An Alternative Method for Characterization and Comparison of Plant Root Shapes

A thesis submitted to the
College of Graduate and Postdoctoral Studies
in partial pulfillment of the requirements
for the degree of Master of Science
in the Department of School of Environment and Sustainability
University of Saskatchewan
Saskatoon

By
Yujie Pei

Contents

Existing Morphological Descriptors for Root Systems

1.1 Background

1.1.1 Importance of Roots

- Mechanical and functional abilities of plant roots
- Plant root plasticity in the resource-limited environment

1.1.2 Importance of Research

1.1.2.1 Demand in crop breeding programs

1.1.2.2 Environment and Sustainability

- reduce the negative impacts of fertilization
- high crop productivity to feed the increasing global population

1.1.3 Promising application in other branching structures

- Trees
- River networks in geography
- Blood vessels in medicine
- Leaf vein networks

1.2 Summary of Existed Descriptors

1.2.1 Metric

1.2.1.1 Basic Geometric Descriptors

- maximum depth
- maximum width
- etc.

1.2.1.2 Compound Descriptors (Computed From the Basic Descriptors)

- Density
- aspect ratio
- etc.

1.2.1.3 Weaknesses

- Rely on the resolution of the images [5]
- Only provide a general view of root morphology [5]
- Difficult to assess the spatial configuration of roots
- Fail to describe the full complexity of root systems

1.2.2 Non-metric

1.2.2.1 Topological Analysis

- Persistent homology [34]
  - Persistence barcode shows the number of branches.
  - Persistence barcode indicates how branched roots connect along the scale of the function (e.g. geodesic distance).
  - Compare the similarity of branching structures by a pair-wise distance matrix using the bottleneck distance method.
- Horton-Strahler index [39]
  - Categorize the topological complexity of the whole branching structure.
  - Provide a numerical measure of connectedness and complexity of the branching at each vertex by a dimensionless ratio: bifurcation ratio.
  - The range of index and the length ratio indicate the size of the branching structure.
- Fractal Analysis [38]
  - Measure the complexity of branching structures.
  - Measure self-similarity of branching structures by fractal dimension of the root systems.

1.2.2.2 Strengths

- Highly complementary to geometric descriptors to characterize how individual roots are connected through branching [10].
- Describe branching structures independent of transformation and deformation.

1.2.2.3 Problems and Weaknesses

- Only analyze the connectedness of branching structures, which is a portion of the complexity of plant root systems.
- Fail to characterize spatial distribution.
- Some biologically topological indices analyze the root growth qualitatively based on line-linked systems [16], but not characterize the mathematical topological properties.
- The fractal analysis aims to describe self-similar structures, which grow by continually repeating simple growth rules [17].

1.3 Problem Statements

...

An Alternative Mathematical Method for Shape Description

2.1 Kac's Idea: Can One Hear the Shape of a Drum? [28]

2.1.1 Interpretations of Kac's Problem

- When the drum vibrates, one can hear the sound, which is composed of tones of various frequencies. How much can shape features be inferred from hearing a discrete spectrum of pure tones produced by a drum?
- If a complete sequence of eigenvalues of the Dirichlet problem for the Laplacian can be obtained precisely, will people determine the shape of a planar?

2.1.2 Problem Statement

- Consider a simply connected membrane Ω in the Euclidean space bounded by a smooth convex curve $\partial\Omega$ (e.g. a drum without any holes)
- Find function ϕ on the closure of Ω, which vanishes at the boundary $\partial\Omega$, and a number λ satisfying $-\Delta\phi = \lambda\phi$.
  - Δ is the Laplace operator. e.g. $\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}$ in Cartesian coordinate system.
  - If there exists a solution $\phi \neq 0$, the corresponding λ is defined as a Dirichlet eigenvalue.
  - For each domain Ω, there has a sequence of eigenvalues $\lambda_1, \lambda_2, \lambda_3, ...$ corresponding to a set of eigenfunction $\phi_1, \phi_2, \phi_3, ...$
  - ϕ_k form an orthonormal basis of $L^2(\Omega)$ of real valued eigenfunctions; the corresponding discrete Dirichlet eigenvalues are positive ($\lambda_k \in \mathbb{R}^+$).
- An important function [21]:

$$h(t) = \sum_{k=1}^{\infty} e^{-\lambda_k t} \tag{2.1}$$

  - It is a Dirichlet series.
  - It is called the spectral function or the heat trace.
  - It is smooth and converges for every $t > 0$.

2.1.3 Summarize the Results of Kac's Idea

$$h(t) = \sum_{k=1}^{\infty} e^{-\lambda_k t} \sim \frac{|\Omega|}{2\pi t} - \frac{L}{4}\frac{1}{\sqrt{2\pi t}} + \frac{1}{6} \tag{2.2}$$

- As $t \to 0^+$, the leading terms of the asymptotic expansion of $h(t)$ imply the geometrical attributes of Ω
  - the total area
  - the perimeter
  - the curvature
- If the domain Ω has the polygonal boundary, the third term shows in the information about the interior angles of the polygon [21].

2.1.4 Conclusion

2.1.4.1 Advantages

- Kac proposed a novel analytical mathematical method for the shape description without using measuring tools, e.g. rulers.
- Other mathematicians extended Kac's idea in exploring the geometrical information of more complex domains with various boundary conditions [32][19][18] [44][37].

2.1.4.2 Limitations

- It is only available for the convex domain, which has a smooth or piecewise smooth boundary.
- Except in very few cases (i.e. rectangular, disk, certain triangles), the complete sequence of eigenvalues λ_k can not be calculated [21].
- Only the first few terms in the asymptotic expansion of $h(t)$ are explicitly available.

2.2 Extended Work of Kac's Idea [11][40]: Heat Content

2.2.1 Fouier's Heat Equation [6]

- Mathematical Formula

$$\frac{\partial u(\boldsymbol{s},t)}{\partial t} = \Delta u(\boldsymbol{s},t) \tag{2.3}$$

- Interpretation

  It is a deterministic model used to characterize the evolution of quantities over the space and time. (e.g. the flow of heat)

2.2.2 Summarize the Idea

- Initial-Boundary Value Problem (IBVP)

  $u(\boldsymbol{s}, t)$ indicates the value of the tempretature at $\boldsymbol{s} \in \Omega$ at time t satisfying Eq. 2.3 and

  - Initial condition: $u(\boldsymbol{s}, t) = f(\boldsymbol{s})$ as $t \to 0$.
  - Dirichlet boundary condition: $u(\boldsymbol{s}, t) = 0$ for $\boldsymbol{s} \in \partial\Omega$

    It is also called the absorbing boundary condition; i.e. any molecule will be instantly absorbed when it touches the boundary $\partial\Omega$;

- A Basic Integration

$$\beta_\Omega(f, g)(t) = \int_\Omega \int_\Omega H_\Omega(\boldsymbol{s}, t|\boldsymbol{s_0}) f(\boldsymbol{s_0}) g(\boldsymbol{s}) d\boldsymbol{s_0} d\boldsymbol{s} \tag{2.4}$$

$$= \int_\Omega u(\boldsymbol{s}, t) g(\boldsymbol{s}) d\boldsymbol{s} \tag{2.5}$$

  - $H_\Omega(\boldsymbol{s}, t|\boldsymbol{s_0})$ is called the heat kernel of Ω describing the density of the heat at $\boldsymbol{s}$ after time t when initially there is only one single hot source at $\boldsymbol{s_0}$.
  - $u(\boldsymbol{s}, t)$ is the general solution to Eq. 2.3, which can be expressed as the convolution of the initial condition with the heat kernel of the domain.
  - $g(\boldsymbol{s})$ is an auxiliary test function for studying the distributional nature of the tempretature function $u(\boldsymbol{s}, t)$ near $\partial\Omega$.

- Heat Content Calculation

  Given

$$g(\boldsymbol{s}) = 1 \tag{2.6}$$

$$Q_\Omega(t) = \int_\Omega \int_\Omega H_\Omega(\boldsymbol{s}, t|\boldsymbol{s_0}) f(\boldsymbol{s_0}) d\boldsymbol{s_0} d\boldsymbol{s} \tag{2.7}$$

$$= \int_\Omega u(\boldsymbol{s}, t) d\boldsymbol{s} \tag{2.8}$$

- Shape Characterization

  - As $t \to 0^+$,

$$Q_\Omega(t) \simeq \sum_{n=1}^{\infty} \beta_n(\Omega) t^{\frac{n}{2}} \tag{2.9}$$

  - Obtain geometrical information of Ω from β_n

* area
* length
* scalar curvature
* mass

2.2.3 Conclusion

- Strengthness
  - Instead of calculating a complete sequence of the Dirichlet eigenvalues for exploring the shape attributes of geometry, the asymptotic expansion of the heat content, defined as integrating the solution to the heat equation over the space-dimension, also implies the geometrical characteristics.
-
  - Only the infinitly differentiable boundary $\partial\Omega$ is considered.
  - Only the first few terms in the asymptotic expansion are explicitly known.
  - Either irregular geometries or discontinuities lead to the complexities, so the explicit solutions $u(\boldsymbol{s}, t)$ are close to non-existed.
  - The numerical evaluation of the analytical $u(\boldsymbol{s}, t)$ and $Q_\Omega(t)$ is usually by no means trivial because they are in the form of infinite series.
  - Similiarly, only the first few coeffients β_n in the asymptotic expansion of $Q_\Omega(t)$ can be expressed as the complicated explicit forms.

2.3 Monte Carlo Simulation for Approximating Heat Content $Q_\Omega(t)$

2.3.1 Background [27]

2.3.1.1 Stachasic Differential Equations (SDEs)

- Stochastic process
- Brownian motion

2.3.1.2 Connection Between SDEs and Heat Equation

- Ito calculus
- Intepretating the heat equation by the probability density of particles undergoing Brownian motion

2.3.2 A Special Case of Heat Content Calculation

- Uniform initial tempretaure distribution in Eq. 2.7

$$f(\boldsymbol{s_0}) = \frac{1}{|\Omega|} \tag{2.10}$$

  - $|\Omega|$ is the area of the domain Ω.
  - Particles's initial positions are distributed uniformly in Ω.

- Intepretation of Eq. 2.6: observing all the Brownian particles unbiasedly.

- Probabilistic Intepretation of Heat Kernel $H_\Omega(\boldsymbol{s}, t|\boldsymbol{s_0})$: conditional probability density function of Brownian particles

- Probabilistic Interpretation of $Q_\Omega(t)$: Survival Probability $S_\Omega(\tau)$

  - First passgae time τ: the time taken by the particle to encounter the absorbing boudary $\partial\Omega$ from the initial position.
  - Derivation of $S_\Omega(\tau)$ based on $H_\Omega(\boldsymbol{s}, t|\boldsymbol{s_0})$.

2.3.3 Monte Carlo Simulation (LRWs) for approximating $S_\Omega(\tau)$

2.3.3.1 Monte Carlo Integration

- Introduction
  - Definition: utilizing the random sampling of a function to compute an estimate of its integral numerically [23].
  - Uniform Sampling Method
- Given an initial position $\boldsymbol{s_0} \in \Omega$, approximating $H_\Omega(\boldsymbol{s}, t|\boldsymbol{s_0})$ by simulating the trajectories of a large number of particles by Lattice Random Walks (LRWs).
- Sampling a lager number of particles, whose initial sites are distributed uniformly within Ω to estimate $S_\Omega(\tau)$.

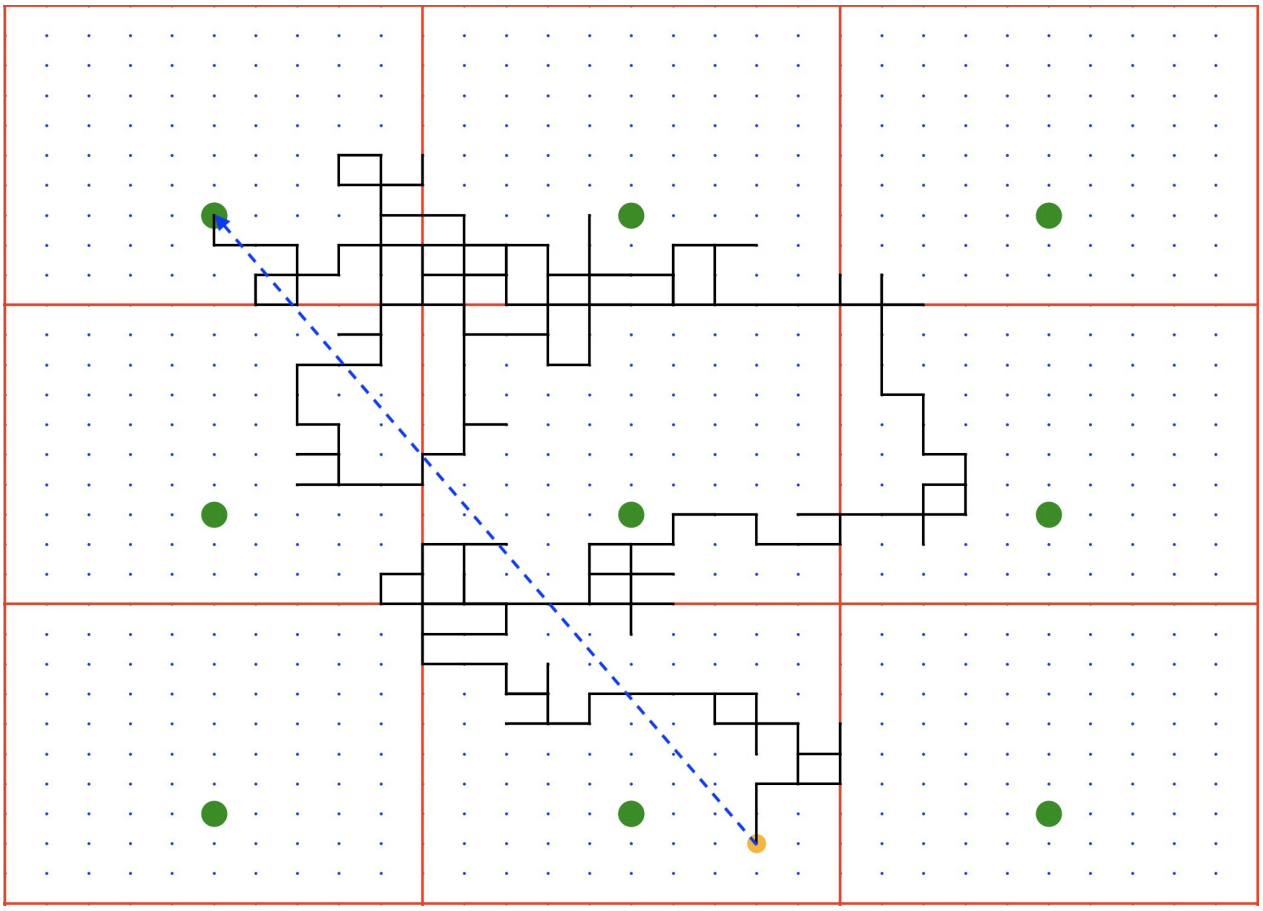


Figure 2.1: This plane is constructed by choosing a primitive cell with four red sides and a green point and replicating it infinitely to tile the whole $2-$ dimensional space. Moreover, there has no overlaps and voids between copies of the cell. A particle initially started LRWs from the orange site and be absorbed by any of the green points. To ensure smoothness and consistency, if the particle leaves the cell through one edge, it will appear in the adjacent cell with the same velocity. The black line segments show the particle's random trajectories, and the length of the blue dotted arrow is defined as its displacement.

2.3.3.2 Design LRWs in the $2-$ dimensional image

- Initial condition: uniform distribution within Ω, which is bounded by the border of the image and the edge of the target object.
- Boundary condition
  - Perodic boundary condition the edges of the image.

    In this thesis, periodic boundary conditions (PBCs) are employed to minimize the influence of images' edges. Fig. 2.1 is a simplest example of implementing PBCs in the Euclidean plane E^2 and tracks the trajectory of a particle undergoing LRWs.
  - Absorbing boundary condition on the boundary of the target shape.

2.3.3.3 Sample Size Determination - Dvoretzky–Kiefer–Wolfowitz (DKW) inequality [13]

- Mathematical formula and interpretation
- Strengths
  - Distribution-free
  - Sample size calculation does not depend on
    - Domain shape and size
    - Target geometry
    - B.C.s

2.3.3.4 Output Analysis

- Kaplan-Meier Estimator [30] [1][**?**]
- Confidence Interval [20] [26] [29][36]
- Two-Sample Statistical Tests (weighted logrank tests) [8] [2] [31] [33] [14] [24]
  - Wilcoxon
  - Tarone-Ware
  - Peto
  - Fleming-Harrington

2.4 Novel Applications of Monte Carlo Simulation for Shape Characterization

Survival analysis is a series of statistical methods dealing with variables that have both a time and event associated with them. It can be applied in a variety of fields [7], including cancer studies in clinical research, event-history analysis in sociology, failure-time analysis in engineering, and so on. As introduced in the last section 2.3, the event we are interested in is whether the particle hit the absorbing boundary or not, and the time refers to the number of steps, n, taken by the particle in LRWs. However, in this section, two novel forms of functions stem from the underlying survival process are developed for characterizing the shape of the object in the image.

2.4.1 Survival Analysis of Displacement $S(d)$

In some survival analysis applications, the origin-destination distance can be an alternative random variable to the time. For example, in the study of transport habits [3] [25], trip distance-based survival analysis can reveal traveller preference and trip patterns and measure the effects of explanatory variables. In Monte Carlo simulations, the displacement of a particle, d, is the shortest distance from the initial to the stop position in the infinite tiling space, which is affected by the number of steps n. Suppose $\mathcal{D}$ is a non-negative random variable representing the displacement moved by the particle from the origin to any sites of absorbing boundary. Like the $S(n)$, for any $d > 0$, we can define $S(d)$ as

$$S(d) = Pr(\mathcal{D} > d) = 1 - F(d) \tag{2.11}$$

where $S(d)$ indicates the probability that the particle has not yet been absorbed when the displacement beyond d, and F is the cumulative distribution function. Moreover, both $S(d)$ and $F(d)$ are right continuous in this thesis.

As mentioned before, the KM [30], or product-limit, can be used to estimate the survival function non-parametrically. Suppose that N particles determined by Dvoretzky–Kiefer–Wolfowitz (DKW) inequality [13] undergo LRWs independently in the simulation with distinct increasing non-negative displacements $d_1 < d_2 < ... < d_j < ... < d_k$. The probability of particles being alive at displacement d_j, $S(d_j)$, is calculated from $S(d_{j-1})$ by

$$S(d_j) = S(d_{j-1})\Big(1 - \frac{\tilde{n}_j}{n_j}\Big) \tag{2.12}$$

where n_j is the number of particles have not hit the target boundary within the displacement d_j, and $\tilde{n}_j$ is the number of particles being absorbed at displacement d_j.

2.4.2 Survival Analysis of Radius S(r)

Another function generated in this thesis for analyzing the spatial point patterns in the image is based on a union of discs of independent random radius centred at the uniform sampling points. The radius, R, of a disc is the shortest distance from an arbitrary initial site, $s \in \boldsymbol{I} \subseteq \mathbb{R}^2$, to the fixed absorbing boundary, $\Gamma \subseteq \mathbb{R}^2$, in the whole tiling of the plane. R is defined as

$$R = \rho(\Gamma, s) = \inf \{\|a - s\| \colon a \in \Gamma\} \tag{2.13}$$

The empty space function [4] of a point process [43] gives us an insight into the distribution of the random variable R. A point process is a collection of points located randomly in some underlying mathematical spaces. It evolved gradually from renewal theory and the statistical analysis of life tables dating back to the 17th century [9]. Point process data has wide applications in various fields [9], including epidemiology, ecology, forestry, mining, astronomy, ecology, meteorology, etc.

1−dimensional point process can be used to model a sequence of random times when a specific event happens. For example, given any period of time, a call center can receive calls at random instants or points of time. Suppose we map the caller locations on a particular day. Then, the map will constitute a random pattern of points in two dimensions because there will have an arbitrary number of such points, and their positions are random. Therefore, a spatial point process can model the random pattern of points in 2−dimensional space. If both the time and locations are recorded, we can consider it as a space-time point process in three dimensions. It is also a simple instance of the marked point process [4] since each location in $\mathbb{R}^2$ is labelled or marked by the time of the call in $\mathbb{R}$.

Let $\boldsymbol{X}$ be a stationary point process in $\mathbb{R}^k$, the empty space function [4], $F(r)$, also named spherical contact distribution function, is the cumulative distribution function of the contact distance, $dist(u, \boldsymbol{X})$, from a fixed point u to the nearest point of random set $\boldsymbol{X}$.

$$F(r) = Pr(dist(u, \boldsymbol{X}) \leq r) = Pr(N(b(u, r)) > 0) \tag{2.14}$$

where $r \geq 0$, and $N(b(u, r))$ is a random variable indicating the number of discs with radius r centered at u. $F(r)$ is independent on u because of the stationarity.

.....how?why?

In the light of the definition of empty space function of a stationary or uniform point process, the survival function $S(r)$ can be expressed as

$$S(r) = Pr(R > r) = 1 - F(r) = 1 - Pr(R \leq r) \tag{2.15}$$

LRWs in Artificial Images

The fixed-time step Monte Carlo simulation, LRWs, has been validated in the annulus in Appendix B by comparing the analytical heat content and numerical survival function. The further validation of LRWs is based on Eq.2.9, which indicates that the geometric features of a region with a smooth boundary in $\mathbb{R}^2$ can be deduced from the asymptotic properties of the heat content. Specifically, the coefficient of the first term is only associated with the total area of the region, and the second one can only be affected by the total length of the boundary of the region. Theoretically, for the equal-area shapes, the larger perimeter causes a faster decay rate of the short-term survival function because of a minus sign in front of the second term in the asymptotic series.

However, unlike simulating LRWs inside of the target region, i.e. the annulus in Appendix B, particles are distributed uniformly outside of the target shape filled with white pixels and undergo LRWs in the unoccupied space filled with black pixels in the $2-$dimensional image. In other words, the geometric attributes and spatial patterns of the target shapes are reflected and implied through exploring the unoccupied space by particles in the image. Therefore, we are interested in acknowledging whether the short-term behaviours of survival functions of the numerical simulation still coincide with the analytical results. Moreover, it is vital to assess the difference between survival functions to test whether they are reliable to compare and distinguish the shapes.

In this chapter, the preliminary step of testing the hypothesis is to generate black-and-white images with the same dimensions as shown in Fig. C.1, Fig. C.2, and Fig. C.3. In the binary images, the objects are convex shapes with the same area or an equal number of white pixels. The next step is to simulate LRWs and estimate survival functions by Kaplan-Meier estimator [30] implemented in lifelines library in Python [**?**]. Then, some commonly applied nonparametric statistical tests, as proposed in section 2.3, are utilized to test the equality of survival functions. Except for investigating how survival functions differ from each other by statistical tests and dissimilarities calculation, and it is meaningful and helpful to reveal why they are distinct by studying the underlying stochastic process.

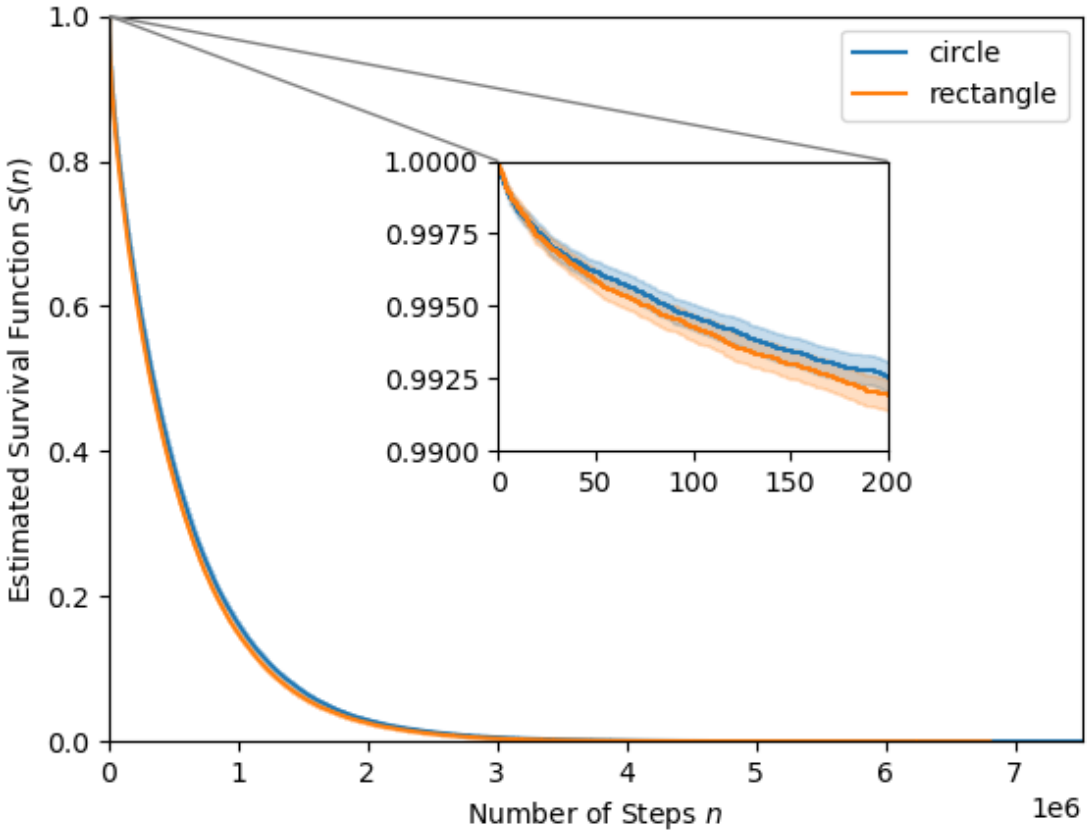


Figure 3.1: Since the survival curves for circle and rectangle overlap and cross, it is hard to detect the difference between them by eyes. However, the inset figure shows the greater details about the survival probabilities within the small number of steps. The area around the survival curve with the pale colour represents 95% confidence interval.

3.1 Circle and Rectangle Yuge

For simplicity, the centroid of shape in Fig. C.1 is located at the center of the image, although periodic boundary conditions can eliminate the boundary and translation effects. Circle and rectangle are regular convex shapes without sophisticated geometric features. Moreover, it is well-known that the rectangle perimeter is longer than the circle circumference when their areas are identical. Therefore, in theory, survival probabilities at short times, i.e. number of steps n, of particles in the image with a rectangle are smaller than a circle.

Fig. 3.1 shows the Kaplan-Meier estimates and the 95% confidence intervals for the survival functions, $S(n)$, of circle and rectangle. The approximated 95% confidence intervals of the survival functions overlay results from the sampling errors. The survival function decays monotonically from 1 to 0. In LRWs, particles' initial positions are distributed uniformly in the unoccupied space in the image, so none of them can be absorbed by the boundary of the target shape. Also, all the particles will finally stop the random walk and

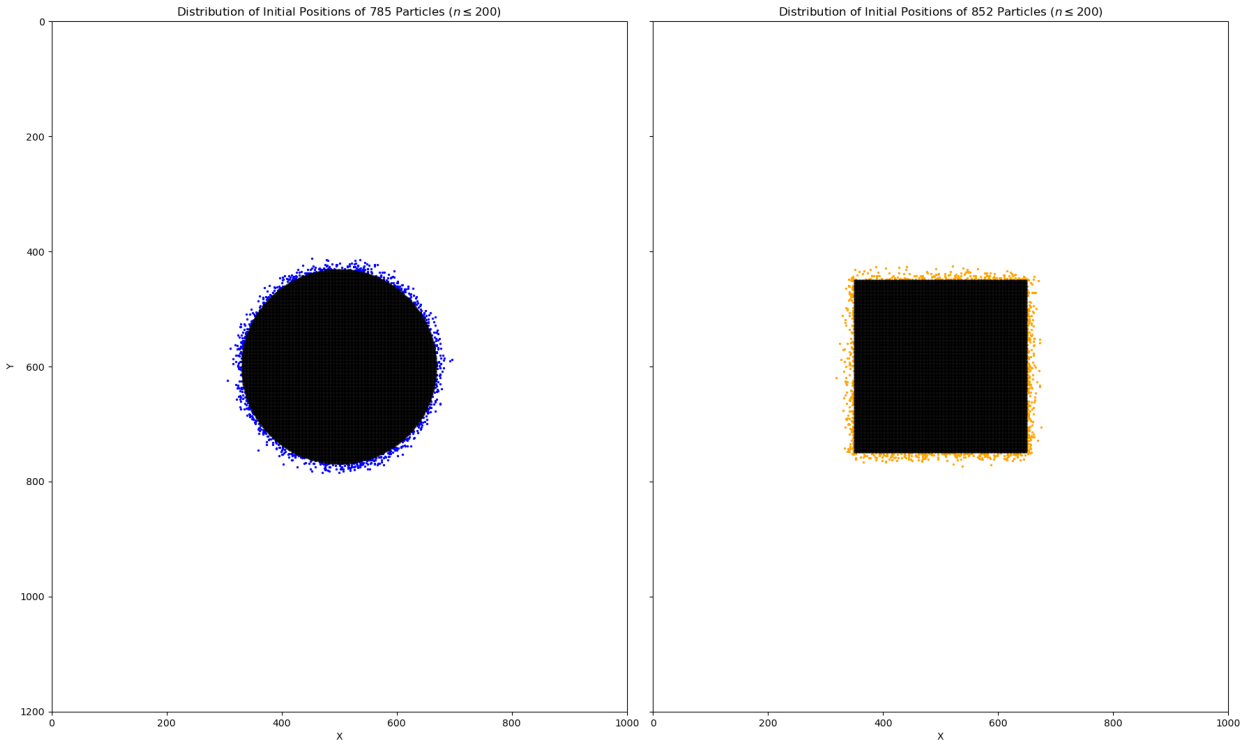


Figure 3.2: The coloured points in the left and right subfigure represent the initial positions of particles in LRWs, whose number of steps less than or equal to 200.

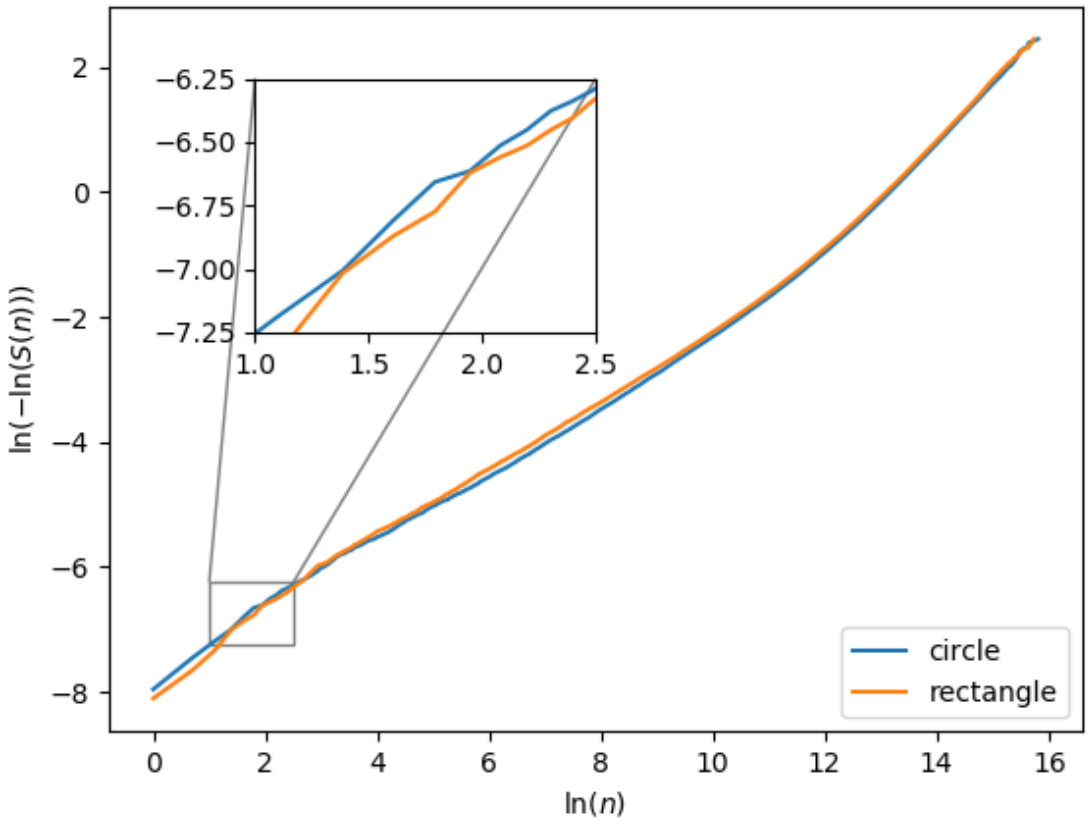


Figure 3.3: It is a graphical method for checking proportionality by looking for parallelism. The lines for circle and rectangle are not parallel since they cross at some points as shown in the inset, and the difference between them changes with logarithmically transformed n.

can not survive because of the absorbing boundary condition.

Overall, the differences between survival functions for the circle and rectangle are not visible. Nevertheless, an inset in Fig. 3.1 helps visually identify distinct short-term behaviours of survival functions for circle and rectangle, which are consistent with the theoretical results. From the concentrated spatial distribution of points illustrated in Fig. 3.2, we can find that if particles start LRWs close to the target object in the image, they will hit the absorbing boundary within a few steps. Alternatively, the short-term survival function describes a region in the shape of an approximate belt next to the target object. Furthermore, the point cloud around the rectangle comprises more particles than the circle resulting from the longer perimeter.

After the short-term study, the entire survival curves of circle and rectangle needed to be compared to validate LRWs further. As introduced in section 2.3, some conventional nonparametric tests in survival analysis assess the null hypothesis that there is no difference in survival probabilities at all number of steps n. Among those tests, the log-rank test results are reliable if the assumption of proportional hazards is satisfied, which can be inspected by looking at the graph Fig. 3.3 showing $\ln(-\ln S(n))$ against $\ln(n)$.

	test_statistic	p
log-rank	137.23	0.0
Tarone-Ware	134.31	0.0
Gehan-Breslow	123.83	0.0
Fleming-Harrington	123.83	0.0

Table 3.1: Survival functions for circle and rectangle are statistically different since p values equal zeros.

The crossing lines with varying shapes in Fig. 3.3 suggest the rejection of the proportional hazards assumption, and the log-rank test may loss power. However, in Tab. 3.1, log-rank, Tarone-Ware, Gehan-Breslow, and Fleming-Harrington tests demonstrate the rejection of the null hypothesis. In other words, there is a statistically significant difference between the survival functions for circle and rectangle. In conclusion, LRWs is an alternative methodology to quantify and distinguish the regular simply connected regions in the $2-$ dimensional image by developing and comparing the corresponding survival functions.

3.2 Complicated Branching Structures

In the preceding section, LRWs are simulated in the 2−dimensional images of equal-area simply connected domains (i.e. without holes in them) with smooth or piecewise smooth boundary. Both short and long-term behaviours of survival functions for the circle and rectangle are analyzed and compared to verify LRWs in the application of shape comparison. However, the plant root shape is an extremely complicated branching structure, consisting of the primary roots, taproots, secondary roots, etc.

Hence, this section is outlined as follows. First of all, some branching structures, as shown in Fig. C.2 and Fig. C.3, are designed and generated in equal size 2−dimensional images. Secondly, without the simulation of LRWs, the spatial pattern and structural features of the branching structures are discussed by the interpretation of the survival function $S(R)$, where R is the nearest distance from an arbitrary point in the space to the fixed boundary of the target object. Thirdly, the other two survival functions, $S(n)$ and $S(d)$, of the branching structures are estimated and compared, where n and d are the numbers of steps taken by the particle from the initial to the stop position and its corresponding displacement in the tiling space, respectively. However, three types of survival functions and their interpretation of the underlying stochastic process in this section are intimately related.

(a) **(b)**

Figure 3.4: (a) and (b) are branch templates with identical area for generating artificial branching structures in G_1 and G_2, respectively. Templates in (a) and (b) have same maximum width (i.e. 150), but different maximum depth (i.e. 1050 and 1200, respectively).

3.2.1 Artificial Branching Structure Construction

The plant root shape can be represented as a spatial structural system consisting of separate branches, and each of them forks out at a specific point, or knot, into at least two branches. Moreover, such branching structure can constantly subjected to tension, compression, flexion, bending, and torsion. In this section, as shown in Fig. C.2 and Fig. C.3, the design of two groups of artificial images builds upon and extends the definition of the perfect binary tree structure [35] in graph theory [42]. A binary tree is a data structure consisting of nodes, each of which has up to two childer named left and right child nodes. It starts with a single topmost node called the root. A perfect binary tree is a specific type of binary tree in which all the interval nodes are connected with two child nodes, and all leaves are at the same level. The root is at level 0, its child nodes are at level 1, and so on.

As shown in Fig. 3.4, the branch templates for G_1 and G_2 are slightly distinct in shapes, connection, the thickness of the left and right child branches, and aspect ratio. According to the concept of perfect binary tree structures, the template can be decomposed into three branches, one main or parent branch and two child branches. The main branch is connected with a left and right child branch equidistantly and symmetrically without voids and overlays. The equal-area artificial branching structures in Fig. C.2 and Fig. C.3 are generated based on the principle of fractal geometry [15] by repeating the branch template at an increasingly smaller scale from top to bottom. The following algorithm summarizes how to construct a

binary branching structure:

Algorithm 1: Generate Artifical Branching Structure

Result: 2−dimensional Artifical Branching Image with Area A

Set number of levels of the branching structure j;

Given a branch template $\widehat{T}$ with a predefined surface area A;

Generate an all-black image Img;

for $l = 0;\ l < j;\ l = l + 1$ **do**

  define l−th level branch template as $\widehat{T}_{j,l}$ by multiplying $\widehat{T}$ with a scaling factor $s_{j,l}$;

  if $l == 0$ **then**

    draw a branch template $\widehat{T}_{j,l}$ in Img;

    find the bottom left and right convex corners of $\widehat{T}_{j,l}$;

  else

    draw and put 2^l branch templates, $\widehat{T}_{j,l}$, at every convex corner;

    find the bottom left and right convex corners of all $\widehat{T}_{j,l}$;

    take a union of previous Img with the collection of smaller-scale branches and redefine Img;

  end

end

if *number of white pixels* $N\ != A$ **then**

  add $A - N$ white pixels around branch templates in Img

else

end

Those artificial structures in Fig. C.2 and Fig. C.3 are iteratively-defined geometries as described in the algorithm **??**, and their complexities grow with the number of hierarchical branching or levels j, where $j = 3, 4, 5, 6$. The large-scale or low-level branches mimic thick roots near the soil surface. In contrast, the smaller-scale branches at the higher level represent the fine roots that gradually become vertical in deeper horizons because of the positive response of gravity in soil.

From the perspective of mathematics, in each image, the region occupied by white pixels and bounded by a piecewise smooth curve is a 2−dimensional simply connected domain. As similar to the methodology validation in the last section, we will compare the short-term survival function behaviours and assess their discrepancy by statistical results. However, the primary purpose is to interpret the survival curves and reveal the spatial pattern information.

		p			
		Logrank	TW	GB	FH
$G_1\ L_3$	$G_1\ L_4$	0.0	0.0	0.0	0.0
	$G_1\ L_5$	0.0	0.0	0.0	0.0
	$G_1\ L_6$	0.0	0.0	0.0	0.0
$G_1\ L_4$	$G_1\ L_5$	0.1773	0.0	0.0	0.0
	$G_1\ L_6$	0.0	0.0	0.0	0.0
$G_1\ L_5$	$G_1\ L_6$	0.0	0.0	0.0	0.0

Table 3.2

		p			
		Logrank	TW	GB	FH
$G_2\ L_3$	$G_2\ L_4$	0.0	0.0	0.0	0.0
	$G_2\ L_5$	0.0	0.0	0.0	0.0
	$G_2\ L_6$	0.0	0.0	0.0	0.0
$G_2\ L_4$	$G_2\ L_5$	0.0	0.0	0.0	0.0
	$G_2\ L_6$	0.0	0.0	0.0	0.0
$G_2\ L_5$	$G_2\ L_6$	0.0	0.0	0.0253	0.0253

Table 3.3

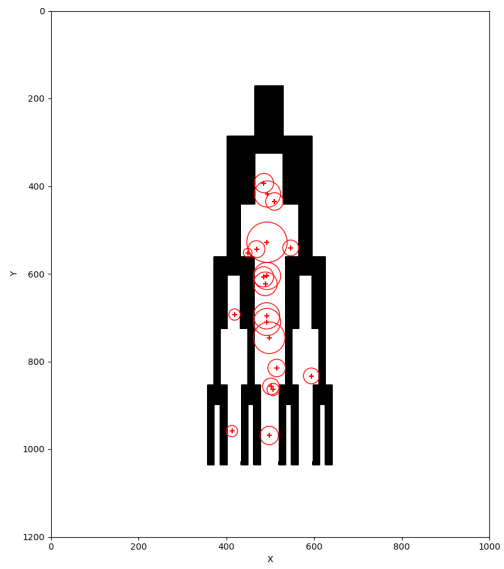
3.2.2 Output Analysis of $S(R)$

3.2.2.1 Interpretation of Survival Curves

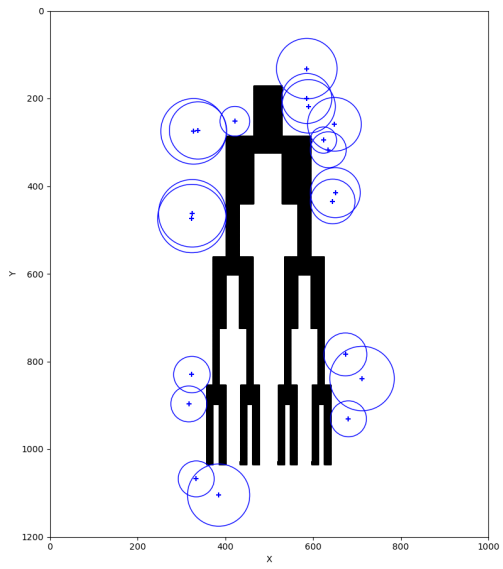
3.2.2.2 Comparison of Survival Curves

3.2.2.3 Distance Matrices

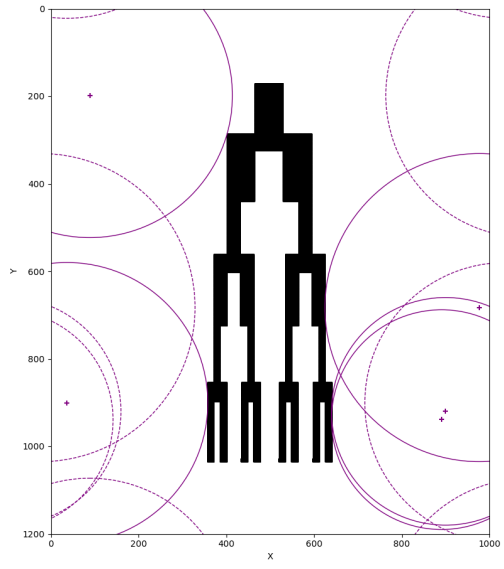
3.2.2.4 Multidimenisonal Scaling



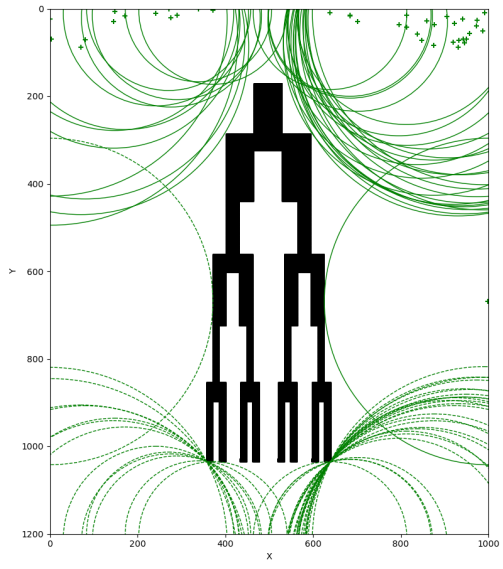
(a)



(b)

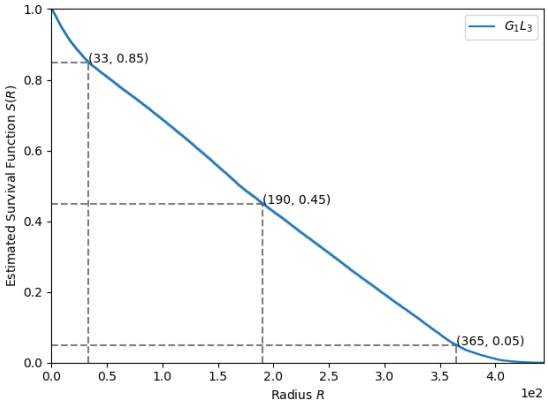


(c)

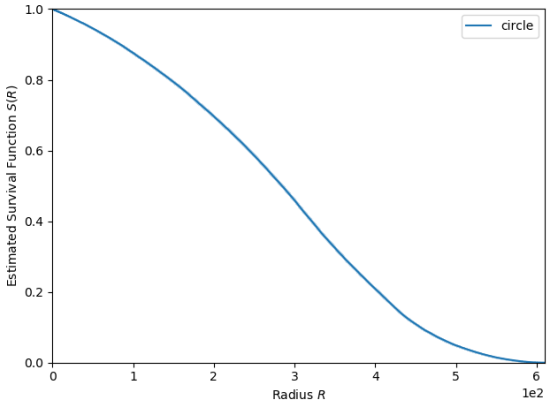


(d)

Figure 3.5: This figure shows four types of the largest circles centred at some colourful plus symbols, +, which are initial positions of particles in LRWs. The radius of a circle is the shortest distance from the initial site to the boundary of the branching structure.

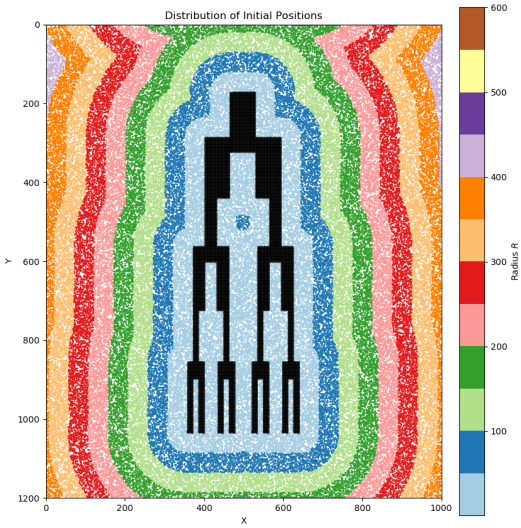


(a)

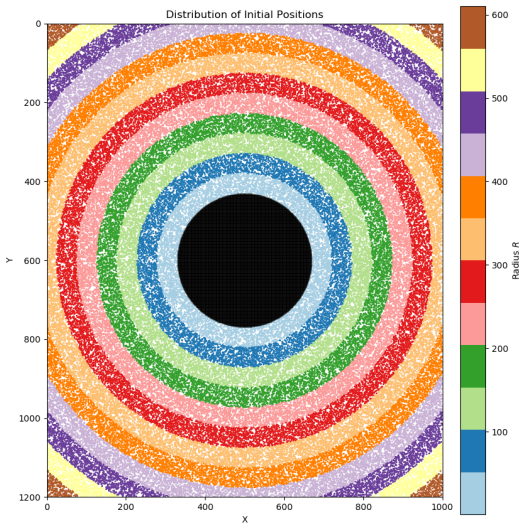


(b)

Figure 3.6

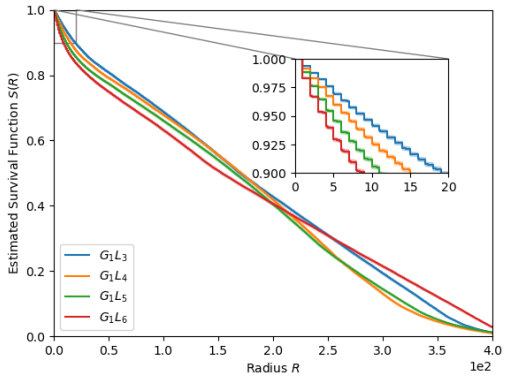


(a)

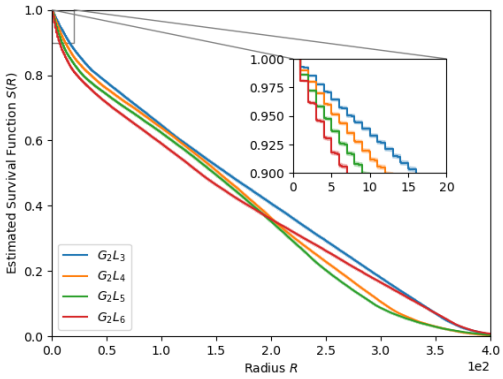


(b)

Figure 3.7

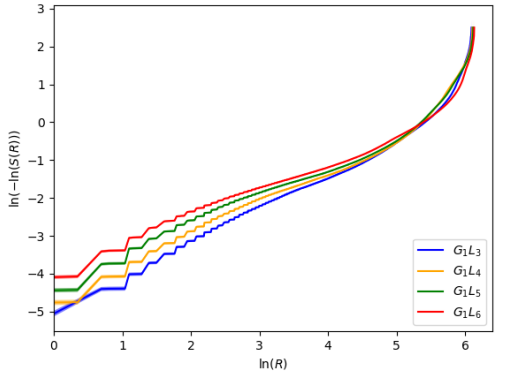


(a)

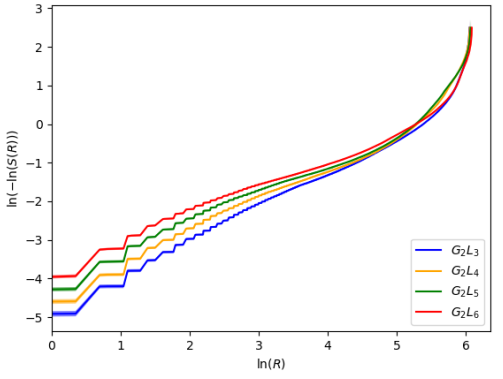


(b)

Figure 3.8



(a)



(b)

Figure 3.9

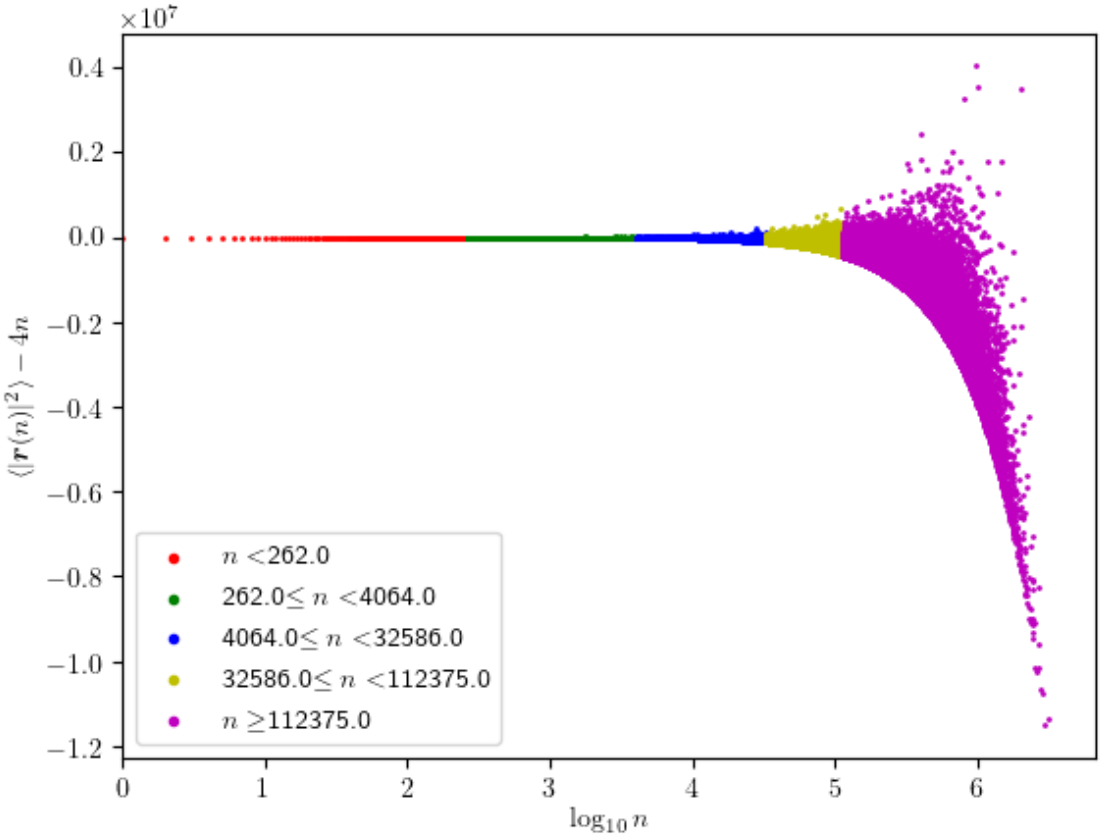


Figure 3.10: Particles in G_1L_3 are divided into several subgroups based on various intervals of steps, and their colours are as identical as the segments in Fig. 3.12a.

3.3 Output Analysis of $S(n)$ and $S(d)$

3.3.1 Relationship Between n and d

In this section, the displacement of a particle, d, is the shortest distance from the initial to the stop position in the infinite tiling plane. In theory, the mean square displacement (MSD) of N Brownian particles at $n-$th step in $2-$dimensional space is defined as

$$MSD = \langle |\boldsymbol{r}(n)|^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{s}_i(n) - \boldsymbol{s}_i(0))^2 = 4Dn \tag{3.1}$$

where the subscript, i, refers to each particle for which the MSD is calculated. $\boldsymbol{s}_i(n)$ and $\boldsymbol{s}_i(0)$ are the $i-$th particle positions at $n-$th step and at the initial time, respectively. D is diffusion coefficient which is related to the variance of the independent displacements of the particle. In the simulation, D equals 1.

Eq. 3.1 indicates a linear relationship between the mean square displacement of the particle and the

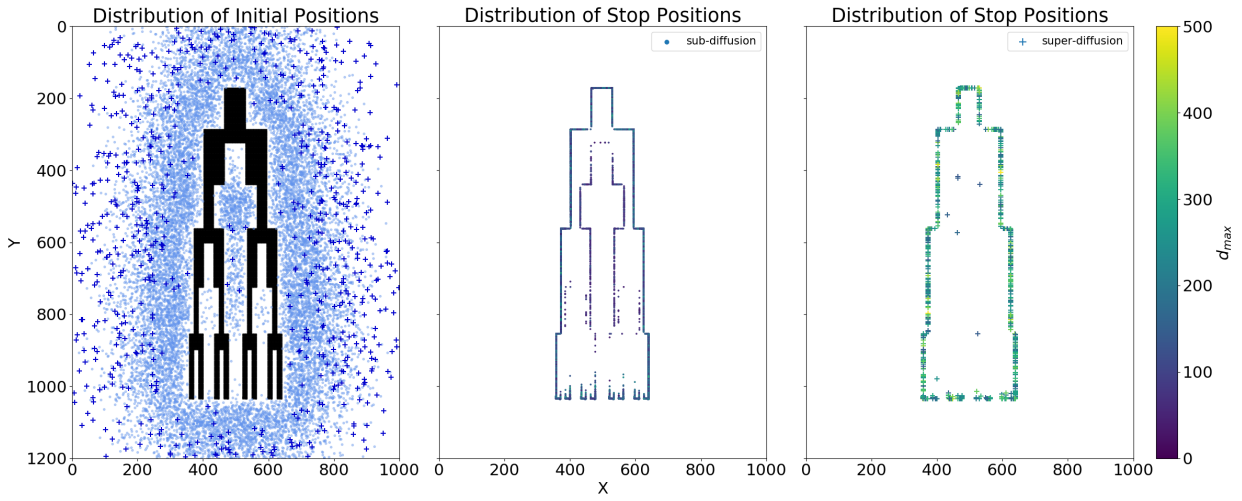


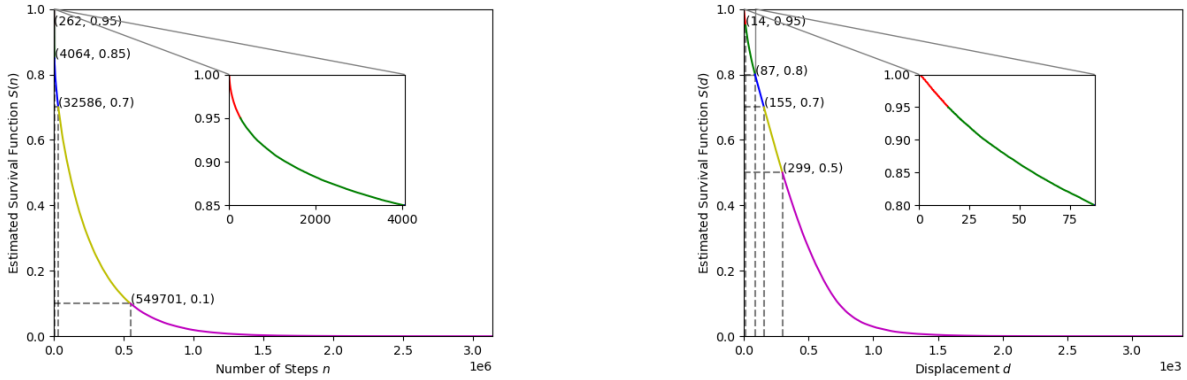
Figure 3.11: In the left figure, 696 dark blue pluses refer to super-diffusion particles, which are distributed closer and more concentrated to the fringe of the branching structure. 15198 pale blue points represent sub-diffusion particles and scatter mainly around the edges of the image. In-between the branches, there has only six super-diffusion particles and an enormous amount of sub-diffusion ones. The middle and right depict the stop positions for super and sub-diffusion particles, respectively, coloured by a perceptually uniform sequential colormap based on their maximum displacement in the tiling space.

number of steps. It is a feature of the normal diffusive behavior. Fig. 3.10 shows how the difference between MSD of the particle and $4n$ varying over $\log_{10} n$ in LRWs. When $n \leq 4064.0$, the variation is not equal to 0 with larger fluctuation, which implies that blue, yellow, and pink particles undergo anomalous diffusion process. In other words,

$$\langle |\boldsymbol{r}(n)|^2 \rangle \propto n^{\gamma} \tag{3.2}$$

where $\gamma \neq 1$. In Fig. 3.12a, negative variation implies $\gamma < 1$ called sub-diffusion process, while positive value denotes $\gamma > 1$ named super-diffusion.

To understand the underlying mechanism of the anomalous-type diffusion process, initial and stop positions of particles in Fig. **??**, whose steps ranged from 4064 to 32586, are illustrated in Fig. 3.11. Suppose particles are trapped in the narrow space in-between the branches or initially start LRWs near the branching structure. In that case, their movements will be restricted because of the nearby absorbing boundary condition, causing the sub-diffusion phenomenon. As shown in the middle subplot of Fig. 3.11, the maximum displacement of sub-diffusion particles is less than 200. There also have exceptional circumstances that 6 particles, in-between the limbs, undergo super-diffusion since they can explore a large portion of space within a predefined range of steps, as shown in the right subfigure of Fig. 3.11. Generally, particles around the edges of the image will be more likely to pass through the periodic boundary, reappear in the adjacent cell, and continue LRWs with the same velocity until hitting the absorbing boundary, which results in large displacement and the super-diffusion process.



(a) **(b)**

Figure 3.12

		p			
		Log-rank	TW	GB	FH
G_1 L_3	G_1 L_4	0.4393	0.0285	0.0005	0.0005
	G_1 L_5	0.0	0.0	0.0	0.0
	G_1 L_6	0.0	0.0	0.0	0.0
G_1 L_4	G_1 L_5	0.0007	0.0	0.0	0.0
	G_1 L_6	0.0002	0.0	0.0	0.0
G_1 L_5	G_1 L_6	0.7223	0.0	0.0	0.0

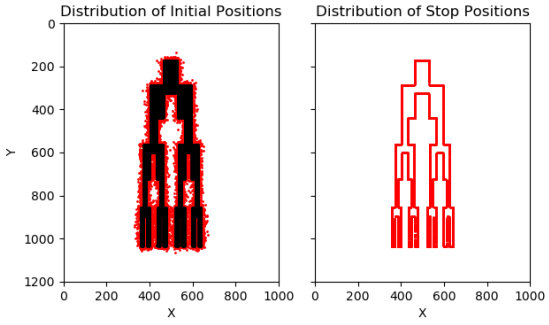
Table 3.4: Two Sample Statistical Tests for $S(n)$ of G_1

3.3.2 Interpretation of Survival Curves

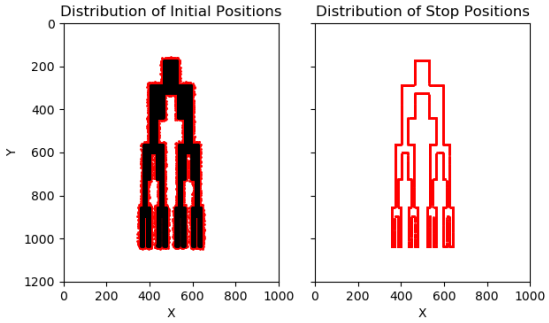
3.3.3 Comparison of Survival Curves

3.3.3.1 Distance Matrices

3.3.3.2 Multidimenisonal Scaling



(a)

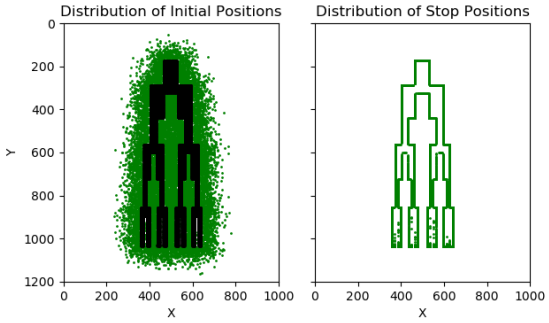


(b)

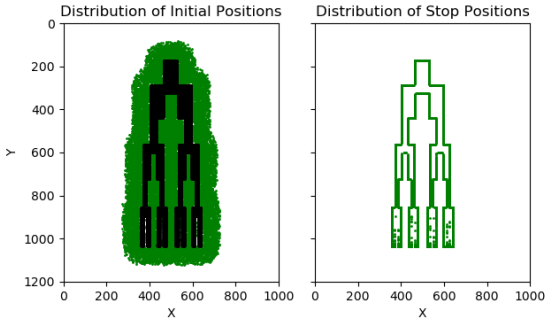
Figure 3.13

		p			
		Log-rank	TW	GB	FH
G_2 L_3	G_2 L_4	0.0	0.0	0.0	0.0
	G_2 L_5	0.0	0.0	0.0	0.0
	G_2 L_6	0.0	0.0	0.0	0.0
G_2 L_4	G_2 L_5	0.0016	0.0	0.0	0.0
	G_2 L_6	0.0004	0.0	0.0	0.0
G_2 L_5	G_2 L_6	0.7199	0.0	0.0	0.0

Table 3.5: Two Sample Statistical Tests for $S(n)$ of G_2



(a)

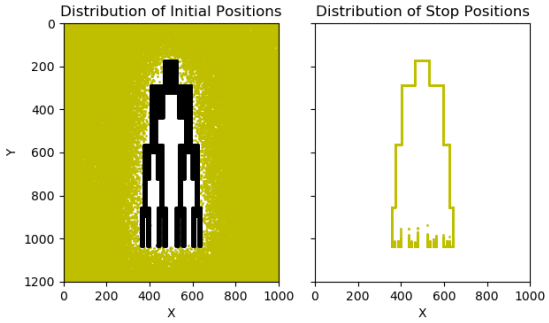


(b)

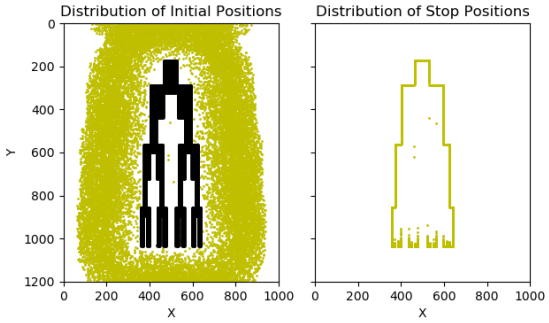
Figure 3.14

		p			
		Logrank	TW	GB	FH
G_1 L_3	G_1 L_4	0.0	0.0	0.0	0.0
	G_1 L_5	0.0	0.0	0.0	0.0
	G_1 L_6	0.0	0.0	0.0	0.0
G_1 L_4	G_1 L_5	0.0072	0.0	0.0	0.0
	G_1 L_6	0.0003	0.0	0.0	0.0
G_1 L_5	G_1 L_6	0.2883	0.0	0.0	0.0

Table 3.6: Two Sample Statistical Tests for $S(d)$ of G_1



(a)

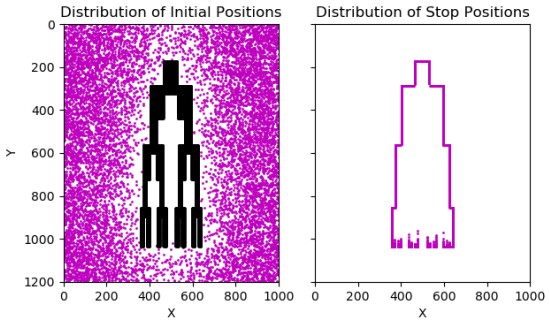


(b)

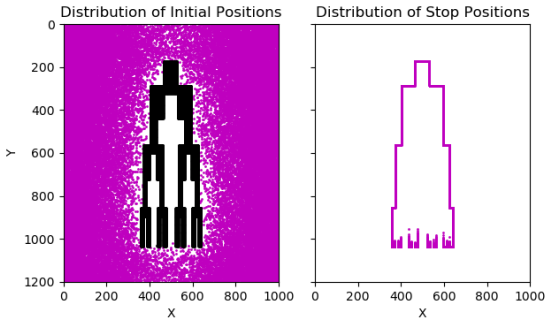
Figure 3.15

		p			
		Logrank	TW	GB	FH
G_2 L_3	G_2 L_4	0.0	0.0	0.0	0.0
	G_2 L_5	0.0	0.0	0.0	0.0
	G_2 L_6	0.0	0.0	0.0	0.0
G_2 L_4	G_2 L_5	0.0001	0.0	0.0	0.0
	G_2 L_6	0.0015	0.0	0.0	0.0
G_2 L_5	G_2 L_6	0.7019	0.0	0.0	0.0

Table 3.7: Two Sample Statistical Tests for $S(d)$ of G_2

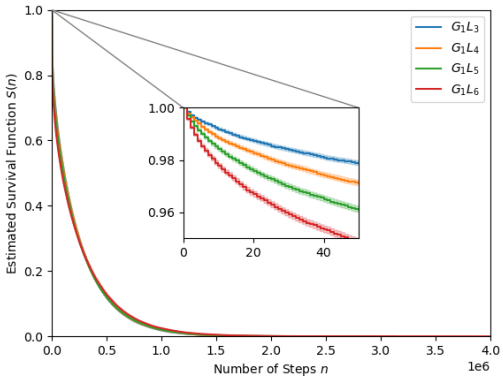


(a)

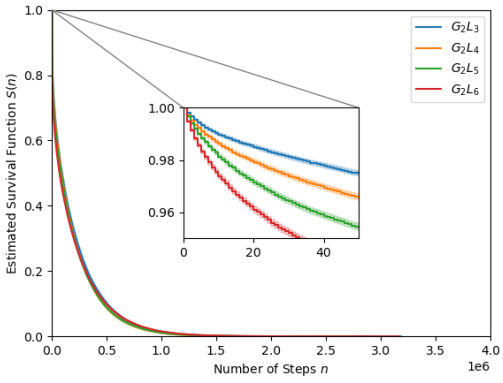


(b)

Figure 3.16

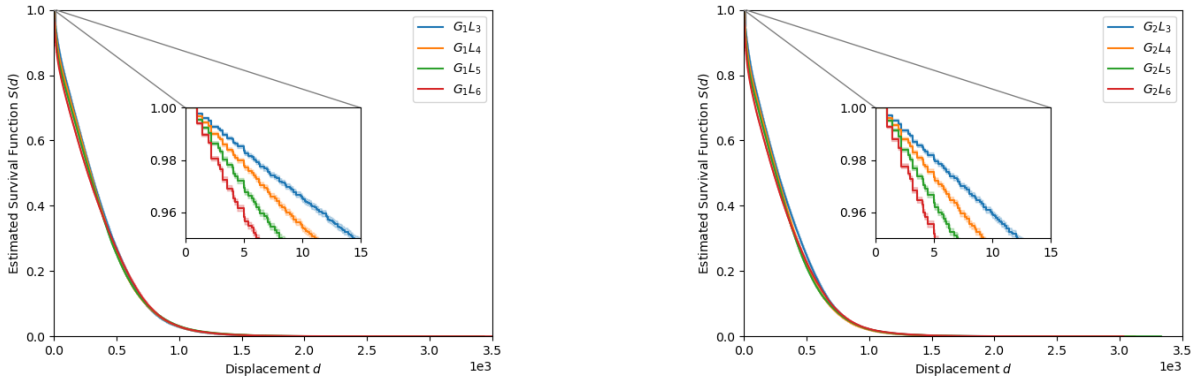


(a)



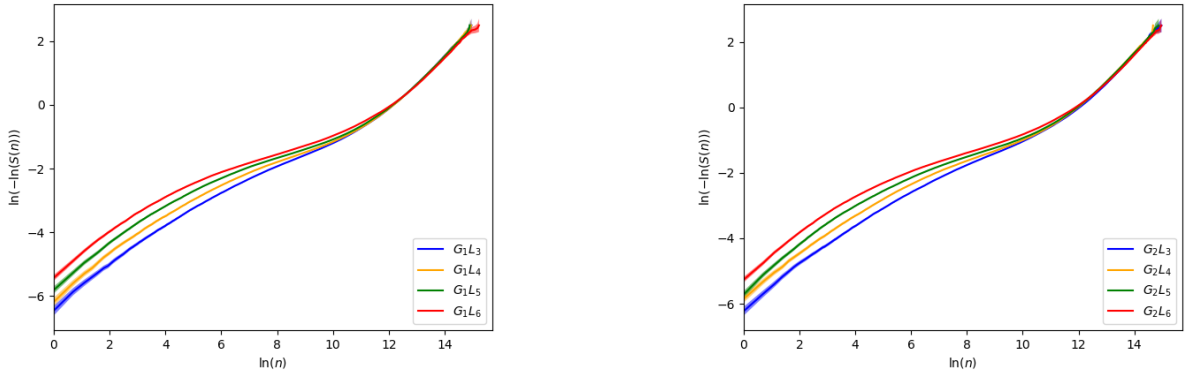
(b)

Figure 3.17: (a) and (b) are survival functions for branching structures in G_1 and G_2, respectively. n is the number of steps taken by the particle from the initial to the stop pixel in LRWs.



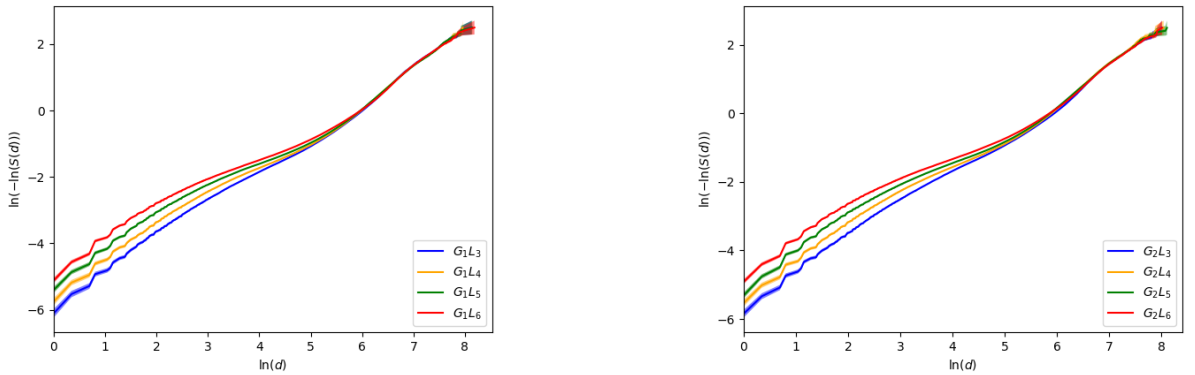
(a) **(b)**

Figure 3.18: (a) and (b) are the estimated survival functions associated with particles' displacement in LRWs in G_1 and G_2, respectively.



(a) **(b)**

Figure 3.19: (a) and (b) are commonly used graphical techniques to check the proportional hazards (PH) assumption of survival data by finding the parallelism. The survival distributions do not support the PH assumption since the hazard ratio in both G_1 and G_2 is not always constant.



(a) **(b)**

Figure 3.20

3.4 Conclusion

- In a short time, the survival function of rectangle decays faster than the circle, which conforms to the analytical results.
- The differences of estimated survival functions between circle and rectangle are statistically significant, which coincides with the real shape dissimilarities.
- Within a same group, when t is small, the more branching the object is, the faster the survival function decays.
- Within a same group, the pairwise survival functions are statistically different.
- The corresponding target structures in G_1 and G_3 are invariant shapes under translation since their survival function are not statistically different. In other words, periodic boundary conditions of the image can eliminate the effect of the locations.
- LRWs can describe and classify the geometries, their spatial configurations, and the unoccupied area in the image.

LRWs in Real Root Images

Conclusion

Future Work

Appendix A

Numerical Methods for Solving Parabolic Partial Differential Equations

A.1 Introduction

- Parabolic PDEs: to characterize time-dependent phenomena
- The intrinsically similar features of the traditional computational techniques are mesh discretization in time and space.

A.2 Summary of Commonly Used Numerical Techniques

A.2.1 Finite Difference Method (FDM) [22]

A.2.2 Finite Element Method (FEM) [45]

A.2.3 Other Tranditional Computational Methods

A.3 Limitation in Practice

Appendix B

Method Validation in Annulus

B.1 Analytical Results

B.1.1 Shape Description

- Problem domain Ω: the region bounded by two concentric circles
- Radius of the larger circle: b
- Radius of the smaller circle: a

B.1.2 Solving Initial-Boundary Value Problem (IBVP)

B.1.2.1 Methods

- Dimensional Analysis: non-dimensional variables
  - $\mu = \frac{b}{a}$
  - $\tau = \frac{t}{a^2}$
  - $\hat{r} = \frac{r}{a}$
- Method of separation of variables

B.1.2.2 Mathematical Equations

- Diffusion equation

$$u_\tau = (u_{\hat{r}\hat{r}} + \frac{1}{\hat{r}} u_{\hat{r}} + \frac{1}{\hat{r}^2} u_{\theta\theta}) \tag{B.1}$$

- Uniform initial condition

$$u(\hat{r}, \theta, 0) = \frac{1}{|\Omega|} \tag{B.2}$$

- Homogenous Dirichlet B.C.

$$u(1, \theta, \tau) = 0 \tag{B.3}$$

- Homogenous Neumann B.C.

$$\hat{r} u'(\mu, \theta, \tau) = 0 \tag{B.4}$$

B.1.2.3 Heat Content Calculation

$$S(\tau) = \int_0^{2\pi} d\theta \int_1^{\mu} \hat{r} d\hat{r} u(\hat{r}, \theta, \tau) \tag{B.5}$$

B.2 Numerical Approximation

B.2.1 Eigenvalues $\lambda_{0,n}$

- Properties
  - $\lambda_{0,n} \in \mathbb{R}^+$, $(n \in \mathbb{N}_+)$
  - Monotonicity and Periodicity
- Estimation
  - $\lambda_{0,n} \in ((n-1)\pi, (n+1)\pi)$ [12]
  - Bisection method [41]

B.2.2 Approximation of $u(\hat{r}, \theta, \tau)$ and $S(\tau)$

- Direct summation
- Series acceleration methods

B.3 Comparison of Numerical and Analytical Results

B.3.1 Sample Size Evaluation

B.3.2 Comparison of $S(\tau)$ and $S(n)$

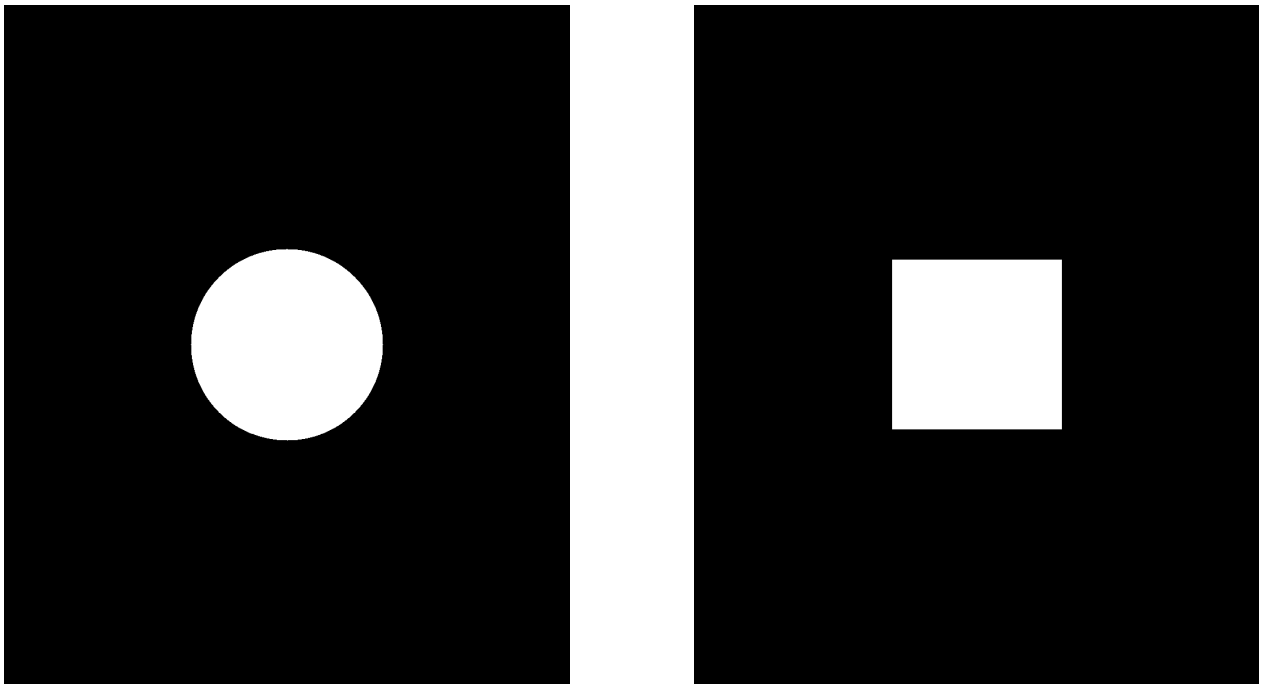
B.4 Conclusion

- The estimated survival function of LRWs is consistent with the analytical result.
- The number of particles in LRWs determined by DKW inequality is large enough to generate reproducible statistical results.

Appendix C

Artificial Images

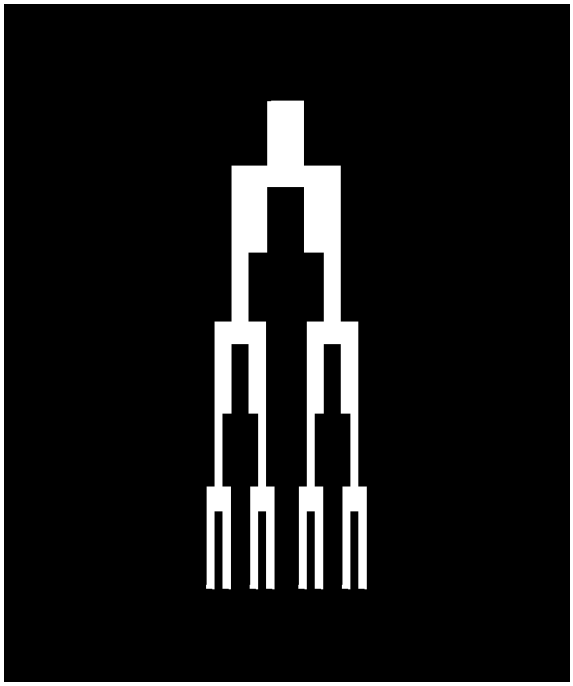
C.1 Simple Shapes



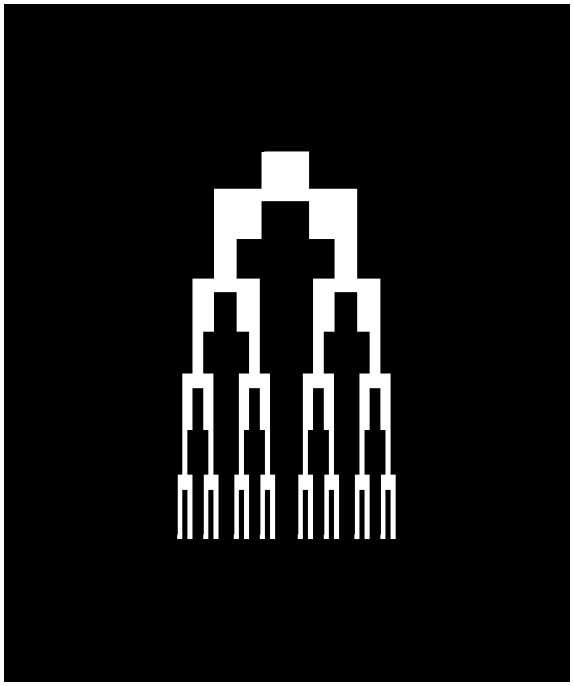
(a) Circle **(b)** Rectangle

Figure C.1: Each image size is 1200 by 1000 pixels with 90000 white pixels.

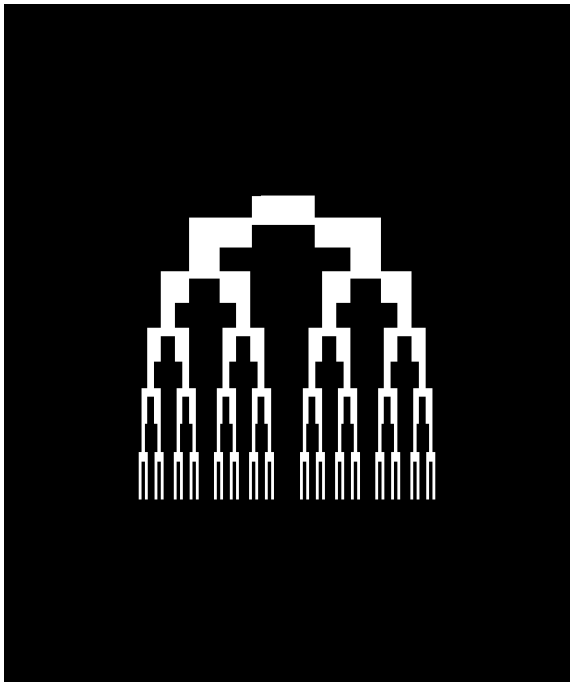
C.2 Complicated Branching Structures



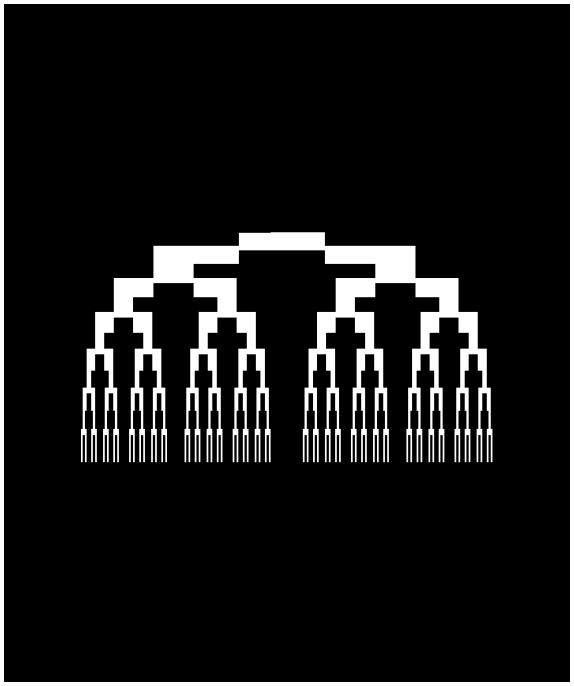
(a) G_1L_3



(b) G_1L_4

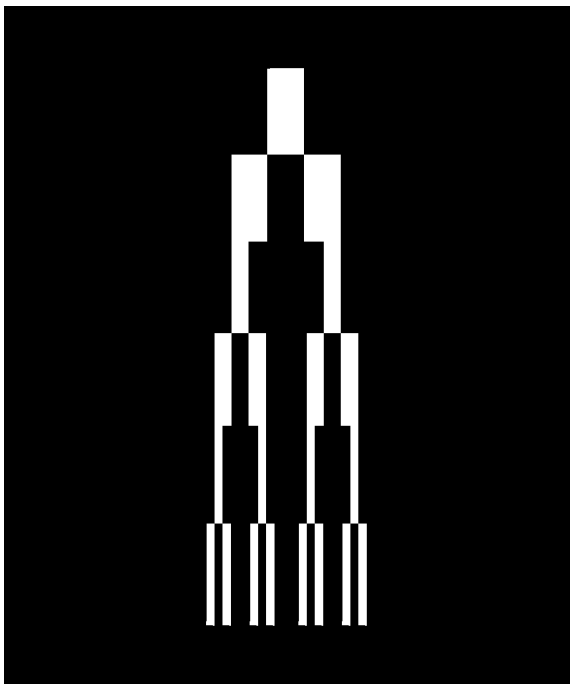


(c) G_1L_5

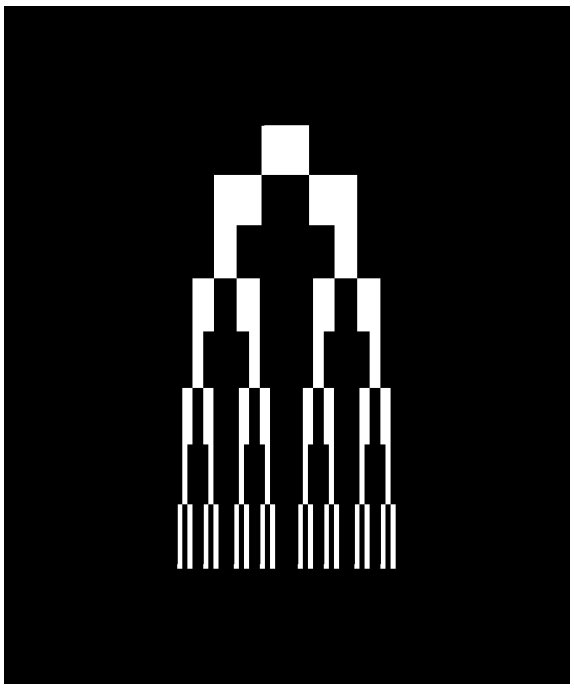


(d) G_1L_6

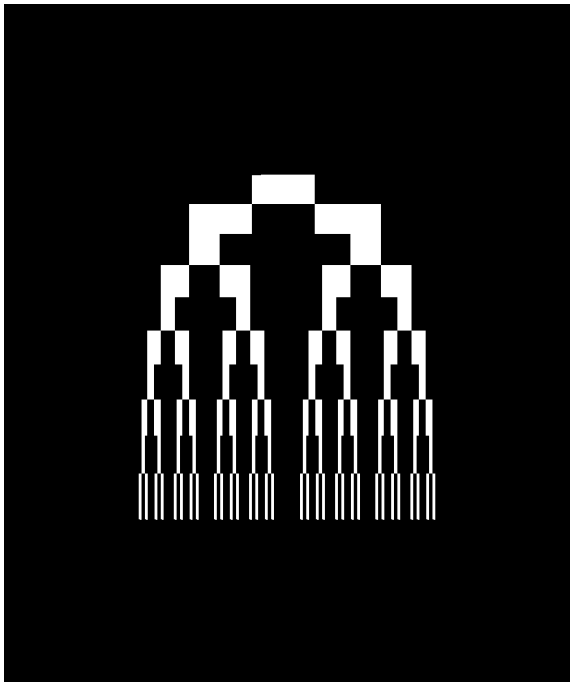
Figure C.2: In the group one, each image size is 1200 by 1000 pixels with 90000 white pixels.



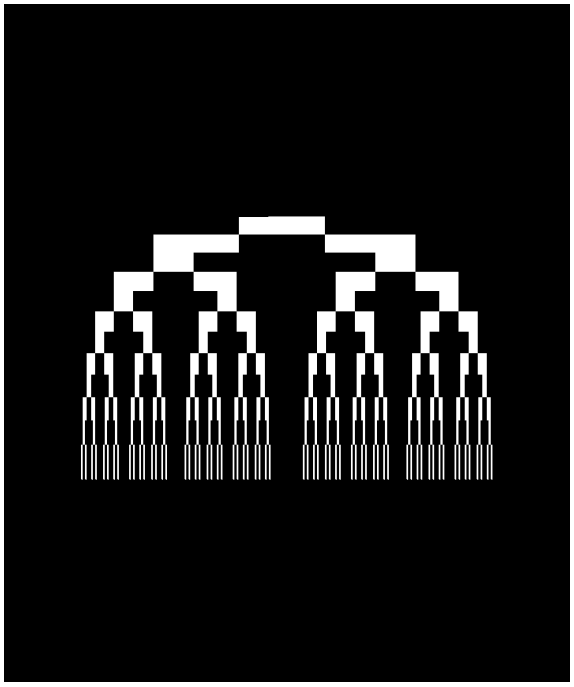
(a) G_2L_3



(b) G_2L_4



(c) G_2L_5



(d) G_2L_6

Figure C.3: In the group two, each image size is 1200 by 1000 pixels with 90000 white pixels.

References

[1] Odd Aalen, Ornulf Borgan, and Hakon Gjessing. *Survival and Event History Analysis: A Process Point of View*. Springer Science & Business Media, 2008.

[2] Girdhar Gopal Agarwal. Statistics for surgeons–understanding survival analysis. *Indian journal of surgical oncology*, 3(3):208–214, 2012.

[3] Panagiotis Ch Anastasopoulos, Grigorios Fountas, Md Tawfiq Sarwar, Matthew G Karlaftis, and Adel W Sadek. Transport habits of travelers using new energy type modes: a random parameters hazard-based approach of travel distance. *Transportation Research Part C: Emerging Technologies*, 77:516–528, 2017.

[4] Adrian Baddeley, Imre Bárány, and Rolf Schneider. Spatial point processes and their applications. *Stochastic Geometry: Lectures Given at the CIME Summer School Held in Martina Franca, Italy, September 13–18, 2004*, pages 1–75, 2007.

[5] Mathilde Balduzzi, Brad M Binder, Alexander Bucksch, Cynthia Chang, Lilan Hong, Anjali S Iyer-Pascuzzi, Christophe Pradal, and Erin E Sparks. Reshaping plant biology: qualitative and quantitative descriptors for plant morphology. *Frontiers in Plant Science*, 8:117, 2017.

[6] Jean Baptiste Joseph Baron Fourier. *The analytical theory of heat*. The University Press, 1878.

[7] Taane G Clark, Michael J Bradburn, Sharon B Love, and Douglas G Altman. Survival analysis part i: basic concepts and first analyses. *British journal of cancer*, 89(2):232–238, 2003.

[8] Ruvie Lou Maria Custodio Martinez. Diagnostics for choosing between log-rank and wilcoxon tests. 2007.

[9] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.

[10] Benjamin M Delory, Mao Li, Christopher N Topp, and Guillaume Lobet. archidart v3. 0: A new data analysis pipeline allowing the topological analysis of plant root systems. *F1000Research*, 7, 2018.

[11] S Desjardins and P Gilkey. Heat content asymptotics for operators of laplace type with neumann boundary conditions. *Mathematische Zeitschrift*, 215(1):251–268, 1994.

[12] *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.26 of 2020-03-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[13] Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pages 642–669, 1956.

[14] İ Etikan, S Abubakar, and R Alkassim. The kaplan-meier estimate in survival analysis. *Biom Biostatistics Int J*, 5(2):00128, 2017.

[15] Kenneth Falconer. *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons, 2004.

[16] AH Fitter. The topology and geometry of plant root systems: influence of watering rate on root system topology in trifolium pratense. *Annals of Botany*, 58(1):91–101, 1986.

[17] AH Fitter and TR Stickland. Fractal characterization of root system architecture. *Functional Ecology*, pages 632–635, 1992.

[18] HPW Gottlieb. Hearing the shape of an annular drum. *The ANZIAM Journal*, 24(4):435–438, 1983.

[19] HPW Gottlieb. Eigenvalues of the laplacian with neumann boundary conditions. *The ANZIAM Journal*, 26(3):293–309, 1985.

[20] M Greenwood. The natural duration of cancer. london: His majesty's stationery office; 1926. *Reports on public health and medical subjects*, (33).

[21] Daniel Grieser and Svenja Maronna. Hearing the shape of a triangle. *Notices of the American Mathematical Society*, 60(11):1440–1447, 2013.

[22] Christian Grossmann, Hans-Görg Roos, and Martin Stynes. *Numerical treatment of partial differential equations*, volume 154. Springer, 2007.

[23] John M Hammersley. Monte carlo methods for solving multivariable problems. *Annals of the New York Academy of Sciences*, 86(3):844–874, 1960.

[24] David P Harrington and Thomas R Fleming. A class of rank test procedures for censored survival data. *Biometrika*, 69(3):553–566, 1982.

[25] Jiahang He, Toshiyuki Yamamoto, Tomio Miwa, and Takayuki Morikawa. Hazard duration model with panel data for daily car travel distance: A toyota city case study. *Sustainability*, 12(16), 2020.

[26] David W Hosmer Jr, Stanley Lemeshow, and Susanne May. *Applied survival analysis: regression modeling of time-to-event data*, volume 618. John Wiley & Sons, 2011.

[27] Kurt Jacobs. *Stochastic processes for physicists: understanding noisy systems*. Cambridge University Press, 2010.

[28] Mark Kac. Can one hear the shape of a drum? *The American Mathematical Monthly*, 73(4P2):1–23, 1966.

[29] John D Kalbfleisch and Ross L Prentice. *The statistical analysis of failure time data*, volume 360. John Wiley & Sons, 2011.

[30] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[31] Pinar Gunel Karadeniz, Ilker Ercan, et al. Examining tests for comparing survival curves with right censored data. *Stat Transit*, 18(2):311–28, 2017.

[32] Mohamed A Khabou, Lotfi Hermi, and Mohamed Ben Hadj Rhouma. Shape recognition using eigenvalues of the dirichlet laplacian. *Pattern Recognition*, 40(1):141–153, 2007.

[33] E Leton and P Zuluaga. Equivalence between score and weighted tests for survival curves. *Communications in Statistics-Theory and Methods*, 30(4):591–608, 2001.

[34] Mao Li, Keith Duncan, Christopher N Topp, and Daniel H Chitwood. Persistent homology and the branching topologies of plants. *American Journal of Botany*, 104(3):349–353, 2017.

[35] Kenneth H Rosen. *Discrete Mathematics & Applications*. McGraw-Hill, 1999.

[36] S Sawyer. The greenwood and exponential greenwood confidence intervals in survival analysis. *Applied survival analysis: regression modeling of time to event data*, pages 1–14, 2003.

[37] BD Sleeman and EME Zayed. Trace formulae for the eigenvalues of the laplacian. *Zeitschrift für angewandte Mathematik und Physik*, 35(1):106–115, 1984.

[38] Jiro Tatsumi, Akira Yamauchi, and Yasuhiro Kono. Fractal analysis of plant root systems. *Annals of Botany*, 64(5):499–503, 1989.

[39] Zoltán Toroczkai. Topological classification of binary trees using the horton-strahler index. *Physical Review E*, 65(1):016130, 2001.

[40] Michiel Vandenberg and Peter B Gilkey. Heat content asymptotics of a riemannian manifold with boundary. *Journal of Functional Analysis*, 120(1):48–71, 1994.

[41] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake Vand erPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1. 0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[42] Douglas Brent West et al. *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River, 2001.

[43] Martina Zähle. Random processes of hausdorff rectifiable closet sets. *Mathematische Nachrichten*, 108(1):49–72, 1982.

[44] EME Zayed. Heat equation for an arbitrary doubly-connected region in $\mathbb{R}^2$ with mixed boundary conditions. *Zeitschrift für angewandte Mathematik und Physik*, 40(3):339–355, 1989.

[45] Miloš Zlámal. On the finite element method. *Numerische Mathematik*, 12(5):394–409, 1968.