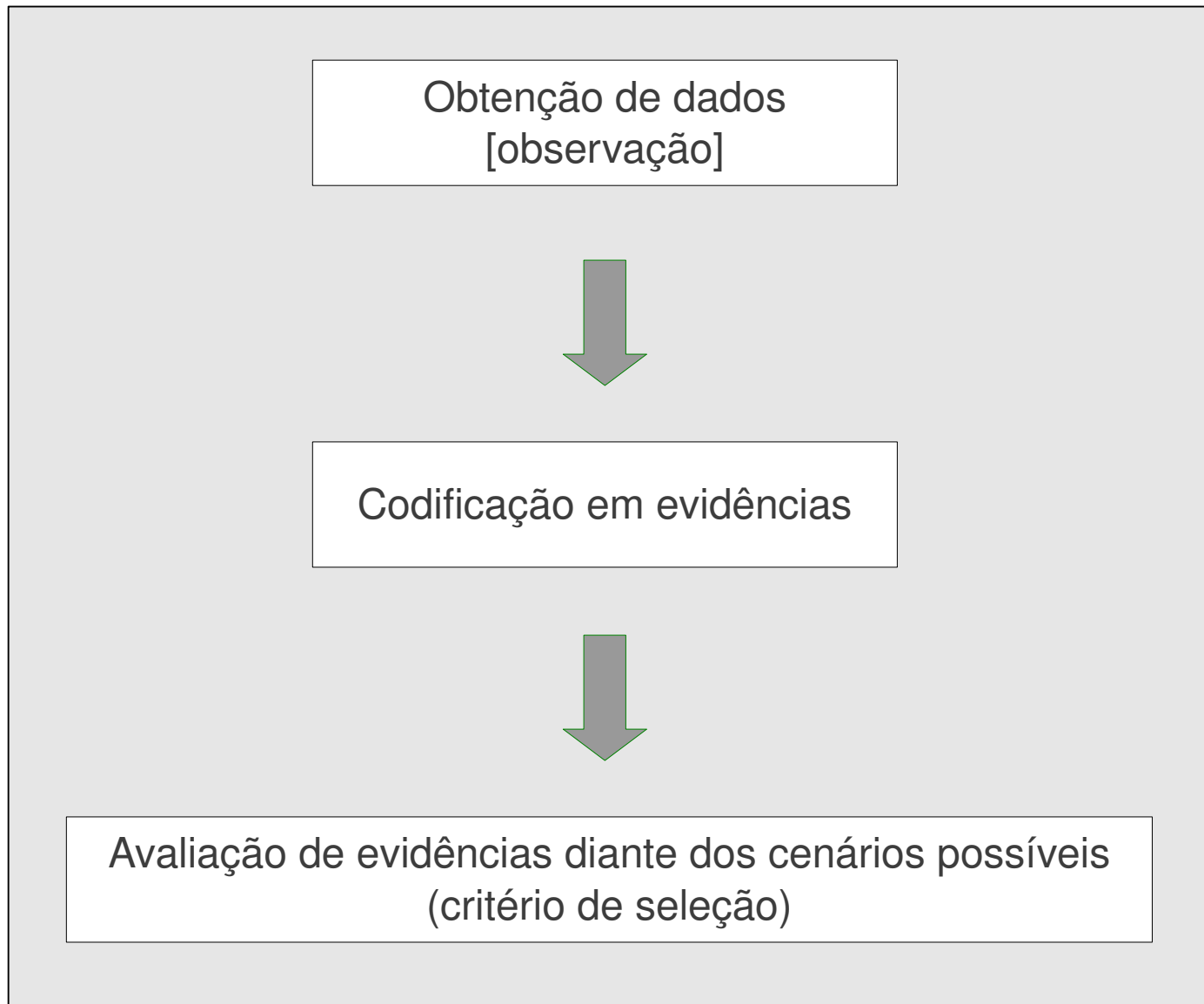


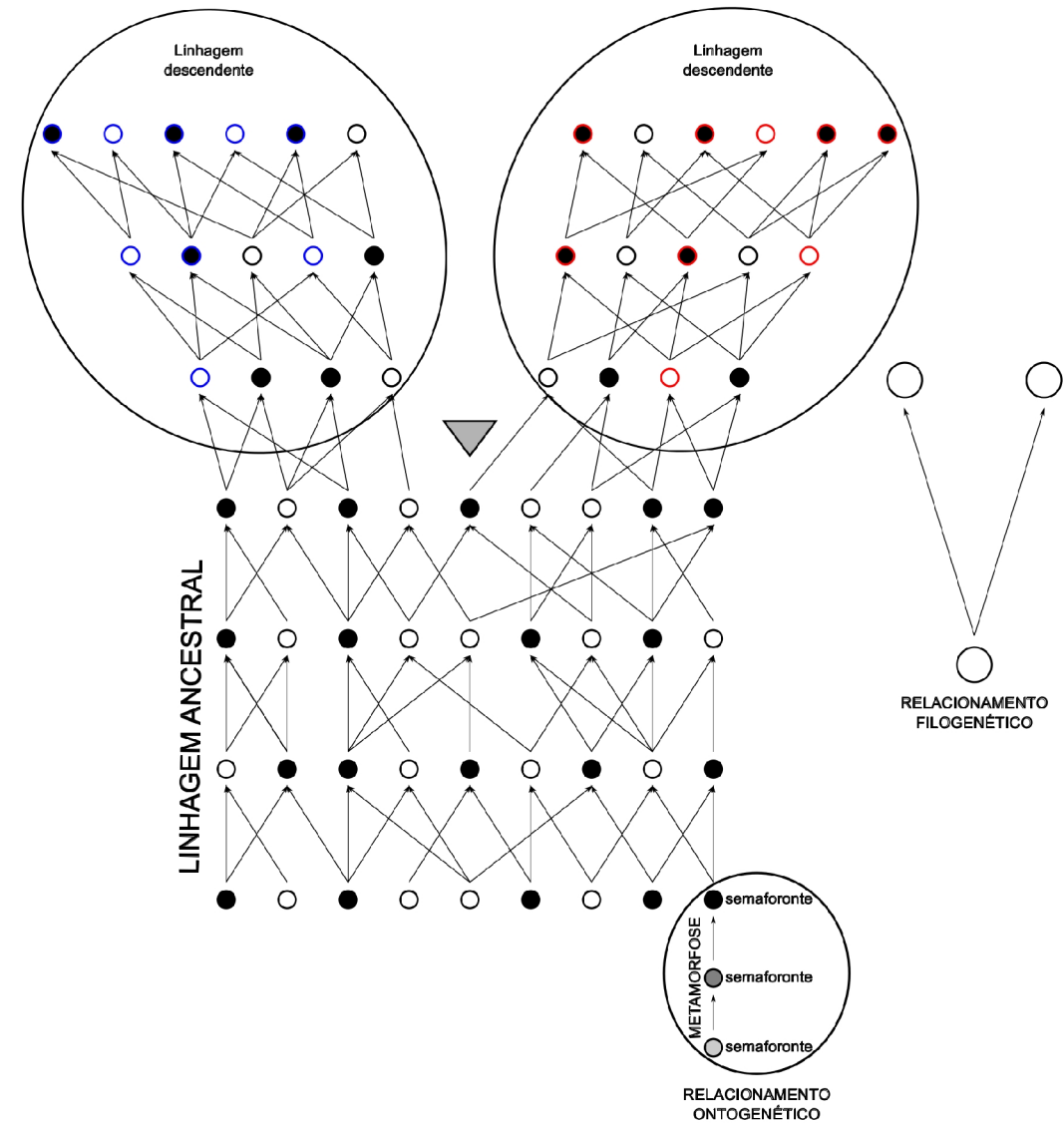
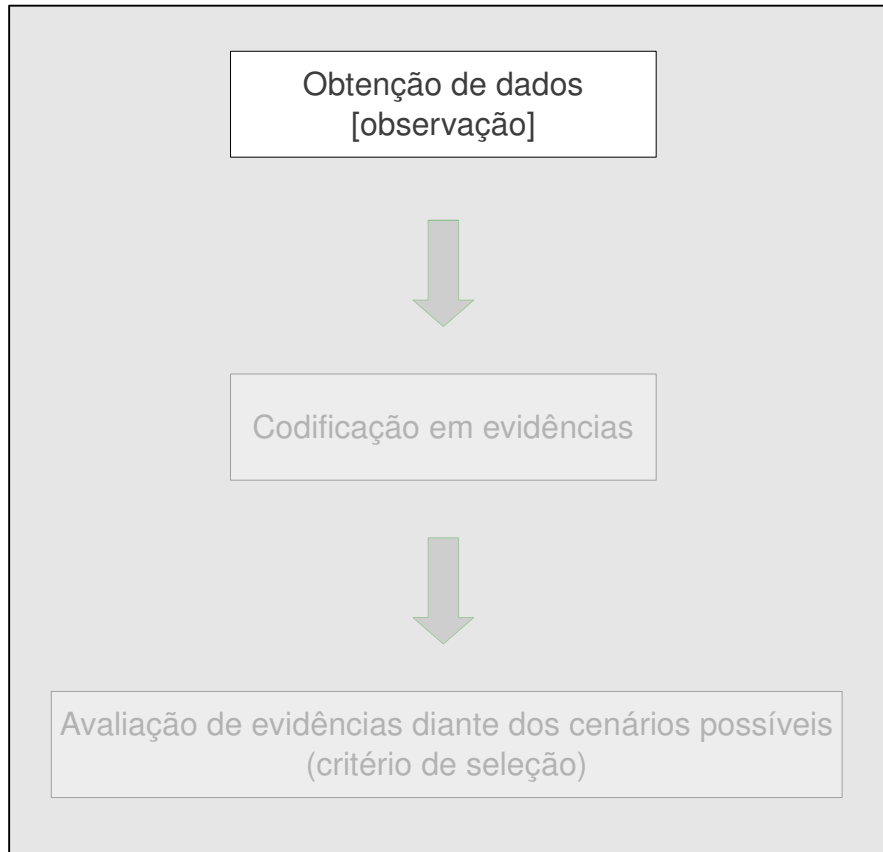
Lógica da inferência filogenética

“Operationally, systematics proceeds by gathering data (observations) from organisms and coding them into evidence to test competing phylogenetic scenarios”

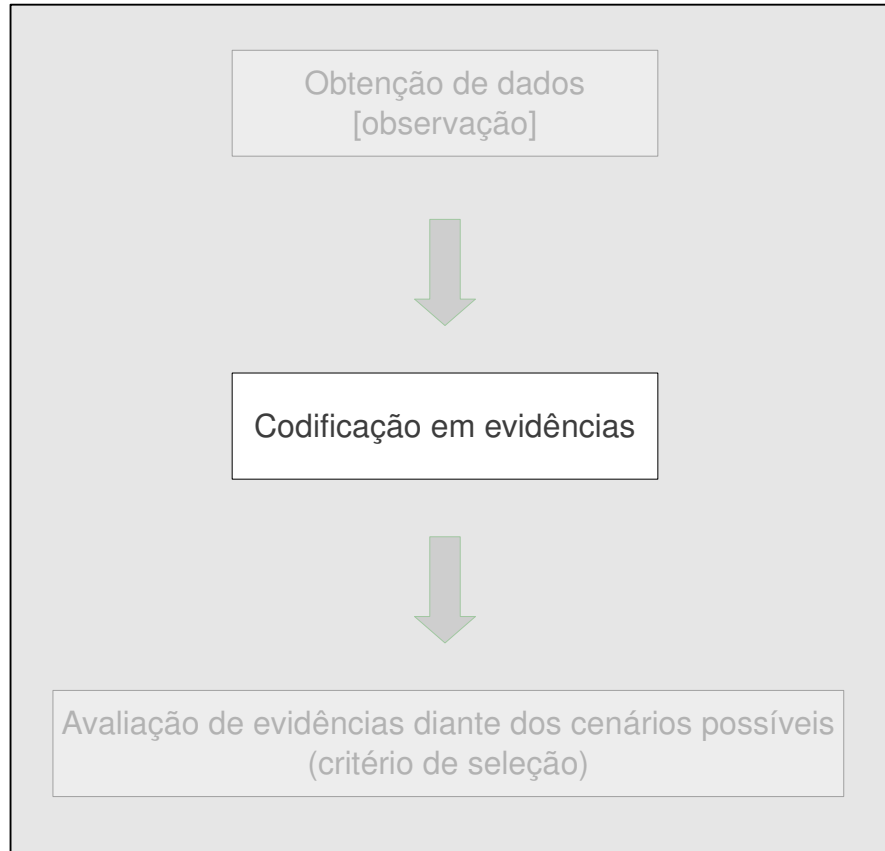
(Wheeler et al., 2006:7)



Lógica da inferência filogenética



Lógica da inferência filogenética



↓ ↓ ↓
 sp.X CTGGCTACGT
 sp.A TGGAGTAAGT
 sp.B CCTAGCAAGT
 sp.C CCTGATTGCA

Fenética:

EVIDÊNCIAS: similaridade global

sp.X CTGGCTACGT
* ** **

sp.A TGGAGTAAGT

sp.X CTGGCTACGT
* * **

sp.B CCTAGCAAGT

CODIFICAÇÃO: matriz de distância

	[X]	[A]	[B]	[C]
X	-			
A	5	-		
B	6	4	-	
C	7	9	7	-

Cladística:

EVIDÊNCIAS: transformação de estados de caracteres

1a. posição: C ↔ T

3a. posição: G ↔ T

10a. posição: T ↔ A

CODIFICAÇÃO: matriz de dados

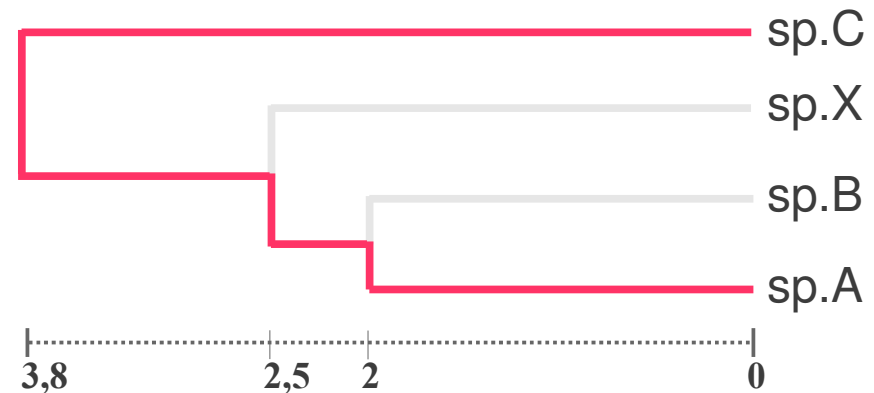
	c ₁	c ₂	c ₃	c ₄	c ₅	c ₆	c ₇	c ₈	c ₉	c ₁₀
sp.X	1	3	2	2	1	3	0	1	2	3
sp.A	3	2	2	0	2	3	0	0	2	3
sp.B	1	1	3	0	2	1	0	0	2	3
sp.C	1	1	3	2	0	3	3	2	1	0

Lógica da inferência filogenética: Fenética

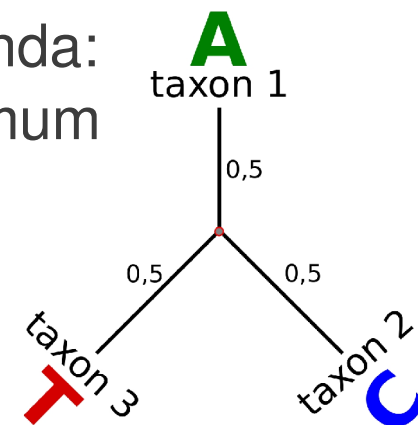
Problemas com o método:

1. Desconsidera que semelhanças decorrem de processos não relacionados com relação de parentesco.
2. Método é incapaz de manter as relações de distâncias originais para matrizes com mais de 4 terminais. Considere:

	[X]	[A]	[B]	[C]
X	-			
A	5	-		
B	6	4	-	
C	7	9	7	-



3. Realismo. Considere o exemplo ao lado e responda: Qual seria o par de base presente no ancestral comum destes terminais?



Lógica da inferência filogenética

Limitações da fenética:

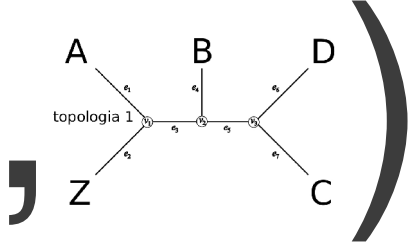
1. Perda de informação
2. Dados heterogêneos tratados da mesma forma
3. Ausência de otimização de eventos de transformação de caracteres distintos

Função objetiva:

$$C = f(D, T)$$

Menor distância fenética = f (

sp.X	1	3	2	2	1	3	0	1	2	3
sp.A	3	2	2	0	2	3	0	0	2	3
sp.B	1	1	3	0	2	1	0	0	2	3
sp.C	1	1	3	2	0	3	3	2	1	0

, 

)

Limitações de um fenograma:

1. Não Permite reconstruções de ancestrais hipotéticos
2. Não permite proposições de homologia
3. Não permite identificar transformações de caracteres

Caracteres & estados de caráter:

Hennig (1966):

CARÁTER



série de transformação

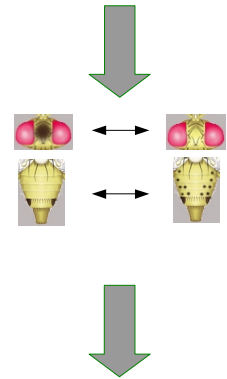
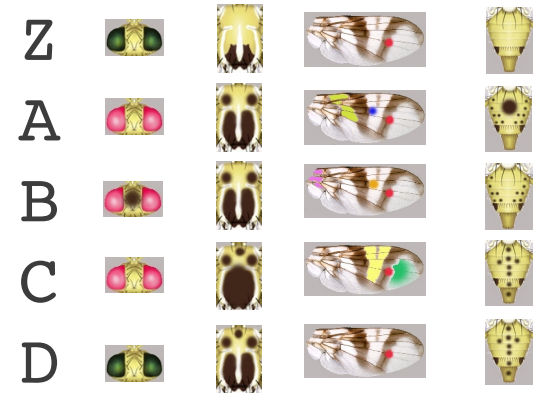
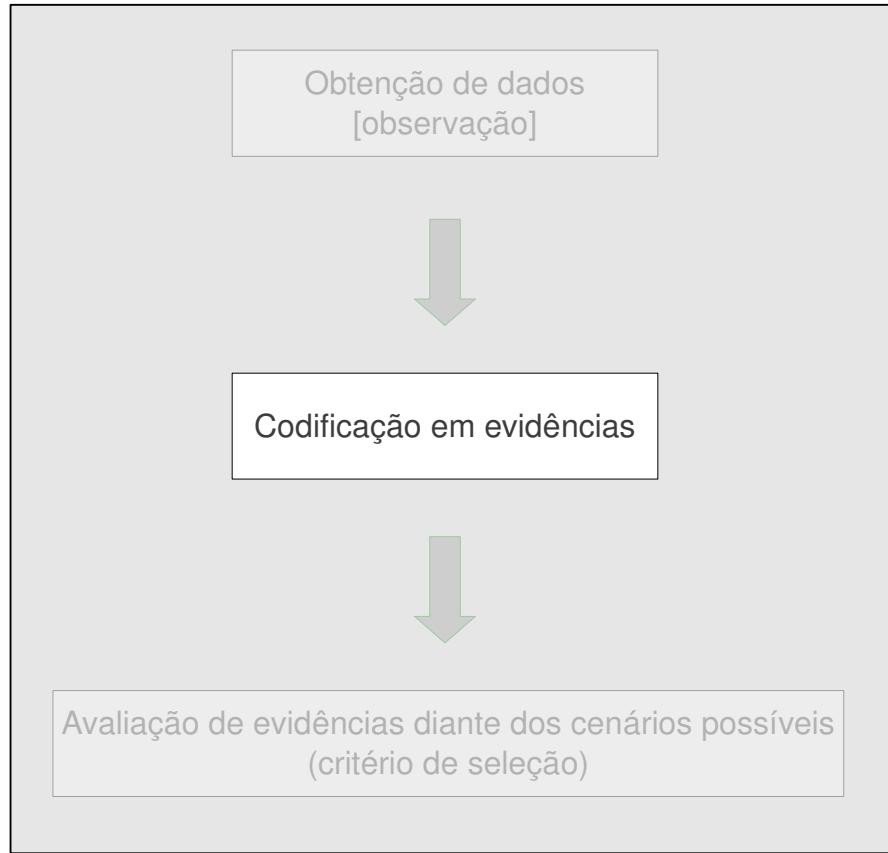


conjunto de condições mutuamente exclusivas

Estados de caráter são conceitualmente definidos como indivíduos históricos menos inclusivos resultado de eventos de transformação hereditários.

É comum definir conceitualmente estados de caráter como elementos puramente observacionais, mas eles são hipóteses de identidade histórica complexas dependente de teoria (descendência com modificação) – não menos hipotéticas que caracteres e outros indivíduos históricos mais inclusivos dos quais fazem parte.

Lógica da inferência filogenética



Compilação de dados observacionais

Definição de caráter e estados de caráter

Codificação de matriz de dados

	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	C ₁₀	C ₁₁	C ₁₂
Z	0	1	0	0	0	0	0	0	0	0	0	0
A	1	1	1	1	0	0	1	0	0	0	0	1
B	1	0	1	1	0	1	0	0	0	0	0	1
C	1	1	2	1	1	0	0	1	1	1	1	0
D	0	1	1	1	1	0	0	0	0	1	1	0



Codificação: caracteres e estados de caráter

Definição de caráter
e estados de caráter






Codificação de
matriz de dados

Caráter 1: coloração do olho

Estados: 0 = verde 
1 = rosa 

Caráter 3: mancha post. torácica





Estados: 0 =  1 =  2 = 

	C_1	C_3
Z	0	0
A	1	1
B	1	1
C	1	2
D	0	1




Assume-se que os estados são eventos de transformação histórica e que sua distribuição está relacionada com ancestralidade.

Codificação: caracteres e estados de caráter

Considere:

Z 
A 
B 
C 
D 

Caráter 3: mancha post. torácica

Estados: 0 =  1 =  2 = 

Caráter 4: mancha ant/lat torácica





Estados: 0 = ausente  1 = presente 

Caráter 5: mancha ant. torácica

Estados: 0 = ausente  1 = presente 

VS.

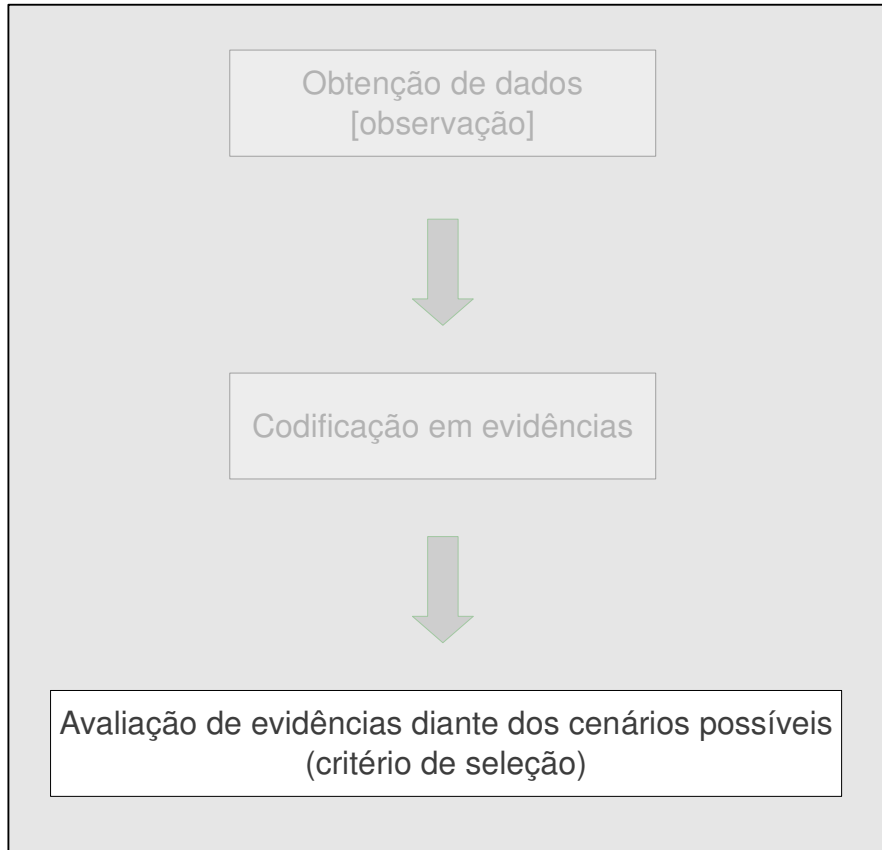
Caráter **X**: Pigmentação torácica

Estados: 0 =  1 =  2 =  3 = 

Observe que o conjunto de estados é abstrato e que cada um implica em uma série de transformação distinta.

Lógica da inferência filogenética

Avaliação e critério de seleção: soluções possíveis



	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	C ₁₀	C ₁₁	C ₁₂
Z	0	1	0	0	0	0	0	0	0	0	0	0
A	1	1	1	1	0	0	1	0	0	0	0	1
B	1	0	1	1	0	1	0	0	0	0	0	1
C	1	1	2	1	1	0	0	1	1	1	1	0
D	0	1	1	1	1	0	0	0	0	1	1	0

Número de Cenários = $(2n-5)!/[2^{n-3} \cdot (n-3)!]$

onde n é igual ao número de terminais (**OTUs**).

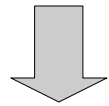
No. de OTUs	No. de Soluções
3	1
4	3
5	15
6	105
...	...

Qual topologia explica melhor a variabilidade hereditária observada?

Critério de otimização:

Topologias como hipóteses:

Teste → Avaliação → Determinação de qualidade relativa



Índices de mérito comparativos

Independente do índice: requer função objetiva

$$C = f(D, T)$$

'Without such a cost, these objects are mere pictures — “tree-shaped-objects” of no use in science'
(Wheeler et al., 2006: Cladistics 12:1-9)

Critério de otimização: Cladística

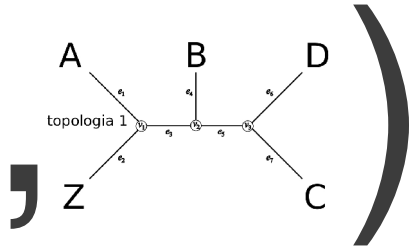
Função objetiva:

$$C = f(D, T)$$

*Menor
distância
patrística* = f (

sp.X	1	3	2	2	1	3	0	1	2	3
sp.A	3	2	2	0	2	3	0	0	2	3
sp.B	1	1	3	0	2	1	0	0	2	3
sp.C	1	1	3	2	0	3	3	2	1	0

;

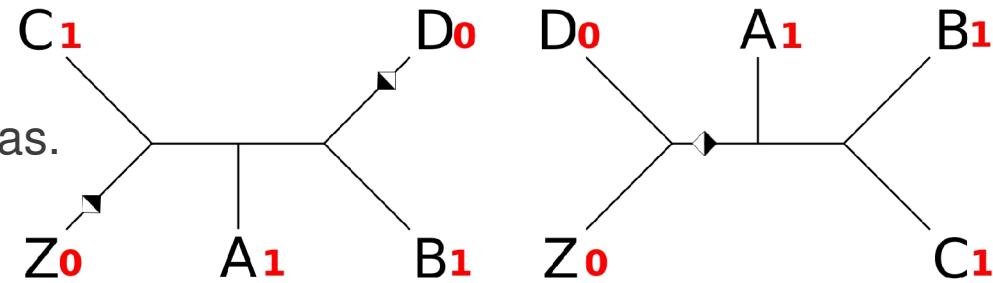


Esta função minimiza o número total de hipóteses de transformação necessária para explicar as observações utilizando o princípio da **parcimônia**.

Lógica da inferência filogenética: critério de seleção

Cladística:

Otimização: refere-se a seleção do melhor elemento de um conjunto disponível de alternativas.



Critério de otimização: **parcimônia**



William of Ockham (c. 1288 - c. 1348): *lex parsimoniae* ou "Occam's Razor"

"entia non sunt multiplicanda praeter necessitatem"

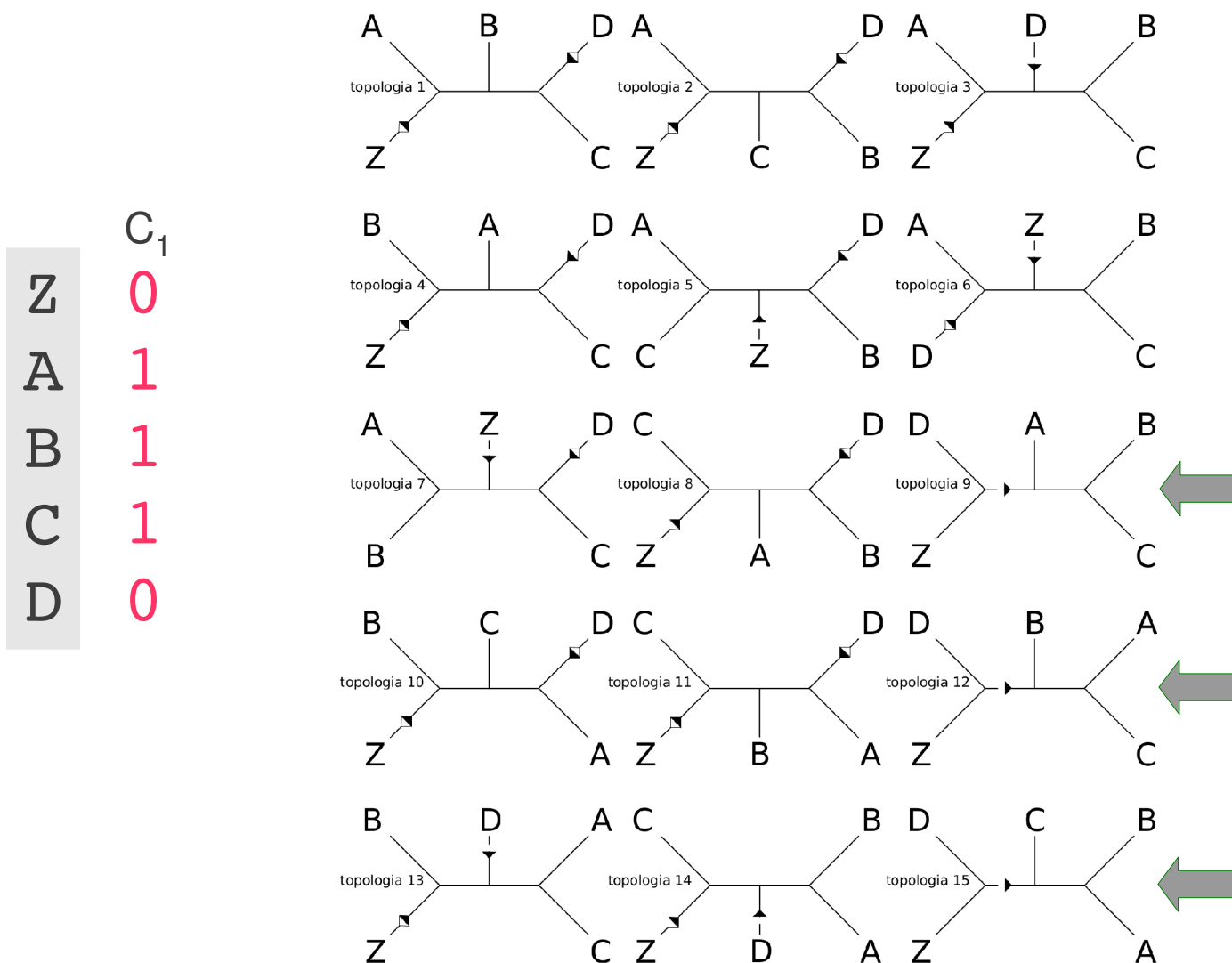
"Entities should not be multiplied unnecessarily."

"when you have two competing theories which make exactly the same predictions, the one that is simpler is the better."

Newton stated the rule: "We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances."

Lógica da inferência filogenética

Otimização e conteúdo informativo

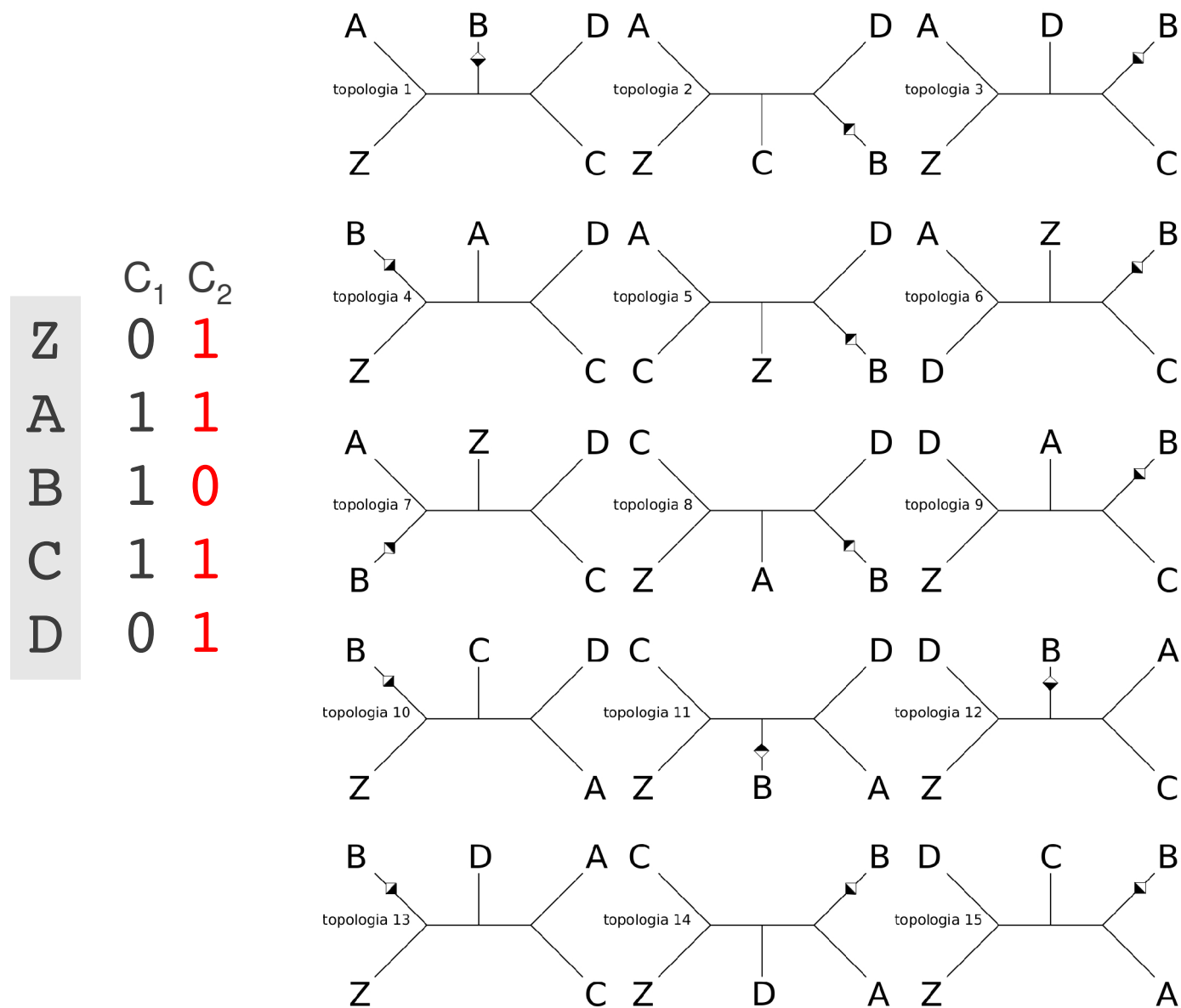


uma transformação
 vs.
 duas transformações

Caráter informativo!

Lógica da inferência filogenética

Otimização e conteúdo informativo

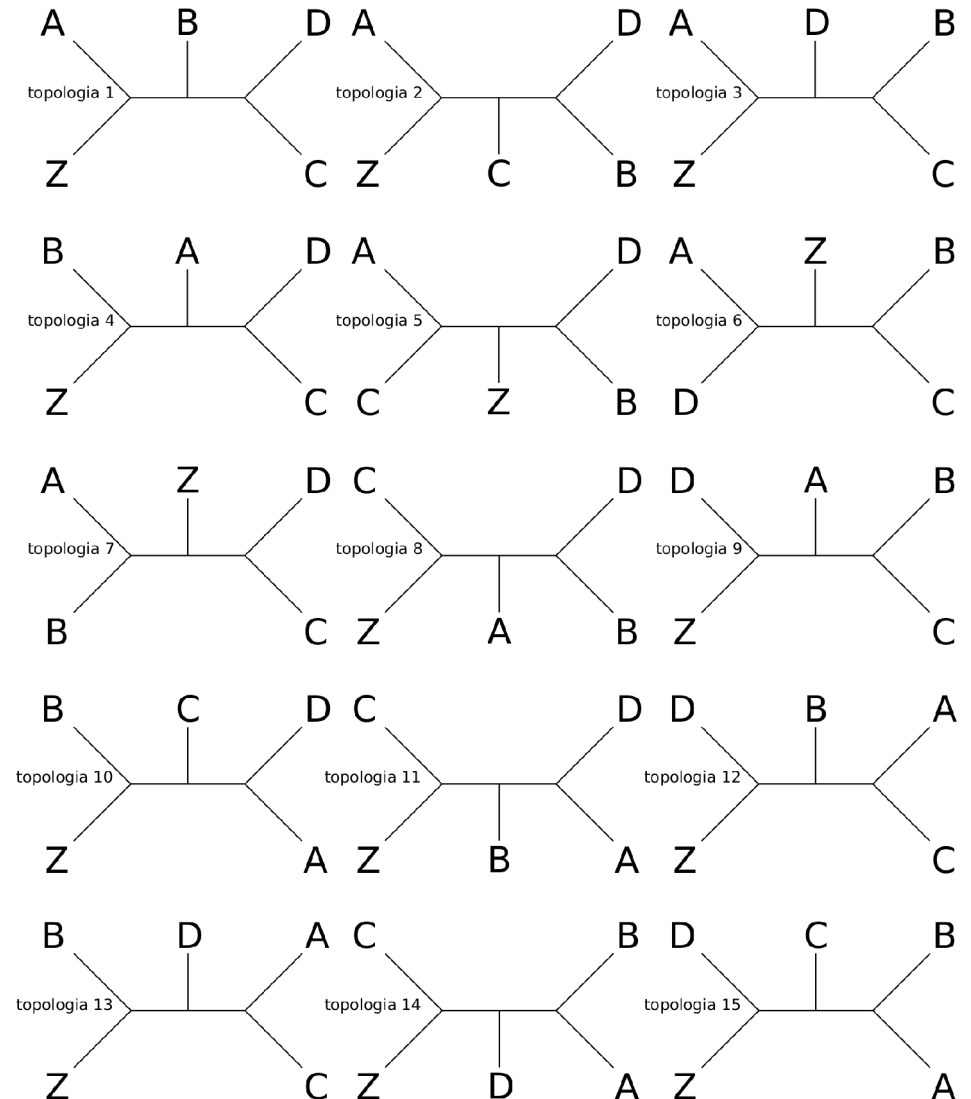
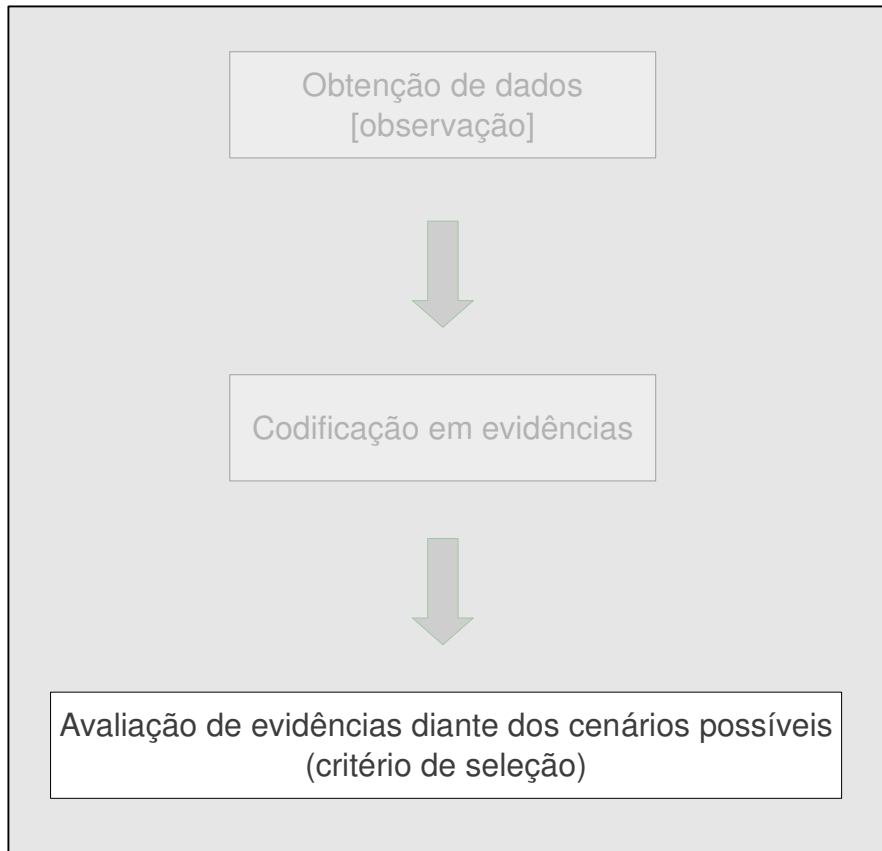


Uma transformação em todos os diagramas.

Caráter não-informativo!

Lógica da inferência filogenética

Avaliação e critério de seleção: soluções possíveis



Lógica da inferência filogenética: critério de seleção

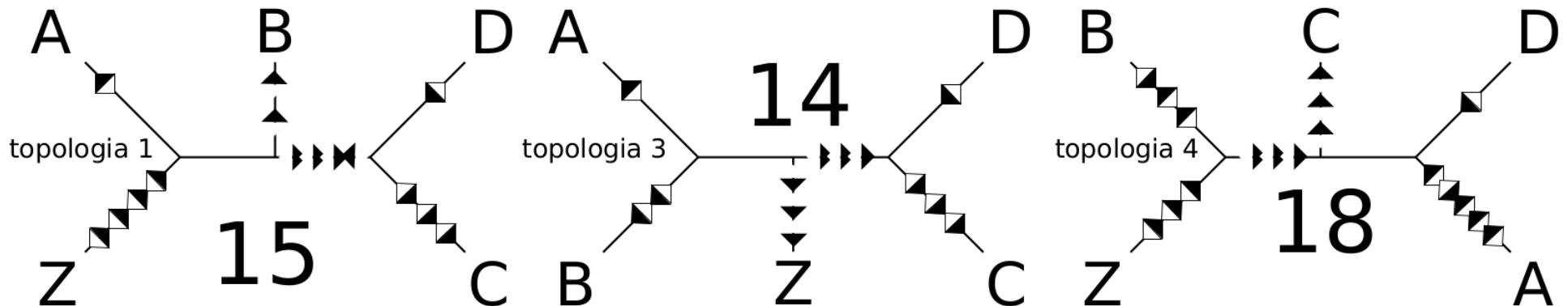
Cladística:

Otimização: refere-se a seleção do melhor elemento de um conjunto disponível de alternativas.

	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	C ₁₀	C ₁₁	C ₁₂
Z	0	1	0	0	0	0	0	0	0	0	0	0
A	1	1	1	1	0	0	1	0	0	0	0	1
B	1	0	1	1	0	1	0	0	0	0	0	1
C	1	1	2	1	1	0	0	1	1	1	1	0
D	0	1	1	1	1	0	0	0	0	1	1	0

Topologia	Top. 1	Top. 2	Top. 3	Top. 4	Top. 5
Tranformações	15	15	14	18	18
Topologia	Top. 6	Top. 7	Top. 8	Top. 9	Top. 10
Tranformações	18	18	18	18	17
Topologia	Top. 11	Top. 12	Top. 13	Top. 14	Top. 15
Tranformações	18	18	17	17	16

Diferentes topologias podem gerar otimizações distintas.

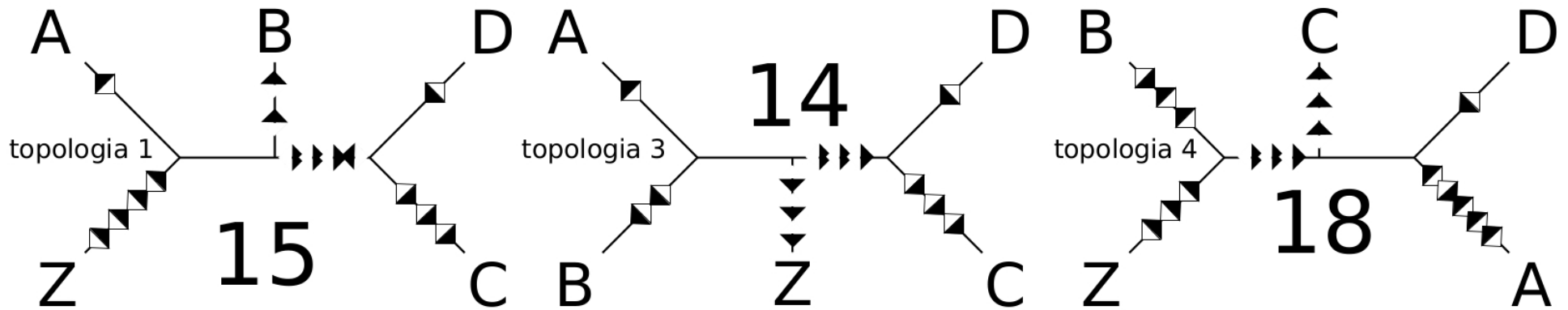


Lógica da inferência filogenética: critério de seleção

Cladística:

Otimização: refere-se a seleção do melhor elemento de um conjunto disponível de alternativas.

Diferentes topologias podem gerar otimizações distintas.



Menor distância patrística = f (

sp.X	1	3	2	2	1	3	0	1	2	3
sp.A	3	2	2	0	2	3	0	0	2	3
sp.B	1	1	3	0	2	1	0	0	2	3
sp.C	1	1	3	2	0	3	3	2	1	0

;

)

A **distância patrística** é a soma dos comprimentos de ramos em uma topologia.

Lógica da inferência filogenética: critério de seleção

*Menor
distância
patrística* = f (

sp.X	1	3	2	2	1	3	0	1	2	3
sp.A	3	2	0	2	3	0	0	2	3	
sp.B	1	1	3	0	2	1	0	0	2	3
sp.C	1	1	3	2	0	3	3	2	1	0

topologia 1

)

Justificativa para Parcimônia:

Prodedimento de inferência \neq modelo de evolução



“Systematic analysis 'must be done under the rules of parsimony, not because nature is parsimonious, but because only parsimonious hypotheses can be defended by the investigator without resorting to authoritarianism or apriorism.”

(Wiley, 1975 in Wheeler 2012)

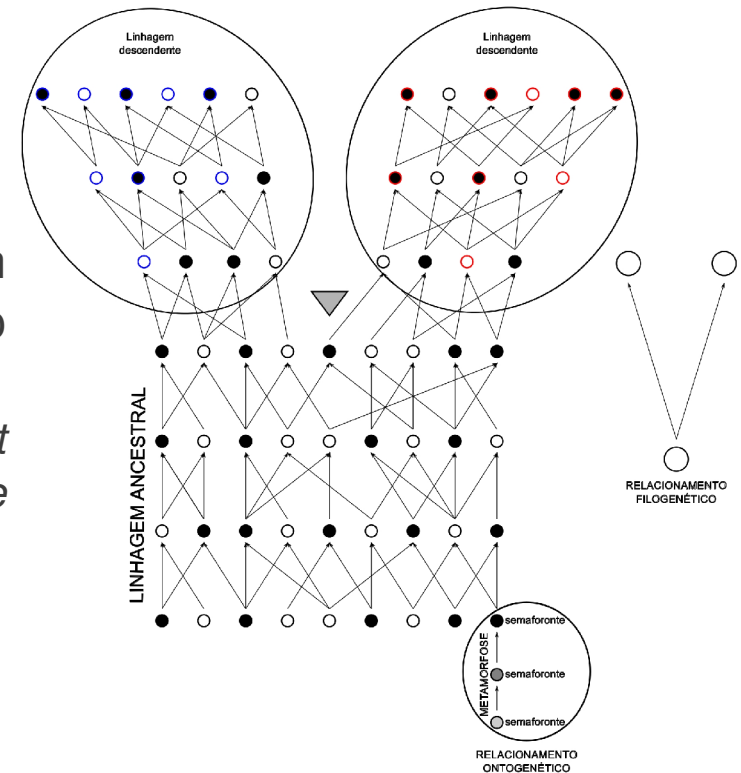
Lógica da inferência filogenética: critério de seleção

Cladística:

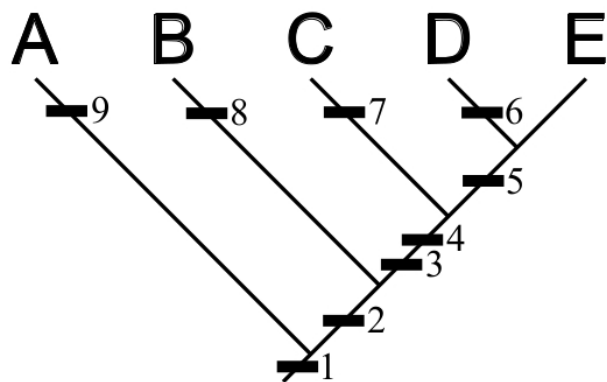
Justificativa para Parcimônia:



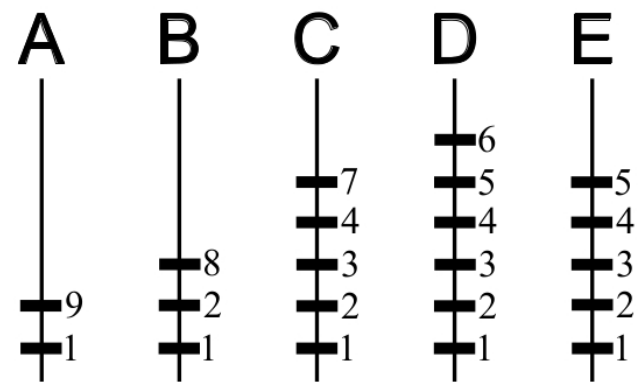
Kluge (2005): descendência com modificação como modelo simplificado de evolução biológica: *"a minimal evolutionary assumption that offspring resemble their parents more than non-parents, but not exactly."* (Wheeler, 2012)



Qual desses cenários seria favorecido pelo critério de otimização?



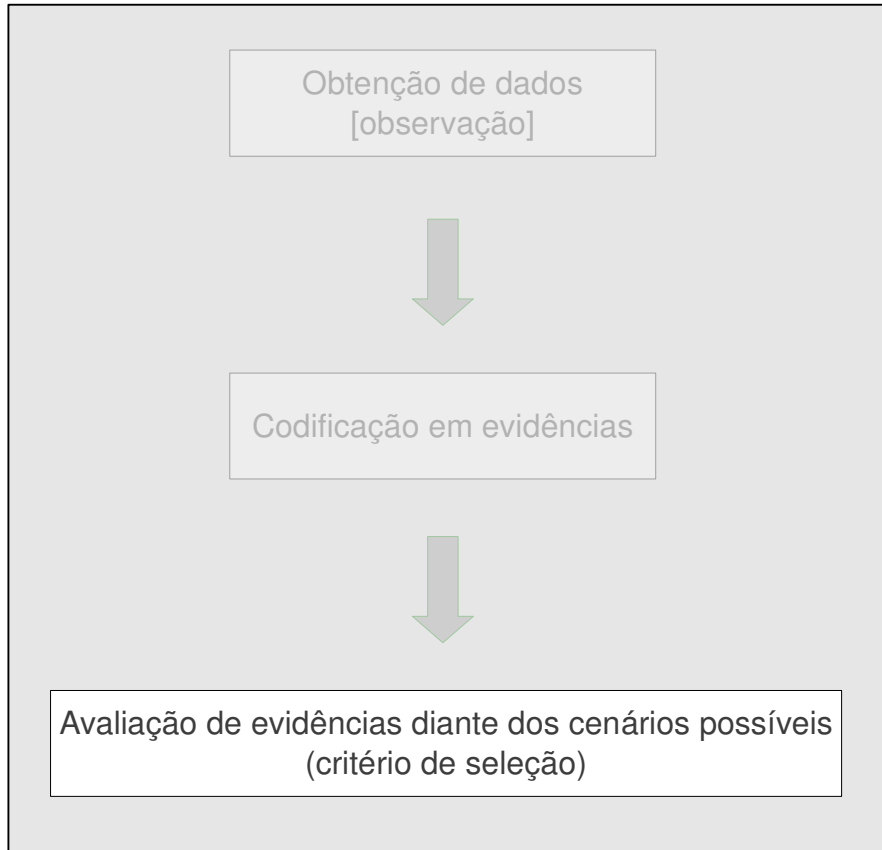
9 transformações



21 transformações

Lógica da inferência filogenética

Avaliação e critério de seleção: soluções possíveis



	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	C ₁₀	C ₁₁	C ₁₂
Z	0	1	0	0	0	0	0	0	0	0	0	0
A	1	1	1	1	0	0	1	0	0	0	0	1
B	1	0	1	1	0	1	0	0	0	0	0	1
C	1	1	2	1	1	0	0	1	1	1	1	0
D	0	1	1	1	1	0	0	0	0	1	1	0

Qual topologia explica melhor a variabilidade hereditária observada?

O algoritmo mais comum:

WAGNER TREE

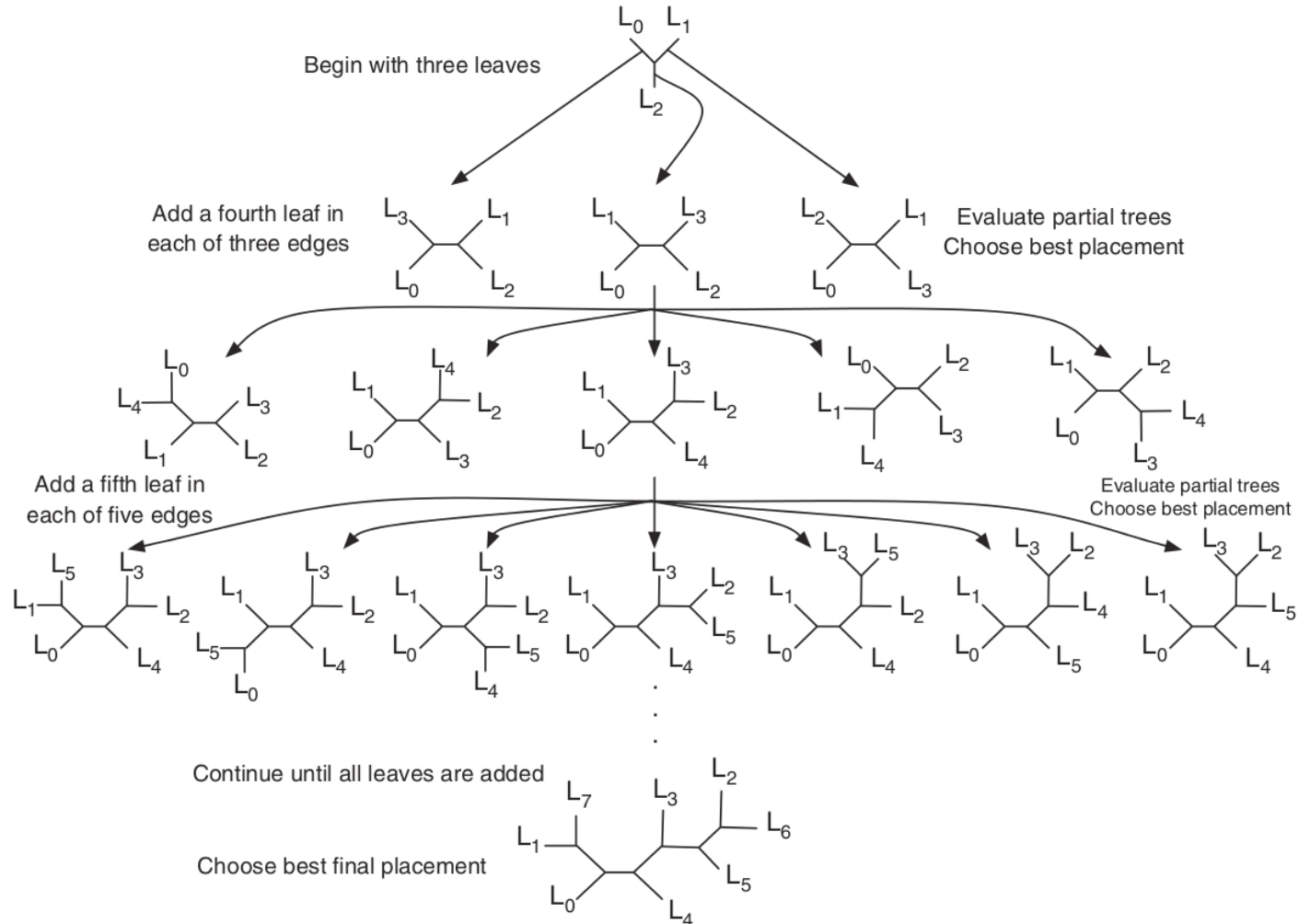
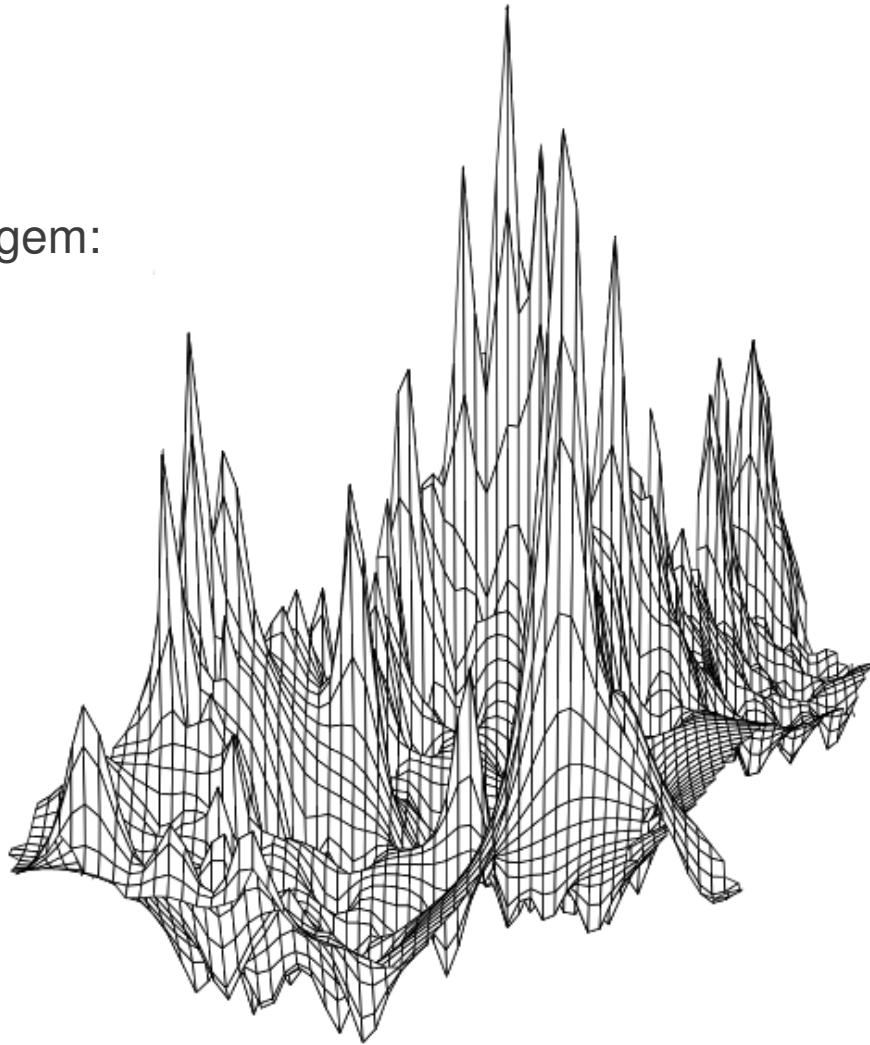


Figure 14.4: Initial tree construction procedure via the Wagner algorithm (Farris, 1970).

Estratégias de busca:

Paisagem:



Número de Cenários = $(2n-5)!/[2^{n-3} \cdot (n-3)!]$
onde n é igual ao número de terminais (**OTUs**).

No. de OTUs	No. de Soluções
3	1
4	3
5	15
6	105
...	...

Buscas exatas vs. Buscas heurísticas

Buscas exatas:

Paisagem:

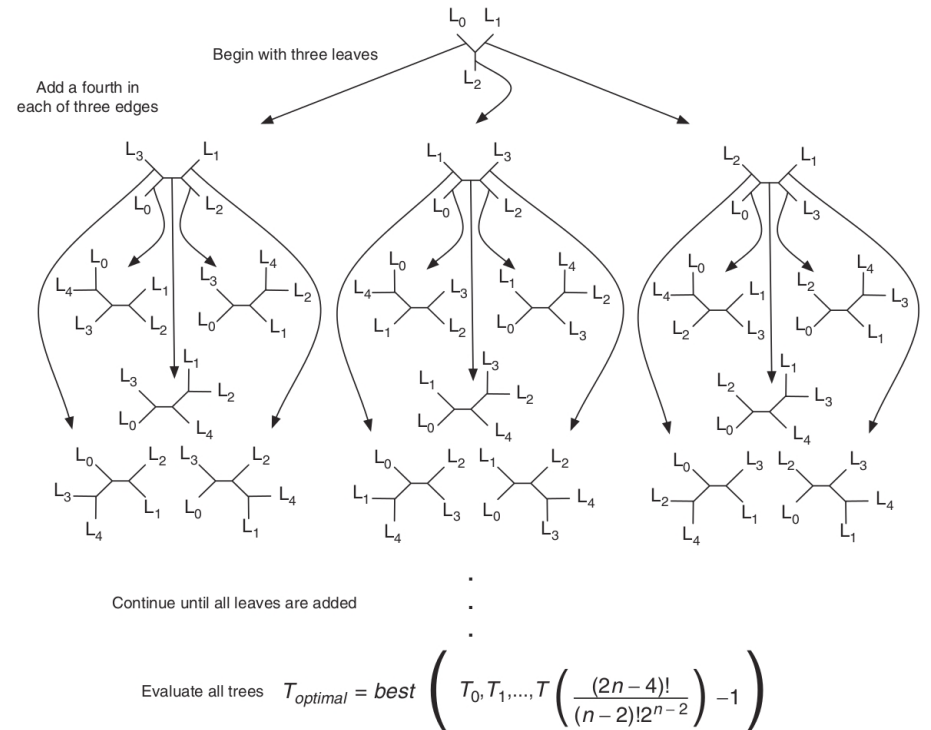
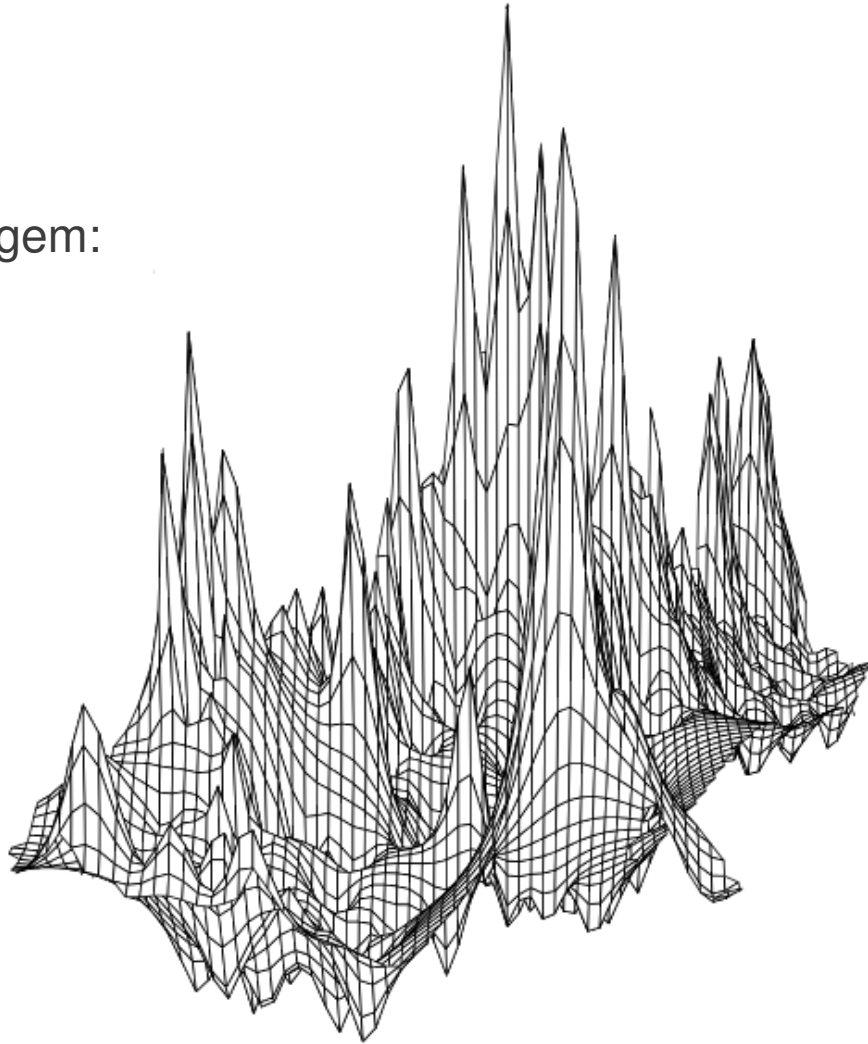
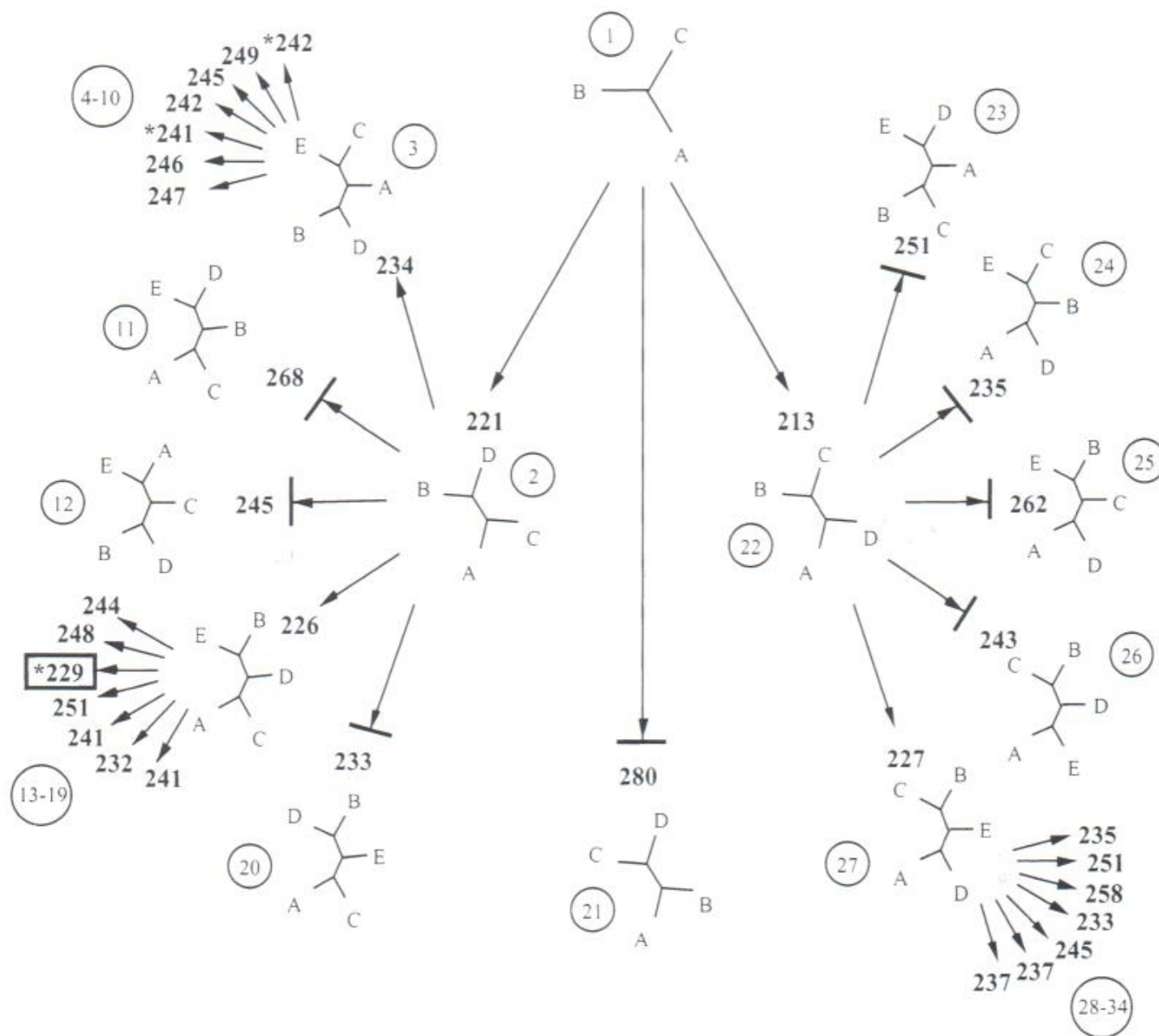
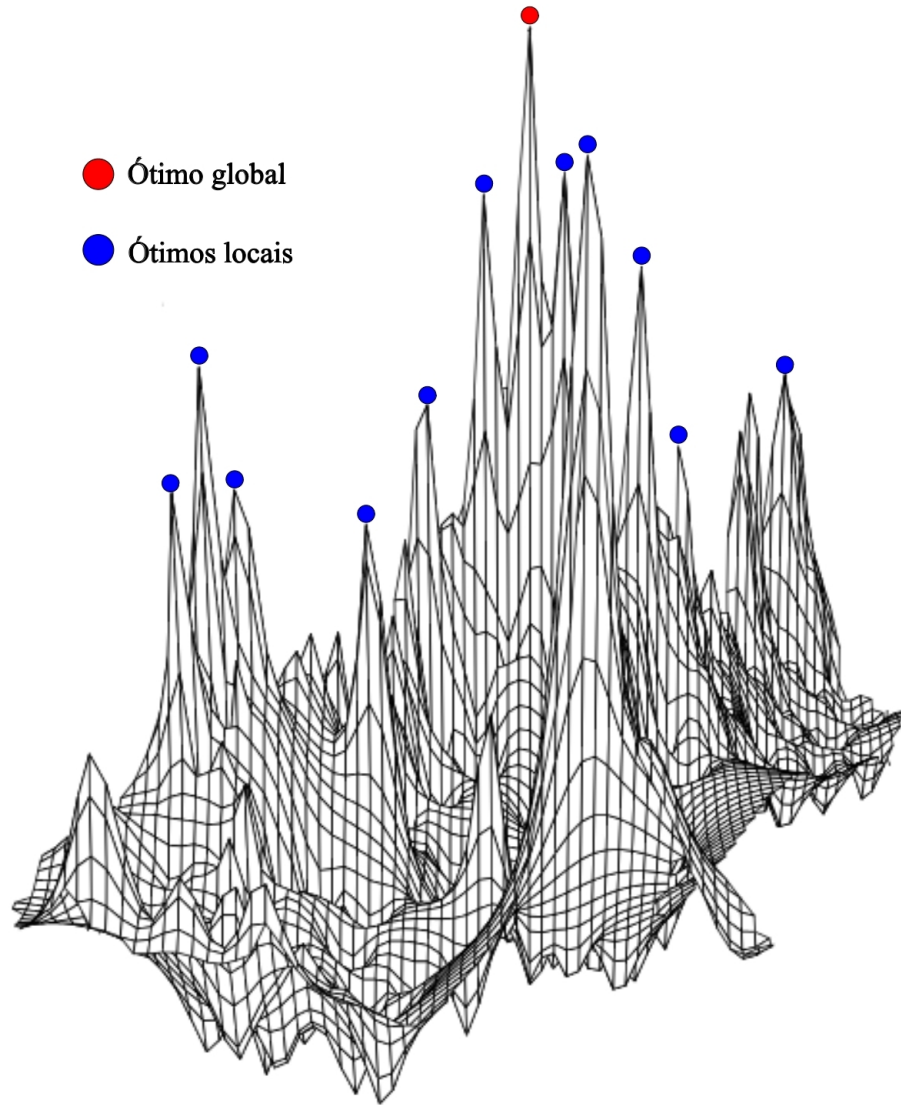


Figure 14.1: Explicit enumeration and evaluation of all trees.

Enumeração implícita



Buscas heurísticas:



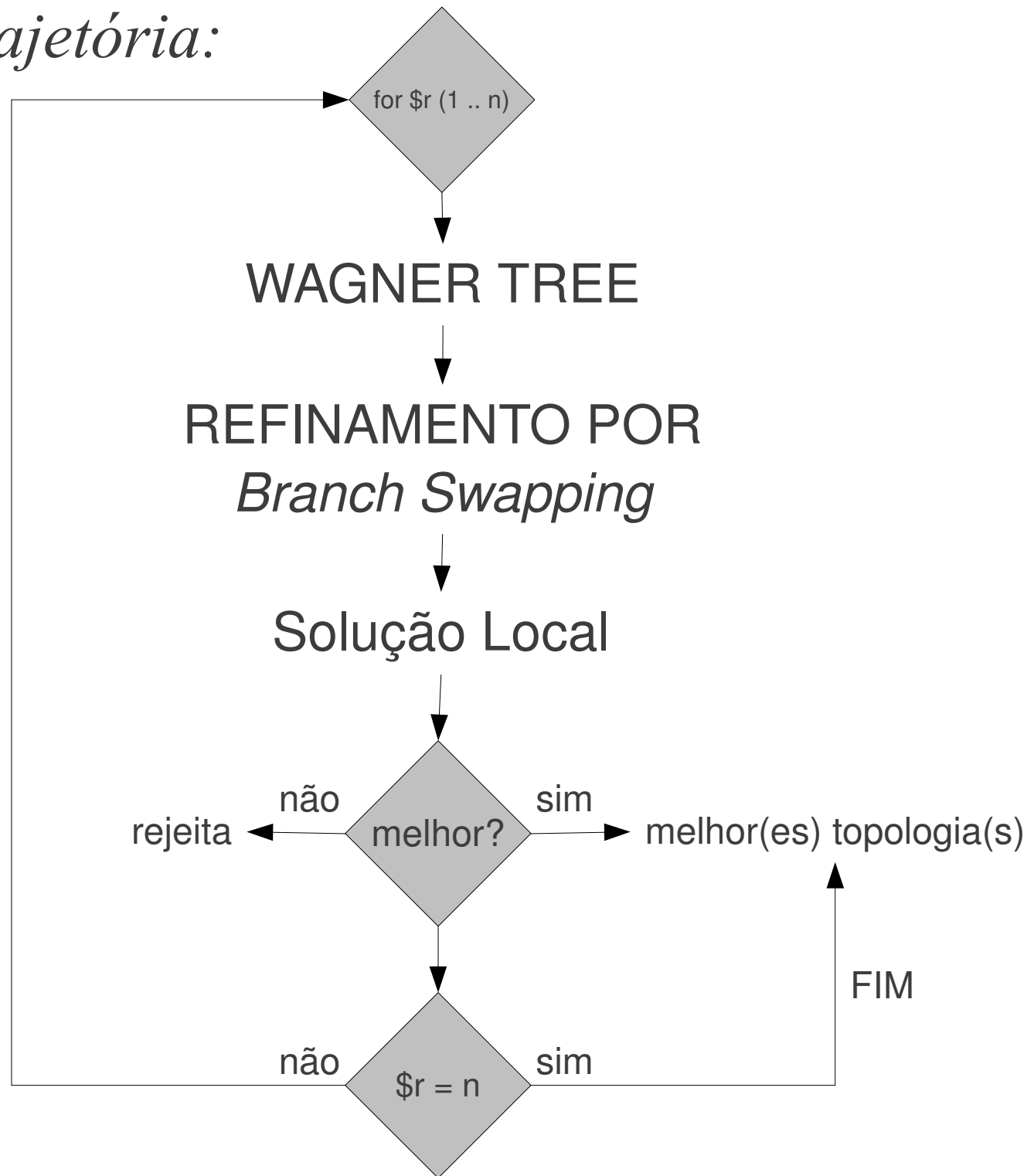
Desafios:

- Explorar adequadamente a paisagem.
- Evitar ótimos locais
- Eficiência computacional

Implementações:

- Buscas por trajetórias
- Aleatorização
- Perturbação
- Recombinação
- Trajetoárias subótimas

Buscas por trajetória:



Buscas por trajetória:

WAGNER TREE

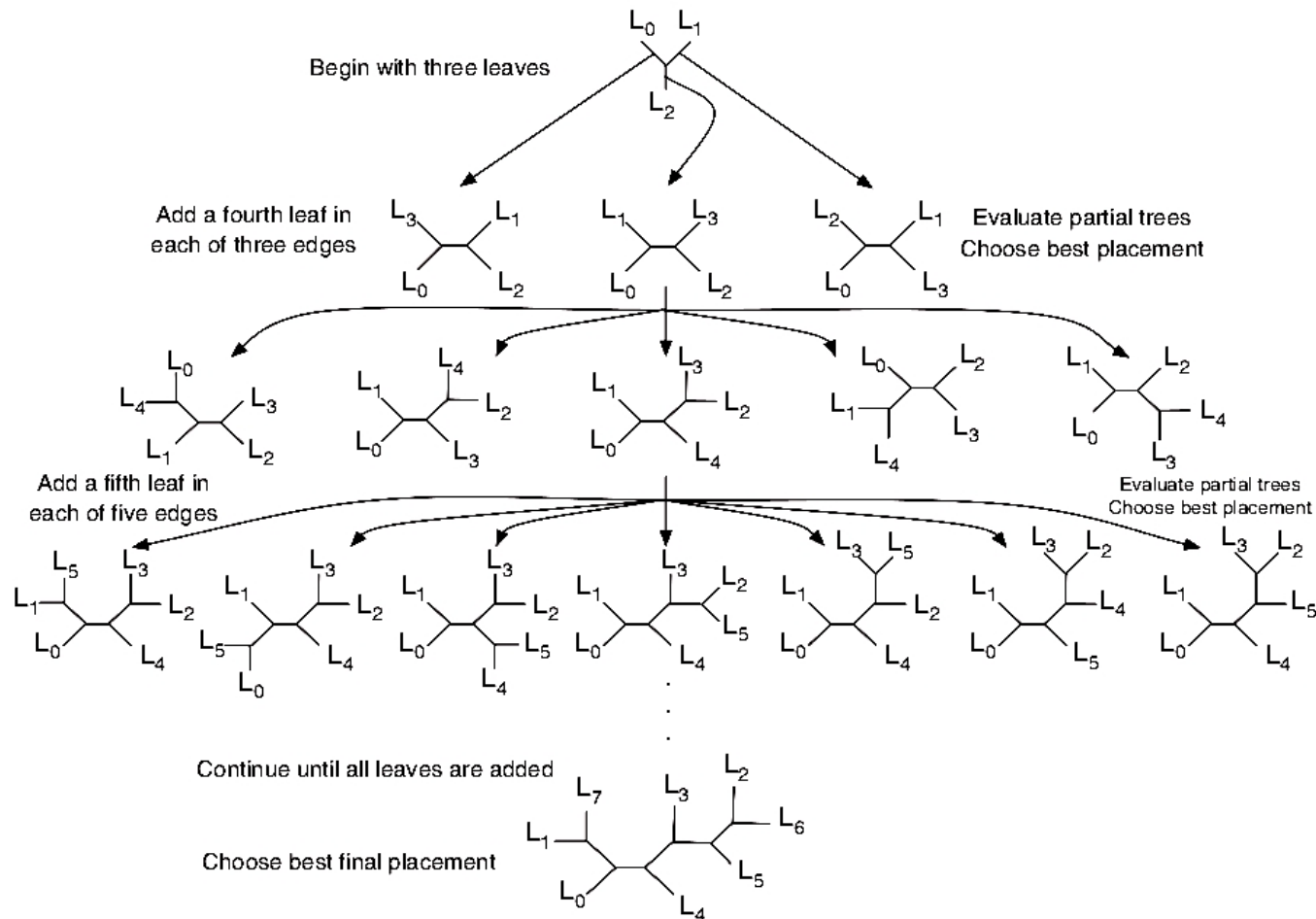
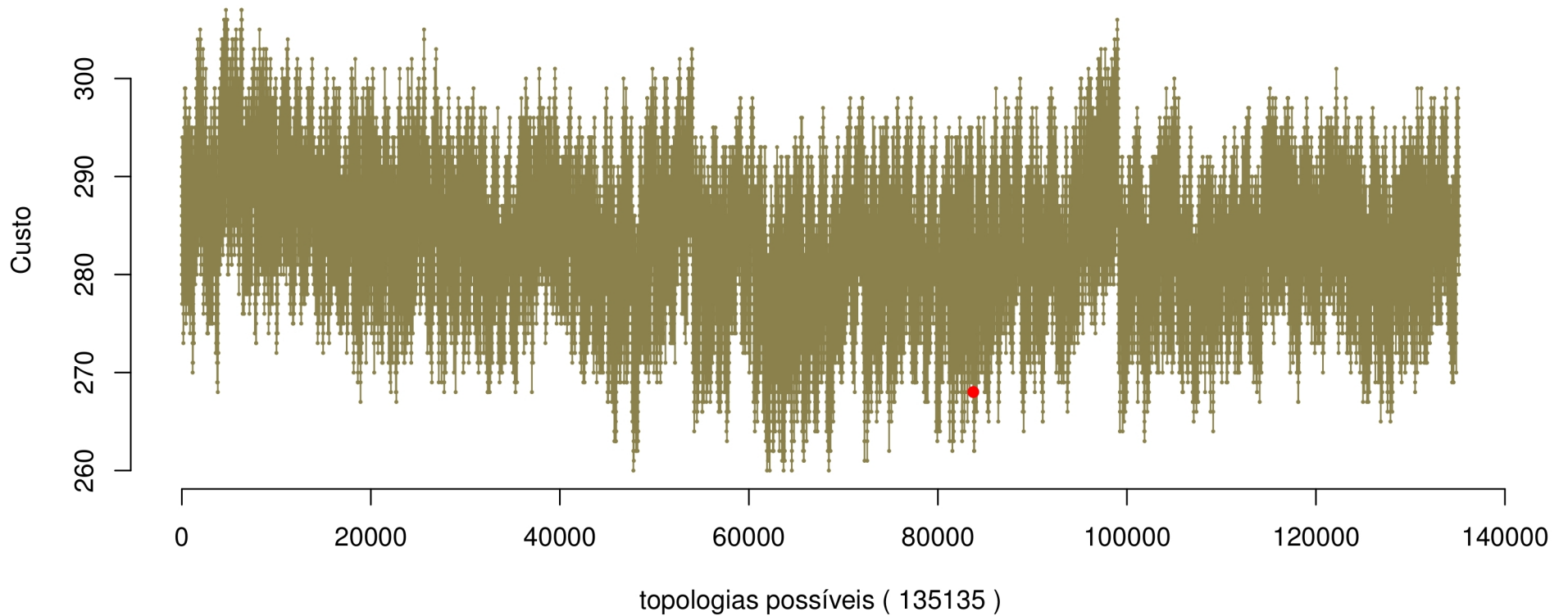


Figure 14.4: Initial tree construction procedure via the Wagner algorithm (Farris, 1970).

Buscas por trajetória: Random addition sequence (RAS)

1 Wagner tree

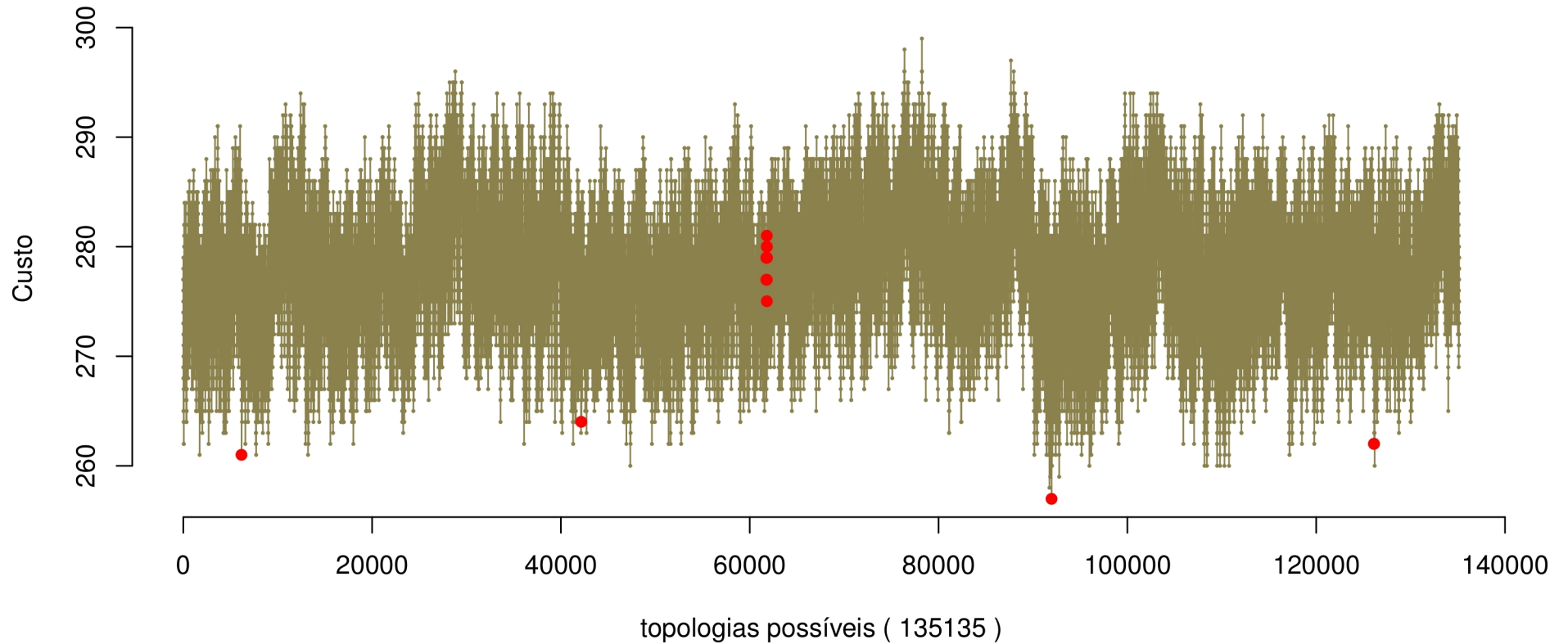
Tree space



Buscas por trajetória: Random addition sequence (RAS)

5 Wagner tree

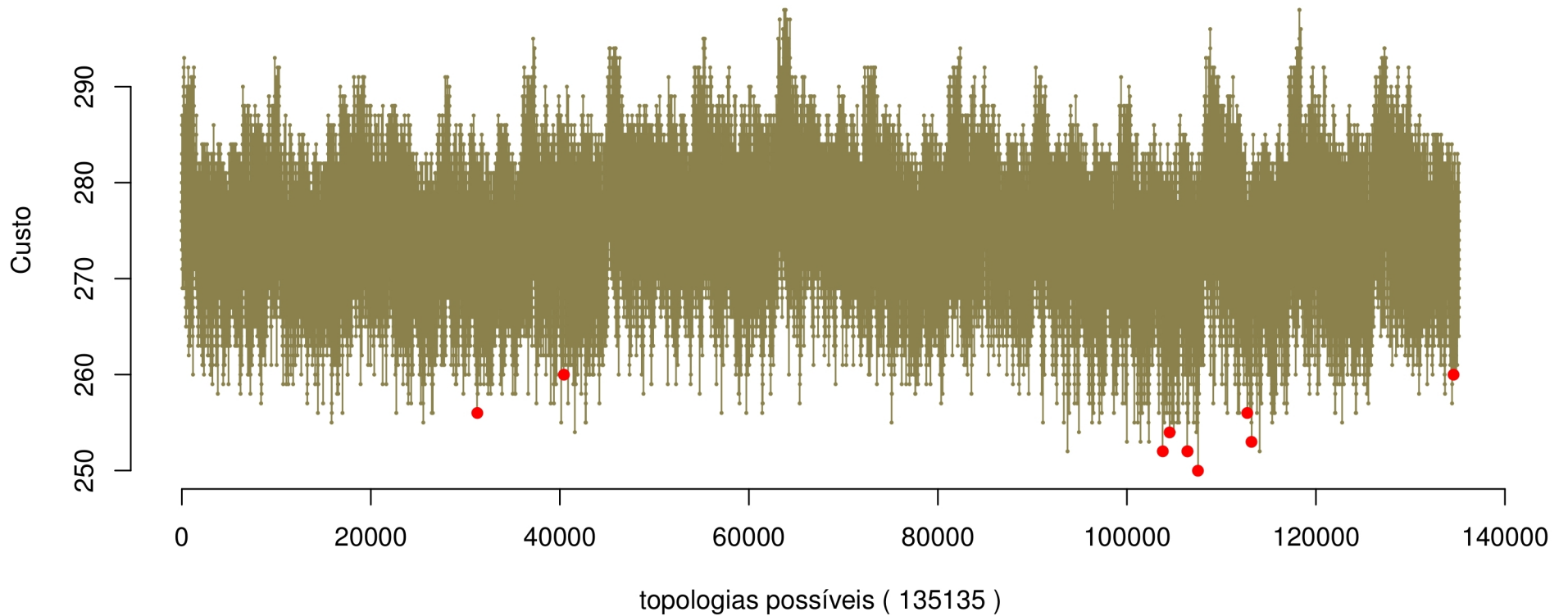
Tree space



Buscas por trajetória: Random addition sequence (**RAS**)

10 Wagner tree

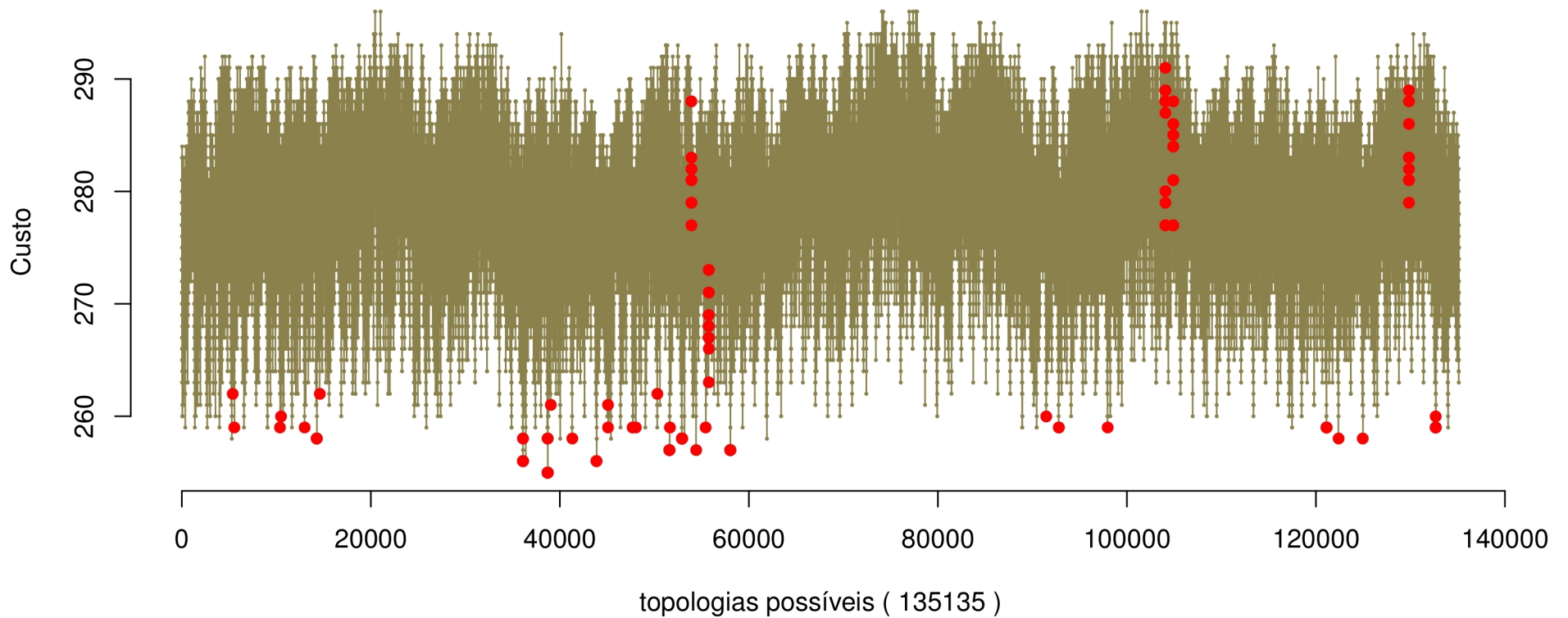
Tree space



Buscas por trajetória: Random addition sequence (RAS)

100 Wagner tree

Tree space



Buscas por trajetória: Refinamento local

Buscas Heurísticas: SPR

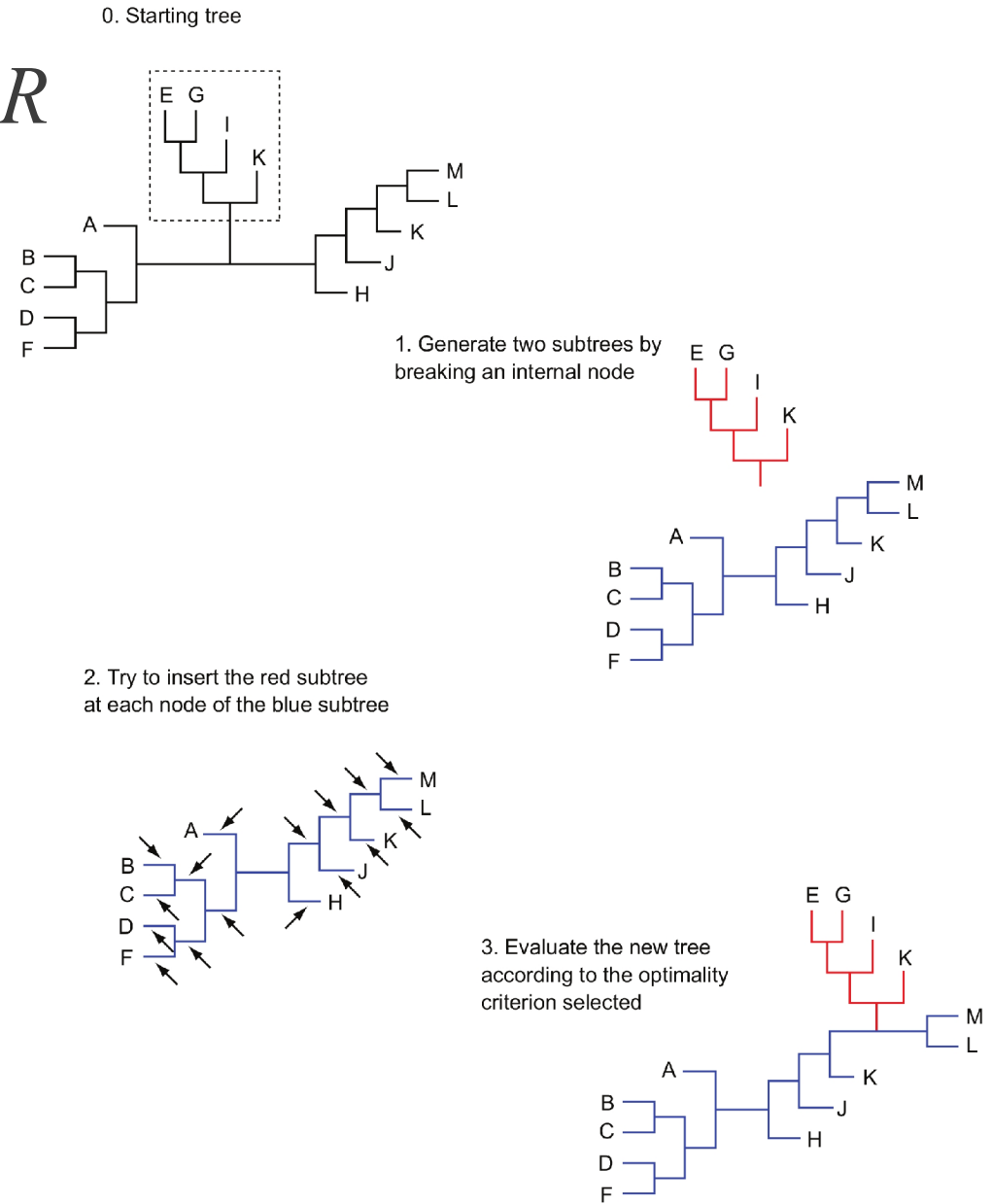
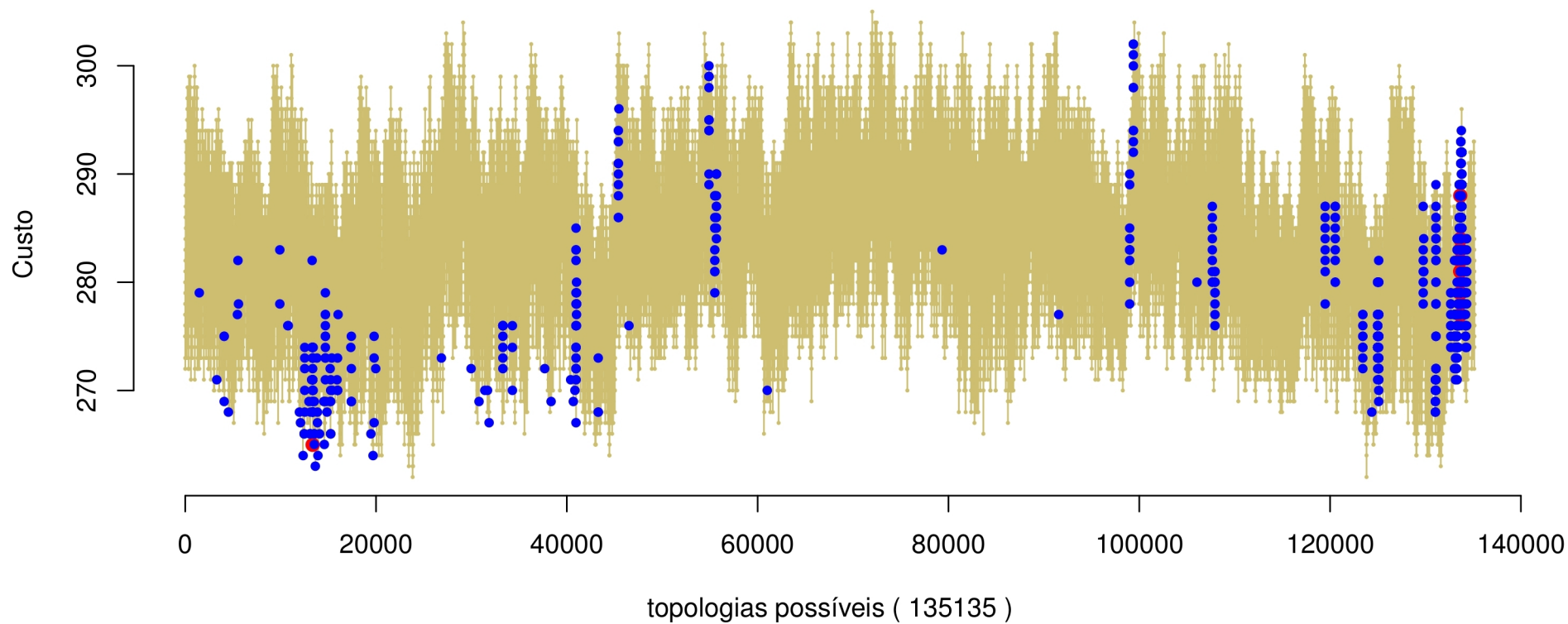


Figure 1. SPR branch swapping. An initial tree (0) gets broken into two subtrees (1). The red subtree is then inserted in each possible branch of the blue subtree (arrows in step 2) and the resulting tree is evaluated (3).

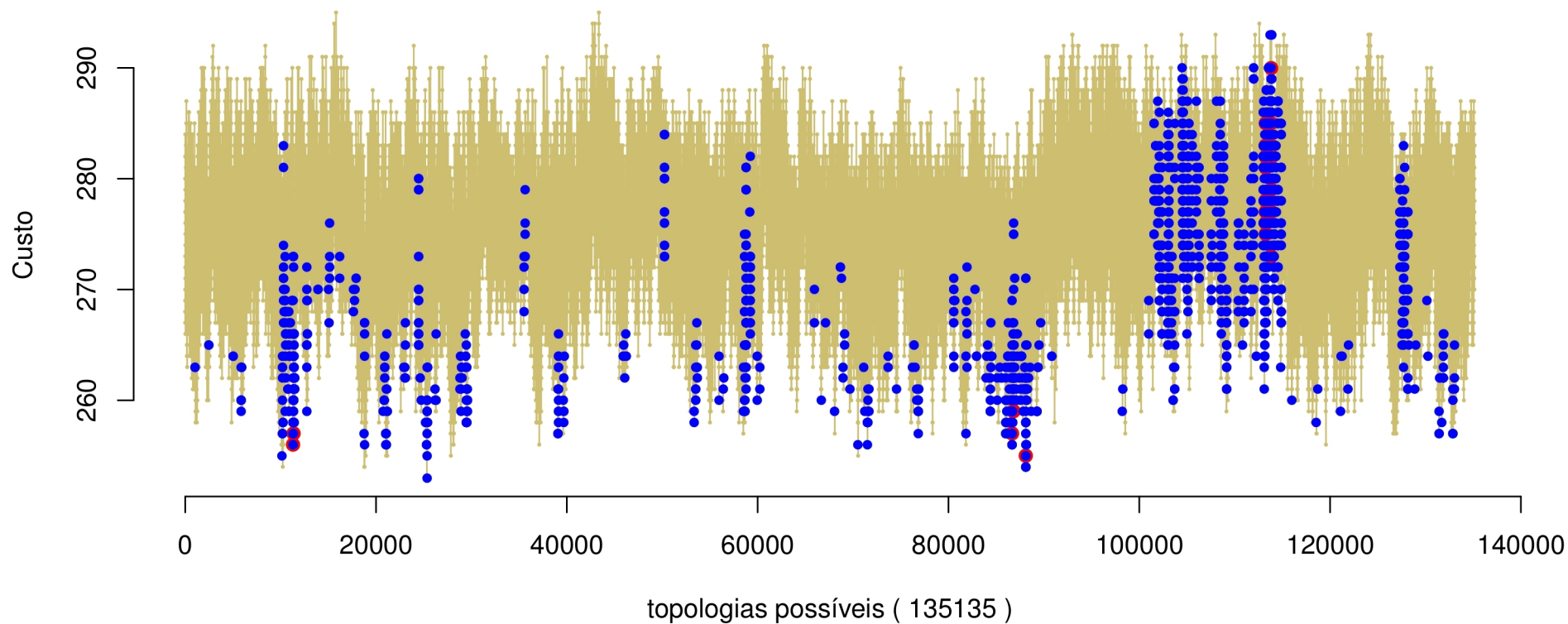
Buscas por trajetória: 1 RAS + SPR

Tree space



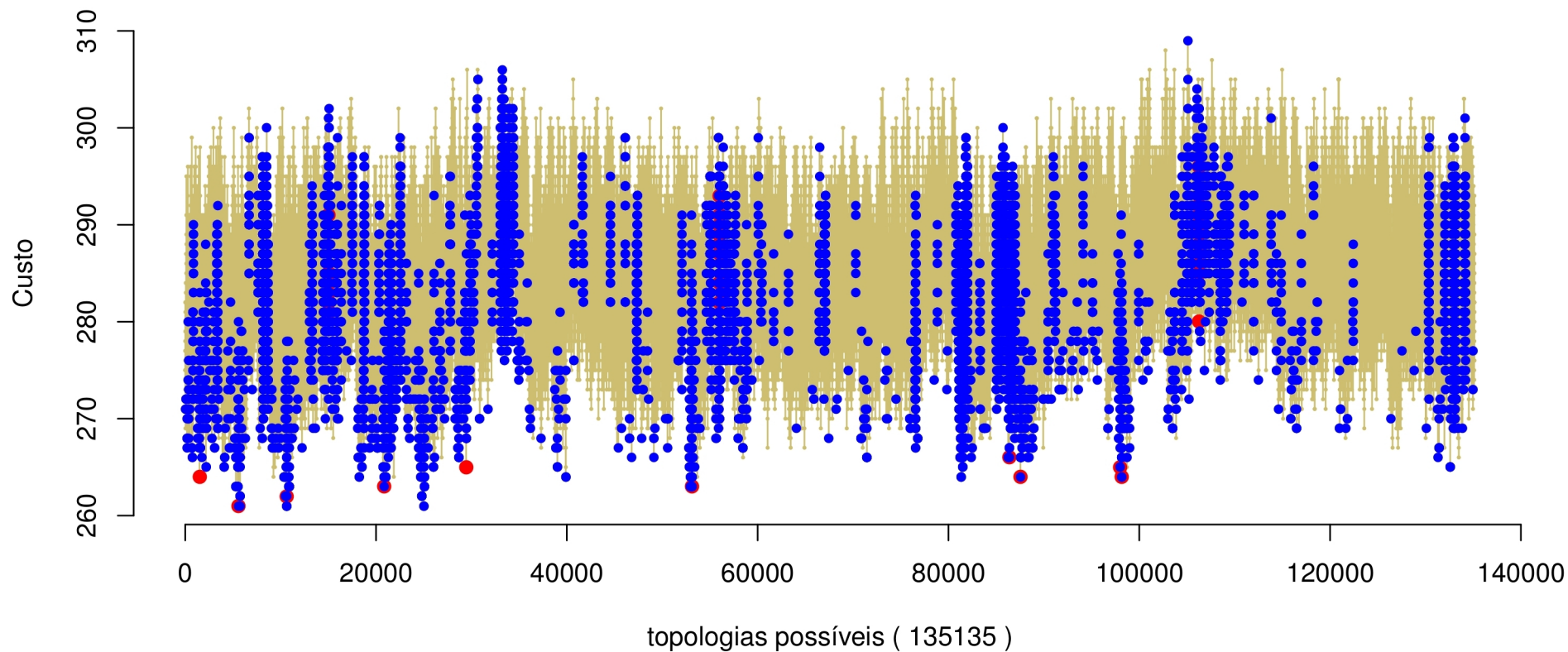
Buscas por trajetória: 5 RAS + SPR

Tree space



Buscas por trajetória: 10 RAS + SPR

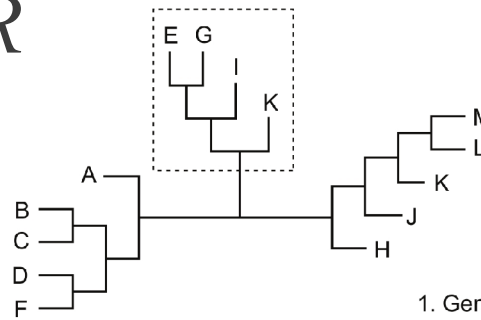
Tree space



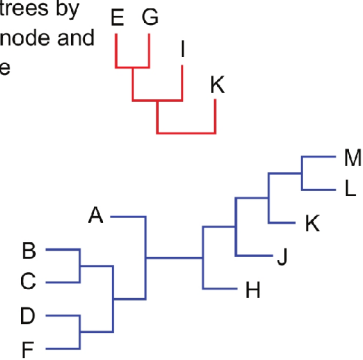
Buscas por trajetória: Refinamento local

Buscas Heurísticas: TBR

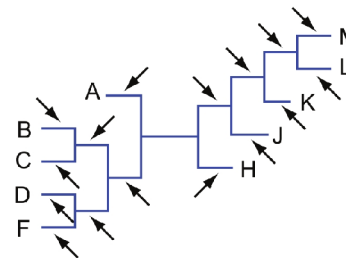
0. Starting tree



1. Generate two subtrees by breaking an internal node and re-rooting the subtree



2. Try to insert all possible rooted red subtrees at each node of the blue subtree



3. Evaluate the new tree according to the optimality criterion selected

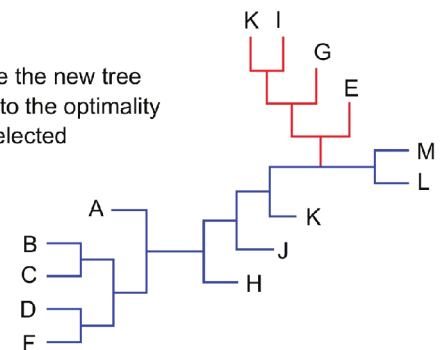
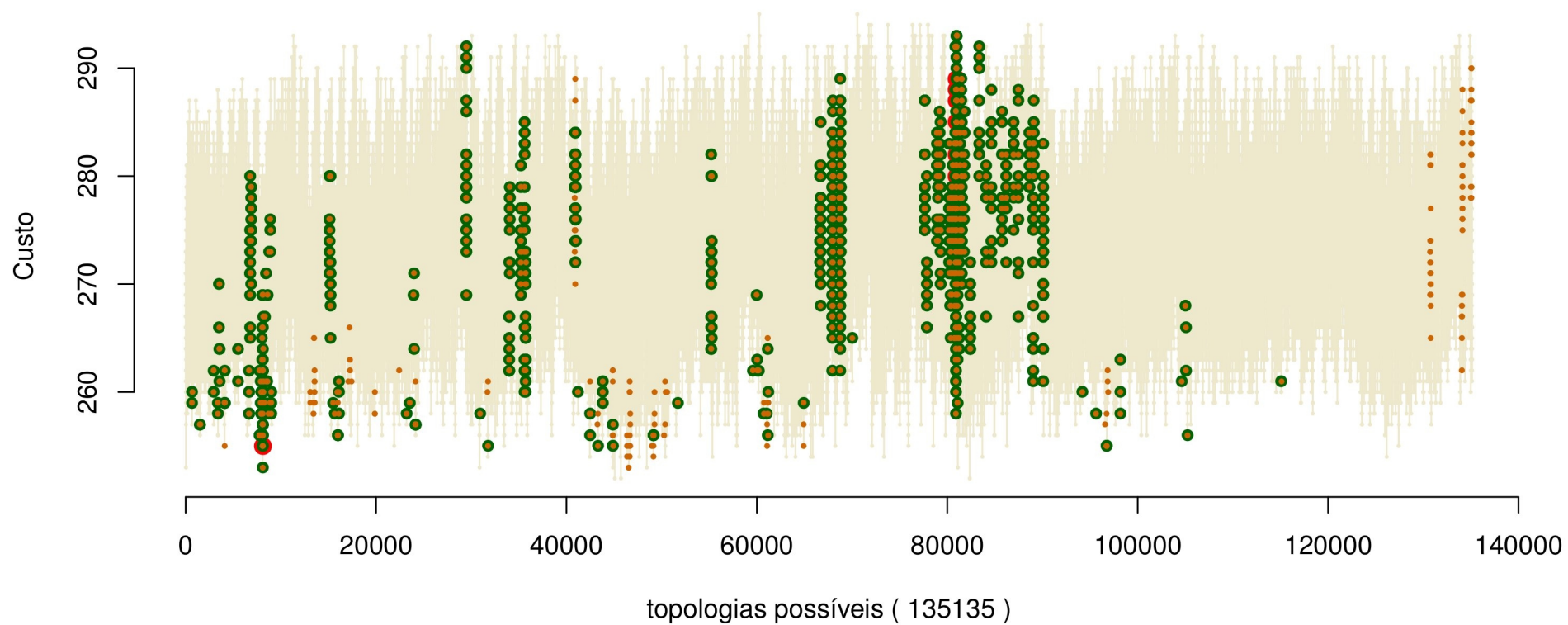


Figure 2. TBR branch swapping. An initial tree (0) gets broken into two subtrees (1). The red subtree is re-rooted on each possible internal branch and inserted in each possible branch of the blue subtree (arrows in step 2) and the resulting tree is evaluated (3).

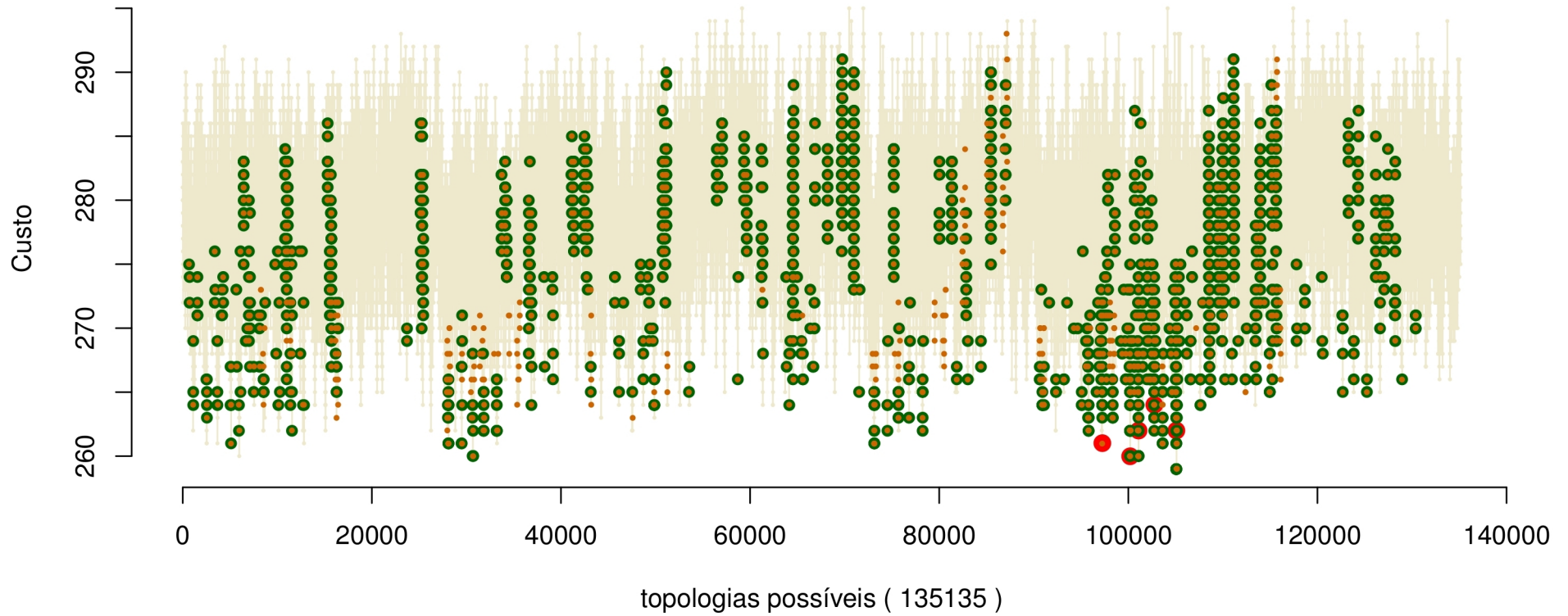
Buscas por trajetória: 1 RAS + TBR

Tree space



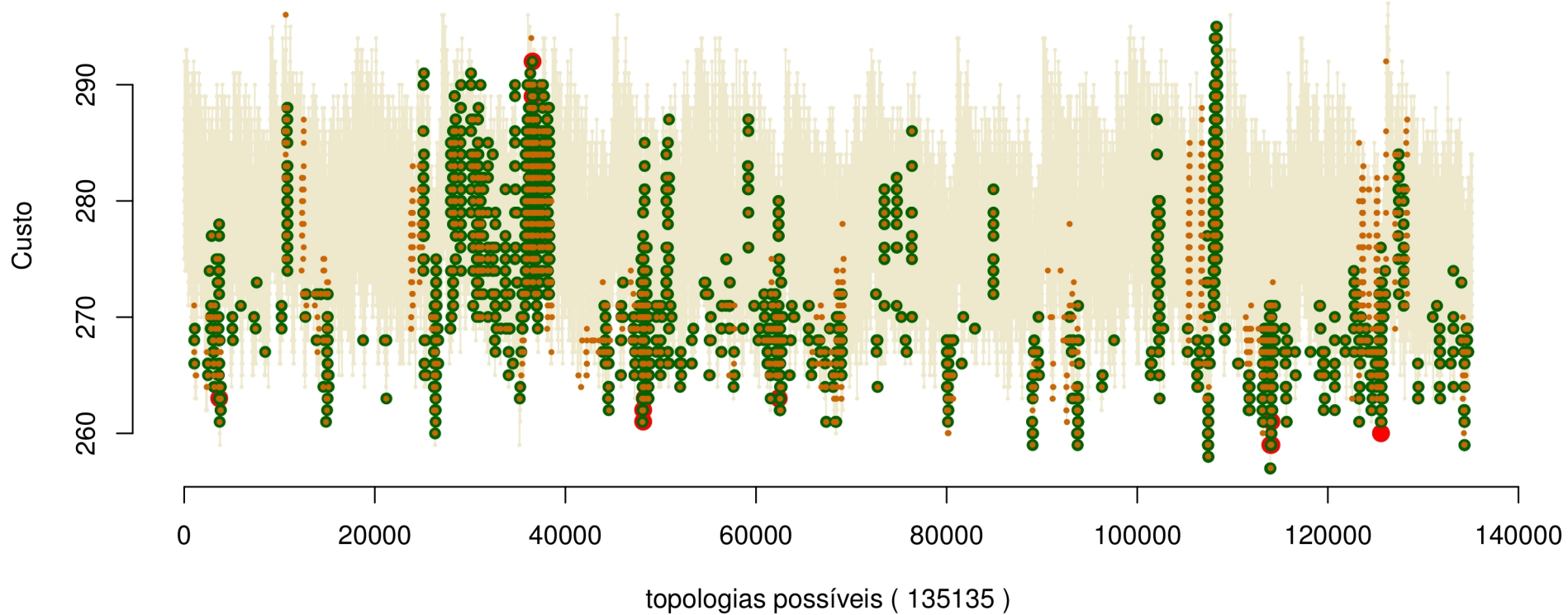
Buscas por trajetória: 5 RAS + TBR

Tree space



Buscas por trajetória: 10 RAS + TBR

Tree space



Buscas Heuristicas: exercício 7.3

mu_hold.run

```
mxram 512;
macro =;
proc zilla.tnt
hold 10000;
log mu_rsa_hold.txt;
loop 1 20
    mu: rep 5 hold #1;
    mu;
stop
log/;
macro -;
```

mu.run

```
mxram 512;
macro =;
proc zilla.tnt
hold 10000;
log mu_rsa.txt;
loop 1 20
    mu: rep #1 hold 5;
    mu;
stop
log/;
macro -;
```

mu.run

```
mxram 512;
macro =;
proc zilla.tnt
hold 10000;
log mu_nested.txt;
loop 1+2 20
    loop 1 20
        mu: rep #1 hold #2;
        mu;
    stop
stop
log/;
macro -;
```

```
grep -P "\d.*" mu_rsa_hold.txt
```