



Topological properties of phylogenetic trees: analysis of tree forms using Horton's laws

石井, 友一郎

(Degree)

博士 (理学)

(Date of Degree)

2025-03-25

(Date of Publication)

2027-03-25

(Resource Type)

doctoral thesis

(Report Number)

甲第9145号

(URL)

<https://hdl.handle.net/20.500.14094/0100496426>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



(別紙様式 3)

論 文 内 容 の 要 旨

氏 名 石井 友一朗

専 攻 惑星学専攻

論文題目 (外国語の場合は, その和訳を併記すること。)

Topological properties of phylogenetic trees: analysis of tree forms using Horton's laws

(系統樹の位相的特性: Horton の法則を活用した樹形解析)

指導教員 山崎 和仁

Part I

Phylogenetic trees visually represent patterns of evolution and diversification. Many studies have focused on the geometric characteristics of phylogenetic trees, such as the number and length of branches (edges). The connection patterns of edges, known as topological properties, represent evolution and diversification patterns. This study quantified the topological properties of phylogenetic trees by utilizing Horton's first law. Specifically, we conducted a quantitative analysis of the topological properties of vertebrate phylogenetic trees and artificial branching objects created using a neutral stochastic branching model and compared the results.

In Chapter 2, the method for analyzing topological properties was presented. First, Horton's first law was introduced. This is a statistical law concerning the morphology of river networks. When Horton's first law is held in a river network with appropriately assigned stream orders, the number of streams of a given order can be expressed as an exponential function. The base of this exponential function is called the bifurcation ratio, denoted as R_b . This study applied Horton's first law to quantify the topological properties of phylogenetic trees. The analysis procedure involved several steps. First, the edges of phylogenetic trees were given orders. This study employed a modified version of Strahler's ordering method. Next, the relationship between order and the number of edges was examined to determine whether Horton's law was satisfied. Finally, the bifurcation ratio was calculated. In this study, the bifurcation ratio was defined as the average of the ratios between the number of j -th and $(j+1)$ -th order edges. This study analyzed and discussed the topological properties of phylogenetic trees using three elements: the bifurcation ratio (R_b), the number of first-order edges ($N(1)$), and the maximum order (n).

In Chapter 3, six vertebrate phylogenetic trees were analyzed: Spiny-rayed fishes, Amphibians, Turtles, Squamata, Aves, and Placental mammals. It was revealed that Horton's first law was well satisfied in all trees. While the number of first-order edges and the maximum order varied significantly between the trees, the bifurcation ratios were similar, with an average value of 3.18 across the six trees. It suggests that phylogenetic trees for different taxa have a common evolutionary mechanism.

In Chapter 4, the analysis of temporal changes in topological properties is conducted. This study analyzed the Aves phylogenetic tree. As a result, two rules of change in topological properties were identified: (1) when the number of first-order edges increased, and the maximum order remained constant, the bifurcation ratio increased; (2) when the number of first-order edges increased and the maximum order increased by one, the bifurcation ratio decreased.

In Chapter 5, artificial branching objects were investigated for their topological properties.

This analysis aimed to examine whether non-neutral elements are in actual phylogenetic trees and to obtain enough topological data. Artificial branching objects were created using a neutral stochastic branching model. In this model, whether terminal edges are bifurcated is determined by a bifurcation probability once per step. The bifurcation probability was constant across all edges and time steps. The bifurcation probability and the number of time steps are parameters. This study obtained 27,517 artificial branching objects through computer simulations. Almost all artificial branching objects satisfied Horton's law. Regarding the topological properties, the data points were distributed in several bands when plotted with the bifurcation ratio on the horizontal axis and the number of first-order edges on the vertical axis. These bands corresponded to the maximum order values, increasing from bottom to top on the graph as $n = 2, 3, \dots, 15$. As the maximum order increased, the range of bifurcation ratio values narrowed. Regardless of the value of the maximum order n , the average bifurcation ratio for each maximum order was approximately three.

In Chapter 6, the topological properties of phylogenetic trees and artificial branching objects were compared by plotting their topological data on the same graph. Data points of six vertebrate phylogenetic trees and artificial branching objects overlapped, and they had similar topological properties. Data points of Aves past trees and artificial branching objects also overlapped, and they have similar topological properties. As the vertebrate phylogenetic trees and artificial branching objects created by the neutral model had the same topological properties, it may be that non-neutral elements do not affect the way edges are connected throughout the phylogenetic tree, even if they change the length or number of edges. Even when the topological properties of the phylogenetic trees changed, they were still the same as those of the artificial branching objects. When the bifurcation ratio increased, some edges branched unevenly, and tree symmetry decreased. When the bifurcation ratio decreased, the edges branched more evenly, and tree symmetry increased. The results of this study demonstrated the similarity in topological properties between vertebrate phylogenetic trees and artificial branching objects generated by the neutral model. However, this does not necessarily imply that the topological properties of phylogenetic trees are unaffected by non-neutral factors. Non-neutral models may also explain the topological properties of vertebrate phylogenetic trees. Further experiments using non-neutral models are needed.

Chapter 7 provides a conclusion of Part I.

Part II

Phylogenetic trees visually represent evolutionary relationships or evolutionary pathways. The geometric properties of phylogenetic trees contain information about biological evolution. Various indices have been proposed to quantify the topology of phylogenetic trees. These

indices generally treat phylogenetic trees as graphs, and the values of these indices are determined based on individual nodes or edges. However, phylogenetic trees express evolutionary relationships of multiple taxa as a connected structure by a tree. Therefore, performing topological analysis while preserving the tree structure is preferable. Ishii (2023) applied Horton's law to phylogenetic trees and discussed their topological properties. This study developed this approach and utilized Horton's law as a tool for quantifying the topological properties of phylogenetic trees. The study also presented a basic method for interpreting the values of indices derived through this approach. This analysis method was applied to vertebrate phylogenetic trees and artificial branching objects as practical examples.

Chapter 9 proposed a method for quantifying and interpreting the topology of phylogenetic trees based on the perspective of Horton's first law. Horton's law is about the morphology of river networks. Focusing on the similarities of river networks and phylogenetic trees, both are branching objects with hierarchical structures. Based on these similarities, this study proposes a method to quantify the topology of phylogenetic trees using the perspective of Horton's first law. The analysis procedure consists of three steps: 1. Assign orders to the edges of the phylogenetic tree to establish a hierarchical structure; 2. Count the number of edges for each order; 3. Calculate the bifurcation ratio. In this method, Strahler's ordering method is employed. The bifurcation ratio is defined as the average of the ratios of the number of j -th and $(j+1)$ -th-order edges. From the perspective of Horton's first law, the fundamental tree form of phylogenetic trees corresponds to trees with a maximum order $n = 2$, representing monophyletic groups with a single staircase-like shape. In this case, the number of steps in the staircase and the value of the bifurcation ratio are identical. More complex phylogenetic trees with a maximum order of three or higher can be understood as hierarchical structures composed of staircase shapes. The ratio of the number of j -th and $(j+1)$ -th-order edges represents the average staircase shape at the hierarchical level formed by these order edges. The bifurcation ratio indicates the average of the average staircase shape across all hierarchical levels. The bifurcation ratio thus represents the overall average topological properties of a phylogenetic tree as the average number of steps in the staircase shape (the fundamental tree form) that constitute the hierarchical structure. This study discussed the topological properties of phylogenetic trees using the bifurcation ratio, the number of first-order edges, and the maximum order. The analytical method proposed in this study calculates topological indices using fundamental tree forms as the basic unit. In this sense, this method treats a phylogenetic tree as a tree. The bifurcation ratio is calculated regardless of whether Horton's law holds.

Chapter 10 presented the analysis targets and methods for examples of analysis using the quantification method for topological properties of phylogenetic trees based on Horton's first

law. The analysis objects were vertebrate phylogenetic trees and artificial branching objects. For each, their topological properties were calculated. Additionally, advanced analyses were conducted, including a comparison of the topological properties of both and an investigation into the temporal change of the topology of the vertebrate phylogenetic trees. Analyzed vertebrate phylogenetic trees were eight types: Actinopterygii, Aves, Chondrichthyes, Amphibia, Mammalia, Reptilia Testudines, Reptilia Crocodilia, and Reptilia Lepidosauria. The terminal edges of these trees represent species, and the edge lengths reflect time. Artificial branching objects were generated using a neutral stochastic branching model. Topological data were compared by plotting data points in the $R_b - \mathcal{N}(1)$ space. The temporal changes in the topology of vertebrate phylogenetic trees were investigated every approximately one million years.

Chapters 11 and 12 present the results and discussion. Concerning the topological properties of vertebrate phylogenetic trees, although the number of first-order edges varied significantly, the differences in bifurcation ratio values were not large. The similarity in bifurcation ratios and differences in the number of first-order edges can be attributed to differences in the maximum order. The average bifurcation ratio across the eight phylogenetic trees was 3.31. This indicates that vertebrate phylogenetic trees have hierarchical structures with staircase shapes of 3.31 steps. The topological properties of artificial branching objects were not clearly distinguished by simulation conditions. However, when the data points were plotted in $R_b - \mathcal{N}(1)$ space, they were grouped into bands corresponding to the maximum order, ranging from $n = 2$ to $n = 11$, from bottom to top. Data points were plotted only in specific regions of the $R_b - \mathcal{N}(1)$ space. It indicates that not all geometrically possible trees are generated under the neutral model. Comparing vertebrate phylogenetic trees with artificial branching objects revealed that the data points of vertebrate trees were plotted within the range of the artificial branching objects' topological properties, and their maximum order values also matched. Similar characteristics were also observed between past phylogenetic trees and artificial branching objects. These results suggest that the topology in phylogenetic trees follows the neutral model. Besides, they indicate the possibility that topological constraints between evolutionary phenomena and phylogenetic trees restrict the diversity of tree forms or evolutionary pathways. Topological constraints refer to the fact that only terminal edges can branch, branching occurs as binary splits, and it is impossible to deterministically determine which edge will branch. When the number of first-order edges increased, four topological change patterns were observed: (1) the bifurcation ratio increases while the maximum order remains constant; (2) the maximum order increases by one, and the bifurcation ratio decreases; (3) the bifurcation ratio decreases while the maximum order remains constant; (4) both the bifurcation ratio and the maximum order remain constant. A

low bifurcation ratio corresponds to staircase shapes with a lower number of steps, indicating a state where all taxa have experienced a similar number of evolutionary events. A high bifurcation ratio corresponds to staircase shapes with a higher number of steps in the staircase. This state is indicated where the evolution events are concentrated in specific taxa. Change pattern (1) can be interpreted as transitioning to a state where taxa with more and fewer evolutionary events coexist. Change pattern (2) can be interpreted as transitioning to a state where evolution is experienced more evenly among taxa. In this change pattern, the number of n -th-order edges increases, and a new hierarchical level is created. As a result, not only staircase shapes but also the hierarchical structure changes. In change pattern (3), the hierarchical structure does not change, but the average staircase shapes change. This change pattern is interpreted as transitioning to a state where evolution is experienced more evenly among taxa. Change pattern (4) does not change the bifurcation ratio, but the average staircase shapes within each hierarchical level change. At least within lower-order hierarchies, evolution is experienced more evenly. Although the resolution of the tree form is limited, the analytical method proposed in this paper has the advantage of enabling the quantification of phylogenetic tree topology easily and providing an intuitive understanding of the overall average tree form based on the bifurcation ratio. Because the bifurcation ratio corresponds to the number of steps in the staircase and a change of one in the bifurcation ratio corresponds to a change of one step in the staircase, it is easy to relate the index value to the actual tree form. This analytical method is also expected to be used in advanced research, such as comparing the forms of sub-trees to study evolutionary characteristics among taxa.

Chapter 13 provides a conclusion of Part II.

氏名	石井 友一朗		
論文 題目	Topological properties of phylogenetic trees: analysis of tree forms using Horton's laws (系統樹の位相的特性:Horton の法則を活用した樹形解析)		
審査 委員	区 分	職 名	氏 名
	主 査	講師	山崎 和仁
	副 査	教授	山本 由弦
	副 査	教授	荒川 政彦
	副 査		印
	副 査		印
要 旨			
本論文は、ホートンの法則に基づいて系統樹の位相的特性について新しい理論的知見を示したものである。第一部と第二部からなり、全17章からなる。 第一部の第1章では、系統樹の概念と、それが示す進化や多様化のパターンを表すトポロジー的特性を紹介している。また、ホートンの法則が系統樹の分析にどのように適用されるかについて論じている。 第2章では、トポロジー的特性を分析する方法について説明している。ホートンの法則はもともと河川のネットワークに関する法則であるが、これを系統樹に適用する方法を提案し、エッジに順序を付け、分岐比 R_b を計算する方法について詳細に述べている。 第3章では、魚類、両生類、カメ、爬虫類、鳥類、胎盤哺乳類の6つの脊椎動物系統樹を分析している。その結果、全ての系統樹がホートンの法則を満たしており、分岐比は平均して 3.18 であることを示している。 第4章では、鳥類系統樹におけるトポロジー的特性の時間的変化について調査している。第一エッジ数 $N(1)$が増加すると分岐比が増加し、最大次数が1増加すると分岐比が減少するという2つの変化の法則が特定される。 第5章では、中立的な確率的分岐モデルを用いて人工分岐パターンを作成し、そのトポロジー的特性を調査している。この結果、人工分岐パターンはホートンの法則をほぼ満たし、分岐比率が常に約3であることが確認される。 第6章では、脊椎動物系統樹と人工分岐パターンのトポロジー的特性を比較している。両者のデータ点は重なり合い、類似したトポロジー的特性を示すことが分かる。これにより、非中立的要因が系統樹のトポロジー構造に与える影響は限られている可能性が示唆されるが、さらなる研究が必要なことも示されている。 第7章では、本研究の結果を総括し、系統樹のトポロジー的特性が中立モデルに基づいて生成された人工分岐パターンと類似していることの解釈が要約されている。 第二部は本研究の白眉である。第9章は第1章と第2章の内容を統合し簡略化している。河川と系統樹は階層構造を持つ分岐パターンであり、この類似性を基に系統樹のトポロジーを定量化する方法を提案している。手順は、1) 系統樹のエッジに順序を割り当てて階層構造を確立する、2) 各順序のエッジ数をカウントする、3) 分岐比を計算する、という三つのステップから成る。系統樹の基本的な形態は、$r=2$ の最大次数を持つ単一の階段形状に対応し、分岐比は階段のステップ数と一致することが示される。 第10章では、ホートンの法則に基づいた系統樹のトポロジー的特性を定量化する方法が実際の例として適用される。分析対象は脊椎動物の系統樹と人工分岐パターンであり、それぞれのトポロジー特性を計算する。さらに、両者のトポロジー特性を比較し、脊椎動物系統樹のトポロジーの時間的変化を調査する。対象となる脊椎動物系統樹は、硬骨魚類、鳥類、軟骨魚類、両生類、哺乳類、爬虫類、カメ、ワニ類、爬虫類の9種類である。人工分岐物体は中立的確率分岐モデルを使用して生成され、トポロジーのデータは $R_b \cdot N(1)$空間にプロットされる。 第11章・第12章では、脊椎動物系統樹のトポロジー特性について、第一順エッジ数に大きな差はあったが、分岐比の差は小さかったことが示されている。分岐比は9つの系統樹で平均 3.31 であり、これにより脊椎動物系統樹は 3.31 段階の階段形状を持つ階層構造を持つことが示される。人工分岐パターンのトポロジー特性はシミュレーション条件によってはっきりと区別されなかったが、$R_b \cdot N(1)$空間にプロットされたデータ点は、最大順序に対応するバンドにグループ化され、特定の領域にのみデータ点がプロットされた。これにより、すべての幾何学的に可能な系統樹が中立モデルの下で生成されるわけではないことが示唆される。			

氏名	石井 友一郎
第 13 章では、第二部の結論を述べる。脊椎動物系統樹のトポロジーは中立モデルに従っていることが示唆され、進化的現象と系統樹の間のトポロジー的制約が樹形の多様性や進化経路を制限している可能性があることが指摘される。また、トポロジーの変化パターンが進化の均等な分布に関する情報を提供することが示される。 第 14 章は図と表、第 15 章は謝辞一覧、第 16 章は参考文献、第 17 章は付録である。 本研究は系統樹について、その位相的特性を研究したものであり、トポロジー的制約が樹形の多様性や進化経路を制限しているについて、重要な知見を得たものとして価値ある集積であると認める。よって、学位申請者の石井友一郎は、博士(理学)の学位を得る資格があると認める。	