

**UNIVERSIDADE FEDERAL DO AMAZONAS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA**

**REDES FILOGENÉTICAS:  
ESTUDO ATRAVÉS DAS REDES COMPLEXAS**

JOÃO VITOR FERNANDES DA GAMA

MANAUS - AM

2025

**JOÃO VITOR FERNANDES DA GAMA**

**REDES FILOGENÉTICAS:  
ESTUDO ATRAVÉS DAS REDES COMPLEXAS**

Dissertação apresentada ao Instituto de Ciências Exatas da Universidade Federal do Amazonas como requisito para a obtenção do título de Mestre.

Orientador: Prof. Dr. Mircea Daniel Galiceanu

Manaus - AM  
Agosto, 2025

## Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

---

G184r      Gama, João Vitor Fernandes da  
              Redes Filogenéticas: Estudo Através das Redes Complexas / João Vitor  
              Fernandes da Gama. - 2025.  
              51 f. : il., color. ; 31 cm.

              Orientador(a): Mircea Daniel Galiceanu.  
              Dissertação (mestrado) - Universidade Federal do Amazonas, Programa  
              de Pós-Graduação em Física, Manaus, 2025.

              1. Redes Complexas. 2. Sistemática Filogenética. 3. Redes Filogenéticas.  
              4. Modelagem Computacional. I. Galiceanu, Mircea Daniel. II.  
              Universidade Federal do Amazonas. Programa de Pós-Graduação em  
              Física. III. Título

---

**REDES FILOGENÉTICAS:  
ESTUDO ATRAVÉS DAS REDES COMPLEXAS**

Dissertação apresentada ao Instituto de Programa de Pós-Graduação de Física da Universidade Federal do Amazonas sob orientação do Prof. Dr. Mircea Daniel Galiceanu como requisito para a obtenção do título de Mestre em Física.

**Banca Examinadora:**

---

Prof. Dr. Mircea Daniel Galiceanu

---

Prof. Dr. Fidel Guerrero Zayas

---

Prof. Dr. Marcus Werner Beims

---

# AGRADECIMENTOS

Meus mais sinceros e profundos agradecimentos à Ellen Cardoso da Gama, minha mãe, pelo suporte material e financeiro e pela ajuda na correção deste trabalho; ao professor Dr. Mircea Daniel Galiceanu pela orientação e pela oportunidade de desenvolver esta linha de pesquisa; ao professor Me. Walter Esteves de Castro Júnior, por me deixar usar sua sala no Departamento de Física e computador sobressalente para estudar e por seu valioso auxílio em diversas questões dentro da universidade; aos demais professores e alunos do Departamento de Física, pelo suporte e pelos diversos auxílios ao longo da graduação; aos colegas Luan Araújo, Lucas Oliveira, e Victor Emanuel Rocha, pelas ideias propostas e debatidas ao longo da elaboração deste trabalho; à Gabriella Guaqueta Lenhardt, pela elaboração da Figura 2.7; à Sabrina Potratz Kreniski, pela elaboração das figuras 2.2, 2.10, 2.12, e 2.13; à Eva Letícia Camara Lacerda e Adrienny Medeiros Moraes, pelo auxílio na correção; ao professor Dr. Paulo "Pirula" Miranda Nascimento pela inspiração inicial desta pesquisa; aos autores das minhas referências bibliográficas, por terem aberto o caminho; e à Ronald James Padavona e colegas, Karsten Brill e colegas, e David Scott Mustaine e colegas, pelo suporte emocional que me manteve firme durante toda a graduação.

---

# RESUMO

Este trabalho focou-se no desenvolvimento de um modelo computacional como uma alternativa para estudar as características de sistemas determinados por relações genealógicas através da teoria das redes complexas. O modelo consiste na construção de uma rede complexa cujos nós são separados em gerações. Os nós de uma geração interagem uns com os outros para gerar a próxima geração em uma taxa de crescimento pré-determinada, e este ciclo se repete até que a rede atinja o limite de número de nós. Foram analisados os coeficientes de Pearson e as distribuições de grau das redes. Os resultados revelaram que o coeficiente de Pearson é quase sempre negativo, com valores oscilantes. A taxa de oscilação está relacionada à taxa de crescimento da rede. Quanto às distribuições de grau, constatou-se a existência de pelo menos dois picos, um sempre em  $k = 2$ , e o outro de valor variável mas sempre maior que 2. O padrão das distribuições de grau não é trivial nem tão pouco foi reportado em literatura científica, assemelhando-se a uma superposição de gaussianas.

**Palavras-chave:** Sistemática Filogenética, Redes Complexas, Redes Filogenéticas, Modelagem Computacional, Distribuição de Grau, Coeficiente de Pearson

---

# ABSTRACT

This work focused on the development of a computational model as an alternative approach to study the characteristics of systems determined by genealogical relationships through the theory of complex networks. The model consists of constructing a complex network whose nodes are separated into generations. The nodes of one generation interact with each other to produce the next generation at a predetermined growth rate, and this cycle repeats until the network reaches the maximum number of nodes. Pearson coefficients and degree distributions of the networks were analyzed. The results revealed that the Pearson coefficient is almost always negative, with oscillating values. The oscillation rate is related to the growth rate of the network. As for the degree distributions, the existence of at least two peaks was observed — one always at  $k = 2$ , and the other at a variable value, but always greater than 2. The pattern of the degree distributions is non-trivial and has not been previously reported in the scientific literature, resembling a superposition of gaussians.

**Keywords: Phylogenetic Systematics, Complex Networks, Phylogenetic Networks, Computational Modelling, Degree Distribution, Pearson Coefficient**

---

## LISTA DE FIGURAS

2.1	Árvore de semelhanças relativas entre as cinco espécies. As semelhanças foram calculadas usando o método de agrupamento hierárquico nativo da biblioteca SciPy do Python. . . . .	7
2.2	Semelhanças entre tubarões, ictiossauros, e golfinhos. . . . .	8
2.3	Relações toco-genéticas ao longo das gerações de uma espécie. Machos são representados por triângulos e fêmeas, por quadrados. As gerações mais antigas são representadas embaixo e as mais novas, em cima. . . . .	10
2.4	Processo de fixação de um caráter em uma população. O caráter é representado pela cor vermelha. . . . .	11
2.5	Processo de cladogênese. A barreira, indicada pela seta, pode tanto ser física quanto genética, e até mesmo cultural. . . . .	12
2.6	Os indivíduos das populações B, à esquerda, e C, à direita, por não mais se reproduzirem entre si, adquirem caracteres diferentes, assim gradualmente se diferenciando. Os caracteres novos são representados pelas cores cinza e preto. . . . .	13
2.7	Relações toco-genéticas de uma população de animais que sofre anagênese, cladogênese, e eventualmente especiação. . . . .	14
2.8	Representação simples da relação entre a população ancestral A e as populações de duas espécies distintas B e C que vieram da população A. . . . .	15
2.9	Cladogramas em formatos diferentes, mas que representam as mesmas relações filogenéticas. Os nós estão destacados dentro de círculos e os grupos-irmãos, dentro das elipses. . . . .	16
2.10	Cladograma do grupo dos vertebrados com caracteres principais de cada sub-grupo. . . . .	17
2.11	Comparação entre grupos parafilético (azul) e monofilético (amarelo) dentro do grupo dos tetrápodes. . . . .	18
2.12	Homologia dos membros anteriores de diferentes tetrápodes. . . . .	20
2.13	Comparação entre asas de insetos, aves, e morcegos. A estrutura óssea das asas de aves e morcegos são homólogas entre si, mas não com o suporte das asas de insetos. Todavia a estrutura <i>asa</i> serve a mesma função de voar nos três grupos, tratando-se portanto de uma analogia. . . . .	22



3.1	Mapa de Königsberg de 1652. . . . .	25
3.2	Representação das sete pontes de Königsberg. . . . .	25
3.3	Representação em forma de grafo do problema das sete pontes de Königsberg. . . . .	26
3.4	Exemplo de caminhos como uma rede complexa. . . . .	28
3.5	Principais tipos de redes complexas acompanhadas de suas matrizes de adjacência. Além das redes não-direcionadas (a), direcionadas (d), e com peso (e), também há um exemplo de rede cuja aresta conecta um nó a ele mesmo (b), um de redes com mais de uma aresta conectando dois nós (c), e um de redes cujos nós estão todos interconectados (f). . . . .	29
3.6	Exemplo de uma rede tipo árvore. . . . .	30
3.7	Mesma rede da Figura 3.6, mas na forma com raiz. . . . .	30
3.8	Distribuição de grau da rede da Figura 3.4 . . . . .	34
3.9	Significados do Coeficiente de Pearson . . . . .	35
4.1	Equivalência entre cladogramas e redes complexas. Os nós e folhas do cladograma se traduzem como nós na rede complexa, e os ramos, como ligações. Tendo mais dados sobre os táxons, é possível usar modelos mais avançados de redes do que uma não-direcionada sem peso. . . . .	36
4.2	Esquematização da simulação do processo toco genético como uma rede complexa. . . . .	40
4.3	Distribuição de graus de 20 redes com 20.000 nós e R fixo. . . . .	41
4.4	Coeficiente de Pearson de redes com R fixo e número total de nós de 1.000 a 20.000 em passos de 1.000. . . . .	42
4.5	Distribuição de graus de 20 redes com 20.000 nós e $N_i$ fixo. . . . .	43
4.6	Coeficiente de Pearson de redes com $N_i$ fixo e número total de nós de 1.000 a 20.000 em passos de 1.000. . . . .	44

---

# LISTA DE TABELAS

2.1	Exemplo simples de uma tabela de características por espécies usada na fenética. . . . .	6
3.1	Tabela de caminhos. . . . .	28

---

# SUMÁRIO

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Sistemática Filogenética</b>	<b>3</b>
2.1	Introdução Histórica . . . . .	3
2.2	Conceitos Básicos da Sistemática Filogenética . . . . .	9
<b>3</b>	<b>Redes Complexas</b>	<b>24</b>
3.1	Introdução Histórica . . . . .	25
3.2	Teoria das Redes Complexas . . . . .	27
3.2.1	Grandezas Estatísticas de Redes Complexas . . . . .	33
<b>4</b>	<b>Redes Filogenéticas</b>	<b>36</b>
4.1	Modelo . . . . .	38
4.2	Algoritmos . . . . .	39
4.3	Resultados . . . . .	40
<b>5</b>	<b>Considerações Finais</b>	<b>46</b>



---

---

# CAPÍTULO 1

---

## INTRODUÇÃO

Relações de parentesco constituem um dos temas mais fundamentais e estruturantes não apenas da sociedade humana, onde são ensinadas desde a mais tenra idade, mas de todos os sistemas biológicos. Na Biologia, essas relações transcendem o aspecto social para se tornarem a base da própria evolução da vida, servindo como arcabouço para compreender os processos de descendência com modificação que moldaram a biodiversidade [1; 2]. A característica determinante das relações de parentesco é a ligação de uma entidade a pelo menos uma outra por meio da relação de origem, onde uma entidade nova é gerada a partir de uma anterior.

Redes complexas emergiram como ferramentas matemáticas poderosas para modelar tais relações de parentesco. Estas estruturas, compostas por nós (vértices) e ligações (arestas), permitem representar quantitativamente as conexões entre entidades biológicas [3; 4]. Enquanto os nós podem representar táxons, espécies ou indivíduos, as ligações podem codificar relações de descendência, similaridades morfológicas ou genéticas, criando assim um modelo matemático adequado para representar as relações filogenéticas [5; 6].

Este trabalho situa-se na intersecção entre duas áreas do conhecimento: a *sistemática filogenética* (ou cladística) e a *teoria das redes complexas*. A cladística, desenvolvida inicialmente por Willi Hennig [7; 8], fornece o marco teórico para inferir relações de parentesco evolutivo entre organismos. Na Seção 2.1 apresenta-se uma contextualização histórica desde os trabalhos fundadores de Linnaeus [9] e Darwin [1] até o desenvolvimento da sistemática filogenética moderna, enquanto a Seção 2.2 detalha os princípios teóricos da abordagem cladística.

Paralelamente, a teoria das redes complexas desenvolveu-se como disciplina matemática independente, com contribuições seminais de Erdős e Rényi [10; 11] e aplicações pioneiras em Biologia por Harary [12; 13]. A Seção 3.1 traça um panorama historiográfico do desenvolvimento desta teoria, desde seus primórdios na teoria dos grafos [14] até suas aplicações contemporâneas em sistemas biológicos [3? ]. A Seção 3.2 apresenta os fundamentos matemáticos das redes complexas, e a Seção ?? ilustra sua aplicação em problemas biológicos através do estudo de redes de disseminação de doenças [15].

A motivação central deste trabalho reside na oportunidade de aplicar o arcabouço teórico e metodológico das redes complexas para abordar questões em sistemática filogenética. Embora ambas as áreas tenham se desenvolvido de forma relativamente independente, sua integração oferece perspectivas promissoras para a análise de dados filogenéticos em larga escala e para a identificação de padrões macroevolutivos [6]. Esta abordagem interdisciplinar alinha-se com a tradição de avanços significativos na Biologia provenientes da aplicação de ferramentas da Física e da Matemática [16; 17].

No Capítulo 4 concretiza-se esta proposta através do desenvolvimento de um modelo baseado em redes para análise de dados filogenéticos (Seção 4.1), com os algoritmos computacionais detalhados na Seção 4.2 e os resultados preliminares apresentados na Seção 4.3. Esta abordagem expande trabalhos anteriores em redes filogenéticas [6] e dá continuidade a investigações iniciais em redes cladísticas [18].

As considerações finais e perspectivas futuras são discutidas no Capítulo 5, destacando o potencial desta linha de pesquisa para contribuir tanto para o avanço do conhecimento em Sistemática Filogenética quanto para o desenvolvimento de metodologias inovadoras na interface entre Biologia e Ciências Exatas.

---

# CAPÍTULO 2

---

## SISTEMÁTICA FILOGENÉTICA

---

O estudo comparativo de animais e plantas surge naturalmente da necessidade de saber as necessidades básicas desses serem na agricultura e pecuária, além de ser um tema de grande interesse para maioria, senão para todos os grupos de seres humanos ao longo da história. É difícil citar um povo que não tem mitos cosmogônicos sobre a criação dos animais, plantas, e seres humanos, ou sobre a grande diversidade de seres vivos.

Vacas, cães, gatos, cabras, e cavalos, por exemplo, claramente têm alguma similaridade aos seres humanos, visto que todos produzem leite e têm pelos; já cobras, lagartos, crocodilos e tartarugas têm pouco em comum conosco, porém muito em comum entre si, mas por que? Esse grande interesse nas relações entre os diversos animais culminou nos estudos de *filogenia* e na **sistemática filogenética**.

### 2.1 Introdução Histórica

Os primeiros estudos sobre a relação entre seres vivos remontam aos antigos gregos, particularmente a Aristóteles, que classificou organismos vivos com base em suas similaridades físicas em um sistema hierárquico e fixo, assim iniciando o campo hoje conhecido como *taxonomia*<sup>1</sup>. Herdando parcialmente o idealismo platônico, Aristóteles propõe que cada ser vivo possui uma *essência* teleológica, ou *animus* (alma) nas palavras dele, que determina as características, comportamentos, e funções no mundo dos seres vivos e

---

<sup>1</sup>A palavra foi criada em 1813 pelo botânico suíço A. P. de Candolle e é formada irregularmente a partir do grego *τάξις*, *taxis* (ordem), e *νόμος*, *nomos* (lei), conectados pela forma francesa -o-; a forma regular seria *taxinomia*, como usada no ressurgimento grego *ταξινομία*

não-vivos. É importante ressaltar que a essência de um ser é imutável, isto é, apesar de que cada ente pode variar um pouco em relação aos seus semelhantes, todos eles são desvios de um arquétipo ideal. Isto baseou, muitos séculos depois, o *fixismo*, ideia ainda mantida viva nos dias de hoje por negadores da ciência de que os seres vivos serão e sempre foram como são hoje.

Quanto ao aspecto teleológico, no que tange à Biologia, Aristóteles propôs que a essência dos seres os atribuía uma função ou objetivo específico que foi projetado para cumprir, e que essa função ou objetivo é uma parte indissociável de sua natureza. Por exemplo, o propósito de uma muda ou fruto é se transformar em uma árvore, e o propósito de um pássaro é voar e se reproduzir. Para Aristóteles, as diversas partes de um ser vivo, como seus órgãos e tecidos, foram projetadas especificamente para cumprir esse propósito, e que elas trabalham de forma coordenada para alcançar esse objetivo.

A teleologia teve um impacto significativo na Biologia nos séculos que se seguiram. Por muitos séculos, cientistas e filósofos acreditaram que os seres vivos foram propositalmente projetados por uma força superior, e que suas diversas estruturas e funções faziam parte de um grande plano. Essa ideia de design proposital e inteligente levou ao desenvolvimento da teologia natural, um campo de estudo que buscava compreender o plano divino por trás do mundo natural.

O sistema de classificação de Aristóteles categorizou organismos vivos em dois grupos principais: plantas e animais. Dentro de cada grupo, os organismos foram divididos em categorias menores com base em suas características físicas. Por exemplo, as plantas foram divididas em árvores, arbustos e ervas, com base em seu tamanho e hábitos de crescimento. Animais foram divididos em grupos como aves, mamíferos e peixes, com base em suas características físicas, com os seres humanos como pináculo da vida.

Tomar o ser humano como o pináculo não se trata de metáfora no caso do sistema aristotélico de classificação. Neste modelo, há de fato uma hierarquia de seres vivos com o ser humano no topo. Tal hierarquia, chamada de *scala naturae*, é organizada pelo quão complexos são os seres. Plantas, fungos, e corais estariam na base dessa hierarquia, enquanto animais, capazes de maior movimentação e de afetar o mundo ao seu redor em grande escala, estariam no topo [19].

No século XVIII, o botânico sueco Carolus Linnaeus desenvolveu o sistema de nomenclatura binomial para nomear e classificar espécies com base em suas características



físicas. Seu sistema foi amplamente adotado e se tornou a base para a taxonomia moderna, sendo usado até hoje. O sistema de Linnaeus é fortemente baseado nas ideias fixistas de Aristóteles [9; 20], ainda sendo usado hoje em dia, apesar de que com muitas modificações para se adaptar ao paradigma da evolução das espécies, descrita por Darwin.

No século XIX, a teoria da evolução por seleção natural de Charles Darwin forneceu um novo quadro para entender as relações entre organismos. Enquanto o primeiro a rejeitar o fixismo e propor que as espécies mudam ao longo do tempo foi o naturalista francês Jean-Baptiste Lamarck, foi Darwin quem entendeu que a ancestralidade das espécies pode ser inferida ao estudar suas semelhanças e diferenças [1]. Enquanto houve grande resistência à teoria evolutiva por parte de pessoas fora da academia, principalmente por motivos religiosos, zoólogos e botânicos em geral reconheceram a validade do darwinismo [21, p.202-226]. Todavia, uma crítica consistente contra a qual Darwin não tinha argumentos era sobre a natureza do mecanismo de transmissão de características, hoje conhecido como genética. Darwin, afinal, não tinha uma compreensão extensa sobre o processo de herança de características, baseando-se principalmente no modelo de *blending inheritance* (herança por mistura). Os primeiros estudos de genética feitos por Mendel, enquanto contemporâneos aos estudos de Darwin, não ganharam projeção até o início do século XX. A união da teoria da evolução com os estudos de Mendel deu origem ao que é conhecido como *Teoria Sintética da Evolução*, ou neo-darwinismo, a qual emergiu somente na década de 1930 [2].

Com a difusão da teoria sintética da evolução, houve uma mudança de paradigma na taxologia. O século XX presenciou o surgimento de três grandes escolas de pensamento nesta área: a escola genética, a escola fenética, e a escola da sistemática filogenética. A escola genética foi a primeira a ser fundada, surgindo por volta da década de 1930, e seus defensores foram justamente alguns dos principais proponentes do neo-darwinismo [22, p.1-7]. A escola fenética por sua vez tem como um dos trabalhos fundadores o livro *Numerical Taxonomy* (Taxonomia Numérica), publicado em 1963 [23]. Por último e mais importante, a escola da sistemática filogenética teve seu primeiro trabalho publicado apenas em alemão ainda em 1950 [7], com os primeiros trabalhos em inglês circulando a partir de 1966, sendo desenvolvida por diversos pesquisadores e se tornando a forma dominante de estudo de taxonomia nos dias de hoje.

A escola genética, também conhecida como escola da sistemática evolutiva, ou escola

gradista, foi fundada pelos pesquisadores George Gaylord Simpson, John B.S. Haldane, Ronald Fisher, Sewall Wright, e Theodosius Dobzhansky. A escola genética enfatizou a importância de usar tanto dados anatômicos quanto moleculares para reconstruir a história evolutiva dos organismos, acreditando que as relações evolutivas poderiam ser inferidas estudando o acúmulo gradual de pequenas mudanças na composição genética das populações ao longo do tempo. Eles também reconheceram a importância da evolução convergente, que é a evolução independente de características semelhantes em organismos não relacionados devido a pressões ambientais similares [24]. Todavia, a escola genética não apresentou uma metodologia concisa e objetiva, fazendo inferências demasiado subjetivas.

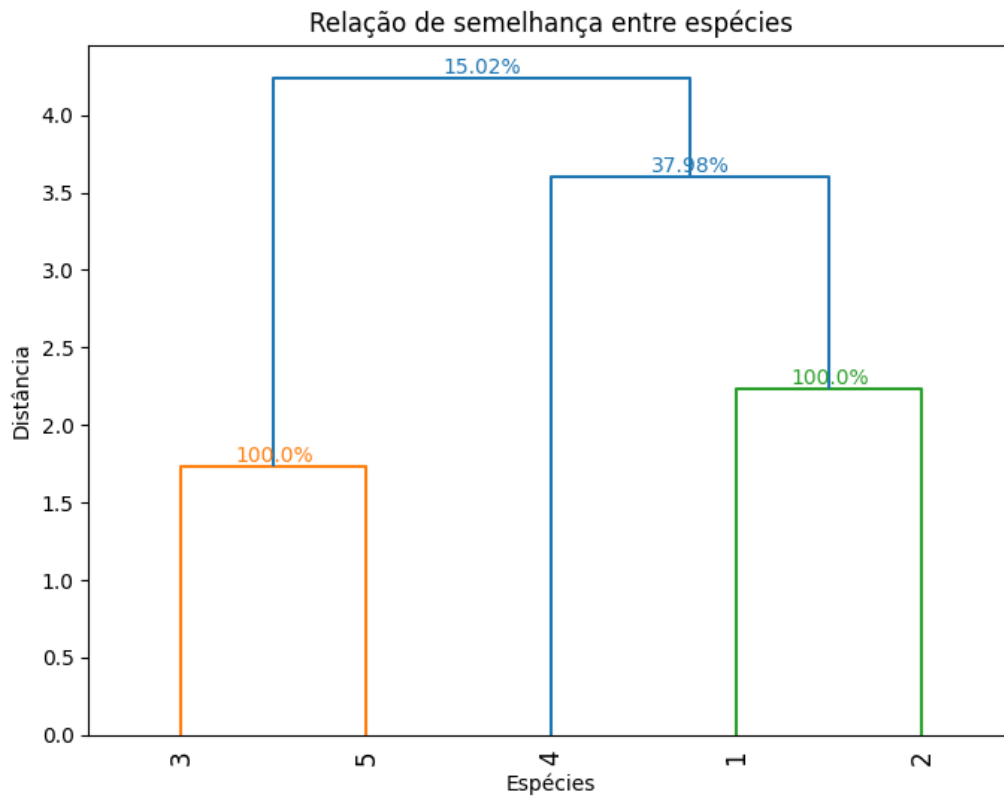
A escola fenética, por outro lado, fundada pelo botânico holandês Pieter Sneath e pelo zoólogo britânico Robert Sokal, enfatizou a importância de usar métodos quantitativos para analisar as semelhanças e diferenças entre organismos, baseados em suas características físicas. Eles defendiam que as relações evolutivas poderiam ser inferidas agrupando organismos com base em sua semelhança geral, sem necessariamente levar em conta sua história evolutiva [23]. O método fenético consistia em fazer tabelas de espécies e características e determinando números para as características de cada espécie. Por exemplo, se a espécie 1 tivesse uma antena com 10 segmentos, a entrada da tabela seria 10, se uma outra espécie tivesse antena com 11 segmentos, seria 11, e caso uma terceira espécie não tivesse antena, 0. Um exemplo ilustrativo é a Tabela 2.1.

**Tabela 2.1:** Exemplo simples de uma tabela de características por espécies usada na fenética.

Características:	A	B	C	D	E
Espécie 1	9	7	8	6	2
Espécie 2	9	7	8	6	2
Espécie 3	7	1	2	8	0
Espécie 4	3	1	9	7	3
Espécie 5	7	1	2	8	0

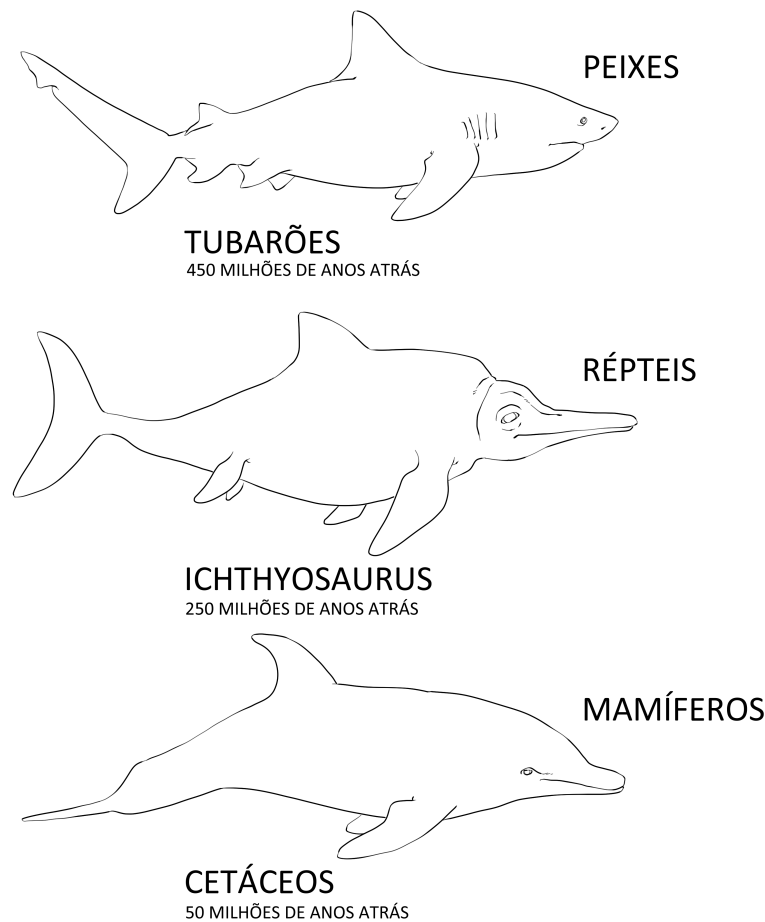
A qual resulta na árvore de semelhanças dada na Figura 2.1.

**Figura 2.1:** Árvore de semelhanças relativas entre as cinco espécies. As semelhanças foram calculadas usando o método de agrupamento hierárquico nativo da biblioteca SciPy do Python.



O problema da escola fenética foi justamente ignorar a teoria evolutiva, visto que há fenômenos que levam espécies diferentes a ter características semelhantes mesmo que sejam evolutivamente distantes. Um equivalente cotidiano seria supor que todas as pessoas de rosto parecido são irmãs, um claro absurdo. Um exemplo clássico de grupos semelhantes, mas cujas semelhanças deram-se apenas devido a pressões ambientais semelhantes, são os grupos dos tubarões, dos golfinhos, e dos ictiossauros. Os três grupos são compostos por espécies aquáticas predadoras, ou seja, o ambiente seleciona os predadores mais eficientes, isto é, os indivíduos com anatomia mais hidrodinâmica. Como os três grupos têm hábitos e comportamentos similares, caçam da mesma maneira, e são da linhagem dos vertebrados, o formato do corpo, ou *bauplan*, das espécies dos três grupos são similares.

**Figura 2.2:** Semelhanças entre tubarões, ictiossauros, e golfinhos.



Fonte: [25]

Em 1950, o entomologista alemão Willi Hennig publicou o artigo fundador da *sistemática filogenética*, ou *cladística*, a qual trazia uma nova abordagem para a classificação sistemática. Nas décadas seguintes, Hennig e aqueles que aderiram à cladística desenvolveram uma teoria sólida e uma metodologia concisa e objetiva. Suas ideias revolucionaram o campo da classificação sistemática e ainda são amplamente usadas hoje.

Hennig argumentou que as escolas gradista e fenética eram falhas, a primeira pela sua falta de uma metodologia objetiva, e a segunda, porque não levavam em conta as histórias evolutivas únicas de cada grupo de organismos. Ele propôs que as relações evolutivas poderiam ser inferidas identificando características compartilhadas, ou *sinapomorfias*, que são características únicas de um grupo particular de organismos e evoluíram de um ancestral comum, em vez de mera semelhança, como fizeram os da escola fenética.

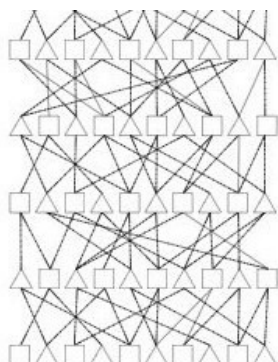
Para tal, adotou-se o método de tabela dos fenéticos, porém com a mudança teórica e metodológica de tratar características não meramente como semelhanças, mas como algo que em determinado momento da evolução do grupo foi uma novidade. A diferença metodológica é explicada em detalhes na seção 2.2.

Apesar de ter surgido entre a fundação das escola gradista e fenética, a abordagem de Hennig representou uma convergência de ideias de ambas as escolas. Como a escola genética, Hennig reconheceu a importância de se basear na teoria da evolução natural, usando tanto dados morfológicos quanto moleculares, à medida que a tecnologia permitiu, para reconstruir as relações evolutivas. Como a escola fenética, Hennig reconheceu a importância de usar métodos quantitativos para analisar as semelhanças e diferenças entre organismos. No entanto, a abordagem de Hennig foi além de ambas as escolas [8].

## 2.2 Conceitos Básicos da Sistemática Filogenética

Hennig começou o seu modelo partindo da dinâmica populacional em animais. Em uma população de animais quaisquer, a principal relação entre os indivíduos é a relação genética entre pais e filhos, ou *tocogenética*, como Hennig nomeou [8, p.17-32]. Ressalta-se novamente que estamos nos limitando a animais por enquanto, visto que a reprodução dialética do reino animal é uma das, senão a forma mais simples de reprodução no que tange a modelos. Plantas, protozoários, fungos, e bactérias todos apresentam auto-inseminação, reprodução assexuada, e outros fenômenos que complicam (apesar de não impossibilitarem, principalmente no contexto de redes filogenéticas) a modelagem. A figura 2.3 representa as relações tocogenéticas em uma população. Nela, fêmeas são representadas por triângulos e machos por quadrados, ou vice-versa. Cada linha representa uma geração, e a relação temporal dá-se de baixo para cima.

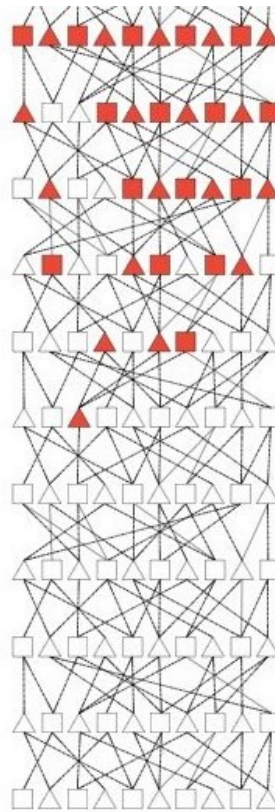
**Figura 2.3:** Relações tocogenéticas ao longo das gerações de uma espécie. Machos são representados por triângulos e fêmeas, por quadrados. As gerações mais antigas são representadas embaixo e as mais novas, em cima.



Fonte: [26]

Gradualmente e aleatoriamente, com o passar das gerações, ocorrem mutações em alguns indivíduos, as quais se traduzem em uma ou mais características. Caso as mutações sejam negativas, esses indivíduos se reproduzem menos ou morrem, mas caso essas mutações sejam vantajosas ou neutras, elas gradualmente vão se fixando na população. Quando uma característica nova se torna dominante na população, estando presente na maioria ou totalidade dos indivíduos, ela recebe o nome de *caráter*.

**Figura 2.4:** Processo de fixação de um caráter em uma população. O caráter é representado pela cor vermelha.



Fonte: [26]

O surgimento de caracteres ao longo de uma linhagem recebe o nome de *anagênese*<sup>2</sup>. A divisão de uma população ocorre frequentemente na natureza, seja pelo surgimento de uma barreira geográfica<sup>3</sup>, seja pela adaptação de diferentes partes da população a diferentes nichos ecológicos<sup>4</sup>, seja por seleção sexual<sup>5</sup>, ou até mesmo por hibridização<sup>6</sup>. A este fenômeno de ramificação dá-se o nome de *cladogênese*<sup>7</sup>.

<sup>2</sup>O termo "anagênese" é derivado das palavras gregas "ana", que significa "para cima" ou "para trás", e "gênese", que significa "origem" ou "criação". O termo foi introduzido pela primeira vez pelo biólogo alemão Ernst Mayr em 1954 em seu livro "Change of genetic environment and evolution".

<sup>3</sup>Isolamento geográfico: Quando uma população de uma espécie é fisicamente separada do restante da espécie, pode ocorrer evolução divergente devido a diferentes pressões seletivas, deriva genética e taxas de mutação. Com o tempo, a população isolada pode acumular diferenças genéticas e se tornar uma espécie distinta.

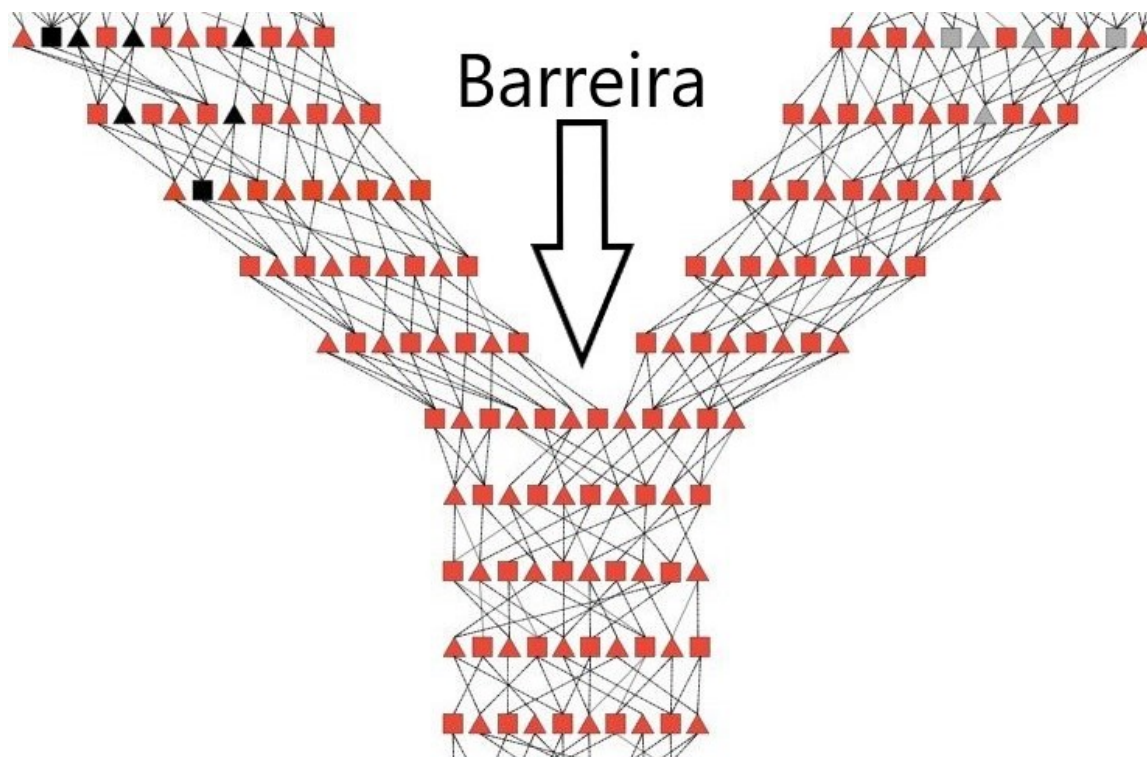
<sup>4</sup>Diferenciação ecológica: Quando diferentes subpopulações de uma espécie se adaptam a diferentes nichos ecológicos, como fontes de alimento ou habitats diferentes, elas podem passar por evolução divergente e eventualmente se tornarem diferentes espécies.

<sup>5</sup>Seleção sexual: Quando a seleção sexual impulsiona a evolução do isolamento reprodutivo entre subpopulações de uma espécie, pode levar à cladogênese. Por exemplo, se as fêmeas preferem machos com certas características, pode ocorrer a evolução de diferentes formas morfológicas de machos em diferentes subpopulações que não são mais capazes de se reproduzir entre si.

<sup>6</sup>Hibridização: Eventos de hibridização podem levar à formação de novas espécies se a prole híbrida estiver reprodutivamente isolada de ambas as populações parentais e puder estabelecer uma população distinta.

<sup>7</sup>O termo "cladogênese" deriva das palavras gregas "klados", que significa "ramo", e "genesis", que significa "origem" ou "criação". O termo foi introduzido pela primeira vez pelo biólogo alemão Ernst Mayr em 1954 em seu livro "Change of genetic environment and evolution".

**Figura 2.5:** Processo de cladogênese. A barreira, indicada pela seta, pode tanto ser física quanto genética, e até mesmo cultural.



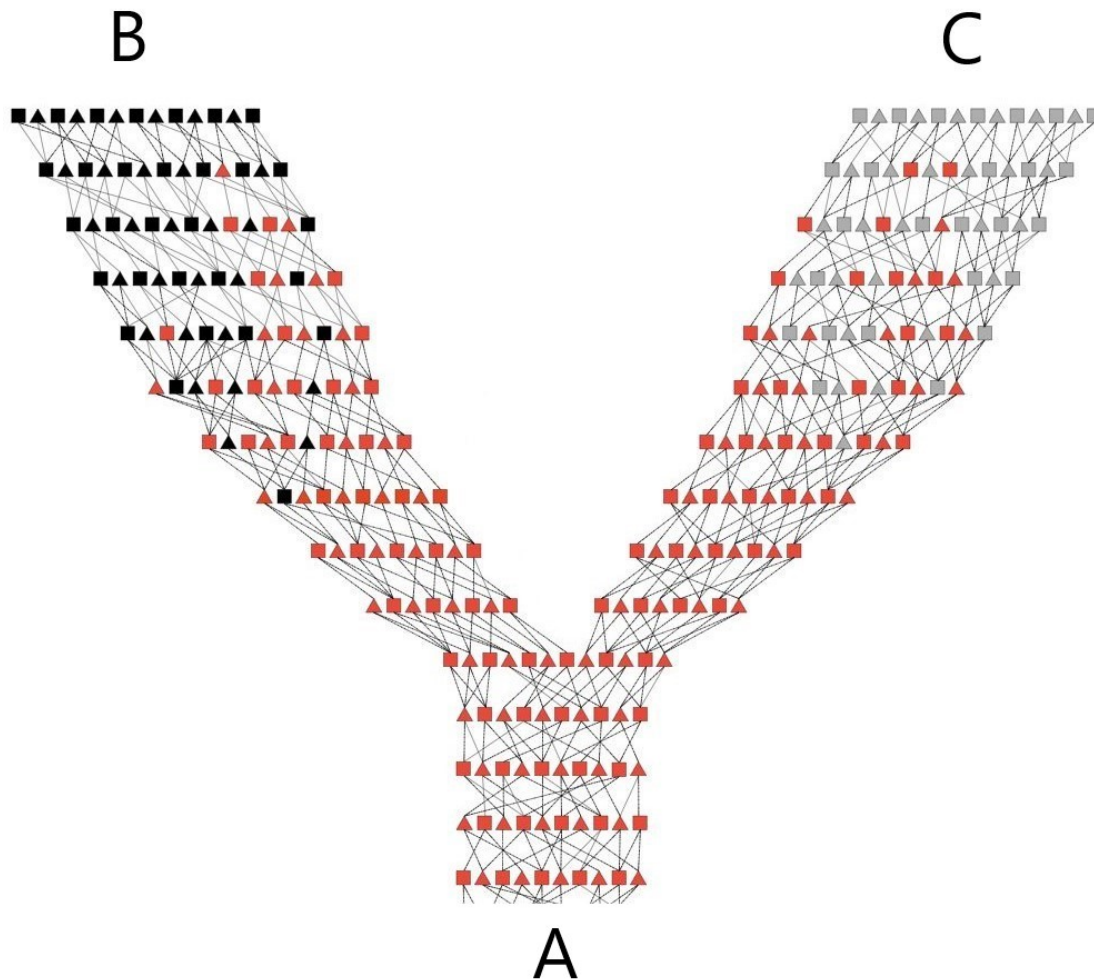
Fonte: [26]

Hennig nomeou a relação de descendência entre a população antes da cladogênese e a que ocorre depois de relação *filogenética*<sup>8</sup>. O processo anagenético então passa a ocorrer nas duas linhagens novas, e caso não interrompido, as duas populações se diferenciarão tanto ao longo das gerações que uma não poderá mais se reproduzir com a outra, ou seja, se tornam espécies diferentes. Ao processo pelo qual novas espécies são formadas a partir de uma espécie ancestral dá-se o nome de *especiação*, e ao processo pelo qual as populações de uma espécie se tornam mais diferentes ao longo do tempo dá-se o nome de *divergência evolutiva*, processo este responsável por toda a biodiversidade do planeta.

<sup>8</sup>O termo "filogenético" vem das palavras gregas "phulon", que significa "raça" ou "tribo", e "genea", que significa "origem" ou "descendência". O termo foi introduzido pela primeira vez em 1866 pelo biólogo alemão Ernst Haeckel em sua obra "Generelle Morphologie der Organismen" (Morfologia Geral dos Organismos) para descrever as relações evolutivas entre diferentes organismos [27].



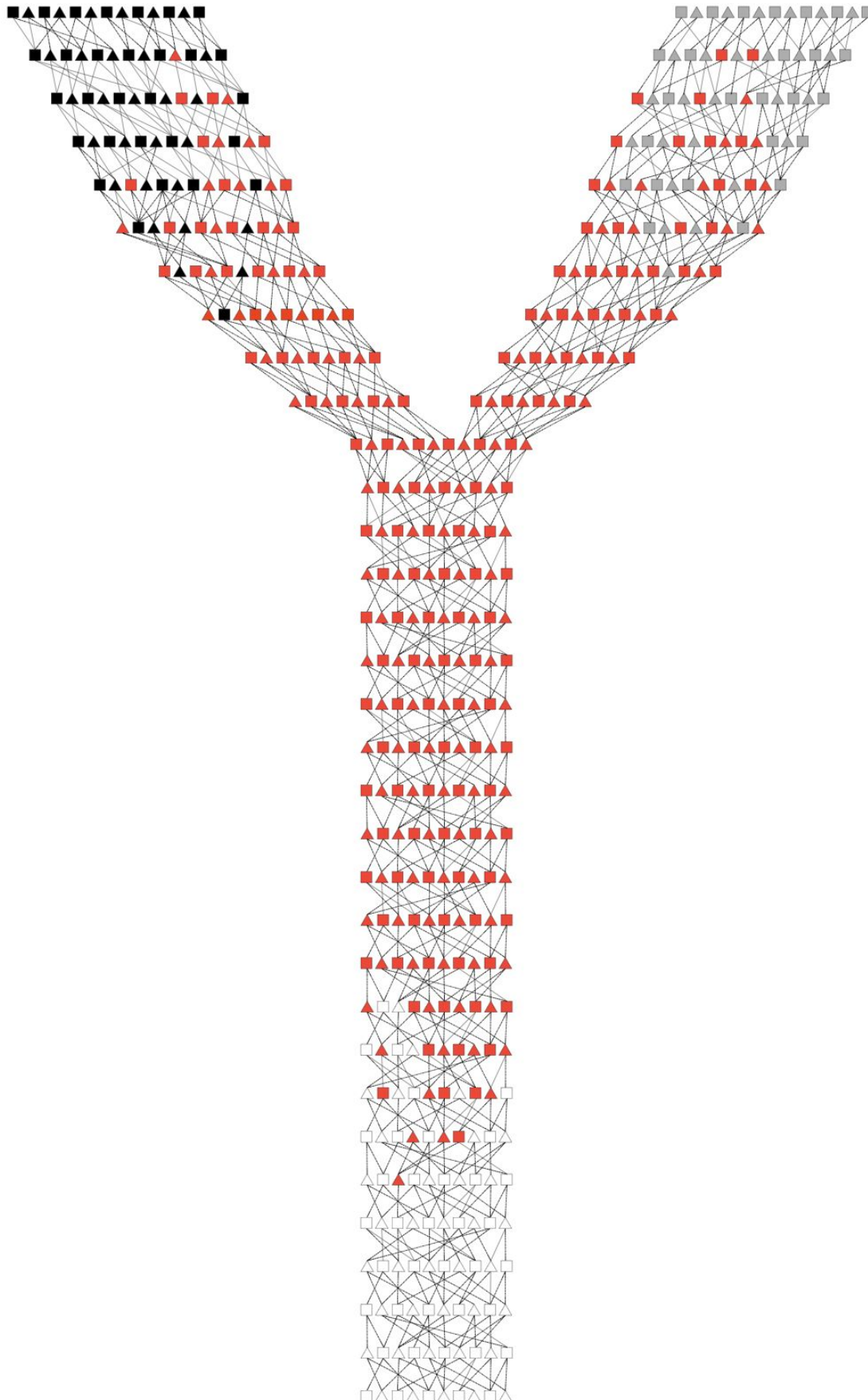
**Figura 2.6:** Os indivíduos das populações B, à esquerda, e C, à direita, por não mais se reproduzirem entre si, adquirem caracteres diferentes, assim gradualmente se diferenciando. Os caracteres novos são representados pelas cores cinza e preto.



Fonte: [26]

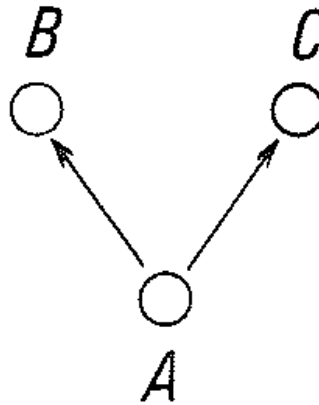
A especiação é um processo que ocorre ao longo de milhares de gerações, levando milhões de anos. Todavia, em organismos de reprodução rápida como bactérias, a dinâmica acima descrita pode ser vista em tempo real. A divisão de uma população em duas, chamada na cladística de dicotomia, é um postulado metodológico, não uma lei imutável das dinâmicas populacionais.

Figura 2.7: Relações tocoenéticas de uma população de animais que sofre anagênese, cladogênese, e eventualmente especiação.



Notemos que a Figura 2.7, que representa as relações toco-genéticas, pode ser reduzida à Figura 2.8, que representa as relações filogenéticas entre as espécies original e descendentes.

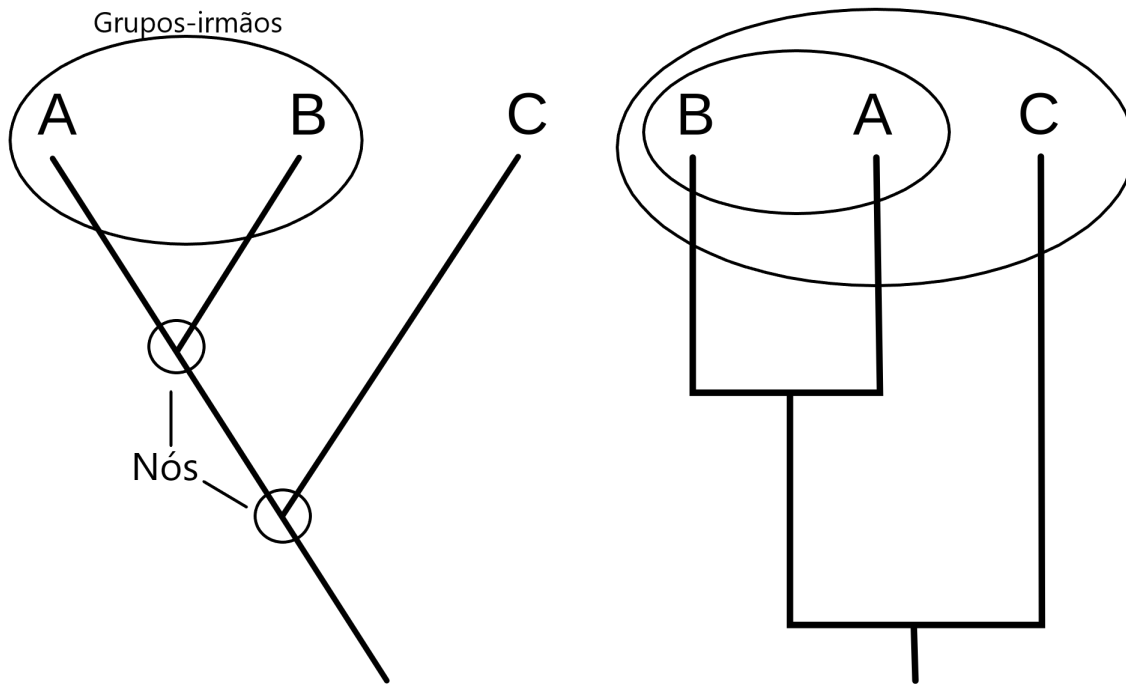
**Figura 2.8:** Representação simples da relação entre a população ancestral A e as populações de duas espécies distintas B e C que vieram da população A.



Fonte: HENNIG, W. *Phylogenetic Systematics*. [S.l.]: University of Illinois Press, 1979. [8, p.18]

A este tipo de representação das relações filogenéticas dá-se o nome de *árvore filogenética*, ou *cladograma*, visto que essencialmente estes gráficos representam as cladogêneses. O formato do cladograma em nada afeta a informação apresentada, desde que a topologia seja conservada, isto é, desde que os pontos de ramificação (ou nós) e as espécies ligadas a esses grupos de ramificações sejam conservados. Os grupos que vêm de um mesmo nó são chamados de grupos-irmãos. Na Figura 2.9, por exemplo, A é grupo-irmão de B, e o grupo de A e B juntos é grupo-irmão de C.

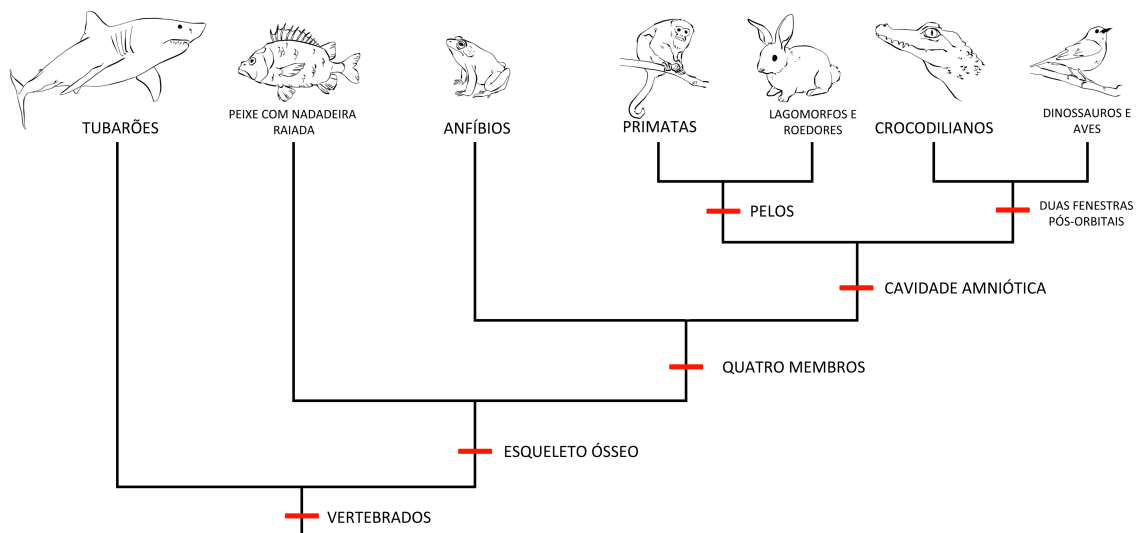
**Figura 2.9:** Cladogramas em formatos diferentes, mas que representam as mesmas relações filogenéticas. Os nós estão destacados dentro de círculos e os grupos-irmãos, dentro das elipses.



Fonte: Elaborada pelo autor com base em [28].

Deve-se ressaltar que as relações filogenéticas, enquanto neste contexto restritas apenas a populações de indivíduos, na verdade são muito mais gerais. Se, ao invés de uma população ancestral e suas descendentes, for trabalhado um grupo de populações ancestrais e seus grupos descendentes, a ideia de relação filogenética ainda se aplica. Por exemplo, o grupo de todas as espécies com vértebras se dividiu em um grupo de espécies com esqueleto cartilagenoso e um com esqueleto ósseo, sendo o caráter de relevância a ossificação do esqueleto. O grupo de espécies com esqueleto ósseo se diversificou ainda mais, como pode ser visto na Figura 2.10.

Figura 2.10: Cladograma do grupo dos vertebrados com caracteres principais de cada sub-grupo.



Fonte: [29]

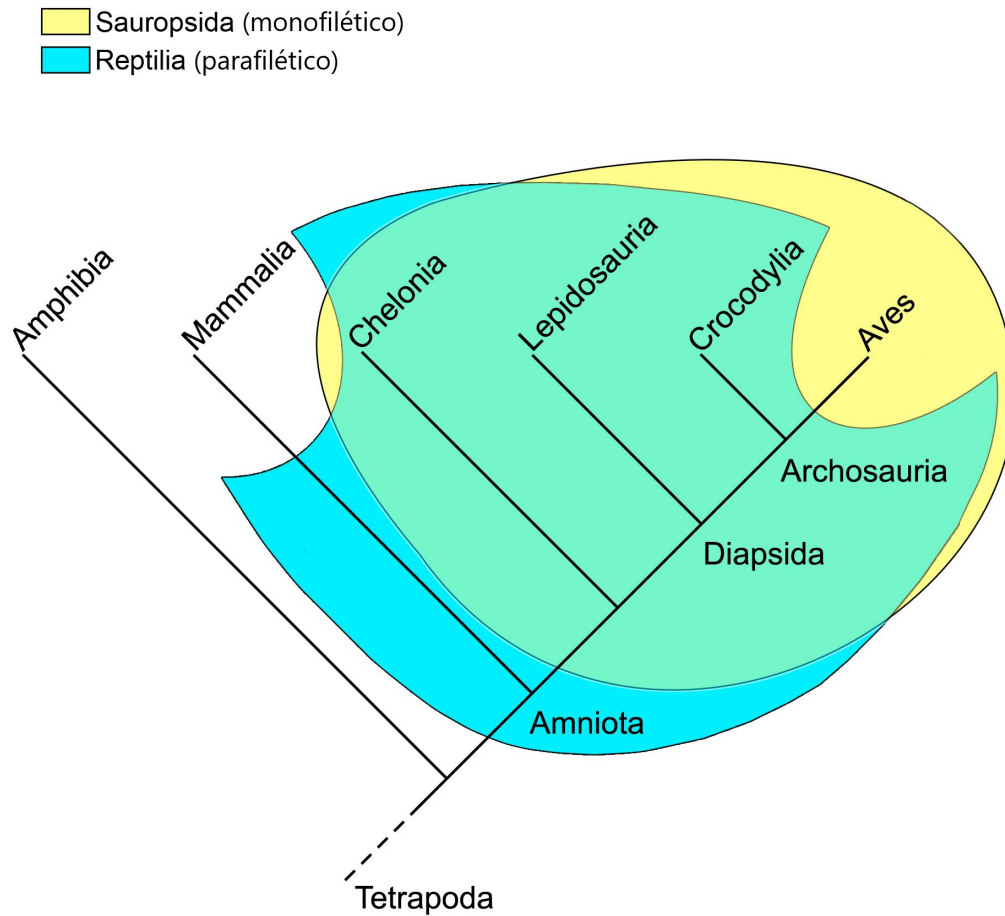
Chama-se, de forma genérica, *grupo* o conjunto de seres vivos que compartilham um ancestral em comum. O grupo dos Osteichthyes, isto é, todos os seres com esqueleto ósseo, tem como *grupo-irmão* o grupo dos Condrictes<sup>9</sup>, grupo-irmão sendo o grupo que compartilha o ancestral em comum com o grupo analisado. Um grupo que inclui o ancestral comum e todos os seus descendentes é chamado de *grupo natural*, ou *grupo monofilético*<sup>10</sup>. Um grupo que deixa de incluir um descendente ou ancestral é dito *parafilético*<sup>11</sup>.

<sup>9</sup>Composto por tubarões, raias, e quimeras.

<sup>10</sup>Do grego antigo "monos", que significa "único" ou "sozinho", e "phylon" (φυλον), que significa "raça", "tribo" ou "linha de descendência".

<sup>11</sup>Formada a partir do prefixo "para-", que significa "ao lado de", "junto a", e do termo "filético", que vem do grego "phylon", e do sufixo "-ético", que denota relação, pertencimento ou afinidade.

**Figura 2.11:** Comparação entre grupos parafilético (azul) e monofilético (amarelo) dentro do grupo dos tetrápodes.



Fonte: [30]. Tradução e edição feita pelo autor.

A diferença metodológica mais importante entre a sistemática filogenética e a fenética vem justamente da consideração do modelo descrito na análise das relações dos seres vivos. Enquanto na fenética as características eram analisadas em um contexto de mera semelhança ou diferença, na cladística elas são tratadas como *caracteres*, ou seja, novidades evolutivas daquela linhagem. É justamente ao atribuir a ideia de novidade ou antigo às características analisadas que se pode fazer uma reconstrução cronológica das relações filogenéticas, assim reconstruindo o processo evolutivo. Enquanto o processo de análise matemática permanece quase inalterado da escola fenética para a filogenética, a base filosófica é diferente, o que leva a resultados distintos. O método de construção de uma matriz de caracteres é:

1. Identifica-se um conjunto de caracteres que sejam relevantes ao estudo. Esses caracte-

res devem ser capazes de distinguir diferentes táxons<sup>12</sup> e devem ser informativos do ponto de vista evolutivo. Tipos comuns de caracteres usados em estudos filogenéticos incluem caracteres morfológicos, moleculares, comportamentais e ecológicos.

2. Uma vez definido o conjunto de caracteres, cada táxon é pontuado para cada característica. Os estados dos caracteres são geralmente codificados como binários (0 ou 1) ou multiestados (0, 1, 2, etc.). Os estados dos caracteres devem ser baseados em características observáveis que possam ser quantificadas.
3. Os caracteres pontuados são organizados em uma matriz, onde cada linha representa um táxon e cada coluna representa uma característica.
4. A matriz de caracteres é avaliada para garantir que seja informativa para o estudo. Isso envolve verificar a completude da matriz, avaliar o grau de homoplasia (a ocorrência de estados de caracteres semelhantes devido à convergência ou paralelismo) e verificar a presença de dados faltantes e outras fontes de erro.

Um caráter ser antigo ou uma novidade evolutiva é algo crucial à análise filogenética. Se um caráter já estava presente no ancestral do grupo, ele é dito *primitivo*, ou uma *plesiomorfia*<sup>13</sup>. Se o caráter é novidade do grupo, não presente nos grupos-irmãos, então ele é dito *derivado*, ou uma *apomorfia*<sup>14</sup>. Se uma plesiomorfia ou apomorfia é compartilhada por vários grupos, então são ditas respectivamente *simplesiomorfia* e *sinapomorfia*<sup>15</sup>.

Em análises filogenéticas, plesiomorfias são ignoradas pois, ao se construir a matriz de caracteres, a plesiomorfia é equivalente a uma coluna inteira constituída apenas de 1 ou 0, daí a necessidade de buscar apenas caracteres apomórficos. Incluir plesiomorfias traria dificuldades ao se construir a topologia da árvore filogenética ou ao tentar encontrar características que dependem do determinante da matriz. Um exemplo dessa situação é a matriz dada na equação (2.1).

$$C_{5 \times 5} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (2.1)$$

<sup>12</sup>Do grego *τάξις* (táxis), que significa "ordem" ou "arranjo", e do sufixo "on", que indica uma unidade ou entidade, táxon é uma unidade taxonômica, isto é, refere-se a um grupo de organismos que são considerados como uma unidade na taxologia, seja a nível de espécie, gênero, família, ordem, classe, filo ou qualquer outro nível hierárquico da classificação biológica.

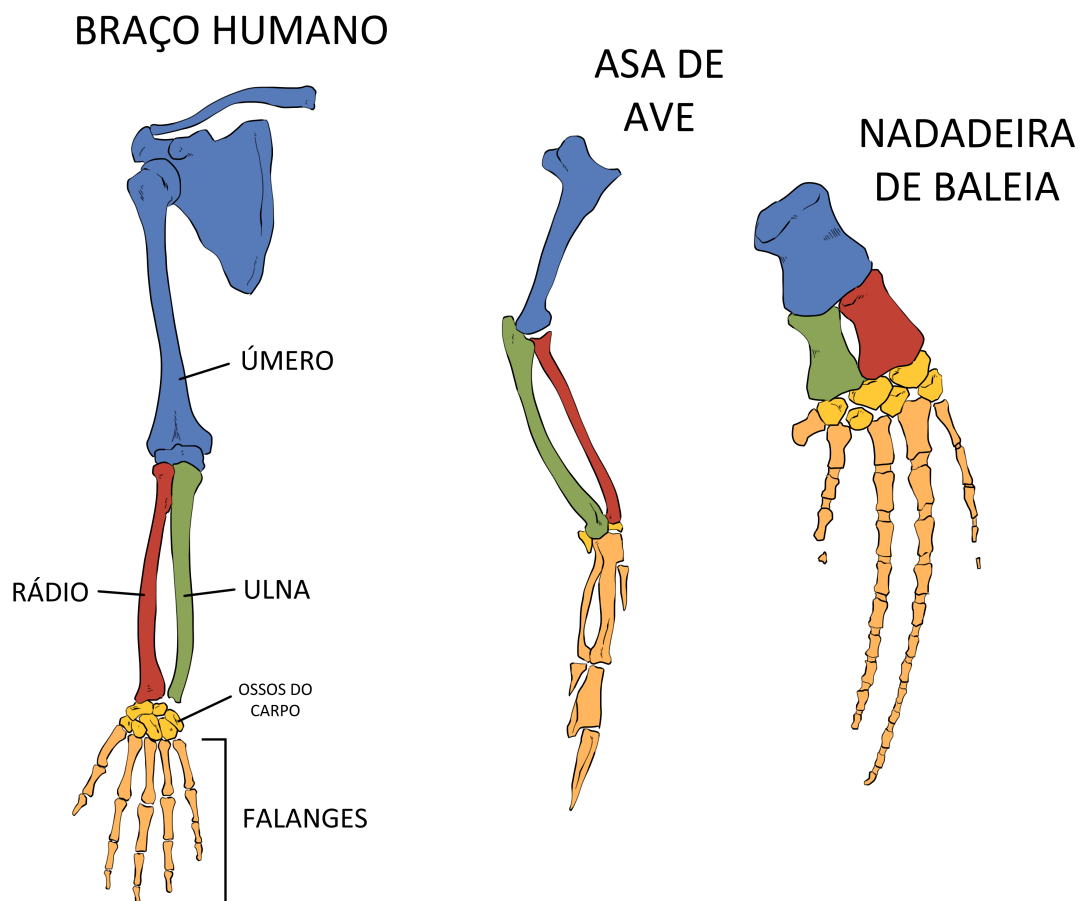
<sup>13</sup>Do grego *πλησιος* (plesios), que significa "próximo", "semelhante", e *μορφή* (morphē), que significa "forma".

<sup>14</sup>Do grego antigo, formada pelos termos "apo", que significa "fora", e "morphē".

<sup>15</sup>O radical "sin" em sinapomorfia vem do grego *συν* (syn), que significa "junto" ou "com".

Quando características de dois ou mais grupos de seres vivos são semelhantes pois têm origem nas características de um ancestral comum, diz-se que essas características são *homólogas*. Um exemplo clássico de homologia trata-se dos membros dos tetrápodos, representado na Figura 2.12.

Figura 2.12: Homologia dos membros anteriores de diferentes tetrápodos.



Fonte: [31]

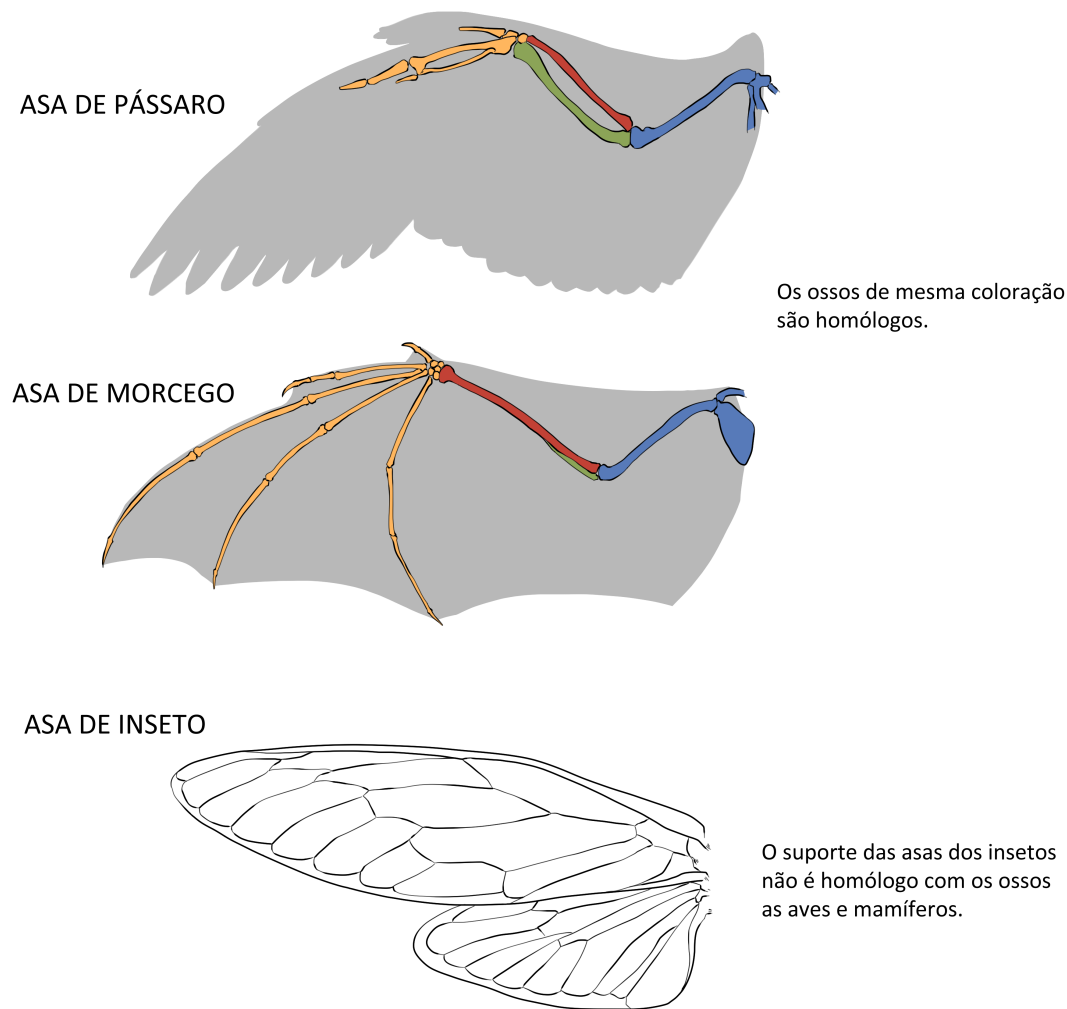
Existem diferentes tipos de homologia. A homologia anatômica refere-se a características estruturais semelhantes, como ossos ou órgãos, que podem ter diferentes funções em organismos distintos. Já a homologia desenvolvimental envolve a semelhança na



sequência de desenvolvimento embrionário entre organismos relacionados. Por fim, a homologia molecular refere-se a semelhanças nas sequências de DNA, RNA ou proteínas entre organismos diferentes.

Nem toda semelhança, no entanto, trata-se de uma homologia. Quando sob mesmas pressões ambientais, seres vivos com ancestrais muito distantes desenvolvem adaptações semelhantes. A este fenômeno dá-se o nome de *convergência evolutiva*. Um exemplo clássico de convergência evolutiva é o desenvolvimento de asas pelo grupo dos morcegos e das aves, cujo ancestral em comum mais recente existiu por volta de 300 milhões de anos atrás [32]. Diz-se que estruturas cuja semelhança é resultado de uma convergência evolutiva são *análogas*. As asas de insetos, por exemplo, são análogas tanto às asas de morcegos quanto às das aves, e claramente não tem origem nenhuma em comum com as asas desses dois grupos.

**Figura 2.13:** Comparação entre asas de insetos, aves, e morcegos. A estrutura óssea das asas de aves e morcegos são homólogas entre si, mas não com o suporte das asas de insetos. Todavia a estrutura asa serve a mesma função de voar nos três grupos, tratando-se portanto de uma analogia.



Fonte: [33]

Em alguns casos, uma característica ancestral que foi modificada em uma linhagem evolutiva é posteriormente restaurada, revertida para uma forma semelhante à do ancestral. Em outras palavras, uma característica que foi perdida em uma linhagem retorna novamente ao longo do tempo evolutivo. A este fenômeno dá-se o nome de *reversão*. Um exemplo de reversão é o kiwi, uma ave não voadora endêmica da Nova Zelândia. Ao longo de sua evolução, o kiwi perdeu a capacidade de voar e suas asas se tornaram

pequenas e não funcionais. No entanto, em algumas populações de kiwis, foram observadas reversões, onde as asas se tornaram maiores e mais estruturalmente semelhantes às de aves voadoras. Embora essas aves ainda sejam predominantemente terrestres e não possam voar como suas contrapartes mais antigas, a reversão parcial na estrutura das asas representa um retorno a uma característica ancestral.

Outro fenômeno importante para o estudo da sistemática filogenética é a *homoplasia*<sup>16</sup>. A homoplasia consiste na ocorrência de características semelhantes em linhagens distintas. A convergência evolutiva é um exemplo de homoplasia causada por pressões seletivas semelhantes, mas nem toda homoplasia é resultado de uma convergência. Há casos quando a reversão de uma característica em uma linhagem faz com que ela fique parecida com outra linhagem, o que também configura homoplasia. Um exemplo de homoplasia é a perda de membros nos grupos das cobras e lagartos sem pernas.

---

<sup>16</sup>O termo homoplasia foi utilizado pela primeira vez por Ray Lankester em 1870. Ele é derivado das duas palavras em grego antigo: *ομός* (homós), que significa "similar, igual, o mesmo", e *πλάσσω* (plássō), que significa "dar forma, moldar"[34].

---

---

# CAPÍTULO 3

---

## REDES COMPLEXAS

Muitos fenômenos e estruturas observados na natureza e na sociedade podem ser abstraídos em um conjunto de pontos, ou *nós*, interligados por linhas, ou *ligações*. Um exemplo simples é o de locais interligados por caminhos, passagens, sendo que cada local é representado por um ponto e cada caminho, por uma aresta. Esta forma de abstrair certos sistemas foi originalmente utilizado por Euler ainda em 1736, estudo este que deu origem à *teoria dos grafos*, mas desde então evoluiu para abranger diversos sistemas, como relações sociais [35], ecológicas [16], e estrutura de polímeros [17], dando origem à *teoria das redes complexas*.

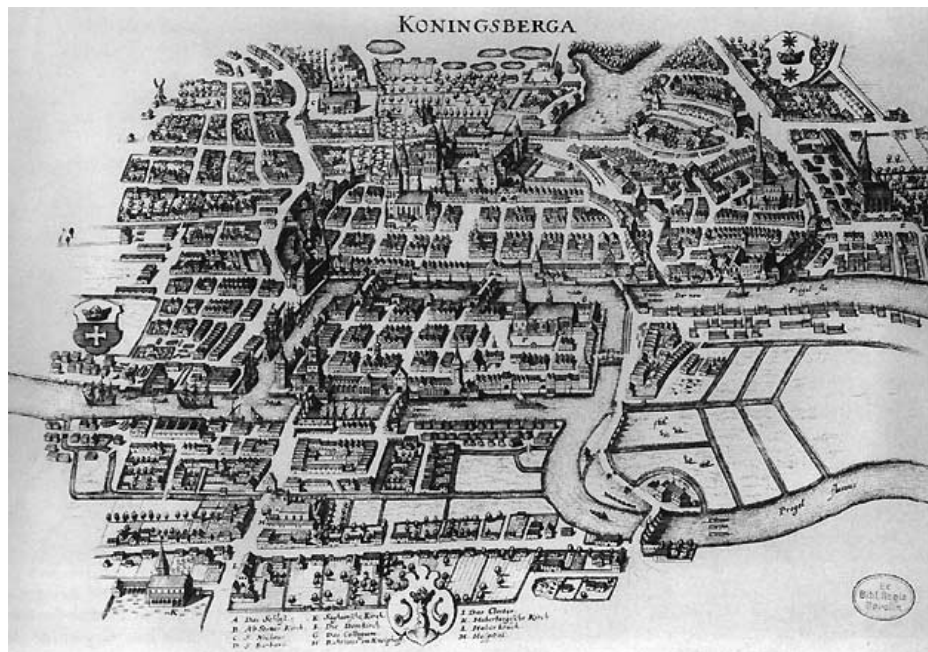
A diferença entre grafos e redes complexas pode ser resumida na abrangência e no objetivo de suas respectivas teorias. A teoria dos grafos estuda todo tipo de grafo, incluindo os completos e triviais, focando-se nas propriedades matemáticas mais do que na interpretação dos mesmos. Já a teoria das redes complexas almeja compreender o sistema estudado a partir da compreensão da estrutura e dinâmica das redes, quase sempre estudando redes obtidas empiricamente, as quais são nada triviais.

O estudo de redes complexas avançou bastante após o desenvolvimento da rede mundial de computadores e dos sites de redes sociais, como Orkut e Facebook, as quais possuem um número gigantesco de elementos [3]. O modelo de rede é perfeito para estas duas estruturas, visto que ambas consistem de pontos ligados entre si por um vínculo, seja este um cabo, no caso dos computadores, seja o vínculo de amizade entre pessoas, no caso de redes sociais.

### 3.1 Introdução Histórica

A raiz do estudo de redes complexas está no estudo de grafos, a qual por si começou com o que hoje pode-se, ironicamente, chamar de uma rede de transporte. Este estudo fundador é um artigo de Leonhard Euler de 1736 no qual o mesmo apresenta uma solução para o *problema das sete pontes de Königsberg* [14].

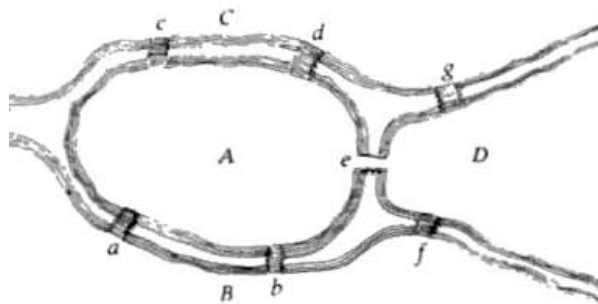
Figura 3.1: Mapa de Königsberg de 1652.



Fonte: [36]

O problema em questão consistia em determinar se era possível sair e voltar para um mesmo ponto da cidade, passando por cada ponte apenas uma vez. Euler simplificou o mapa da cidade para:

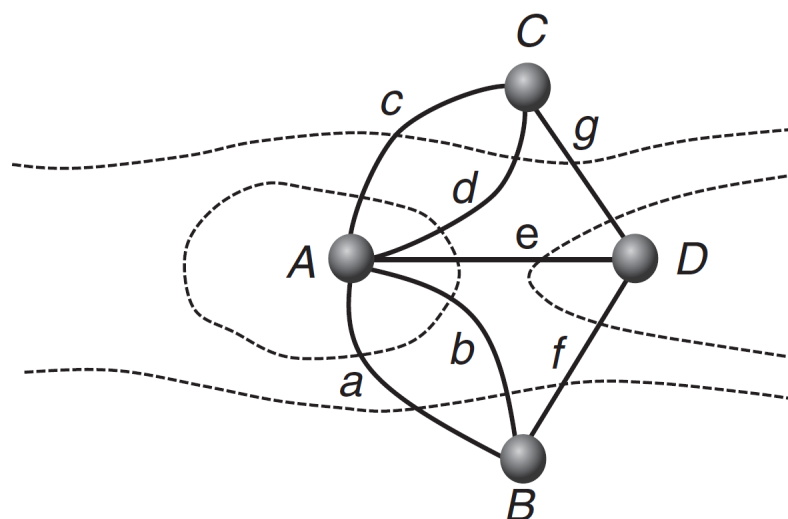
Figura 3.2: Representação das sete pontes de Königsberg.



Fonte: Graph Theory 1736-1936 [14]

Euler, para encontrar a solução do problema, entreteve a ideia de fazer uma lista de todas as possíveis rotas, mas além de perceber que este seria um trabalho tedioso e que muitas rotas seriam redundantes, ele também notou que, para problemas similares com maior número de ilhas e pontes, tal método tornar-se-ia demasiado trabalhoso. Euler portanto decidiu mudar sua metodologia, atribuindo a cada ilha uma letra maiúscula, e a cada ponte, uma minúscula, de forma tal que o ato de ir da ilha A para a B, por qualquer uma das duas pontes, é escrita simplesmente como AB, e fazer a rota AB e BD é denotada por ABD, e assim por diante para rotas mais longas. Pode-se visualizar, de forma abstrata, este método da seguinte forma:

**Figura 3.3:** Representação em forma de grafo do problema das sete pontes de Königsberg.



Fonte: The Structure of Complex Networks: Theory and Applications [4]

Com este método, Euler demonstrou que o número de vezes que uma ilha poderia aparecer na rota correspondia ao número de pontes que levavam a tal ilha, e como o número de ilhas é par e o de pontes, ímpar, uma ponte necessariamente precisará ser atravessada mais de uma vez. A este tipo de problema geométrico ao qual não interessa nenhuma medida de distância ou área, Euler se referiu como *geometria de posição*<sup>1</sup>, a qual veio a se desenvolver na topologia moderna.

A metodologia de Euler foi formalizada por Carl Hierholzer, cujo artigo foi publicado postumamente pelo seu colega Christian Wiener (com ajuda de Jacob Lüroth) em 1873 [14, p.10]. Nestes 142 anos, diversos matemáticos e físicos, como Kirchhoff e Cayley, já haviam contribuído para com o que veio a ser conhecida como **teoria de grafos**. A palavra

<sup>1</sup>Geometria de posição iniciou-se com G. W. Leibniz [14, p.3].

*grafo* vem do inglês *graph* e foi introduzida por James Joseph Sylvester em 1878. Ela é abreviação de *graphic formula* (fórmula gráfica), termo amplamente usado em artigos de química no século XIX [14, p.64-65], sendo atualmente conhecido como *fórmula estrutural*. Fórmulas estruturais são usadas até hoje na química para diferenciar isômeros.

Já a primeira implementação de métodos estatísticos para o estudo de grafos deu-se somente em 1959 [37, p.xii], com o artigo *Graph Theory and Probability* (Teoria de Grafos e Probabilidade) [10], de Pál Erdős, trabalho o qual foi desenvolvido em 1960 por Erdős em parceria com Alfréd Rényi em *On the Evolution of Random Graphs* (Sobre a Evolução de Grafos Aleatórios) [11].

O estudo de redes (networks) começou pelo menos em 1939, bem antes dos resultados de Erdős e Rényi [5]. A aplicação de teoria de grafos no estudo de redes, no entanto, começou a partir dos trabalhos de Frank Harary, que ainda em 1953 começou a aplicar a teoria matemática iniciada por Euler às ciências sociais [12], como o próprio Harary expõe em um artigo publicado em 1985 [5]. Frank Harary foi um dos principais responsáveis pelo desenvolvimento das aplicações da teoria de grafos às mais diversas áreas do conhecimento humano [13; 38], de Física (aplicação em redes elétricas, por exemplo) à Ciência de Computadores, e até Biologia [4, p.111], publicando mais de 700 artigos ao longo de sua vida [39].

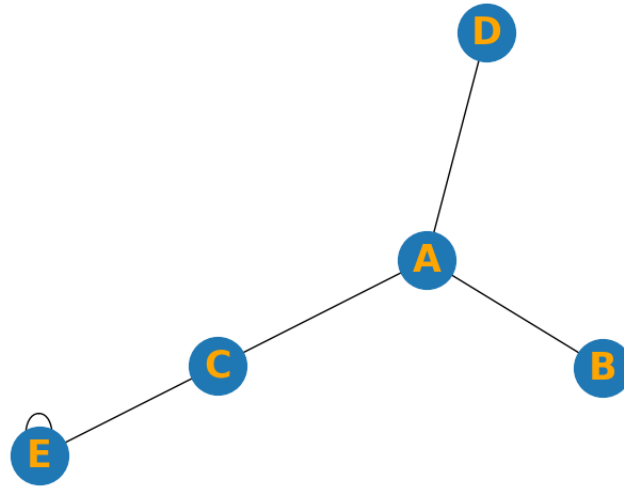
### 3.2 Teoria das Redes Complexas

Matematicamente, redes complexas são idênticas a grafos, havendo apenas uma variação de nomenclatura. Diz-se *nó* ao invés de vértice, e *ligação* ao invés de aresta. Naturalmente também se diz *rede* ao invés de grafo. A principal diferença entre grafos e redes não está no objeto matemático em si, mas no arcabouço teórico e metodológico utilizado para estudar tais objetos, e na atribuição de significado dos componentes das redes.

Sabendo conceitualmente o que são redes complexas e tendo visto suas origens históricas, resta saber como podemos estudá-las matematicamente. Tomemos um exemplo simples de rede de transporte com cinco nós, isto é, locais no caso das redes de transporte, como ponto de partida. Pode-se listar os cinco nós em uma lista ordenada de cinco entradas, cada um correspondendo a um local diferente, lista esta a qual naturalmente contém todos os possíveis destinos. Chamemos estes locais de A, B, C, D e E, e digamos que há uma passagem de ida e volta de A para B, C e D, há uma passagem de C para E, e

há um caminho de E que volta para o ponto de partida. A visualização destes caminhos é simplesmente:

**Figura 3.4:** *Exemplo de caminhos como uma rede complexa.*



Listemos na coluna de uma tabela todos os pontos de partida e na linha, todos os pontos de chegada. Se houver um caminho do  $i$ -ésimo ponto da linha para o  $j$ -ésimo ponto da coluna, por exemplo A e B, então o elemento  $ij$  da tabela é 1, caso contrário é 0. Tem-se:

**Tabela 3.1:** *Tabela de caminhos.*

	A	B	C	D	E
A	0	1	1	1	0
B	1	0	0	0	0
C	1	0	0	0	1
D	1	0	0	0	0
E	0	0	1	0	1

A forma final da tabela implica imediatamente na utilização de matrizes para representar redes:

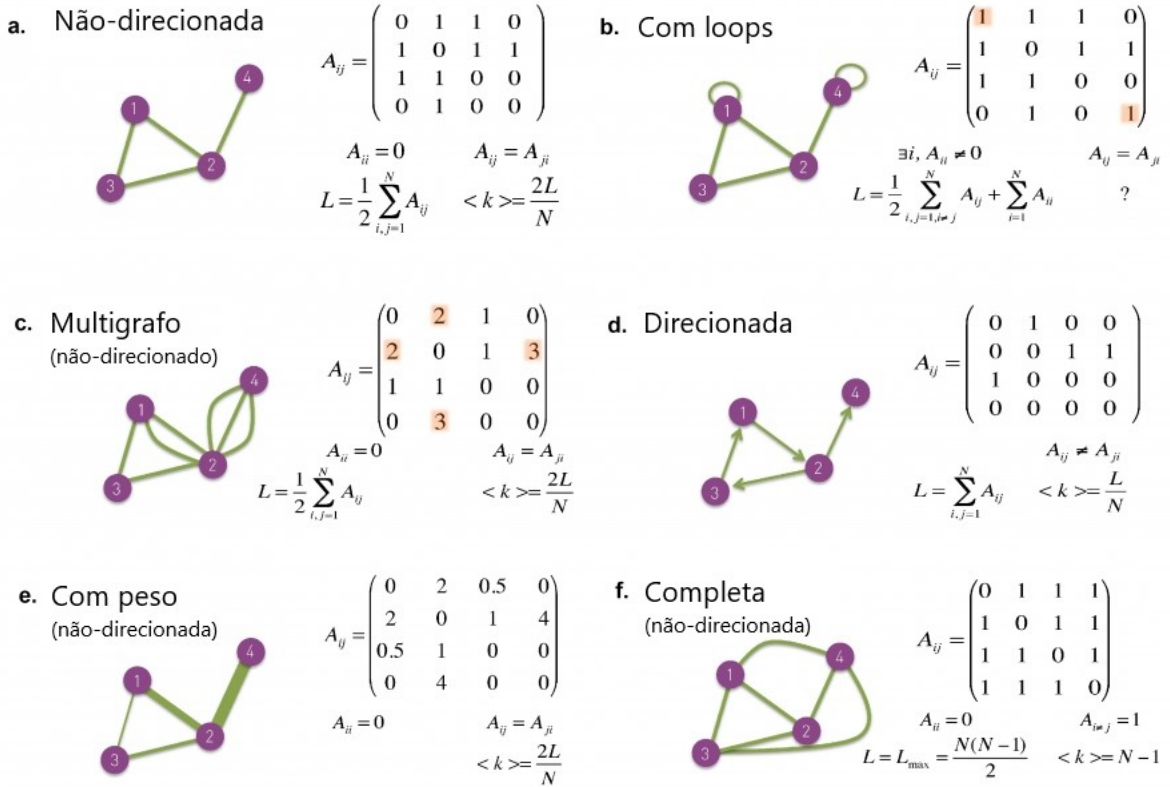
$$A_{5 \times 5} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}. \quad (3.1)$$

A matriz descrita é chamada *matriz de adjacência* da rede. Se a rede de caminhos tivesse, por exemplo, dois caminhos diferentes que levassem de A até B, os elementos  $1 \times 2$  e  $2 \times 1$



seriam 2 ao invés de 1, isto é um exemplo de *multi-grafo*. E se quiséssemos expressar a distância de um ponto ao outro, poderíamos colocar a distância no lugar de 1, este sendo um exemplo de *rede com peso*. Por fim, se um caminho fosse somente de ida, digamos de A a B, então o elemento 1x2 seria 1, mas o 2x1 seria 0, este sendo um exemplo de *rede direcionada*. Redes não-direcionadas, direcionadas, com ou sem peso, são os quatro tipos principais de redes complexas estudadas. Redes podem ser não-conectadas também, isto é pode haver um subconjunto de nós que não tem ligação alguma com o restante dos nós, como uma rede fechada de computadores sem conexão com internet [3].

**Figura 3.5:** Principais tipos de redes complexas acompanhadas de suas matrizes de adjacência. Além das redes não-direcionadas (a), direcionadas (d), e com peso (e), também há um exemplo de rede cuja aresta conecta um nó a ele mesmo (b), um de redes com mais de uma aresta conectando dois nós (c), e um de redes cujos nós estão todos interconectados (f).



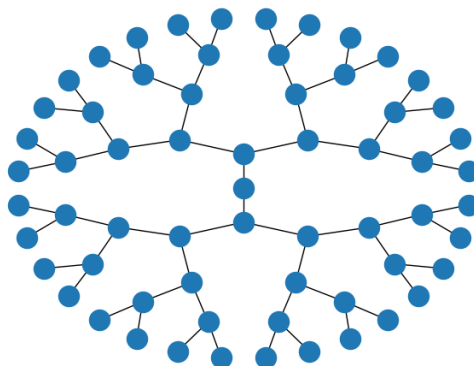
Fonte: [40]. Tradução do inglês para o português feita pelo autor.

Um grupo importante de redes, cujo estudo foi iniciado por Arthur Cayley ainda na teoria de grafos no século XIX, é o grupo de redes acíclicas, ou *redes de tipo árvore*. Estas redes são caracterizadas pela ausência de ciclos, ou caminhos fechados, em suas estruturas. Um exemplo de rede de tipo árvore é a própria rede da Figura 3.4 se for ignorada a ligação do nó E consigo mesmo. Se todos os componentes de uma rede não-conectada,

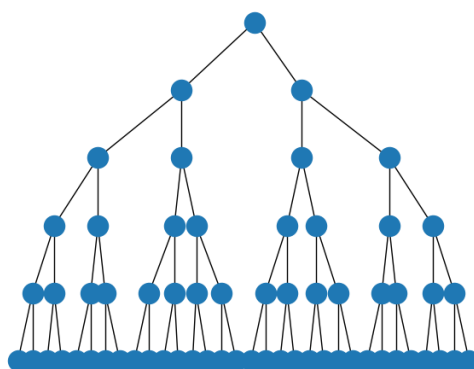
isto é, as sub-redes da rede não-conectada, forem do tipo árvore, então a rede como um todo é dita uma *floresta*.

Redes do tipo árvore podem ser representadas com ou sem raiz. Enquanto a diferença é puramente estética, não afetando em nada a matemática, representar árvores com ou sem raiz pode ser fundamental dependendo da interpretação dada à rede. Por exemplo, todo cladograma é representado como uma árvore com raiz.

**Figura 3.6:** Exemplo de uma rede tipo árvore.



**Figura 3.7:** Mesma rede da Figura 3.6, mas na forma com raiz.



Os nós com apenas uma ligação da rede tipo árvore com raiz, isto é, os da periferia, são chamados de *folhas*. O número de ligações do nó é chamado de *grau* do nó, e é uma das propriedades mais fundamentais da teoria de redes complexas. Voltando ao exemplo da rede da Figura 3.4, o grau  $k$  do  $i$ -ésimo nó é dado pela soma de todos os elementos da linha ou da coluna. Para uma rede qualquer de  $N$  nós, a equação é:

$$k_i = \sum_{j=1}^N A_{ij}. \quad (3.2)$$

Caso a rede fosse direcionada, diferencia-se grau de entrada de grau de saída. E caso as ligações tivessem peso, bastaria normalizar o peso, ou seja, dividir o peso por ele próprio. Pode-se representar com apenas um objeto o grau de cada nó através da *matriz de grau*:

$$D_{N \times N} = [k_i]. \quad (3.3)$$

A matriz de grau da rede da Figura 3.4 é por exemplo:

$$D_{5 \times 5} = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}. \quad (3.4)$$

A diferença entre a matriz de graus e a de adjacência corresponde à *matriz laplaciana*, a qual surge naturalmente em diversos problemas das redes complexas, como a difusão sobre uma rede, que será trabalhado posteriormente.

$$L_{N \times N} = D_{N \times N} - A_{N \times N}. \quad (3.5)$$

Em uma rede não-direcionada, um nó corresponde a uma unidade de grau para dois nós diferentes. Por exemplo, se A está ligado a B, então o A e B tem grau 1, um total de dois graus, mas a rede tem somente uma ligação. Tem-se portanto que o número  $m$  de ligações de uma rede é:

$$m = \frac{1}{2} \sum_{i=1}^N k_i. \quad (3.6)$$

Já o grau médio  $\langle k \rangle$  dos graus da rede é dado, naturalmente, pela média dos graus:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i. \quad (3.7)$$

Mas segue da equação (3.6) que  $\sum_{i=1}^N k_i = 2m$ , logo  $c$  é:

$$\langle k \rangle = \frac{2m}{N}. \quad (3.8)$$

Já o máximo número  $M$  de ligações possíveis de uma rede é dado pela combinação dois-a-dois dos  $N$  nós da redes:

$$M = \binom{N}{2} = \frac{1}{2}N(N-1). \quad (3.9)$$

Com esta grandeza define-se a *conectância* da rede, ou simplesmente densidade. A conectância  $\rho$  é dada pela razão entre  $m$  e  $M$ , e é útil no estudo da dinâmica de redes.

$$\rho = \frac{m}{M} = \frac{\frac{1}{2}N\langle k \rangle}{\frac{1}{2}N(N-1)} = \frac{\langle k \rangle}{N-1}. \quad (3.10)$$

Se uma rede cresce com o tempo, mas a sua conectância tende a um número finito, a rede é dita *densa*. Caso a conectância tenda a 0 conforme  $N$  tende ao infinito, a rede é dita *esparsa*. A densidade da rede é uma propriedade fundamental em estudos de disseminação de doenças, por exemplo. Tendo uma rede social de  $N$  indivíduos, se a densidade for baixa, a doença naturalmente terá dificuldades para se espalhar pela população, mas se a densidade da rede for alta, a doença se difundirá com facilidade.

Quanto ao exemplo da difusão de uma doença, pode-se fazer um modelo acurado da mesma sobre a rede [3, p.152]. Sendo a doença do indivíduo  $i$  uma função  $\psi_i(t)$  tal que todos os nós adjacentes ao  $i$  venham a contrair a doença, tem-se que a derivada temporal de  $\psi_j$ , vizinho do indivíduo  $i$ , é:

$$\frac{d\psi_j}{dt} = C \sum_i^N A_{ij}(\psi_i - \psi_j). \quad (3.11)$$

Sendo  $C$  a constante de difusão específica ao problema, e  $A_{ij}$  o elemento da  $i$ -ésima linha e  $j$ -ésima coluna da matriz de adjacência. Notemos que este modelo de difusão não se limita ao problema de doenças, mas pode ser aplicado a qualquer fenômeno no qual uma grandeza é transmitida de um nó aos seus vizinhos, como transporte de carga, de uma mensagem ou ideia, notícias, e assim por diante. A forma de transmissão da função  $\psi$ , isto é, a parte entre os parênteses  $\psi_i - \psi_j$  varia conforme o fenômeno escrito, como será visto no capítulo 4, a forma mais usual de difusão é a apresentada na equação (3.11), cujo desenvolvimento é:

$$\begin{aligned}
\frac{d\psi_j}{dt} &= C \sum_i^N A_{ij}(\psi_i - \psi_j) \\
&= C \sum_i^N A_{ij}\psi_i - C \sum_i^N A_{ij}\psi_j \\
&= C \sum_i^N A_{ij}\psi_i - C k_j \psi_j \\
&= C \sum_i^N (A_{ij} - k_j \delta_{ij}) \psi_i \\
&= C \sum_i^N (A_{ij} - k_j \delta_{ij}) \psi_i.
\end{aligned} \tag{3.12}$$

Definindo um vetor  $\vec{\psi}$  cujas entradas são as funções  $\psi_j$ , pode-se re-escrever a equação (3.12) em forma matricial:

$$\begin{aligned}
\frac{d\vec{\psi}}{dt} &= C(\vec{A} - \vec{D})\vec{\psi} \\
&= -C\vec{L}\vec{\psi}.
\end{aligned} \tag{3.13}$$

Sendo  $\vec{A}$  e  $\vec{D}$  as matrizes de adjacência e grau, respectivamente. Ressalta-se que a equação resultante é a equação diferencial de Laplace, porém com o sinal do laplaciano invertido graças à definição histórica da matriz laplaciana.

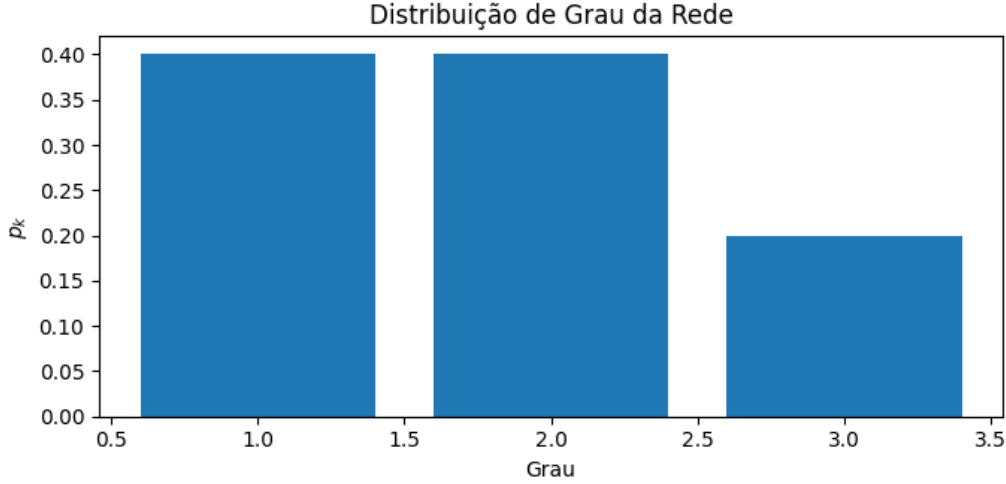
### 3.2.1 Grandezas Estatísticas de Redes Complexas

Estando definidas as grandezas básicas de uma rede, resta definir as grandezas que de fato trazem informação sobre os sistemas estudados através da teoria das redes complexas. O estudo estatístico de redes é vasto e um tanto complexo, e portanto limitar-nos-emos às gradezas e definições relevantes ao estudo de cladogramas, aqui tratados como redes não-direcionadas. As grandezas relevantes são a *distribuição de grau* e o *coeficiente de Pearson*.

Tendo uma rede real, uma pergunta imediata é, escolhendo um nó aleatório, qual a probabilidade deste nó ter grau  $k$ ? A resposta para essa pergunta é uma distribuição de probabilidade, chamada de distribuição de grau, que como o nome indica é a distribuição estatística dos graus dos nós da rede [3, p.243-247]. Mais especificamente, trata-se da

distribuição da grandeza  $p_k = \frac{n(k)}{N}$ , sendo  $n(k)$  o número de nós com o grau  $k$  e  $N$  o número de nós da rede. Os  $p_k$  representam a probabilidade de, escolhendo um nó aleatoriamente da rede, este nó ter grau  $k$ . Voltando ao exemplo da Figura 3.4, a sua distribuição de grau é:

Figura 3.8: Distribuição de grau da rede da Figura 3.4



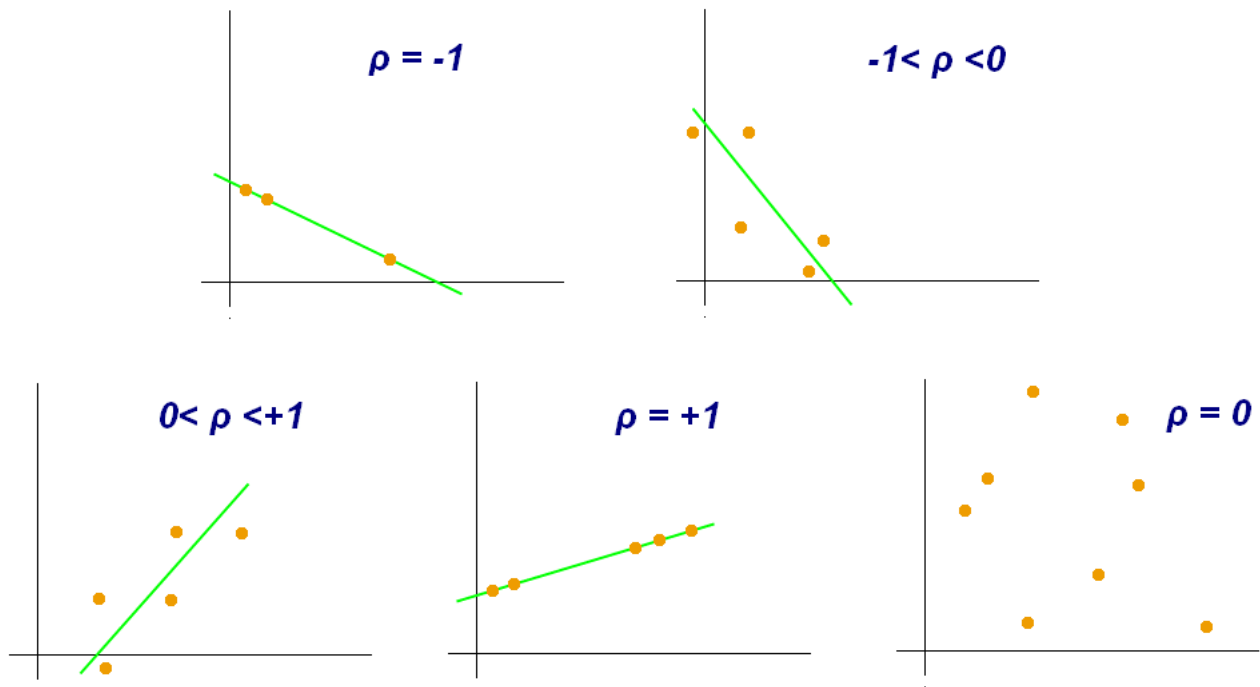
Saber se nós de grau alto se conectam mais com outros nós de grau baixo ou alto, isto é, a correlação entre o grau de um nó e o grau de seus vizinhos, é outra medida importante no estudo das redes. A correlação entre dois conjuntos de dados é denotada matematicamente pelo *coeficiente de Pearson* [3, p.214-215], que para correlação entre o grau dos nós  $i$  e  $j$  e o grau dos seus  $k$  vizinhos é:

$$\begin{aligned}
 \sum_k A_{ik} A_{jk} - \frac{k_i k_j}{N} &= \sum_k A_{ik} A_{jk} - \frac{1}{N} \sum_k A_{ik} \sum_l A_{jl} \\
 &= \sum_k A_{ik} A_{jk} - N \langle A_i \rangle \langle A_j \rangle \\
 &= \sum_k (A_{ik} - \langle A_i \rangle) \cdot (A_{jk} - \langle A_j \rangle),
 \end{aligned} \tag{3.14}$$

que, normalizado, torna-se o coeficiente de Pearson:

$$r_{ij} = \frac{\sum_k (A_{ik} - \langle A_i \rangle) \cdot (A_{jk} - \langle A_j \rangle)}{\sqrt{\sum_k (A_{ik} - \langle A_i \rangle)^2} \sqrt{\sum_k (A_{jk} - \langle A_j \rangle)^2}}. \tag{3.15}$$

Figura 3.9: Significados do Coeficiente de Pearson



Outras duas grandezas comumente analisadas são o *diâmetro* e o *coeficiente de agrupamento*. Tendo dois nós de uma rede, se estes nós estão conectados direta (por uma simples ligação) ou indiretamente (ligados a outros nós conectados entre si), define-se a distância entre eles como a soma das ligações que conectam esses nós. A maior destas medidas é para todos os possíveis pares de nós da rede é chamado de diâmetro. Já o coeficiente de agrupamento é uma medida de quantos caminhos fechados há em uma rede, isto é, a razão entre o número de trios de nós interconectados e o número total de trios conectados. Ambas medidas foram ignoradas neste estudo, pois o diâmetro não tem significado físico relevante e o coeficiente de agrupamento é nulo visto que todas as redes são árvores.

---

---

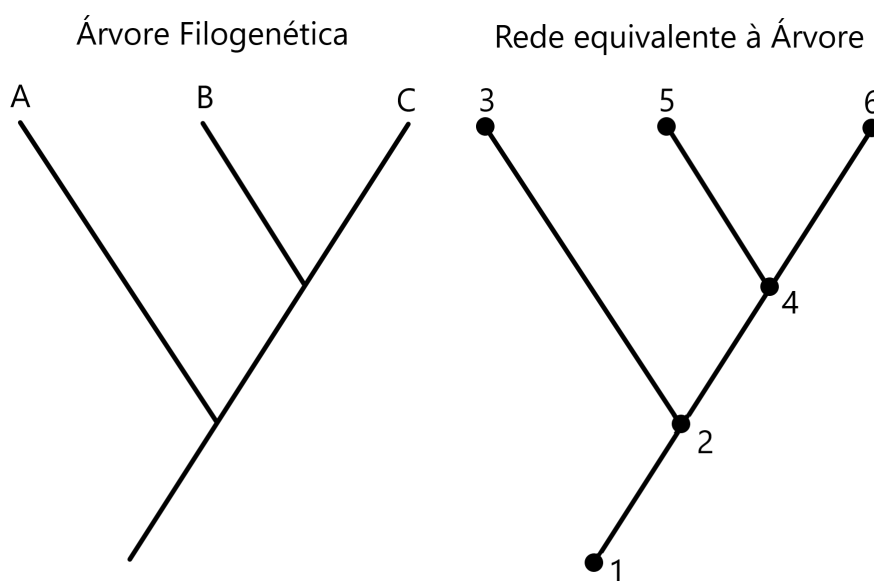
# CAPÍTULO 4

---

## REDES FILOGENÉTICAS

Munido do conhecimento de sistemática filogenética e de redes complexas, uma pergunta naturalmente surge: é possível estudar árvores filogenéticas através do arcabouço teórico das redes complexas? Afinal, um cladograma e uma rede complexas tem estrutura similar:

**Figura 4.1:** *Equivalência entre cladogramas e redes complexas. Os nós e folhas do cladograma se traduzem como nós na rede complexa, e os ramos, como ligações. Tendo mais dados sobre os táxons, é possível usar modelos mais avançados de redes do que uma não-direcionada sem peso.*



Se um cladograma pode ser representado como rede, há algum padrão ou informação relevante na distribuição de grau? E na conectividade? A transmissão de características



pode ser descrita usando a difusão? As outras propriedades de redes complexas representam algo de relevante sobre a evolução do grupo? Estas perguntas motivaram o estudo de árvores filogenéticas no contexto das redes. A este tipo de rede foi dado o nome de *rede filogenética*.

Estudos similares já foram realizados, mas limitados ao escopo da teoria de grafos. Um notável exemplo é o livro *Phylogenetic Networks: Concepts, Algorithms and Applications* publicado em 2011 e de autoria de Daniel Huson, Regula Rupp, e Celine Scornavacca [6]. Outro relevante estudo, realizado entre 2019 e 2020, dos cladogramas de fungos do gênero *Aspergillus*, serviu de base para esta tese [18].

O estudo da cladística de fungos iniciou-se na construção do cladograma das espécies depositadas na micoteca DPUA/UFAM e, apesar da falta de dados mais robustos para garantir confirmação, sugere que cladogramas se comportam como redes livres de escala muito pequenas ou redes livre de escala generalizadas com o grau máximo permitido sendo pequeno. As redes livre de escala são caracterizadas pelo grande número de hubs<sup>1</sup> e pelo fato de que sua distribuição de grau segue uma lei de potência [18].

A grande dificuldade em obter redes filogenéticas de milhares, dezenas, ou até mesmo centenas de milhares de espécies diferentes impossibilitou um estudo mais aprofundado sobre as grandezas estatísticas das redes filogenéticas. Para contornar esta limitação, foi proposto simular as redes com base nas relações tocogenéticas, as quais alicerçam as relações filogenéticas.

No modelo proposto, para determinar o cladograma que melhor aproxima a evolução das espécies, simula-se um ensemble estatístico de cladogramas possíveis e faz-se comparação com os dados reais, buscando minimizar as diferenças. Pode-se também simular as relações tocogenéticas sob uma pressão ambiental específica, o que permite a previsão das possíveis adaptações de organismos de reprodução rápida, como bactérias e vírus em uma pandemia. Isso permitiria otimizar a utilização de recursos e o tempo investido na pesquisa e desenvolvimento de vacinas e remédios, por exemplo. Outra aplicação seria no combate a pragas agrícolas.

---

<sup>1</sup>Hubs, cuja tradução mais precisa ao português possível é "eixo", no contexto de redes complexas são nós que acumulam um grande número de vizinhos. Em redes de transporte, portos e aeroportos movimentados são um exemplo de hub; em redes sociais, uma pessoa com muitos contatos (alguém famoso, por exemplo) é um exemplo de hub.

## 4.1 Modelo

Optou-se por, ao invés de fazer um único código-fonte que simule todo e qualquer tipo de rede filogenética em sua totalidade de fenômenos, dividir o modelo em módulos separados, ou *casos*, que simulam um fenômeno de cada vez, para assim gradualmente se construir um código único geral. Todos os casos possuem uma estrutura em comum: os indivíduos são simulados como nós, e as relações de parentesco serão representadas pelas ligações. Desta forma constrói-se primeiramente a rede toco-genética para depois obter a filogenética.

Cada indivíduo pode se reproduzir com outros da sua geração, com antepassados, com indivíduos da geração seguinte, sua prole, e até ele mesmo (em caso de plantas, por exemplo), sendo a autofecundação permitida, ou não, caso os nós possam ser ligados a eles mesmos. Todas essas possibilidades de reprodução são consideradas em casos diferentes a princípio. Herança de características, passagem do tempo, ciclo da vida, morte, localização geográfica, diferenciação de machos e fêmeas, e seleção natural e sexual são todos casos separados. Os casos são:

- Caso ideal;
- Casos de dinâmica populacional (PD);
- Casos de herança de característica (HC);
- Casos infraestruturais;
- Casos de características biológicas (BC).

Os cinco tipos de casos diferentes podem ser categorizados em dois grupos distintos: casos-teste e casos efetivos. Os casos-teste, o ideal, os de dinâmica populacional, e o de herança de característica, foram desenvolvidos como base para estudar a forma mais efetiva de modelar as redes toco-genéticas, isto é, desenvolver as funções necessárias para redes mais complexas uma a uma. Os casos-teste também testam as grandezas mais importantes das redes: distribuição de grau e coeficiente de Pearson (casos ideal e de controle populacional), e difusão (caso HC). Os casos efetivos são aqueles nos quais fenômenos mais complexos, posição espacial e passagem do tempo por exemplo, são modelados. Para simular geografia, tempo, e as dinâmicas sexuais das populações que

surgem graças ao espaço-tempo, tem-se de mudar a forma como as redes são geradas, daí a diferença dos grupos. Maiores detalhes são dados na seção 4.2.

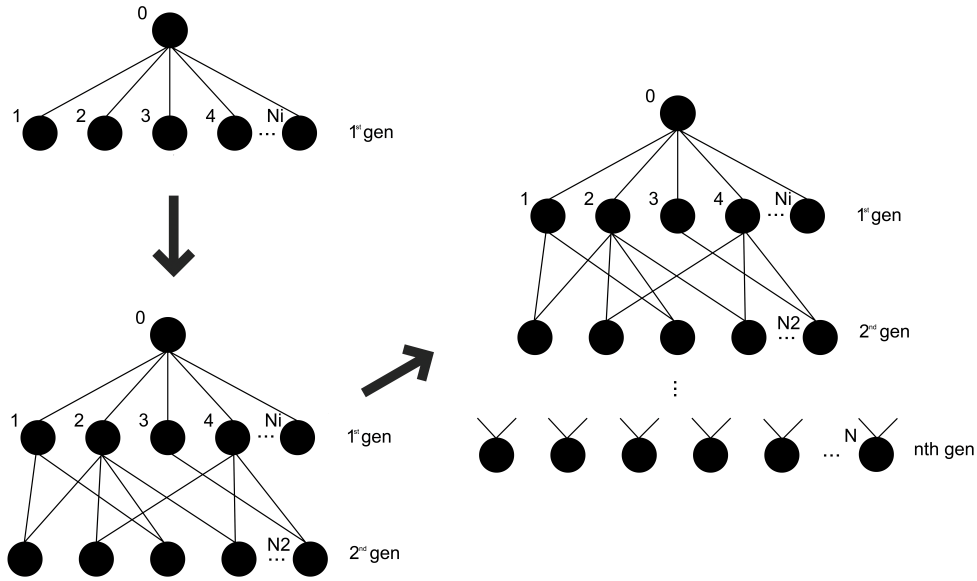
O caso ideal se caracteriza pela separação de indivíduos em gerações, pelo número de nós-filhos de um indivíduo ser aleatório, por todos os indivíduos serem exatamente iguais e hermafroditas, e por não haver cladogênese. O caso ideal foi desenvolvido com a finalidade de servir de piloto para casos complexos, disto isto este estudo se limitou ao caso ideal graças à riqueza de resultados obtidos do mesmo.

## 4.2 Algoritmos

O caso ideal, como dito na seção anterior, serviu de base para todos os casos-teste. Seu algoritmo se resume em:

1. Cria-se o nó 0, chamado de *ancestral primordial*;
2. Cria-se a primeira geração, cujos indivíduos todos estão ligados ao ancestral primordial;
3. É feita uma lista de todos os nós dividida por geração;
4. Cria-se os nós da segunda geração, cujo parentesco é escolhido aleatoriamente;
5. Atualiza-se a lista de indivíduos por geração e cria-se os nós de uma nova geração, cujo parentesco é escolhido aleatoriamente;
6. O item 5 é repetido até a rede chegar ao limite de nós imposto pelo usuário.

A criação do ancestral primordial serve dois propósitos: evitar quaisquer problemas no cálculo dos autovalores da rede, visto que na primeira geração pode ocorrer de um indivíduo não ter filhos, e representar as gerações anteriores à simulação. A importância do segundo ponto dá-se principalmente quando é necessário descrever características. A escolha aleatória de maternidade é válida aqui pois todos os indivíduos são iguais. O limite de número de nós da rede foi imposto pois esta é a forma mais simples de garantir que a simulação seja feita em tempo hábil. O número de nós por geração cresce de forma exponencial conforme o número  $t$  de gerações e com base na população da geração anterior:



**Figura 4.2:** Esquemática da simulação do processo toco-genético como uma rede complexa.

$$n_{t+1} = n_t \cdot e^{R \cdot t+1} \quad (4.1)$$

Alguns dos demais casos foram desenvolvidos e brevemente estudados, porém o caso ideal sozinho gerou resultados interessantes que necessitaram atenção integral da pesquisa.

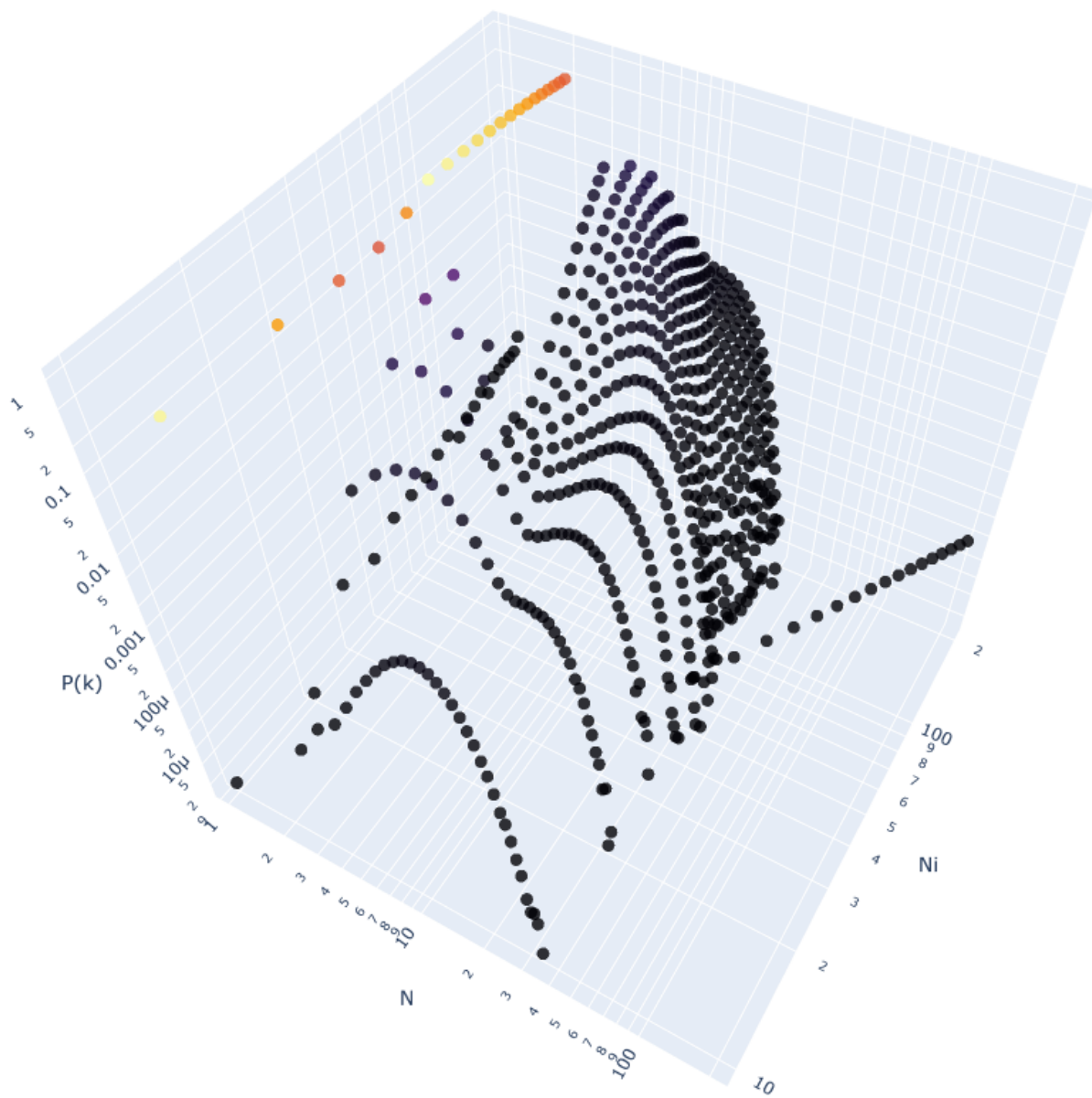
### 4.3 Resultados

Foram simuladas quarenta redes para cada variável, sendo os resultados aqui apresentados a média dos dados dessas simulações. Essas médias são basicamente redes virtuais cujas medidas são a média de cada uma das medidas das quarenta redes simuladas. O número 40 foi escolhido de forma a garantir que a regra prática do teorema central do limite fosse válida.

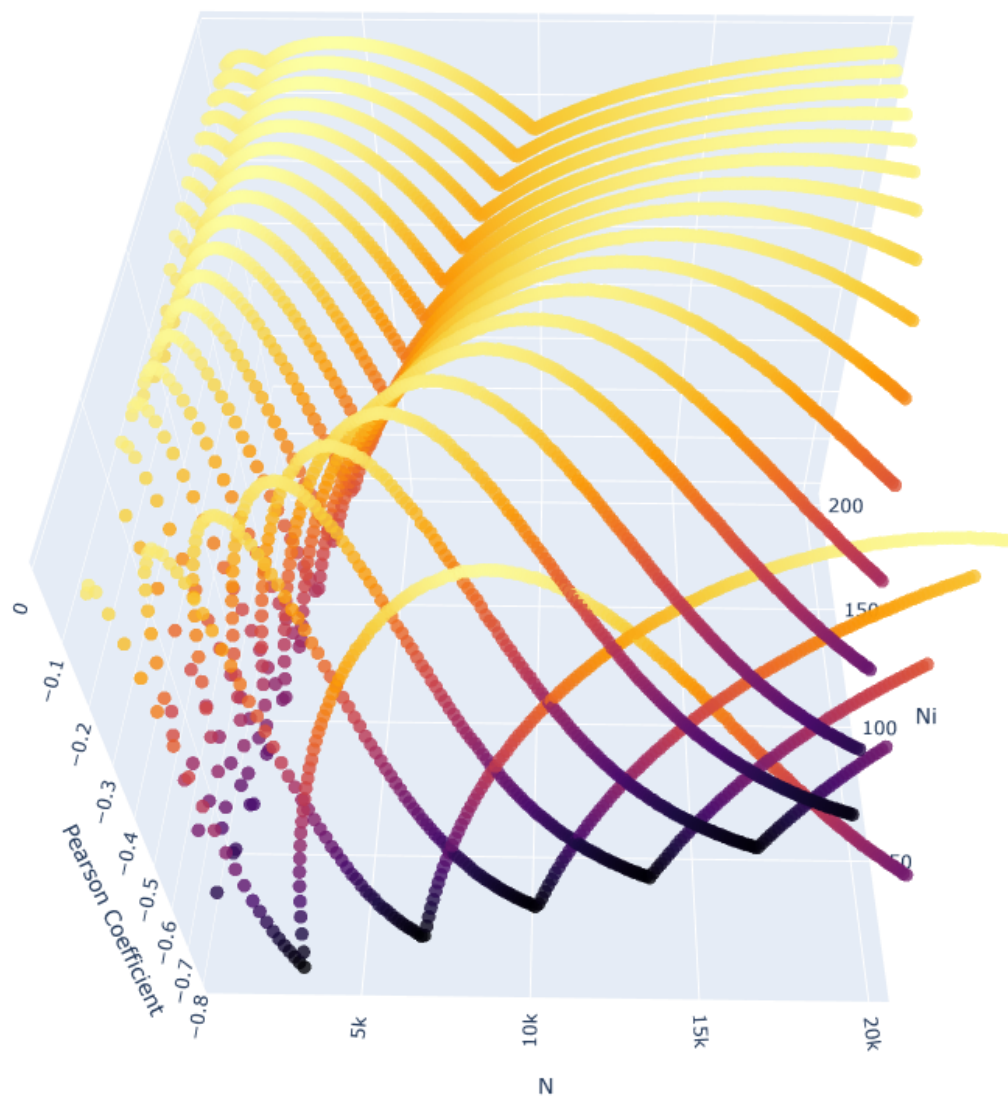
A partir das redes (as árvores genealógicas), duas medidas foram feitas: a distribuição de graus e o coeficiente de Pearson. O coeficiente de Pearson dessas redes indica o quanto há uma tendência na população de que pais com vários filhos tenham descendentes que, por sua vez, não tenham tantos filhos; enquanto a distribuição de graus indica o número médio de relações genéticas diretas que um indivíduo possui, isto é, filhos mais os dois pais. Foram criados dois conjuntos de redes: um com taxa de crescimento constante  $R = 0,5$  e variando  $N_i$  entre 10 e 200 em passos de 10, e outro com  $N_i = 200$  e variando  $R$

entre 0,1 e 0,9 em passos de 0,1.

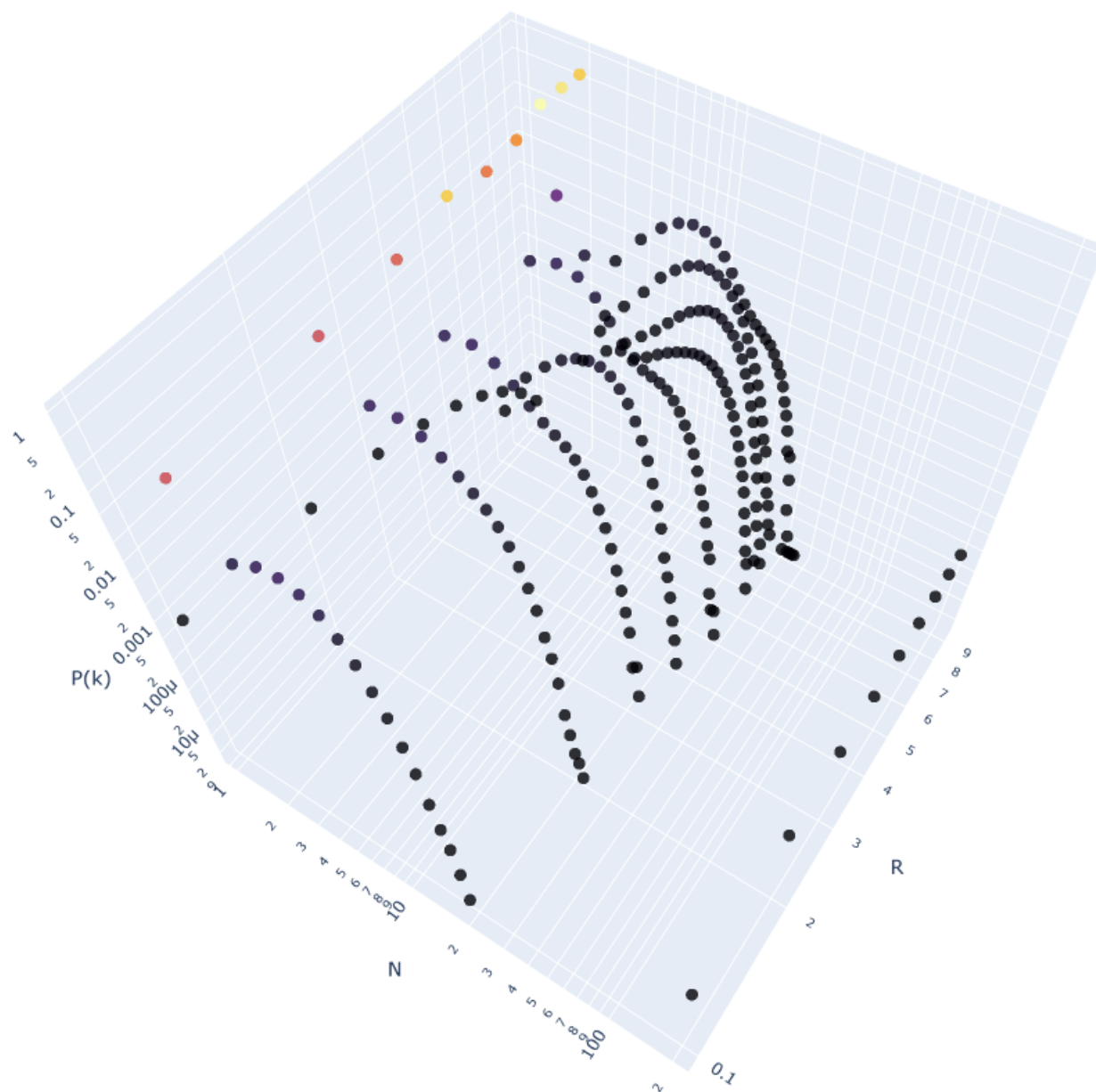
**Figura 4.3:** Distribuição de graus de 20 redes com 20.000 nós e  $R$  fixo.



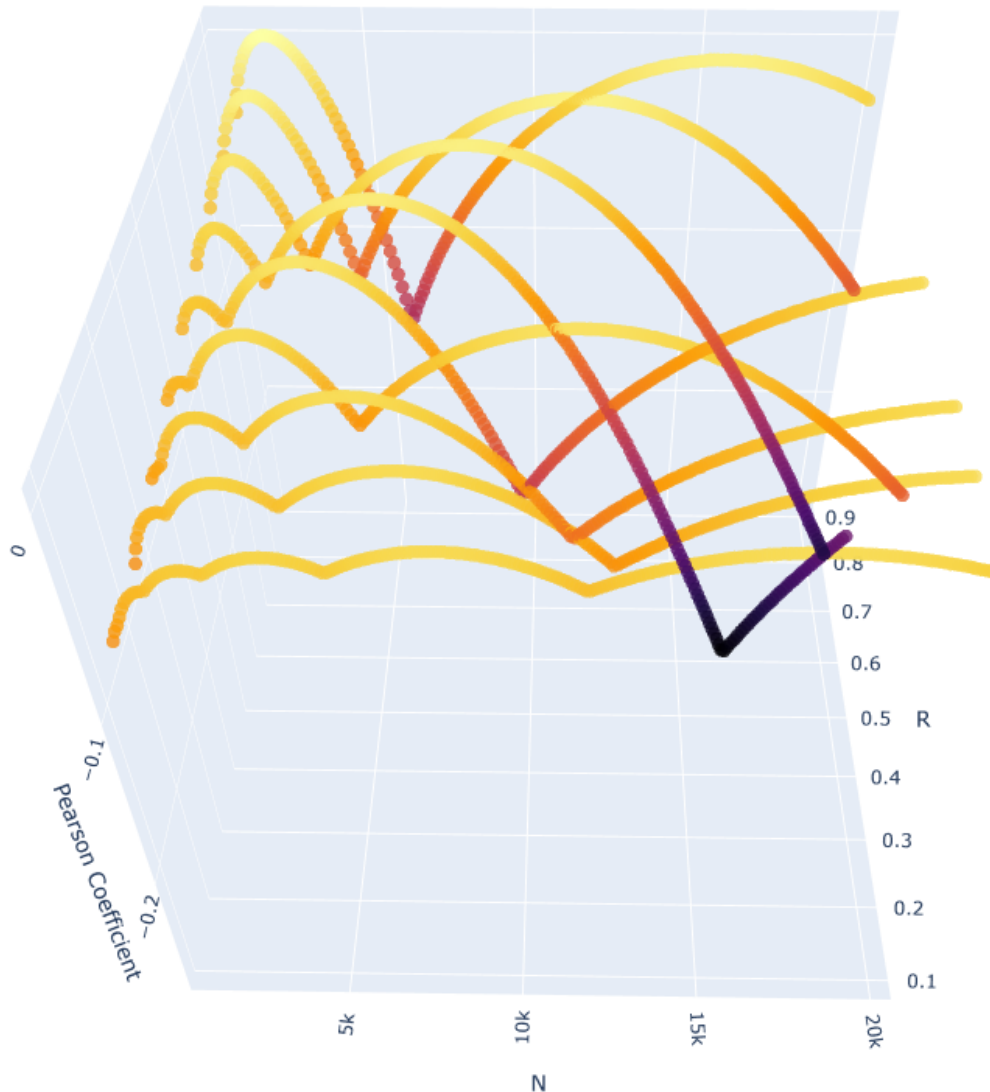
**Figura 4.4:** Coeficiente de Pearson de redes com  $R$  fixo e número total de nós de 1.000 a 20.000 em passos de 1.000.



**Figura 4.5:** Distribuição de graus de 20 redes com 20.000 nós e  $N_i$  fixo.



**Figura 4.6:** Coeficiente de Pearson de redes com  $N_i$  fixo e número total de nós de 1.000 a 20.000 em passos de 1.000.



As cores representam o valor do eixo z, sendo o mapeado para preto, e a cor tendendo ao branco para maiores valores de z. As Figuras 4.3 e 4.5 possuem duas semelhanças. A primeira é que o nó mais frequente em todas as redes é 2, o que mostra que a maioria dos indivíduos possui apenas dois parentes diretos: mãe e pai. A segunda é que a frequência dos graus oscila à medida que o valor no eixo y ( $N_i$  ou R) aumenta. As Figuras 4.4 e 4.6 também compartilham esse comportamento oscilatório, embora a taxa de crescimento tenha um impacto maior no coeficiente de Pearson.

A informação mais importante das distribuições de graus não é o pico principal ( $k = 2$ ), mas sim o segundo em cada curva, que indica o número mais frequente de filhos que os pais possuem. Com isso, seria possível, por exemplo, maximizar os esforços de repopular



espécies ameaçadas em um novo ambiente. Os coeficientes de Pearson mostram o padrão de estagnação da reprodução máxima em uma população. Ou seja, quando  $r$  se aproxima de -1, a taxa de nascimentos é máxima, e quando se aproxima de 0, a população tende a manter a taxa de nascimentos.

A curvas da distribuição de grau na figura 4.3 se assemelham a, senão são de fato, composições de gaussianas. A primeira gaussiana tem pico em  $k = 2$ , e segunda, por volta de  $k = 10$ . Conforme  $N_i$  cresce, o pico tende a valores mais baixos. Isso demonstra que há dois cenários mais frequentes nas redes: o primeiro é dos indivíduos não terem filhos, e o segundo, muitos. Outrossim, a frequência dos indivíduos com muitos descendentes é notavelmente menor, fato evidenciado pelos picos primários (em cores quentes) e secundários (em frias). Em redes maiores, a ocorrência de curvas terciárias ainda mais raras é esperada.

Ambos os gráficos 4.4 e 4.6 apontam o comportamento oscilatório do crescimento das redes. Estritamente falando, quanto mais negativo o coeficiente de Pearson dessas rede, mais extrema é a dicotomia entre indivíduos solitários ( $k = 2$ ) e indivíduos prolíficos ( $k \gg 2$ ). As curvas do gráfico 4.6 se provam informativas aqui, pois ao comparar o número de mínimos das curvas das redes com menores taxas de crescimento com as das redes de maiores taxas de crescimentos, é óbvio que as primeiras tem número de mínimos maiores. Em outras palavras, para populações crescerem na taxa imposta, o fenômeno de super ninhadas faz-se necessário.

Mesmo essas populações simuladas simples apresentam padrões exclusivos não encontrados em outros tipos de redes. O segundo pico nas curvas não apenas sugere que ter super ninhadas é vantajoso para uma espécie em ambientes sem restrições, mas também que a população inicial pode desempenhar um papel importante no tamanho dessas super ninhadas nas gerações futuras.

---

---

# CAPÍTULO 5

---

## CONSIDERAÇÕES FINAIS

A Distribuição de Grau Revela uma Estratégia Reprodutiva Bimodal: A observação mais significativa foi que a distribuição de grau não segue uma lei de potência simples, mas sim uma composição de distribuições, com picos principais em  $k = 2$  (indivíduos sem filhos) e secundários em valores mais altos ( $k = 10$ ). Isto não é um artefato do modelo, mas sim sua principal descoberta. Ele indica que, em uma população sem pressões seletivas, a estratégia mais comum é não deixar descendentes, mas uma fração significativa da população adota uma estratégia de "super-ninhadas". Isto sugere que, em ambientes sem restrições, a maximização da prole por um subconjunto de indivíduos é uma característica emergente que impulsiona o crescimento populacional. O fato do pico secundário diminuir com o aumento de  $N_i$  sugere que populações fundadoras maiores diluem a necessidade ou a vantagem destas super-ninhadas.

O comportamento oscilatório e predominantemente negativo do coeficiente de Pearson é outra descoberta crucial. Um coeficiente negativo indica que a rede é assortativa negativamente: nós de alto grau (indivíduos prolíficos) tendem a se conectar a nós de baixo grau (indivíduos com poucos ou nenhum filho). Isto é exatamente o que o modelo gera: um "pai" com muitos filhos (alto grau) está conectado a "filhos" que, naquela geração, ainda não reproduziram (baixo grau, frequentemente  $k = 2$ ). As oscilações observadas nas Figuras 4.4 e 4.6 refletem os ciclos de crescimento geracional. O aumento no número de mínimos para taxas de crescimento ( $R$ ) menores indica que populações que crescem mais devagar passam por mais desses ciclos de "privação e abundância" para atingir o tamanho final, tornando a dicotomia entre reprodutores e não-reprodutores mais frequente e

pronunciada.

A simples transposição de um cladograma para uma rede complexa, como proposto inicialmente, mostrou-se insuficiente para capturar a dinâmica subjacente. No entanto, o modelo tocogenético mostrou-se uma ferramenta poderosa para gerar redes filogenéticas sintéticas cujas propriedades podem ser investigadas. O modelo mostrou que redes filogenéticas, mesmo as mais simples, exibem propriedades topológicas únicas (bimodalidade, assortatividade negativa) que as distinguem de outros tipos de redes complexas conhecidas.

É importante reconhecer as limitações deste estudo. O Caso Ideal, por definição, ignora pressões evolutivas cruciais como seleção natural, deriva genética, mutações e limitações espaciais. Os resultados aqui apresentados representam, portanto, a "assinatura de rede" de uma população em seu estado mais fundamental e sem restrições.

O trabalho futuro reside em executar o plano modular já delineado: introduzir os casos de dinâmica populacional, herança de características e infraestrutura para investigar como essas assinaturas são modificadas por pressões seletivas. O potencial aplicativo, como na simulação de evolução de patógenos ou pragas, torna-se assim uma possibilidade tangível.

Em suma, esta dissertação demonstrou com sucesso que a framework das redes complexas é não apenas aplicável, mas também profundamente reveladora para o estudo da evolução. Ao modelar a tocogenia, foi possível identificar assinaturas topológicas específicas – a bimodalidade reprodutiva e a assortatividade negativa – que descrevem matematicamente a dinâmica intrínseca do crescimento populacional. Estes resultados estabelecem uma base sólida e quantitativa para a construção de modelos filogenéticos mais realistas e preditivos, abrindo um novo caminho para unificar a teoria das redes complexas com a biologia evolutiva.

---

# REFERÊNCIAS BIBLIOGRÁFICAS

- 1 DARWIN, C. **A Origem das Espécies**. [S.l.]: Edipro, 2018.
- 2 MAYR, E. **The Growth of Biological Thought: Diversity, Evolution, and Inheritance**. [S.l.]: Belknap Press of Harvard University Press, 1982.
- 3 NEWMAN, M. **Networks: An Introduction**. 1. ed. [S.l.]: Oxford University Press, USA, 2010.
- 4 ESTRADA, E. **The Structure of Complex Networks: Theory and Applications**. [S.l.]: Oxford University Press, 2016.
- 5 BARNES, J.; HARARY, F. Graph theory in network analysis. **Social Networks**, v. 5, n. 2, p. 235–244, 1983.
- 6 HUSON, D. H.; RUPP, R.; SCORNAVACCA, C. **Phylogenetic Networks: Concepts, Algorithms and Applications**. [S.l.]: Cambridge University Press, 2011.
- 7 HENNIG, W. **Grundzüge einer Theorie der phylogenetischen Systematik**. [S.l.]: Deutscher Zentralverlag, 1950. (Deutschen Entomologischen Institut).
- 8 HENNIG, W. **Phylogenetic Systematics**. [S.l.]: University of Illinois Press, 1979.
- 9 LINNAEUS, C. **Systema naturae, per regna tria naturae : secundum classes, ordines, genera, species cum characteribus, differentiis, synonymis, locis**. Vindobonae [Vienna]: Laurentius Salvius, 1758. Published by Typis Ioannis Thomae from 1767-1770. Disponível em: <<https://www.biodiversitylibrary.org/bibliography/559>>.
- 10 ERDÖS, P. Graph theory and probability. **Canadian Journal of Mathematics**, v. 11, p. 34 – 38, 1959.
- 11 ERDOS, P.; RENYI, A. On the evolution of random graphs. **A Magyar Tudományos Akadémia. Matematikai Kutató Intézetének Közleményei**, v. 5, p. 17–61, 1960.
- 12 HARARY, F.; NORMAN, R. Z. **Graph theory as a mathematical model in social science**. [S.l.]: University of Michigan, Institute for Social Research Ann Arbor, 1953.
- 13 HARARY, F. **Graph Theory and Theoretical Physics**. [S.l.]: AP, 1967.
- 14 BIGGS, N. L.; LLOYD, E. K.; WILSON, R. J. **Graph Theory 1736-1936**. [S.l.]: Oxford University Press, 1999.
- 15 GALICEANU, M. *et al.* Mechanisms to decrease the diseases spreading on generalized scale-free networks. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, v. 31, n. 3, p. 033131, 03 2021. Disponível em: <<https://doi.org/10.1063/5.0038631>>.

- 16 BELLINGERI, M.; BODINI, A. Food web's backbones and energy delivery in ecosystems. *Oikos*, v. 125, 08 2015.
- 17 JURJIU, A.; GALICEANU, M. Dynamics of a polymer network modeled by a fractal cactus. *Polymers*, v. 10, n. 7, 2018. ISSN 2073-4360.
- 18 GAMA, J. V. Estudo estatístico da rede cladística de fungos. Relatório final de programa de iniciação científica (PIBIC) orientado pelo professor Dr. Mircea Daniel Galiceanu. Código do projeto: PIB-E/0138/2019. 2020.
- 19 ARISTOTLE; BARNES, J. **The Complete Works of Aristotle: The Revised Oxford Translation, One-Volume Digital Edition**. [S.l.]: Princeton University Press, 2014.
- 20 LINNAEUS, W. T. C. **Systema Naturae**. London: Lackington, Allen, and Co., 1806. Translated from the 13th edition. Disponível em: <<http://linnaeuslink.org/records/record/2643a>>.
- 21 BOWLER, P. J. **Evolution: The History of an Idea**. 3rd. ed. [S.l.]: University of California Press, 2003.
- 22 MAYR, E. **Animal Species and Evolution**. [S.l.]: Harvard University Press, 1963.
- 23 SOKAL, P. H. A. S. R. R. **Principles of numerical taxonomy**. [S.l.]: W. H. Freeman, 1963. (A Series of books in biology).
- 24 SIMPSON, G. G. **Principles of Animal Taxonomy**. [S.l.]: Columbia University Press, 1990.
- 25 KRENISKI, S. B. P. **Semelhanças entre Tubarões, Ictiossauros, e Golfinhos e a Convergência Evolutiva**. 2023. Figura elaborada para este trabalho.
- 26 NASCIMENTO, P. M.; LENHARDT, G. G. **Relações Tocogenéticas de uma População Animal**. 2022. Figura extraída do vídeo "Cladística - reconstruindo a Evolução (#Pirula 94)" por Paulo Miranda Nascimento, em 2014, e refeita em 2023 por Gabriella Guaqueta Lenhardt para esta tese.
- 27 CAMPBELL, N. A. *et al.* **Campbell Biology**. 9th. ed. [S.l.]: Benjamin Cummings / Pearson, 2010.
- 28 COMMONS, W. **File:Identical cladograms.svg — Wikimedia Commons, the free media repository**. 2022. [Online; accessed 16-June-2023].
- 29 KRENISKI, S. B. P. **Cladograma do Grupo dos Vertebrados com Caracteres Principais de cada Sub-grupo**. 2023. Figura elaborada para este trabalho.
- 30 COMMONS, W. **File:Reptilia-Sauropsida.jpg — Wikimedia Commons, the free media repository**. 2021. [Online; accessed 30-June-2023].
- 31 KRENISKI, S. B. P. **Homologia dos Membros Anteriores de Diferentes Tetrápodas**. 2023. Figura elaborada para este trabalho.
- 32 MANN, A. *et al.* Reassessment of historic 'microsaurs' from joggins, nova scotia, reveals hidden diversity in the earliest amniote ecosystem. *Papers in Palaeontology*, v. 6, n. 4, p. 605–625, 2020. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/spp2.1316>>.
- 33 KRENISKI, S. B. P. **Comparação entre Asas de Insetos, Aves, e Morcegos**. 2023. Figura elaborada para este trabalho.
- 34 LANKESTER, E. R. II.— **On the use of the term homology in modern zoology, and the distinction between homogenetic and homoplastic agreements**. Zenodo, 1870. Disponível em: <<https://doi.org/10.1080/00222937008696201>>.
- 35 BARABASI, A.-L.; FRANGOS, J. **Linked: The New Science of Networks**. [S.l.: s.n.], 2002.
- 36 MERIAN-ERBEN, S. von. **Königsberg**. 1652. [Online]. Disponível em: <[https://www.preussenchronik.de/bild\\_jsp/key=bild\\_kathe2.html](https://www.preussenchronik.de/bild_jsp/key=bild_kathe2.html)>. Acesso em: 02 de julho de 2023.

- 37 BOLLOBÁS, B. **Random Graphs**. 2. ed. [S.l.]: Cambridge University Press, 2001.
- 38 HARARY, F. **Graph Theory**. [S.l.]: CRC Press, 1994.
- 39 RANJAN, D. **Frank Harary 1921-2005**. Originalmente publicado na rede interna de Columbia University ([URL](#)), estando agora apenas disponível no arquivo da Internet Wayback Machine ([URL](#)).
- 40 BARABASI, A.-L. **Set of Elementary Network Characteristics**. Imagem 2.7 ([URL](#)) retirada do site do capítulo 2 ([URL](#)) em 20 de junho de 2023.
- 41 PENROSE, M. **Random Geometric Graphs**. [S.l.]: OUP Oxford, 2003. (Oxford Studies in Probability).
- 42 MASUDA, N.; LAMBIOTTE, R. **A Guide to Temporal Networks**. [S.l.]: World Scientific, 2020. (Complexity Science).
- 43 HOLME, P.; SARAMÄKI, J. **Temporal Network Theory**. [S.l.]: Springer International Publishing, 2019. (Computational Social Sciences).