Simple indicators for Lorentzian causets

Tommaso Bolognesi - CNR/ISTI - Pisa

Alexander Lamb - Princeton University

June 17, 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 ($x \not\prec x$) 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 [10, 16]).

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, by 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 [11] 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 [13]:

"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 shall focus on edge counts for the *transitive reduction* of the considered graphs. This appears to be the correct setting for addressing 'Lorentzianity' issues. Beside analyzing node degree distributions, 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 apply the introduced longest/shortest path plots to some example causets obtained by the deterministic (as opposed to stochastic) causet construction techniques that one of the authors has explored in recent years [5, 6, 7, 8], in which some simple 'strategies' for obtaining the non-locality feature are at work and emerge quite visibly. In Section 6 we introduce a new type of 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 7 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: $E = \{(p, q) \in S^2 \mid L^2(p,\ q) \geq 0 \wedge t_p < t_q\}$. 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 exactly 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., [12]) 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 [9].) 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 $\delta * \text{Area(S)} = \delta * 2\pi\text{Cosh}(\tau)d\tau$. 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:

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

The Minkowski coordinate t of these points is itself a random variable 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 t is constant:

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

('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 $t_{\max} = \text{Sinh}(\tau_{\text{Max}})$,
- $r = \sqrt{t^2 + 1}$
- 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 $p << 1$). Note that a percolation causet may be disconnected, and may include several source nodes. An 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 [12], 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, and by deriving edges as done with sprinkled causets in Lorentzian manifolds, while using the Euclidean distance in place of the Lorentz distance.

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, thus 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 $\{1, ..., 55\}$). Note that we shall always refer to the *out*-degrees.

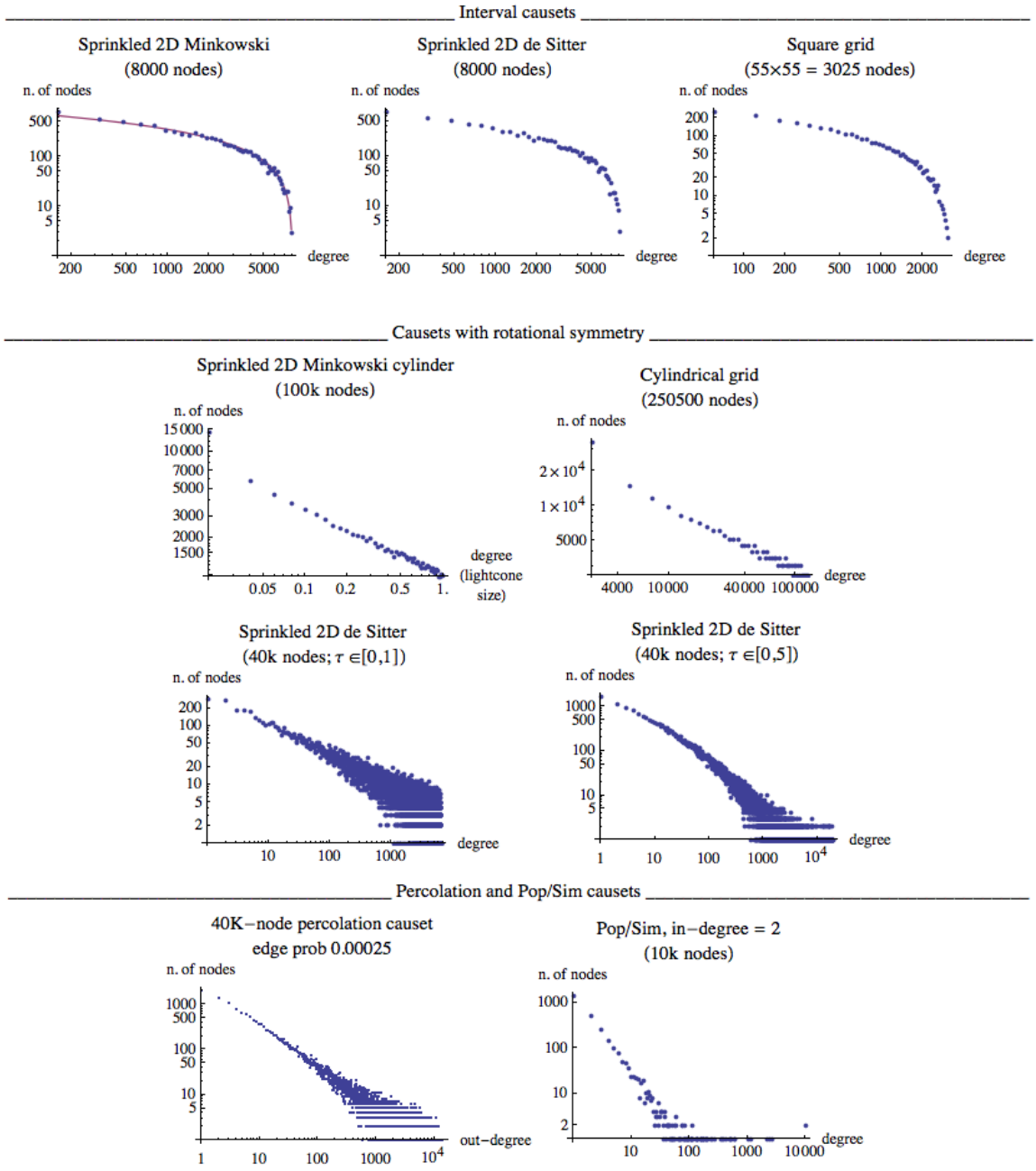

Figure 1 - Node degree distributions for transitively closed causets

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.) $r=x*y$, where r.v.'s x and y have uniform densities in the unit interval: $f_x(z) = 1$, $f_y(z) = 1$, $z \in [0,1]$. The reason is that, by a -45 degree rotation, we can represent the sprinkled Minkowski interval as a unit square [0, 1]◊[0, 1] in Euclidean 2D space, and model the degree of the generic sprinkled point $p(1\text{-}x, 1\text{-}y)$ as the area $x*y$ of rectangle $R[(0, 0), (x, y)]$ (identified by its lower-left and upper-right vertices). The distribution function for r is: $F_r(z) \doteq \text{Prob}[x*y \le z] = z(1\text{-Log}(z))$, yielding, by derivation, the density function $f_r(z) = \text{-Log}(z)$. 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 ("Minkowski cylinder"). The out-degree of a node is computed as the area of the future lightcone of that node. The distance h of the generic node p from the upper border of the cylinder is a r.v. with uniform density: $f_h(x) = 1$, 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. $A = h^2$. By applying well known results on functions of r.v.'s, we obtain $f_A(y) = \frac{1}{2} y^{-1/2}$, $y \in [0, 1]$ for the density of r.v. 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 $\tau_{max} = 1$ (central 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 [12]; 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.

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 [12] 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 [12] 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, grid, pop/sim.

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 [14, 17]:

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

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.

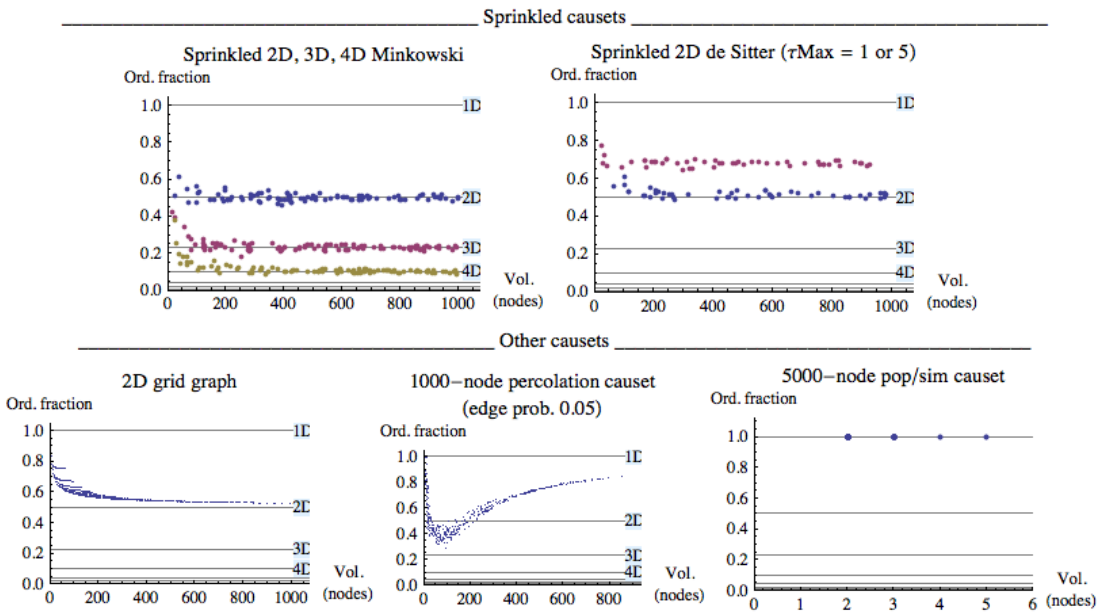

Figure 2 - Ordering fraction spectra for causet classes

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 similar diagrams - we call them *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:

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

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 [12], 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 [15]:

> *"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; we can only do it via the notion of embedding, i.e. by asking whether or not C could be faithfully embedded in a Lorentzian manifold. Indeed, our 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. Note also that the quote above refers, implicitly, to causet *links*. 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 valencies (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. Unless otherwise stated, in this section we shall always refer to links, for example when considering node degrees.

The plot in Figure 3-right, illustrates the *logarithmic* growth of the degree of a generic node from a sprinkled, transitively reduced 2D Minkowski causet, as the density of the sprinkling increases. The smoothness of the function is obtained by averaging over 100 samples. That we should expect a logarithmic growth can be intuitively understood by the following rough analysis (see Figure 3-left). For convenience we rotate lightcones by -45 degrees so that intervals become rectangles. Let A_1 be the set of d points that we sprinkle in the unit square R_1 (an interval) with source $s(0, 0)$ and sink $t(1, 1)$. Let us estimate the number of links outgoing from s. If $p_1(x_1, y_1)$ is the point of A_1 with smallest x-coordinate, we may statistically expect x_1 to be around $1/d$ and y_1 to be around $1/2$. Edge (s, p_1) must necessarily be a link because the rectangle of which it is a diagonal is empty by construction. The other links can be found by iterating the procedure, starting now from rectangle $R_2((1/d, 0), (1, 1/2))$, identified by its diagonal endpoints, the set A_2 of points in it, and point $p_2(x_2, y_2)$ of A_2 with smallest x-coordinate, which identifies the next link (s, p_2). In light of the progressive halving of the rectangle heights y_i's (which is the element of 'roughness' in the argument) and the uniform density d of the sprinkling, we may expect interdistance $x_{i+1} - x_i$ to double up at each step, and to be $2^i/d$. Thus, for the last, rightmost link found, (s, p_k), we can write $(2^k - 1)/d \approx 1$, yielding $k \approx \mathrm{Log}_2(d)$. (Note that the fitting function in Figure 3 is the *natural* logarithm.)

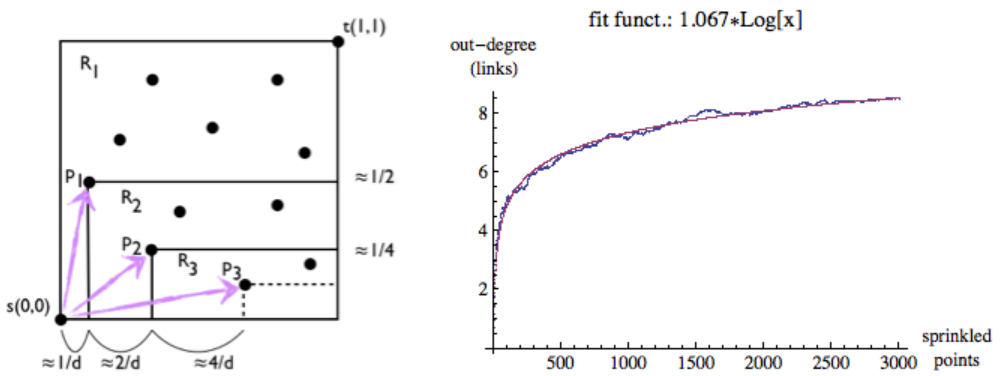

Figure 3 - Logarithmic growth of the degree of node (0, 0) of a causet obtained by sprinkling d points in the 2D-Minkowski unit interval with source (0, 0), as a function of d

In conclusion, it seems natural to propose that a necessary (certainly not sufficient) 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 node degrees that grow logarithmically with the number of nodes.

The case of a regular grid is trivial: node degrees do not grow with the number of nodes. What about percolation causets? In this case, and at the largest scale, the node degree growth turns out to be bounded. This is related to the phenomenon of 'posts', which was first studied in [3]. A 'post' is a node x that creates a bipartition of the *whole* set of causets nodes into the future and the past lightcone of x. In other words, *all* causet nodes are related to x. Reference [3] proves that posts almost surely keep appearing undefinitely, yielding a suggestive '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.

Let us now focus on two other types of statistical indicator, related to counting links: *node degree distributions*, and an ad-hoc indicator which combines information on longest and shortest paths between nodes, that we conveniently call *longest/shortest path plot*.

4.1 Node degree distribution for transitively reduced causets

Having checked that a given causet construction technique guarantees the unbounded growth of node degrees, we may want to further assess it by analyzing the actual distribution of node degrees that it yields. Figure 4 shows, on the left, the node degree density for a 10k-node transitively reduced causet obtained by sprinkling in a 2D Minkowski interval, and, on the right, a similar plot for a 5k-node causet obtained by sprinkling in a de Sitter interval spanning the time range (0, 5). In the first case the density values (dots) are matched by a Poisson distribution (solid line) $\frac{H\lambda^k}{e^\lambda k!}$, where H =10k and λ = 7.831; in the second case the same matching yields parameters H = 5k and λ = 6.945, and it is clear that the approximation is less satisfactory. This is an effect of the curvature of de Sitter spacetime, which gets worse as we increase the time span.

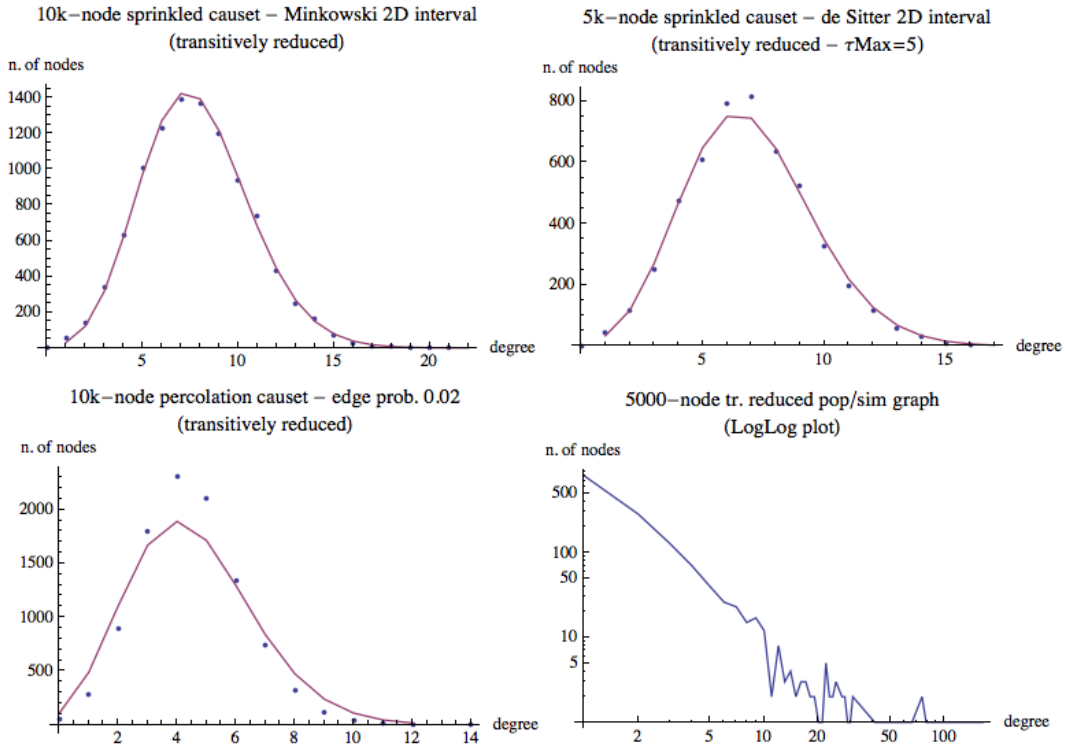

Figure 4 - Node degree densities for transitively reduced causets. Upper-left: 10k-node causet built by sprinkling in a Minkowski 2D interval, and Poisson density fitting. Upper-right: same plot and fitting for 5k-node sprinkled causet in a de Sitter interval with time span (0, 5). Lower-left: 10k-node percolation. Lower-right: 5k pop/sim causet (LogLog plot of node degree density)

The regular directed square grid that we have considered earlier is not accounted for in this figure: transitive reduction leaves the graph unaffected, and node degree distribution is trivially constant, since all nodes have as many outgoing edges as the dimensionality of the grid.

In the lower-left plot of Figure 4 we show the node degree density for a 10k-node percolation causet with edge probability 0.02; the best fit by a Poisson density $\frac{H\lambda^k}{e^\lambda k!}$, where H =10k and λ = 4.530, is even less satisfactory than for the de Sitter case. Finally, the lower-right diagram is a LogLog plot of the node degree density for a transitively reduced 5k-node popularity/similarity causet, which radically departs from the previous cases and appears to follow a power law.

Unfortunately, also when moving from 2D to 3D sprinkled Minkowski intervals, node degree distributions (not shown here) depart from

Poisson. Thus, looking at the node degree distribution for a (transitively reduced) causet *not* obtained by sprinkling, and for which we may even lack a dimensionality estimation or curvature information, does not appear as a practically useful tool for assessing Lorentzianity.

4.2 Longest/shortest path plot

In this subsection we introduce a statistical indicator meant to reflect Lorentzian non-locality more explicitly than those of the previous subsection. 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 two segments of light rays forming a $\pi/2$ angle and connecting p_1 and p_2 via one intermediate point p_x.

When a causet C is derived by sprinkling in Minkowski space M, we can approximate the Lorentz distance between two points *p1* and *p2* in timelike relation in M by the length (number of edges) of the *longest* chain P between them in C [14], which therefore 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 $C_{\text{red}}[s, t]$ with source s and sink t, and is built as follows. For each event x in $C_{\text{red}}[s, t]$ 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 5 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 existence 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'.

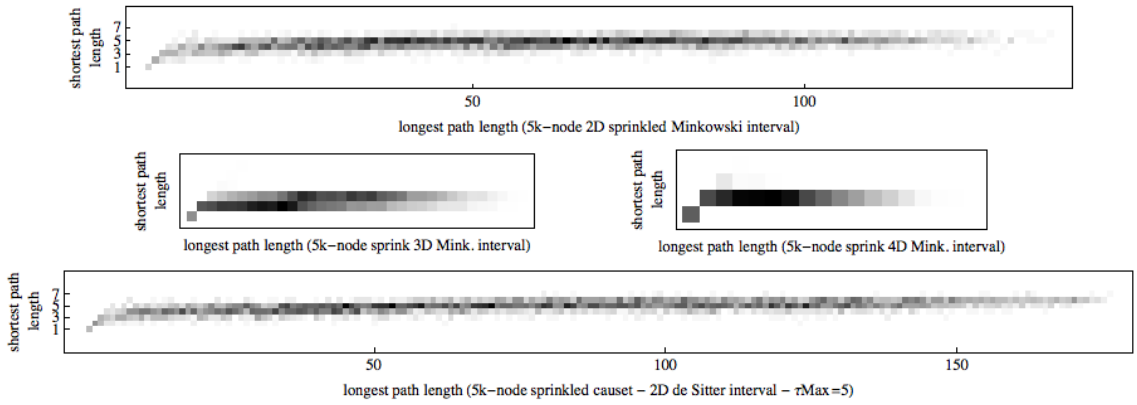

Figure 5 - Longest/shortest path plots for 5k-node causets obtained by sprinkling in 2D (upper), 3D, 4D (central row) Minkowski intervals; similar plot for 5k-node sprinkled causet from 2D de Sitter interval, with time span (0, 5).

Of course it would be interesting to obtain an analytical account for the slow growth of the above longest/shortest path plots. For doing this, we express analytically functions *lpl*(d) and *spl*(d), defining the *average* 'longest' and 'shortest path length' from source to sink of a sprinkled 2D Minkowski interval as a function of the sprinkling density d, and then eliminate d, finding the direct dependence *sdp*(*lpd*). Function *lpl*(d) is O ($d^{1/2}$), as shown by the fit in Figure 6.

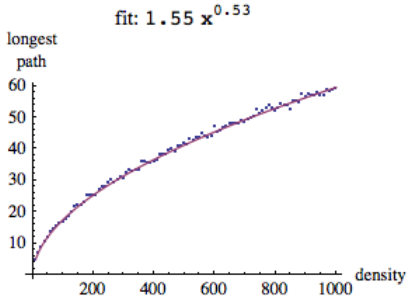

Figure 6 - The longest path length of a transitively reduced sprinkled Minkowski 2D causet has an $O\left(d^{1/2}\right)$ growth.

Our rough justification for this growth function is as follows. We rotate the sprinkled interval $I(s, t)$ by -45 degrees, as done in Figure 3-left, so that links - the edges of the transitively reduced DAG - become diagonals of empty rectangles, pointing up and to the right. Consider the *links* outgoing from the root (or any other node), such as (s, p_1), (s, p_2), (s, p_3) in Figure 3-left. The endpoints p_1, p_2, p_3 approximately lie on a curve $xy = k$, for some constant k, since links approximate units of Lorentz distance, and the locus of points at equal Lorentz distance from a given point in Minkowski space is an hyperbola. As done before, we may grossly assign to point p_1, the leftmost point at the right of the root, the coordinates $(1/d, 1/2)$, which allows us to estimate $k = (2\,d)^{-1}$. A link-path from s to t then corresponds to a sequence of empty rectangles chained by their vertices. For maximizing the number of elements in the path we need to minimize, for each rectangle, the maximum between its horizontal and vertical dimensions, which means that the rectangles tend to be squares of side $z = k^{1/2} = (2\,d)^{-1/2}$. For connecting s to t across the *unit* box, we need $(2\,d)^{1/2}$ such 'squares', which is a fair approximation of the fitting function in Figure 6.

Unfortunately the analysis of $spl(d)$ is harder. We have experimental evidence that the function exhibit a less than logarithmic growth, and a natural candidate seems to be function *iterated logarithm*(d), or *Log** ('log star'), which counts the number of times one has to take the (base 2) logarithm, starting from the argument d, for getting below 1 (see, for example, http://www.gabrielnivasch.org/fun/inverse-ackermann). Log*(x) assumes value 4 for x in the interval $(16, 2^{16}]$, and value 5 in $(2^{16}, 2^{(2^{16})}]$, hence a convincing experimental verification of this conjecture for values beyond 5 appears out of reach. In any case, by eliminating d from $lpl(d) = (2\,d)^{1/2}$ and $spl(d) = \text{Log*}(d)$, we obtain

$$spl\,(lpl) \;=\; \text{Log}^{*}\left(lpl^{\,2}/2\right) \tag{4}$$

This function has the same growth rate of Log*(lpl), and on the large domain it is almost always greater by one unit than Log*(lpl) (if not, they are equal). However, in the range of longest path lengths 100+, function (4) yields value 4, which is consistent with the values in the upper plot of Figure 5.

It was not possible to effectively discriminate between a regular grid and a sprinkled causet by looking at transitive closures, or by 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 coincide, is shown in Figure 7-left.

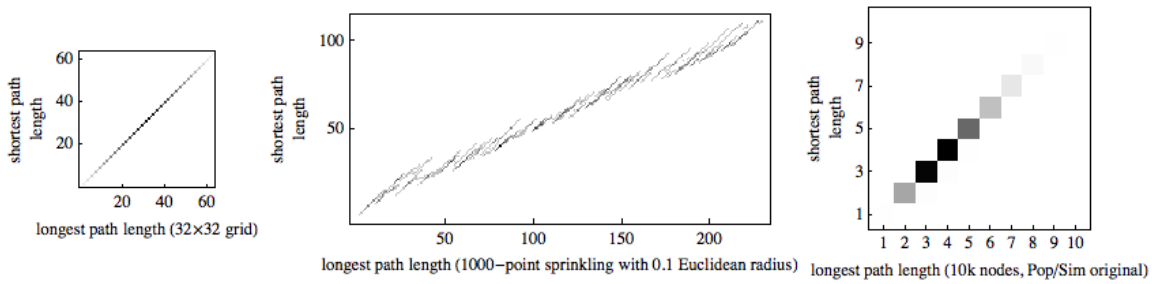

Figure 7 - Longest/shortest path plots for three causets. Left: 32◊32-node grid. Center: graph obtained by sprinkling 1000 points in a unit square, where edges connect points at Euclidean distance smaller than 0.1. Right: 10k-node pop/sim causet.

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 can also consider 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 7-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 7-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 [12], our analysis reveals, again, a remarkable difference between the two causet types, at the finite scale, with pop/sim causets unable to develop very short paths as alternatives to very long paths to the same node. Note that the plot in Figure 7-right refers to a 10k-node pop/sim causet, while that in Figure 5-bottom refers to a sprinkled de Sitter causet with half the number of nodes. No clue of non-locality seems to emerge in pop/sim causets.

On the contrary, the longest/shortest path plots for graphs from percolation dynamics (Figure 8) bears some similarity with those for

sprinkled causets, in the sense that high ratios between longest and shortest path lengths can be achieved in both cases, although the diagrams for percolation manifest a more pronounced tendency to grow.

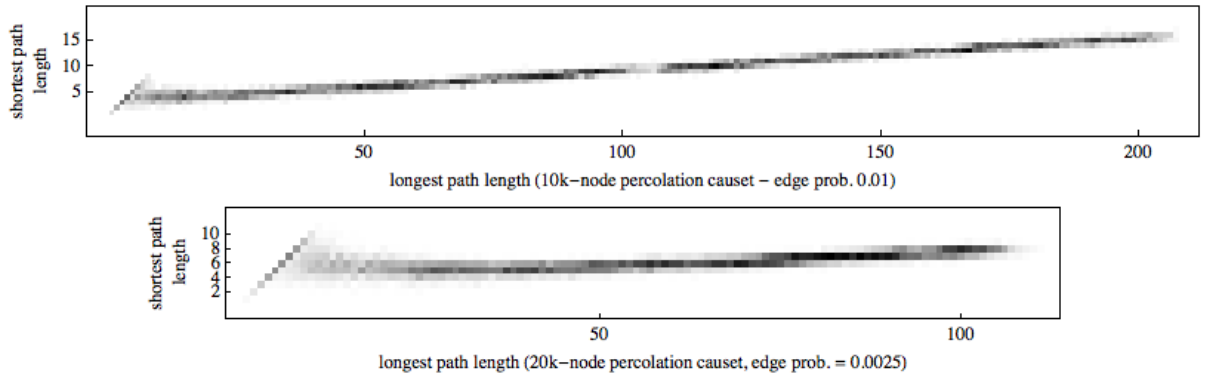

Figure 8 - Longest/shortest path plots for two causets from percolation dynamics. Upper: 10k-node causet with edge probability 0.01. Lower: 20k-causet with edge probability 0.0025.

5. Longest/shortest path plots for deterministic causets

In our recent work [5, 6, 7] we have investigated *algorithmic causets*, intended as DAGs built by deterministic procedures, following the conjecture that Nature and spacetime have a fundamentally algorithmic dynamics, and the observation that algorithmic behaviours can indeed exhibit a much richer variety of emergent properties than stochastic ones, as widely illustrated by Wolfram with cellular automata and other models of computation [XXX]. Having introduced the longest/shortest path plot as a tool for investigating DAG 'Lorentzianity', or non-locality, we are naturally led to put our algorithmic causets under this new lens, and to verify the extent to which they can emulate sprinkled, Lorenzian graphs. We are also interested in revealing the 'tricks' by which some algorithmic causets might succeed in yielding relatively flat longest/shortest path plots.

Out of the several algorithmic models, and thousands of computations that we have examined, we have selected here a few interesting cases.

In *Trinet mobile automata* [4], a variant of network mobile automata [XXX], a control unit moves on an *undirected*, planar, trivalent graph, that we call 'trinet', while modifying and growing it at each step by simple graph rewrite rules. A rather direct way to turn these algorithmic, undirected trinets into DAGs, exploiting the fact that nodes get progressively numbered, consists in orienting each newly introduced edge so that it always points from older to more recent node. Two examples of transitively reduced, *directed* trinets obtained in this way, and their respective longest/shortest path plots, are shown in first two rows of Figure 9.

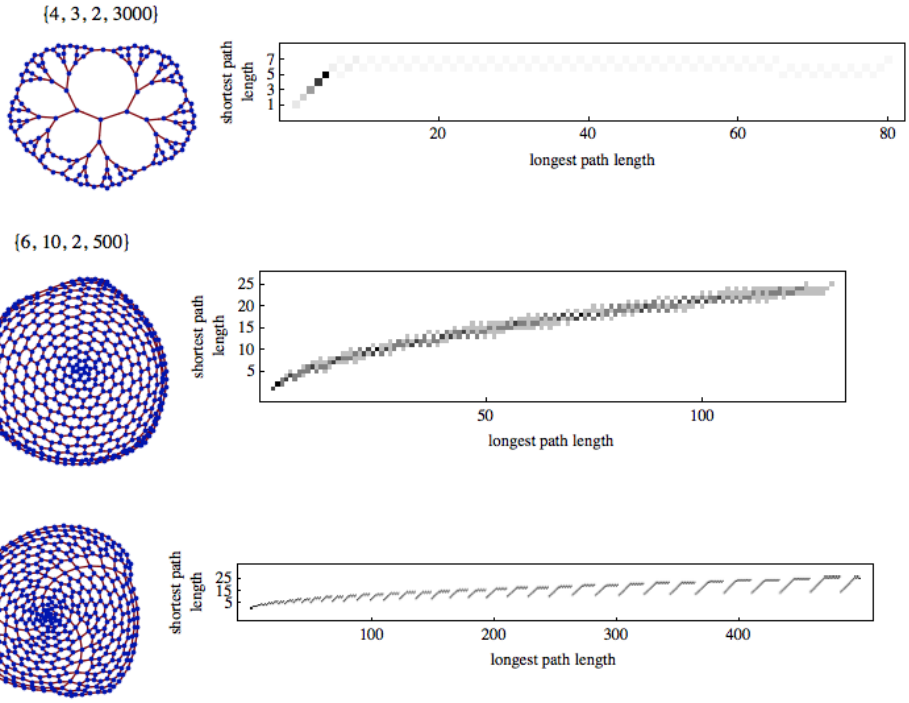

Figure 9 - Trivalent, planar, transitively reduced DAGs obtained from 3000-step (upper row) and 500-step (central row) trinet mobile automaton computations with parameters as indicated in the numeric labels (left), and corresponding longest/shortest path plots (right). Lower row: the 'raw', totally ordered causet obtained by the technique described in [6] from the mobile automaton computation producing the trinet just above, and corresponding longest/shortest path plot.

The directed acyclic trinet in the upper row of Figure 9 grows by a spiralling process, with nodes and edges progressively added at the border. Arc directions are not shown: they point outwards, and are oriented counter-clockwise on the border. During the growth, the graph keeps assuming highly symmetric shapes, like the one shown, interspersed among less regular patterns. The symmetric case is essentially a binary tree with a ternary root *r* in the middle, and leaves interconnected by a cyclic path. When *x* is internal, longest and shortest paths from the root to *x* coincide, as reflected by the staircase-like initial segment of the longest/shortest path plot. When *x* is on the border, the shortest path reaches the node directly from *r*, and has length equal to the tree depth *d* or *d+1*, while the longest possible path traverses the whole cycle of leaves before reaching *x*, and has therefore length $d + 3*2^d - 1$; this is reflected by the remaining, essentially flat portion of the plot.

The example in the central row in Figure 9 is built by a (directed) trinet mobile automaton using different parameters. The DAG is again planar, mostly with hexagonal faces, and three pentagonal faces at the center, where the root is found. In this case the shortest path to generic node *x* is essentially radial, while the longest path follows a spiral - there are indeed three concentric spirals. Roughly, the size of the spiral grows like the square of the radius, which explains why the *spl* growth appears to be $O(\sqrt{lpl})$, where *spl* and *lpl* denote the shortest and the longest path lengths.

The example in the lower row of Figure 9 is obtained by a different approach. The causet on the left is built by applying the technique described in [5] to the computation that produces the trinet just above it; with this technique, each causet node corresponds to a computation step of the trinet mobile automaton, i.e. to one particular stage of trinet growth. One peculiarity of this approach is that, when applied to trinet mobile automata, it yields *totally ordered* DAGs (for a discussion see [6]). Their transitive reduction would be just a sequence of edges, and the longest and shortest paths to any node would then coincide, yielding uninteresting plots identical to that of the regular square grid. Thus, in this case we have kept the *raw* causet, without transitive reduction; it consists of a hexagonal grid, in which the presence of three radial 'shortcuts' reduces the lengths of the shortest paths, inducing a slower growth of the plot.

The reader may correctly object that the last example contradicts our stipulation that longest/shortest path plots be computed only for transitively reduced causets, in order to preserve their value as indicators of non-locality/Lorentzianity. But we can easily obtain non trivial, transitively reduced DAGs from totally ordered graphs as follows. Consider again the totally ordered causet at the bottom of Figure 9 - call it G. Its edges can be partitioned into *links* (forming a linear path touching all nodes) and *non-links* (the remaining edges, each connecting two nodes that are not directly connected by one link). By splitting each directed non-link $h \to k$ into two directed edges $h \to n$ and $n \to k$, where *n* is a new node, we obtain a new graph GG quite similar to G, but not totally ordered. All edges of GG turn out to be links, e.g. the transitive reduction GG is identical to GG itself, as the reader can easily verify. The longest/shortest path plot for GG is essentially equivalent to that of G, since the structure of the two graphs is basically the same.

In Figure 10 we present, without much detail, three totally ordered causets still obtained by the technique described in [5], and resulting from some experiments with genetic algorithms on trinet mobile automata [8]. As in the last example, the structure of these DAGs is such as to achieve fairly high ratios between longest and shortest path lengths, yielding quite interesting longest/shortest path plots.

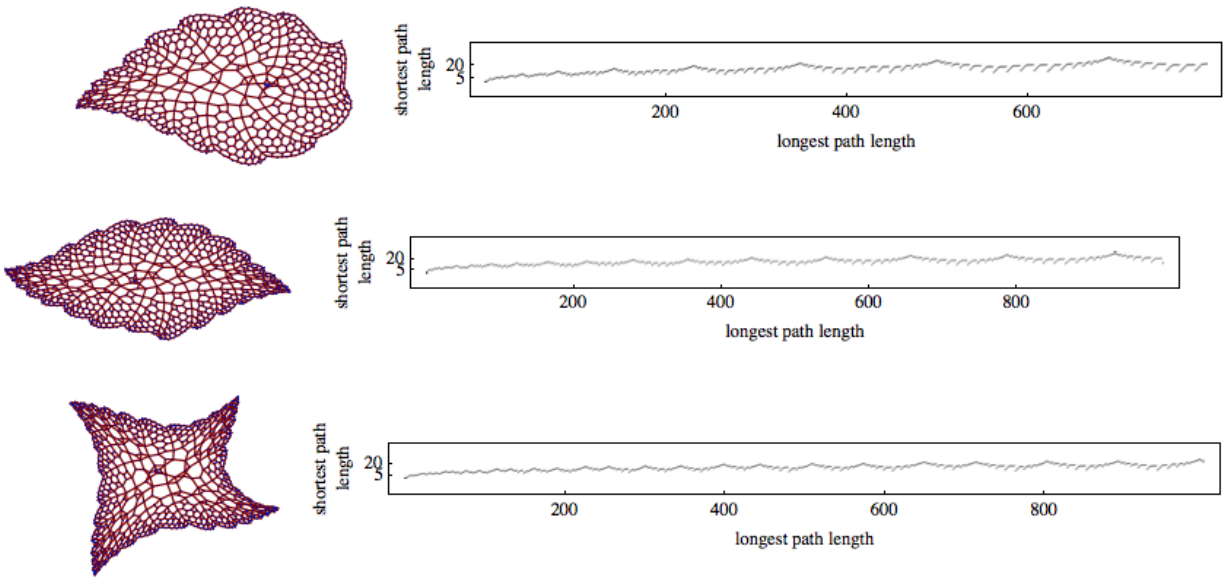

Figure 10 - Totally ordered causets obtained from the application of genetic algorithms to trinet mobile automata, and their longest/shortest path plots.

In Figure 11 we present a totally ordered causet obtained, by the same technique, from the computation of a Turing machine with 3 states and a tape alphabet with 3 symbols (machine n. 61786677050, according to the Wolfram numbering scheme implemented by *Mathematica* function 'TuringMachine[] ').

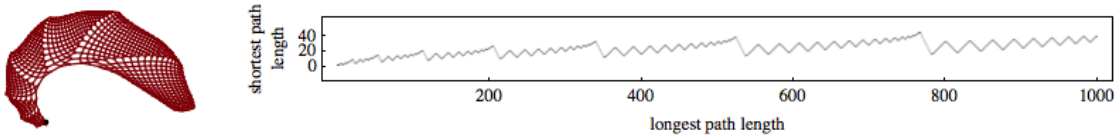

Figure 11 - Totally ordered causet obtained from a computation of a Turing machine, and associated longest/shortest path plot.

By using the above described idea of splitting non-links into two edges, the totally ordered DAGs of the last two figures can be readily turned into non totally ordered, transitively reduced causets whose longest/shortest path plots are similar to those in the two figures.

Some causets derived from computations of *binary string rewrite systems* are illustrated in [5]. Unlike those from computations of trinet mobile automata and Turing machines, these causets are, in general, *not* totally ordered. The transitive reduction of one of them is shown in Figure 12, and for it we have computed the longest/shortest path plot. Again, high values of the ratio between longest and shortest path lengths are obtained.

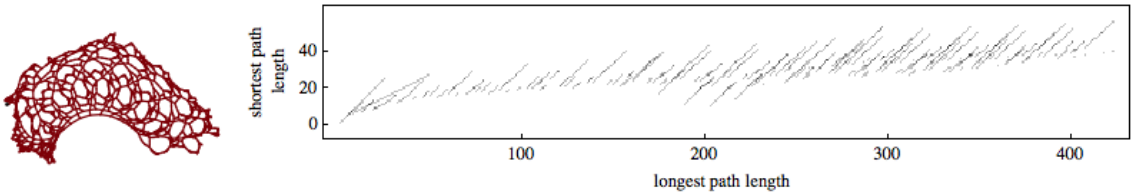

Figure 12 - A transitively reduced causet obtained from a 1500-step string rewrite system based on rewrite rules {100→0111, 1011→010, 0→0011}, starting from string 101, and associated longest/shortest path plot.

In conclusion, our examples show that in some cases - definitely a minority - causets from simple, deterministic models of computation, both regular and less regular, achieve the special property of 'non-locality', intended as the ability to connect, via a short path, even nodes found at a large, longest-path distance.

6. New deterministic 2D causet construction techniques

The deterministic causet construction techniques considered in the previous section have been deliberately conceived as completely independent from the sprinkling technique, which heavily depends on an underlying manifold. In this section we introduce a new family of deterministic techniques that tries to mediate between sprinkling and the purely abstract algorithmic approach, in an attempt to improve the quality of the longest/shortest path plots -- our indicator of Lorentzian non-locality -- while retaining the benefits of the deterministic approach. The additional benefit that we expect, over stochastic techniques, is the emergence of structure, in one form or another.

For doing so, 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, like reading, writing, comparing and swapping cells, adding new ones. At each step the cell array A of length n contains a permutation π 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 [XXX], we are interested in enumerating and exploring exhaustively (as much as possible) the complete space of instances of each automaton. If all reactions are potentially applicable to any situation, the size of this space is $|R|^{|S|}$.

6.1 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 cell array A, which keeps permutation π as described above.

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

Reaction - coded by 4 bits: $b_1 \ldots b_4 \Rightarrow$ 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 bits per situation). We have simulated and inspected all of them, starting from an initial configuration in which $A_{init} = (1, 2)$ and the ant is in state $s_{init} = 0$, positioned at cell 1. Note that when the reaction fails - the ant attempting to access non-existent 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.

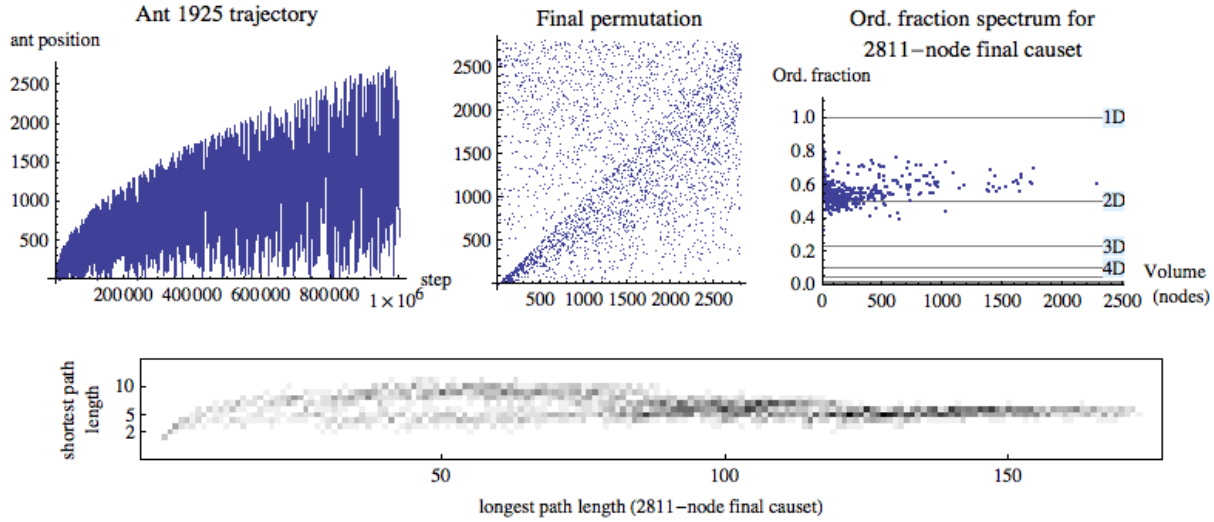

Figure 13 - Stateful PA automaton n. 1925. Ant trajectory, final permutation, ordering fraction spectrum, longest/shortest path length.

The first diagram plots the positions occupied by the ant on the growing permutation 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 5).

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

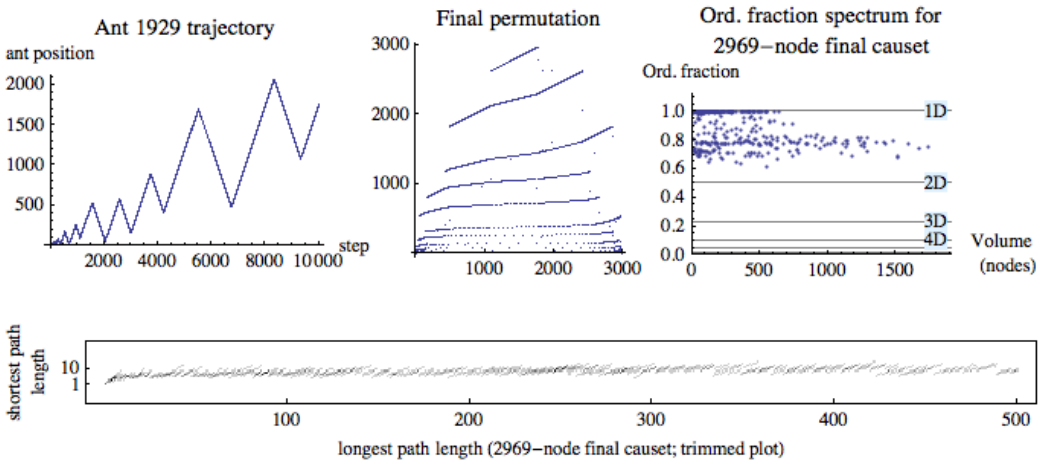

Figure 14 - Stateful PA automaton n. 1929. Ant trajectory, final permutation, ordering fraction spectrum, longest/shortest path length.

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, revealing a somewhat more structured causet, and 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). Again the plot looks quite satisfactory.

6.2 Stateless PA automaton

The second type of algorithm that we consider is the *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), π(i)), where bit(i) is a bit and π(i) is the permutation element, as before.

Situation - coded by 2 bits: b_1 and b_2 ⇒ 4 cases.

b_1 = bit(c). Cell c is where the ant is currently positioned.
b_2 = bit(c-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 ⟹ Write b_2 in cell c and swap cells c and c-1.
b_1 = 1 ⟹ 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 position).

Reasoning as in the previous case, we now obtain $8^4 = 4096$ different automaton instances. We have simulated and inspected all of them, starting from an initial configuration in which A_{init} = ((0, 1), (0, 2)) and the ant is 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.

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.

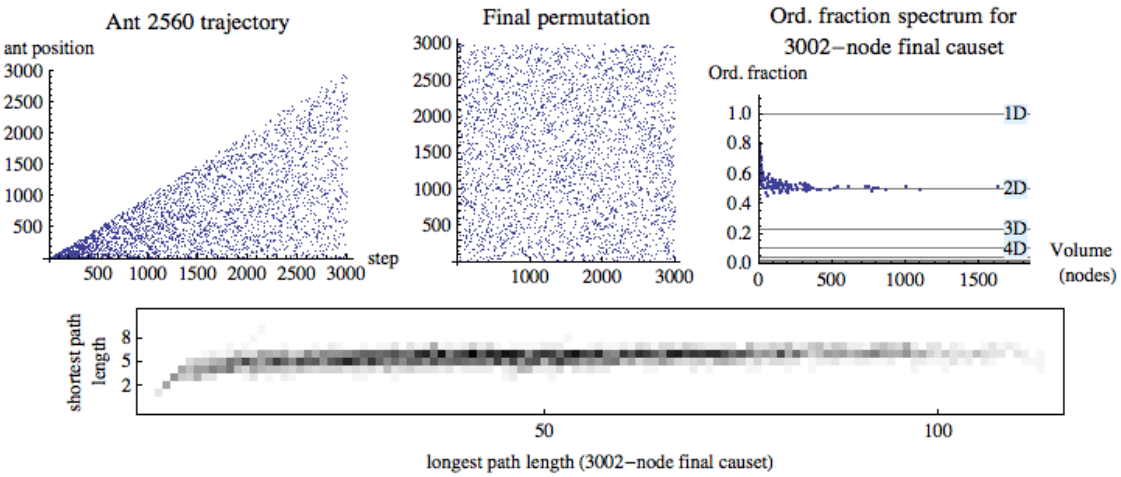

Figure 15 - Stateless PA automaton n. 2560. 3000-step computation. Ant trajectory, final permutation, ordering fraction spectrum, longest/shortest path length.

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. 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. Based on these observations, we can 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 16 just the ant trajectory for a second instance of the model, namely n. 3593.

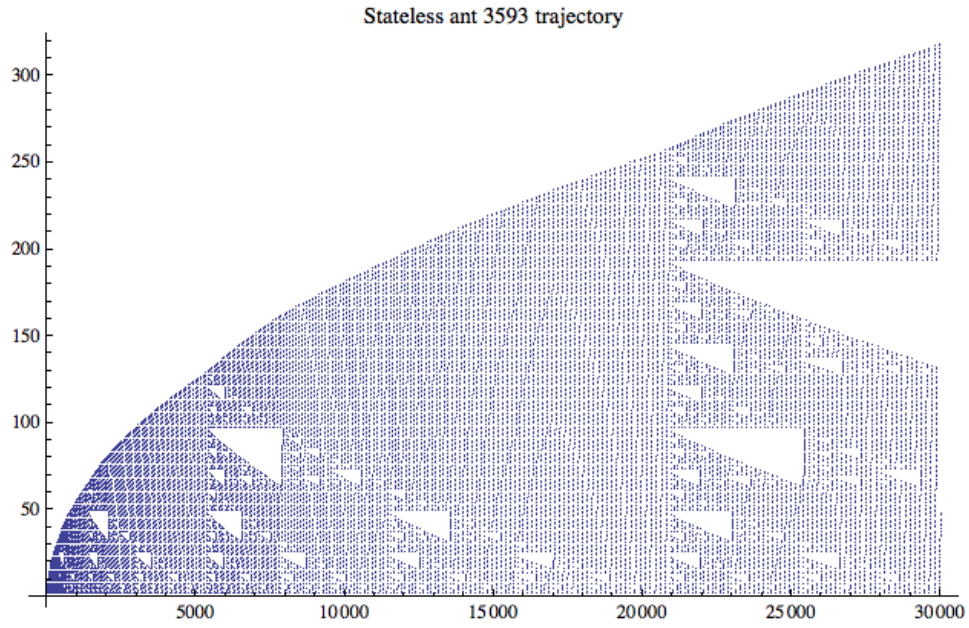

Figure 16 - Ant trajectory for stateless PA automaton n. 3593

Unfortunately, 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 [18]: it is precisely this richness of emergent phenomena, from regular to pseudo-random patterns, that makes the study of algorithmic causets so attractive.

7. Conclusions

In this paper we have introduced a number of 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 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, showing that, in spite of their usefulness for Myrheim-Meyer dimension estimation, the former cannot characterize "Lorentzianity", failing to discriminate between cases as diverse as the highly non-local causets obtained by sprinkling in Minkowski space, and highly local, directed graphs such as a regular square 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 [12] 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 [12], the analysis is inevitably bound to address only the transitively closed version of the causet, with the limitations that we have just indicated.

Considering transitively reduced causets proves more useful, although, unfortunately, transitive reduction is computationally costly [2], being O(|V|*|E|) for a directed acyclic graph G(V, E) (this has been the main computational bottleneck of our investigation). We have experimentally verified that the node degrees of a transitively reduced sprinkled Minkowski interval causet grow logarithmically with the node density (Figure 3), not polynomially as suggested in [15]; and that these degrees closely follow a Poisson distribution, departing from it in the presence of curvature (Figure 4).

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 typical, almost flat shape that these plots assume for sprinkled Minkowski and de Sitter intervals is not observed with the other considered stochastic causet classes.

We have also considered deterministic, algorithmic causets, and identified explicitly some of the structural graph properties that favour the appearance of 'well-behaved' longest/shortest path plots. We have introduced two new classes of 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, completely indistinguishable from a sprinkled Minkowski 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, would prevent 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, in particular, for cases in which an optimal behaviour in terms of longest/shortest path plots combines with rich internal structure.

It would also be interesting to spot analogous cases, in nature or even in engineering, 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, "The Structure of Random Graph Orders", SIAM J. Discrete Math., Vol. 10, No. 2, pp. 318-335 (1997).

4 T. Bolognesi, Planar Trinet Dynamics with Two Rewrite Rules. In: Complex Systems, vol. 18 (1) pp. 1 – 41. Complex Systems Publications, Inc, 2008.

5 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.)

6 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.

7 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.

8 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.

9 T. Bolognesi, "Inflation in 2D de Sitter Spacetime", http://demonstrations.wolfram.com/InflationIn2DDeSitterSpacetime/Wolfram Demonstrations Project, Published: February 28, 2013.

10 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.

11 Joe Henson, Constructing an interval of Minkowski space from a causal set]

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

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

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

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

16 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.

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

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