

The effect of Poisson sprinkling methods on causal sets in 1+1-dimensional flat spacetime

Arjun Deshpande^{1*}, Robert Pitu^{2*}, David D. Reid³

- ¹ Rockwoods International School, Udaipur, Rajasthan, India
- ² "Mircea cel Batran" National High School, Constanta, Constanta, Romania
- ³ Department of Physics, University of Chicago, Chicago, Illinois
- * These authors contributed equally to this work.

SUMMARY

The causal set theory (CST) is a theory of the smallscale structure of spacetime, which provides a discrete approach to describing quantum gravity. Studying the properties of causal sets requires methods for constructing appropriate causal sets. The most commonly used approach is to perform a random sprinkling. However, there are different methods for sprinkling, and it is not clear how each commonly used method affects the results. We hypothesized that the methods would be statistically equivalent, but that some noticeable differences might occur, such as a more uniform distribution for the subinterval sprinkling method compared to the direct sprinkling and edge bias compensation methods. We aimed to assess this hypothesis by analyzing the results of three different methods of sprinkling. For our analysis, we calculated distributions of the longest path length, interval size, and paths of various lengths for each sprinkling method. We found that the methods were statistically similar. However, one of the methods, sub-interval sprinkling, showed some slight advantages over the other two. These findings can serve as a point of reference for active researchers in the field of causal set theory, and is applicable to other research fields working with similar graphs.

INTRODUCTION

The causal set theory (CST) is one of several approaches used for the development of a quantum theory of gravity (1). Contrary to our everyday experience, where we perceive objects to have continuity, such as a table that feels rigid and uniform to the touch, CST supposes the discreteness of spacetime. This is similar to how a table, on an atomic scale, begins to look less and less continuous. Additionally, CST postulates that events within a finite region of spacetime can be modelled by a finite number of elements of a partially ordered set. This partially ordered set is called a causal set. This is an arrangement such that, for certain pairs of elements one precedes the other. The theory allows for a non-continuous structure that maintains local Lorentz invariance meaning the laws of physics stay the same for all observers (2).

A causal set is a set of partially ordered events for which a relation < is defined by obeying three distinct properties (3):

- 1. transitivity: if x < y and y < z then x < z;
- 2. non-circularity: if x < y and y < x then x = y;

3. finitarity: there cannot be an infinite number of events between any two fixed events.

The causal relation in spacetime satisfies the first two conditions. The third condition introduces spacetime discreteness.

The physical meaning of the causal relation comes from the fact that the maximum speed at which information can travel between events in spacetime is the speed of light, c. We therefore introduce the concept of a light-cone, as the region of spacetime that can be reached by photons traveling through spacetime. Our study used 1+1-dimensional flat spacetime (1 spatial and 1 time dimension), with units such that c=1. With these choices, the lightcone of an event is formed by $\pm 45^{\circ}$ lines that intersect at the event associated with the lightcone. The two ways we represented the lightcones are shown in **Figure 1**.

Although the ultimate goal of CST is to provide a model of spacetime on small scales (quantum gravity), it must also work as a model of spacetime on large scales (special and general relativity). The macroscale properties of causal sets, particularly with respect to special relativity, were the focus of our work.

Given a causal set, it is difficult to determine whether it is an appropriate model of spacetime, which means that it can be embedded into a spacetime (3). Therefore, to produce such a causal set, one can select points in spacetime and use those points to generate the set. This selection of points in a spacetime is called sprinkling. The causal relations between the sprinkled points in the spacetime determines the ordering of the elements in the causal set. To maintain Lorentz invariance, the sprinkling is done randomly using a Poisson distribution (4). This essentially means that each of the selected points are independent of each other. The distribution expresses the probability of a given number of selected events in a region of spacetime if these selections occur with a known constant mean rate. The distribution follows the distinct form:

$$P(N) = \frac{(\rho V)^N}{N!} e^{-\rho V},$$

where P(N) is the probability of selecting N points in a region of volume V if the selection is carried out with a uniform density ρ .

There are multiple methods of performing Poisson sprinklings. However, we found no studies showing how various commonly used methods differ in their results. Our

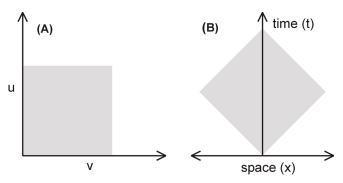


Figure 1: The spacetime region in lightcone and Cartesian coordinates. (A) shows the original (u,v) light cone coordinate system used to generate the sprinkling. (B) shows the transformed causal diamond in traditional spacetime coordinates (t,x).

purpose was to perform such a study, where we hypothesized that different methods, performed appropriately would be statistically equivalent, but that some noticeable differences might still occur. To conduct our study, we used three different methods that are used in current research to observe whether there are statistical differences between them (**Figure 2**). The three methods we used were:

- Direct sprinkling (DS): A simple selection of points within a given region of spacetime.
- Subinterval sprinkling (SBI): Divide the region into 9 equal parts and proportionally sprinkle points into each subregion.
- Edge bias compensation (EBC): Implement measures to remove an endpoint bias inherent in many random number generators.

In addition to a direct comparison of the methods, our analysis also included a few quantities of current use in CST. These quantities were distributions of the longest path length, interval size, and paths of various lengths (*k*-paths) for each of the sprinkling methods. These quantities were chosen because (a) they are of current interest in CST for calculating structures of physical interest such as space-like and time-like distances and estimating spacetime dimensions among other things and (b) they are comparatively easy to compute by computer.

RESULTS

The purpose of this study was to make a direct comparison of the three sprinkling methods mentioned previously. A feature of a Poisson sprinkling into a spacetime is that the number of points selected within a region should be directly proportional to the size of the region. We used this property to perform sprinklings by the different methods and performed statistical comparisons of the methods and results produced by these methods for the three quantities mentioned in the preceding paragraph.

To conduct this analysis, we ran Poisson sprinklings with expected sizes $\langle N \rangle$ up to 500 in the region of spacetime (**Figure 1**). For each value of $\langle N \rangle$ we ran 50 sprinklings. We divided the region into four bins of equal size. For each individual sprinkling of size $N_{,}$, we counted the number of points in each bin and compared them to the expected number, $N_{,}/4$, using a chi-squared goodness-of-fit test (**Figure 2D**). For this test the data hovered roughly around the value of 3. We made the standard choice of $\alpha = 0.05$ for the significance

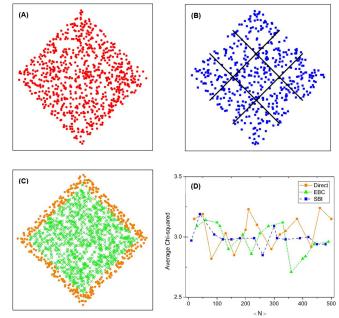


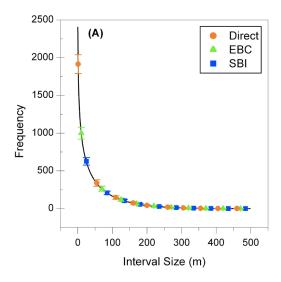
Figure 2: Examples of the sprinkling methods and how they compare. (A) shows a sprinkling of points by the direct method. (B) shows a sprinkling of points by the subinterval method (SBI). The lines show the boundaries of the subintervals. (C) shows a sprinkling of points by the edge-bias compensation method (EBC). The central (green) crosses are the accepted points; the outer (orange) dots are the rejected points. The spacetime size of the causal diamond containing only the accepted points is the same as those in (A) and (B). (D) shows the results of the chi-squared test of the sprinkling methods

level which gave a critical value of χ^2_{crit} = 7.815 given that our test had three degrees of freedom. We also determined the standard deviations of all three methods which were σ_{SBI} = 0.092, σ_{DIR} = 0.115 and σ_{EBC} = 0.125.

The interval size distribution for a causal set calculates the number of intervals of varying size m contained within the set. An interval in a causal set is the set of all elements between two related elements. So, two elements x < y define an interval consisting of all elements q such that x < q < y. The interval size is then the number of such elements q. This distribution has recently been used to study things like how to define local neighborhoods in causal sets and the estimation of spacetime dimensions (5,6).

In total, we calculated the interval size distributions for causal sets in the range $\langle N \rangle = [10-500]$ and performed a chi-squared calculation for each method. All the results were based on averages of 50 sprinklings except for $\langle N \rangle = 500$ which was based on 100 sprinklings. We plotted the interval size distributions for each sprinkling method for $\langle N \rangle = 500$ (**Figure 3**). For clarity, we display this data as a linear plot and a semi-log plot. It is clear from the figure that the data all match the analytical curve within the standard deviation. Average interval sizes for the same values of m for the different methods were also within each other's standard deviations, showing that the results were consistent with each other.

Our chi-squared results for the interval size distribution tell an interesting story. No single method produced the best fit to the theoretical result for all cases. For the smallest causal



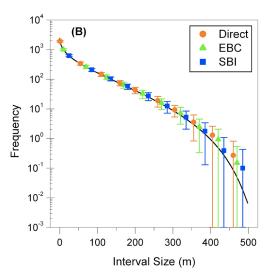


Figure 3: The interval size distribution from each sprinkling method for <N> = 500. (A) shows a linear plot. Due to the large frequency range, the spread in the values at large interval sizes is suppressed. (B) shows a semi-log plot of the same data to spread out points and better visualize how close the data was to the expected curve. The solid curve is the theoretical prediction of Glaser and Surya (5). The error bars are the standard deviations. (A) and (B) only show a representative subset of data points from each method. This was done because the large number of overlapping points of the entire data set would make it impossible to distinguish points from the different methods. We used the full dataset in our analysis.

sets of expected sizes 10, 20, and 50, each method gave the best fit in one of the cases. However, the larger sizes, 100 – 500, showed consistently that the subinterval sprinkling method has the best fit, followed by edge bias, then direct sprinkling.

For the longest path distribution recall that a causal set is partly defined by its ordering (or precedence) relation <. A chain in a causal set is a subset consisting only of related elements, such as, x < y < ... < z. Two related elements are linked if they are "nearest neighbors" with respect to <; that is, if x < z and there is no element y such that x < y < z. A path in a causal set is a chain consisting only of linked elements. The length of a path is the number of links it contains. One reason paths are important in CST is because the length of the longest path, ℓ_{max} , corresponds to an important property of spacetime known as the proper time (1). We calculated the average longest path length $<\ell_{max}>$ of 500 sprinklings as a function of the size of the causal set < N> (Figure 4). We found that all the data for each sprinkling method were within each other's standard deviations. Our results were also compared to a numerical fit performed by Rideout and Wallden and also proved to be consistent with their result (7).

The preceding discussion focused on the longest path between two related elements in a causal set, but there will also be shorter paths of different lengths between the same two elements. This distribution of path lengths is also of interest and holds the promise of being an important descriptor for determining whether a causal set is like a spacetime (8).

We calculated the distribution of the number of paths of varying lengths k, the k-path distribution using each sprinkling method (**Figure 5**). As indicated by the error bars the number of k-paths is subject to large statistical fluctuations. The results for each method were within the (large) standard deviations of the results from the other methods. For this case agreement with the relevant analytical curve was also

good. However, the chi-squared analysis indicated that these results had less agreement with the theoretical result than the other calculations for the larger causal sets. Furthermore, these best fit results for the four size cases did not provide a clear distinction between the sprinkling methods because different methods produced the best fit for different cases.

DISCUSSION

In this work, we studied three sprinkling methods by

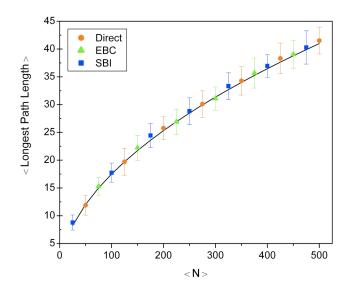


Figure 4: The longest path distribution from each sprinkling method. The error bars are the standard deviations. The solid curve is from Rideout and Wallden (7). The graph only shows a representative subset of data points from each method. This was done because the large number of overlapping points of the entire data set would make it impossible to distinguish points from the different methods. We used the full dataset in our analysis.

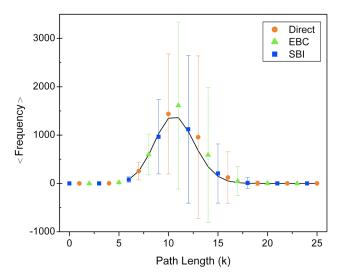


Figure 5: The path-length distribution for the different methods for <N> = 75. The vertical axis is the average k-path frequency. The solid curve is the theoretical prediction of Aghili et al. (6). The error bars are the standard deviations. The analytical result has been cut off for the first four values due to numerical error. This numerical issue was also observed by Aghili et al. (8). We only show a representative subset of data points from each method. This was done because the large number of overlapping points of the entire data set would make it impossible to distinguish points from the different methods. We used the full dataset in our analysis.

comparing them using four numerical calculations and theoretical expectations where available. We expected that the methods would be statistically equivalent and that is what we found. We also wanted to see if we could identify any trends to suggest that one method might be better than the others.

For the sprinkling methods the fact that the chi-squared statistic values for each method were consistently less than the critical value indicates that all three sprinkling methods were consistent with a Poisson sprinkling. However, the subinterval sprinkling method showed the least fluctuation as supported by its smaller standard deviations. This shows that points sprinkled using SBI tend to be spread more evenly, creating a more consistent density of points.

We also expected the chi-squared values to decrease as <*N*> increased due to improved statistics. This trend is slightly noticeable for the case of subinterval sprinkling. However, statistical fluctuations are too large to definitely say that we observed this effect. A larger simulation, such as a larger size and more sprinklings, is probably needed to adequately study this issue.

For the interval size distribution our results suggest that all methods are statistically equivalent because results from each method was within the standard deviations of the other two. However, the fact that the results of the subinterval method were better correlated with the theoretical expression provides a reason to give slight preference to SBI.

The longest path distribution produced the best correlation between the various methods. For this distribution, no exact theoretical expression is known for finite causal sets. Details of the sprinkling method used to generate the smooth curve were not provided (7). As indicated by the results all three

sprinkling methods matched the Rideout-Wallden curve very well. For the causal set sizes we studied, the longest path distribution does not seem to be sensitive to any of the sprinkling methods.

Finally, for the *k*-path distribution the three sprinkling methods also seem to be statistically equivalent, but with large fluctuations. This calculation also did not agree as well with the theoretical expectation as the other metrics. This comparatively worse agreement is most likely a result of the large fluctuations. Also, neither method consistently produced the best fit. For the *k*-path distribution it appears that the effects of statistical fluctuations dominate over any effects from the sprinkling methods.

Overall, our results slightly suggest that the subinterval method does a better job at providing a uniform distribution of points. Because of this improved distribution, there may be quantities for which the subinterval method will give better results. Our results suggest that the interval size distribution may be such a quantity. However, it is important to note that due to the statistical nature of the Poisson sprinkling process, results may vary depending on sample size, spacetime dimension, and number of sprinklings.

The main caveat of our study is that it was small in scope which limits our ability to reach general conclusions. A study using much larger causal sets, higher dimensions, a wider variety of theoretical descriptors, and sprinklings into curved spacetimes in addition to flat would be more revealing. A computationally intensive effort of this magnitude was beyond the time and resources available to us for this study.

Despite the limited scope of this effort, the calculations and methods used in this study followed standard practice in current causal set research. Our conclusions can inform future research when creating 2-dimensional causal sets by sprinkling. This work, and possible extensions of it could provide a base template that can be further optimized to create an industry standard for causal set research. Furthermore, work such as ours can be used to study randomness generators, perform calculations in graph theory and combinatorics, and in other fields that require a random discretization of a continuous space.

MATERIALS AND METHODS

The Sprinkling Process

All of our sprinklings were performed in an interval of spacetime between two causally connected events. The interval between two events in spacetime, p and f (where p precedes f), is defined to be the region of spacetime common to the future of p and the past of f. To simplify the sprinkling process, we represented the interval in light cone coordinates (u,v). This is a special coordinate system where each coordinate combines both space and time. In these coordinates our 1+1-dimensional region is just a square (**Figure 1A**). The light signals that intersect at a spacetime point trace out a double cone called a light cone.

The sprinkling, and all other calculations, were programmed in Python, using Google Collaboratory. To perform the sprinkling for a desired average number of causal set elements < N>, we used the Numpy library's random poisson function to obtain a sequence of causal set sizes N_i and the numpy random function to separately select u and v coordinates for each point until we reached the desired number (9).

After sprinkling in light cone coordinates, we performed the

coordinate transformation,

$$t = \frac{u+v}{\sqrt{2}} \ , \ \ x = \frac{u-v}{\sqrt{2}} \, ,$$

to the traditional spacetime coordinates (t,x) (Figure 1B). Because of the light cone structure of spacetime, in (t,x) coordinates an interval of this type is also called a causal diamond.

This basic sprinkling process was used to conduct random sprinklings by the three methods we discuss next. A direct comparison of these methods is discussed in the results section.

To describe the direct sprinkling method, let the lower left corner of our spacetime interval be (u,v)=(0,0) and the upper right corner be $(u,v)=(u_{\max},v_{\max})$. To sprinkle N points in this region, for each point i we got its u_i coordinate by selecting a random number in the range $[0, u_{\max})$ and its v_i coordinate by selecting another random number in the range $[0, v_{\max})$. This process was repeated until we got coordinates for all N points. Note that the random number selection was inclusive of the 0 but exclusive of u_{\max} and v_{\max} . A sprinkling performed by this method is shown in **Figure 2A**.

For the subinterval method, the main interval was divided into nine equal subintervals. A sprinkling of N points in the main interval was conducted by sprinkling approximately N/9 points into each subinterval by the direct sprinkling method described above. The purpose of this approach is to help ensure a uniform sprinkling throughout the region that should be consistent with a Poisson process (**Figure 2B**).

The fact that the random number selections were exclusive of the upper endpoint of the range introduces a slight undesirable bias in the sprinkling at the edge of the spacetime region. This full or partial exclusivity of the endpoints is a common feature of many pseudo-random number generators. The edge bias compensation method handles this bias by using the rejection method. To sprinkle *N* points into the main interval, we completely enclosed the interval into a larger region. Then we used the direct sprinkling method to select points in this larger region, rejecting any points that fell outside the desired interval. With this method it is possible to select points on the outer edge of the region, eliminating the edge bias. We continued to select points in this way until we got *N* points within the desired interval (**Figure 2C**).

Once the points were sprinkled, the relations matrix, R_{ij} was created. This is an $N \times N$ two-dimensional array, storing the relation of each pair of points as a Boolean value. For any pair of points (i,j), $R_{ij} = 1$ if point i exists in the future of point j (both u and v coordinates of i are greater than those of j), otherwise, $R_{ij} = 0$. Once the relations matrix is known, there is no longer any need to refer to events in spacetime, but only the partially ordered set of elements governed by the relations matrix. This set of elements is the causal set.

From the relation matrix, the link matrix L_{ij} was formed. As previously described, a link is an irreducible relation. The link matrix is also an $N \times N$ two-dimensional array, storing the information about which elements are linked as a Boolean value. Therefore, $L_{ij} = 1$ if $R_{ij} = 1$ and there are no other elements q such that both $R_{qj} = 1$ and $R_{iq} = 1$, otherwise $L_{ij} = 0$. Using the relations and link matrices we were able to calculate several important properties of the causal sets.

Calculational Methods

As discussed in the results section, we computed three quantities for the purpose of comparing the influence of the sprinkling methods on calculations in CST.

Recall that two elements x < y define an interval consisting of all elements q such that x < q < y. The interval size m(x,y) is then the number of such elements. For example, an interval of size of 0 is given by two elements that are linked. The interval size distribution was calculated by considering all such related pairs and counting the number of elements between them: $m(x,y) = \sum_{q} R_{qx} \times R_{yq}$. Their distribution, over all of the generated sprinklings gave us their standard deviation, and average frequency.

An analytical expression for the interval size distribution for faithfully embeddable causal sets in flat spacetime has been worked out by Glaser and Surya (5). Their result, when specialized to the two-dimensional case is,

$$\langle N_m \rangle = \frac{m!}{(m+2)!} \frac{N^{m+2}}{\Gamma(m+3)} \frac{1}{2} F_2 \begin{pmatrix} 1+m, 1+m, \\ 3+m, 3+m, \end{pmatrix} - N$$

where $< N_m >$ is the average number of intervals of size m in a causal set interval of size N. The function $\Gamma(\cdot)$ is the gamma function and ${}_{2}F_{2}(\cdot)$ is the hypergeometric function.

The longest path distribution computes how the average longest path between the defining elements of a causal set interval changes with respect to the expected size of the interval. To compute this quantity, we used Clough's *DAGOLOGY* code to calculate the longest path matrix, an *N×N* two-dimensional array storing the maximum path length between each pair of related elements (10). Therefore, the maximum value found in this matrix gives the value we seek. We determined average longest paths for causal sets generated from 500 sprinklings. We also calculated the standard deviation and variance arrays using the Numpy library's numpy.std and numpy.var functions, respectively (9).

The analytical curve to which we compared our simulations is from a numerical fit performed by Rideout and Wallden,

$$\langle \ell_{\text{max}} \rangle \approx \left(2 - (0.79) \times N^{-\frac{1}{4}} \right) \times \sqrt{N},$$

where $<\ell_{max}>$ is the average longest path length, and N is the size of the causal set (7).

We also calculated the distribution of paths of varying lengths k. This k-path distribution describes how many paths of a specified length occur between the defining elements of a causal set interval where k represents the number of links in the path. The k-path distribution was calculated using a modified version of the Depth-First Search algorithm (11). Using this recursive algorithm, all possible paths between the endpoints are "walked" and their respective lengths (number of links) are stored in memory. Their distribution, over all the generated sprinklings gave us their standard deviation, and average frequency.

The average number of k-paths has large statistical fluctuations. To help combat these fluctuations, we used the largest number of sprinklings for these calculations. Our results were based on averages over 10,000 sprinklings

except for $\langle N \rangle$ = 100, which used 500 sprinklings because of the long computer time needed to run that calculation.

An analytical expression for the *k*-path distribution for faithfully embeddable causal sets in flat spacetime has been worked out by Aghili et al. (6). Their result, when specialized to our two-dimensional case is.

$$\langle n_k \rangle = \frac{N!}{(N-k+1)!} \bigg(\frac{\Gamma(3)}{2} \bigg)^{k-1} \times \sum_{i=0}^{N-k+1} \binom{N-k+1}{i} \frac{(-1)^i \Gamma(i+1)}{[\Gamma(i+k)]^2} f_{i,k} \,,$$

where $\langle n_k \rangle$ is the average number of k-paths, $\Gamma(\cdot)$ is the gamma function, and the f_{ik} are defined by,

$$f_{i,1} = \Gamma(i+1)$$

for k = 1, $0 \le i \le N$, and

$$f_{i,k} = \sum_{i=0}^{i} \Gamma(i-j+1) f_{i,k-1}$$

for $2 \le k \le N+1$, and $0 \le i \le N-k+1$.

Our primary method for comparing the simulated data to expected analytical results was by a chi-squared goodness of fit test. The chi-squared calculation follows the formula:

$$\chi^2 = \sum_{i=1}^{B} \frac{(x_i - c_i)^2}{c_i},$$

where χ^2 is the chi-squared value, the c_i are the calculated values and the x_i are the expected values given by the analytical expressions or what's expected for a Poisson process, and B is the total number of bins or data points. For interpreting the results, we adopted a significance level of 0.05 and got critical chi-squared values from a standard chi-squared distribution table (12). Our codes can be found on GitHub at the following link: github.com/Arjundeshpande/CausalSets/tree/main.

ACKNOWLEDGMENTS

For help with understanding the fundamental physics of causal set theory, we give special thanks to Mr. Ajay Singh (Rockwoods International School) and to Mr. Lucian Oprea (Mircea cel Batran National High School).

Received: October 06, 2023 Accepted: April 03, 2024 Published: February 14, 2025

REFERENCES

- Bombelli L., et al. "Space-Time as a Causal Set." Physical Review Letters, vol. 59, no. 5, Aug. 1987, https://doi.org/10.1103/PhysRevLett.59.521.
- Dowker, F., et al. "Quantum gravity phenomenology, Lorentz invariance and discreteness." Modern Physics Letters A, vol. 19, no. 24, Aug. 2004, https://doi.org/10.1142/ S0217732304015026.
- 3. Reid, D. D. "Discrete quantum gravity and causal sets."

- Canadian Journal of Physics, vol. 79, no. 1, Jan. 2001, https://doi.org/10.1139/p01-032.
- Christ, N. H., et al. "Random lattice field theory: General formulation. "Nuclear Physics B, vol. 202, no.1, Jul. 1982, https://doi.org/10.1016/0550-3213(82)90222-x.
- Glaser, L., and Surya, S. "Towards a definition of locality in a manifoldlike causal set. "Physical Review D, vol. 88, no.12, Dec. 2013, https://doi.org/10.1103/PhysRevD.88.124026.
- 6. Aghili, M., L., et al. "Discrete spacetime: a web of chains." *Classical and Quantum Gravity*, vol. 36, no.18, Aug. 2019, https://doi.org/10.1088/1361-6382/ab30e3.
- Rideout, D., and Wallden, P. "Spacelike distance from discrete causal order." Classical and quantum gravity, vol. 26, no.15, Jul. 2009, https://doi.org/10.1088/0264-9381/26/15/155013.
- Aghili, M., et al. "Path length distribution in two-dimensional causal sets." The European Physical Journal C, vol. 78, no. 9, Sep. 2018, 10.1140/epjc/s10052-018-6229-7.
- Harris, C.R., et al. "Array programming with NumPy." Nature, vol. 585, no. 7825, Sep. 2020, https://doi.org/10.1038/s41586-020-2649-2.
- "DAGOLOGY." Clough, J., github.com/JamesClough/ dagology. Accessed 10 Nov. 2022.
- Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein, "Introduction to Algorithms," 3rd ed., The MIT Press, 2009, pp. 603-12.
- Spiegel, M. R., "Mathematical Handbook of Formulas and Tables," Schaum's Outline Series, McGraw-Hill, New York, 1969, p. 259.

Copyright: © 2025 Deshpande, Pitu, and Reid. All JEI articles are distributed under the attribution non-commercial, no derivative license (http://creativecommons.org/licenses/by-nc-nd/4.0/). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.