wolfram's A New Kind of Science*, H. Zenil (editor), Springer, 2013.

10 T. Bolognesi, 'Inflation in 2D de Sitter Spacetime', http://demonstrations.wolfram.com/InflationIn2DDeSitterSpacetime/Wolfram Demonstrations Project, Published: February 28, 2013.

11 T. Bolognesi, M. Tarini, 'Shortest paths in sprinkled Minkowski causal sets', draft.

12 Luca Bombelli, Johan Lee, David Meyer, and Rafael D. Sorkin. 'Space-time as a causal set.', *Phys. Rev. Lett.*, 59(5):521–524, Aug. 1987.

13 Luca Bombelli, Johan Lee, David Meyer, and Rafael D. Sorkin, Reply to Comment on "Space-Time as a Causal Set", *Phys. Rev. Lett.*, 60(7):656, Feb. 1988.

14 G. Brightwell, R. Gregory, 'The Structure of Random Discrete Spacetime', *Phys. Rev. Lett.*, 66(3):260–63, Jan.. 1991.

15 A. Eichhorn, S. Mizera, 'Spectral dimension in causal set quantum gravity, arXiv:1311.2530v1 [gr-qc] Nov. 11, 2013.

16 Joe Henson, Constructing an interval of Minkowski space from a causal set, arxiv:gr-qc/0601069 v1, 17 Jan 2006. http://arxiv.org/abs/gr-qc/0601069

17 D. Krioukov, M. Kitsak, R. S. Sinkovits, D. Rideout, D. Meyer, M. Boguña, 'Network Cosmology', arXiv:1203.2109v2 [gr-qc], Nov. 2012

18 F. Markopoulou, The Computing Spacetime, arXiv:1201.3398v1 [gr-qc], 2012.

19 Jan Myrheim, Statistical Geometry CERN preprint Ref.TH.2538-CERN, 1978.

20 D. Rideout, Homepage, http://www.phy.syr.edu/~rideout/ (valid in oct. 2013).

21 David P. Rideout and Rafael D. Sorkin. Classical sequential growth dynamics for causal sets. *Phys. Rev. D*, 61:024002, Dec 1999. arXiv:gr-qc/9904062v3.

22 Rafael Sorkin, 'Causal Sets: Discrete Gravity', Notes for the Valdivia Summer School, Jan. 2002.

23 S. Wolfram, *A New Kind of Science*, Wolfram Media, Inc., 2002.