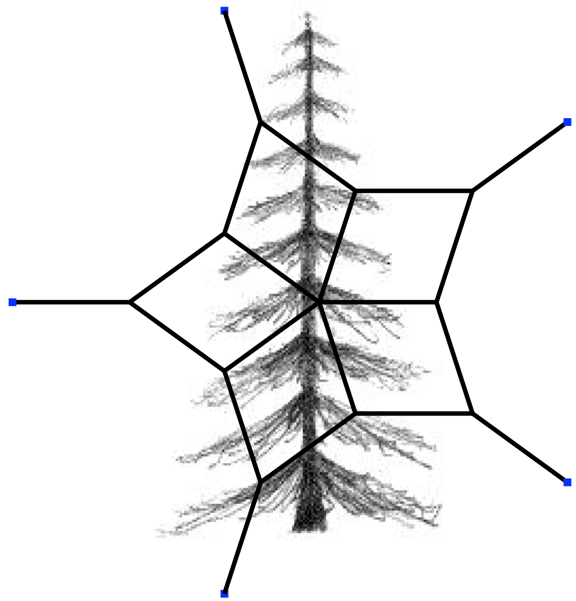
Eine überraschende Anwendung der Geometrie in der phylogenetischen Analyse



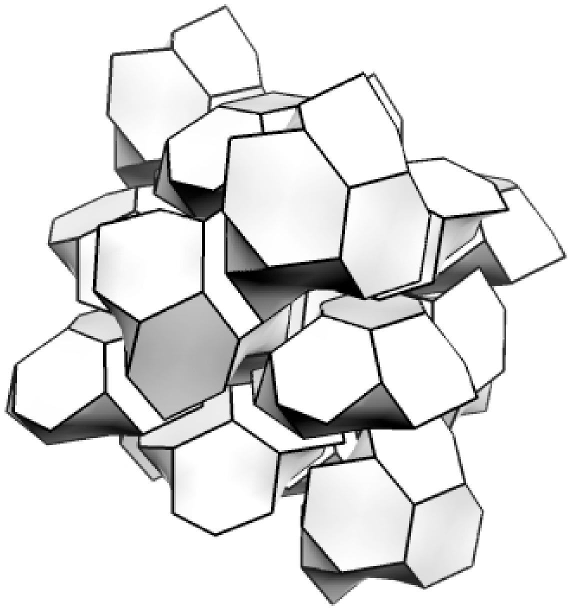
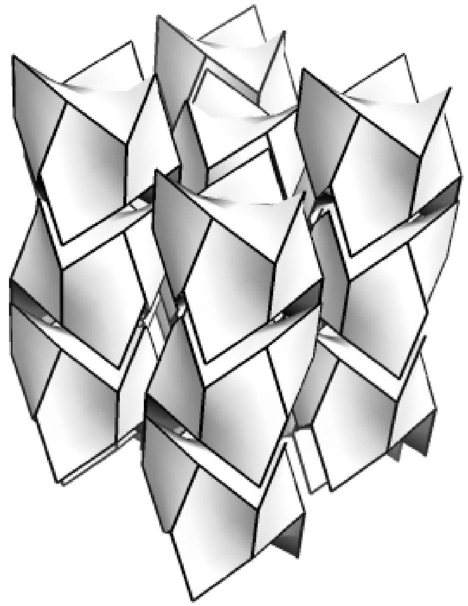
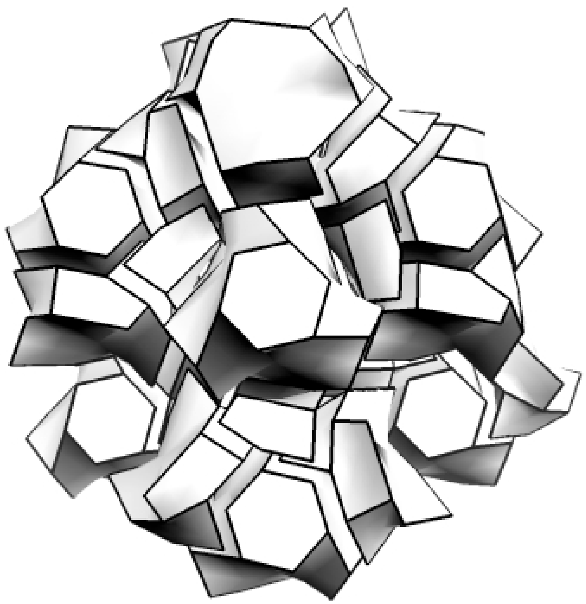
Daniel H. Huson
PACM, Department of Mathematics
Princeton University

huson@math.princeton.edu

Geometrie

- Klassifizierung und Konstruktion aller periodischen Pflasterungen in allen drei zweidimensionalen Geometrien (u.a. mit O. Delgado)
- Anwendung: Design von Reifenprofilen (mit Continental AG Hannover)
- Klassifikation und Konstruktion periodischer Pflasterungen des dreidimensionalen Euklidischen Raumes (mit O. Delgado)
- Delaunay Komplexe für Kristallstrukturen (O. Delgado und N. Dolbilin)
- Anwendung: Aufzählung von Kristallstrukturen, z.B. Zeolite (mit O. Delgado, A. Dress, J. Klinowski und A.L. Mackay)
- Orbifolds und zweidimensionale Symmetriegruppen (mit J.H. Conway)
- Fibrifolds und einen topologischen Zugang zur Klassifikation dreidimensionaler kristallographischer Gruppen (mit J.H. Conway, O. Delgado und W.P. Thurston)

3D Pflasterungen

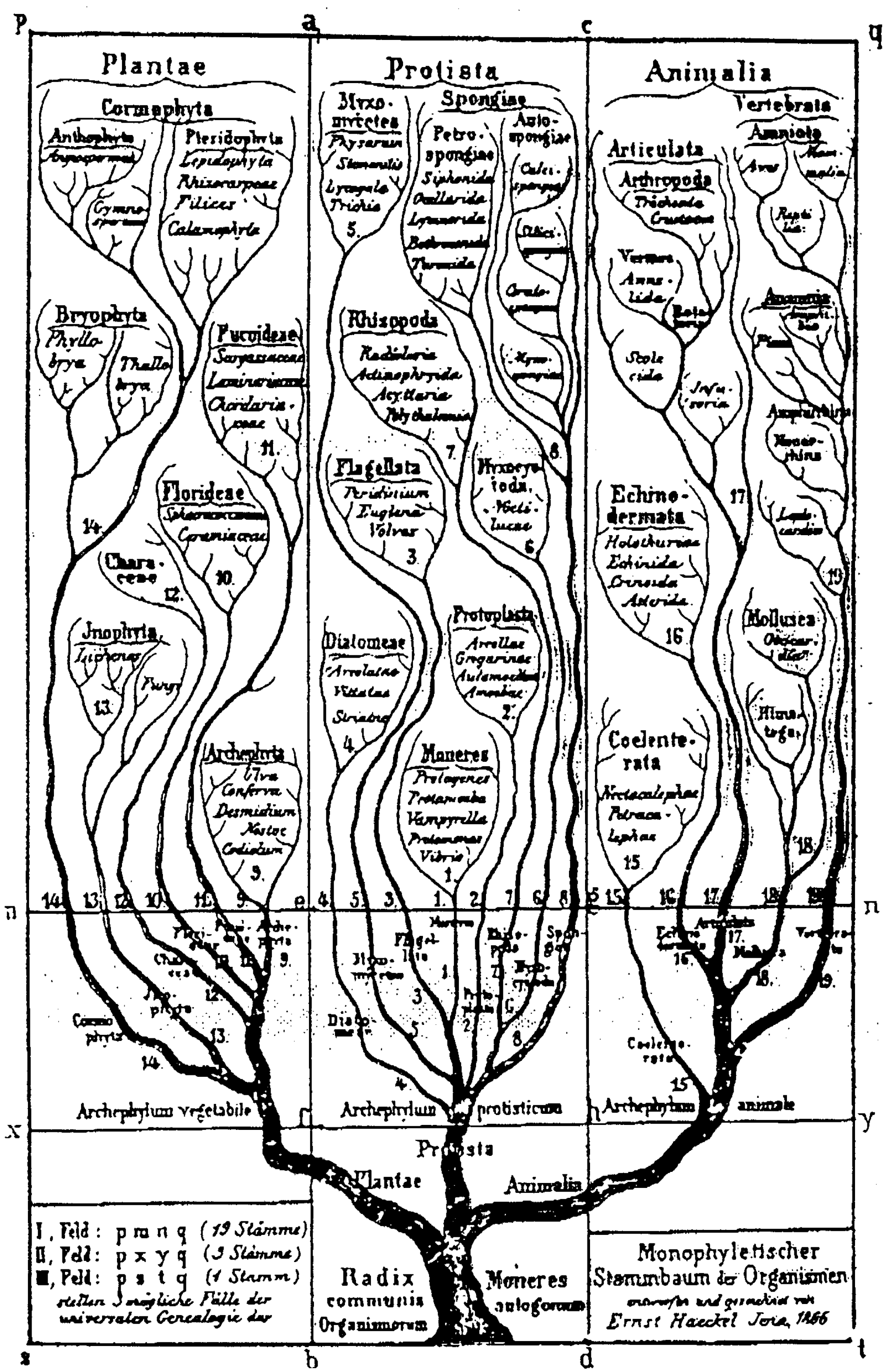




Bio-Mathematik

- Splitzerlegungstheorie und Anwendung in der Biologie (u.a. mit H.-J. Bandelt und A. Dress)
- Divide-and-Conquer Methoden zur Rekonstruktion grosser Bäume (mit L. Vawter und T. Warnow)
- Baumverfeinerungstechniken (mit T. Warnow)
- Strategien zur Überwindung des “Sättigungsproblemes” bei der phylogenetischen Distanzberechnung (mit K. Smith und T. Warnow)
- Anwendungen der Kombinatorik und Geometrie in der medizinischen Bildverarbeitung (mit A. Dress und W. Schubert)
- Untersuchung der Verbreitung von alpinen Pflanzen (u.a. mit P. Lockhart und M. Steel)

Überblick

$$\begin{tikzcd}
\text{Phylogenetische Bäume} \arrow[r] \arrow[d] & \text{kompatible Splits} \arrow[d] \\
\text{Phylogenetische Netzwerke} \arrow[r] \arrow[d] & \text{schwache-kompatible Splits} \\
\text{Realisierung} & \text{Penrose Pflasterungen} \arrow[l]
\end{tikzcd}$$

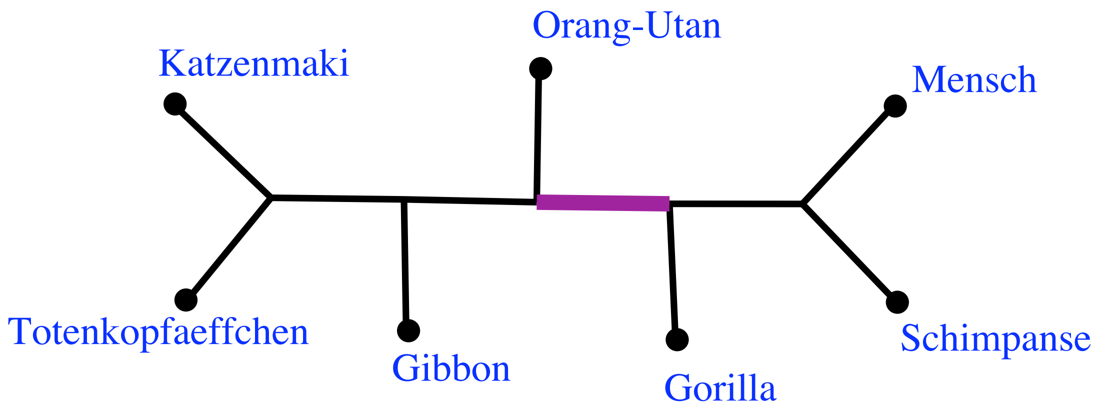


Baum des Lebens

"Monophyletischer Stammbaum der Organismen"

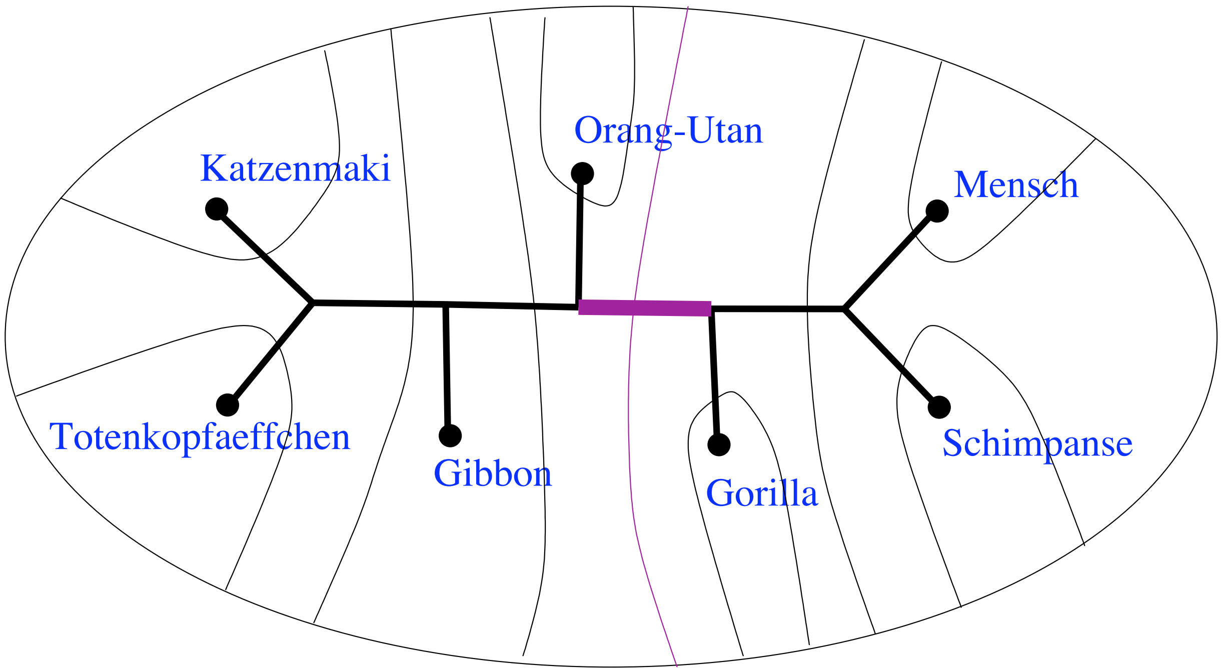
Ernst Haeckel
Jena, 1866

Kompatible Splitsysteme

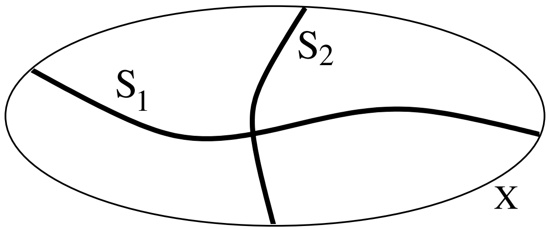


Jede Kante e induziert einen *Split* $A \dot{\cup} B = X$

Kompatible Splitsysteme



Zwei Splits sind *kompatibel*, falls einer der vier möglichen Schnitte leer ist:



Bäume und Splitsysteme

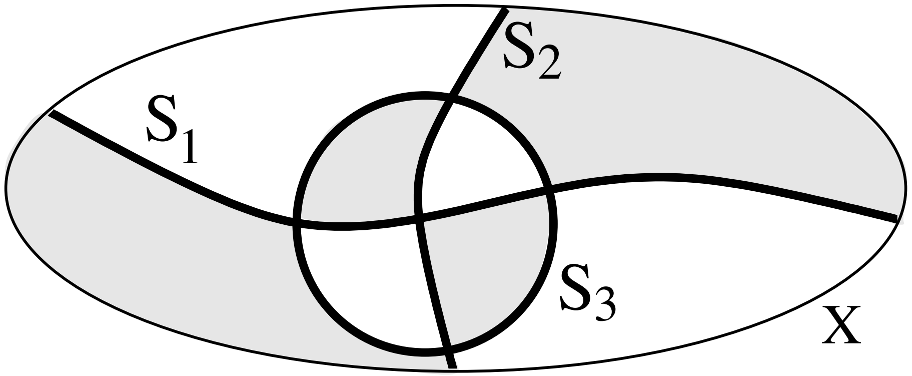
Satz (Buneman 1971)
"Phylogenetischer Baum $T \equiv$ kompatibles Splitsystem Σ"

Problem: Phylogenetische Daten sind nicht immer "baumhaft"
(Rück- und Parallelmutationen, Sättigung, Rekombination u.s.w.)

Thema des Vortrages:
Definition, Berechnung und Darstellung phylogenetischer Netwerke

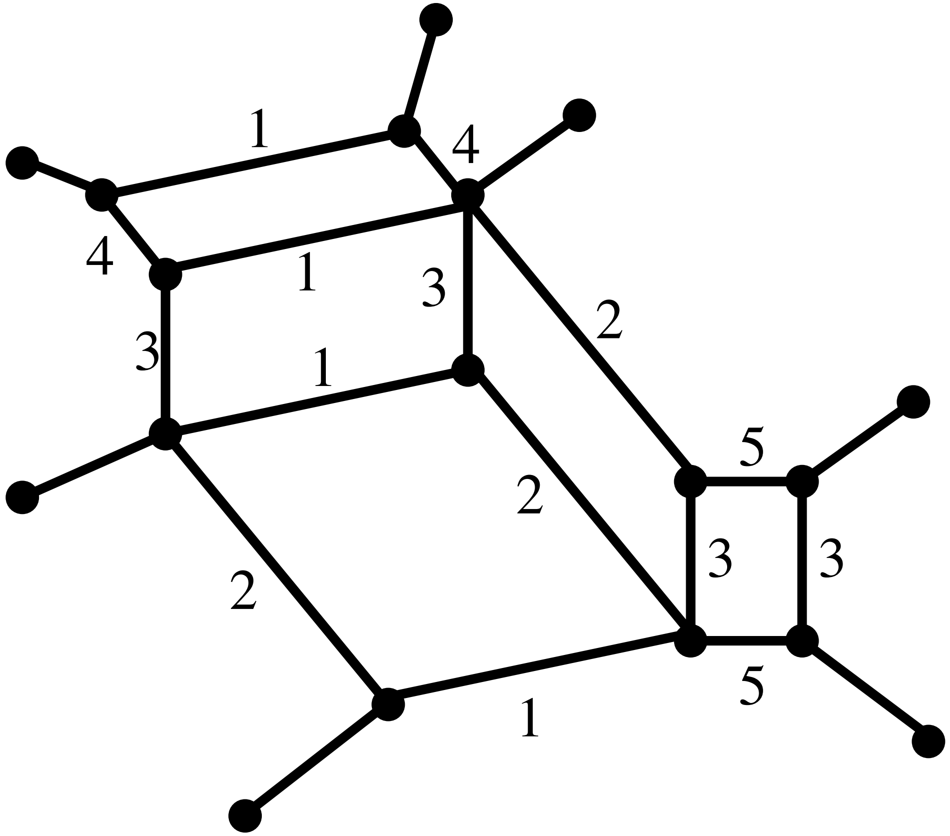
Schwach-kompatible Splitsysteme

H.J. Bandelt und A. Dress (1992): Drei Splits S_1, S_2 und S_3 heissen *schwach kompatibel*, falls hier mindestens ein weisses und ein graues Feld leer ist:



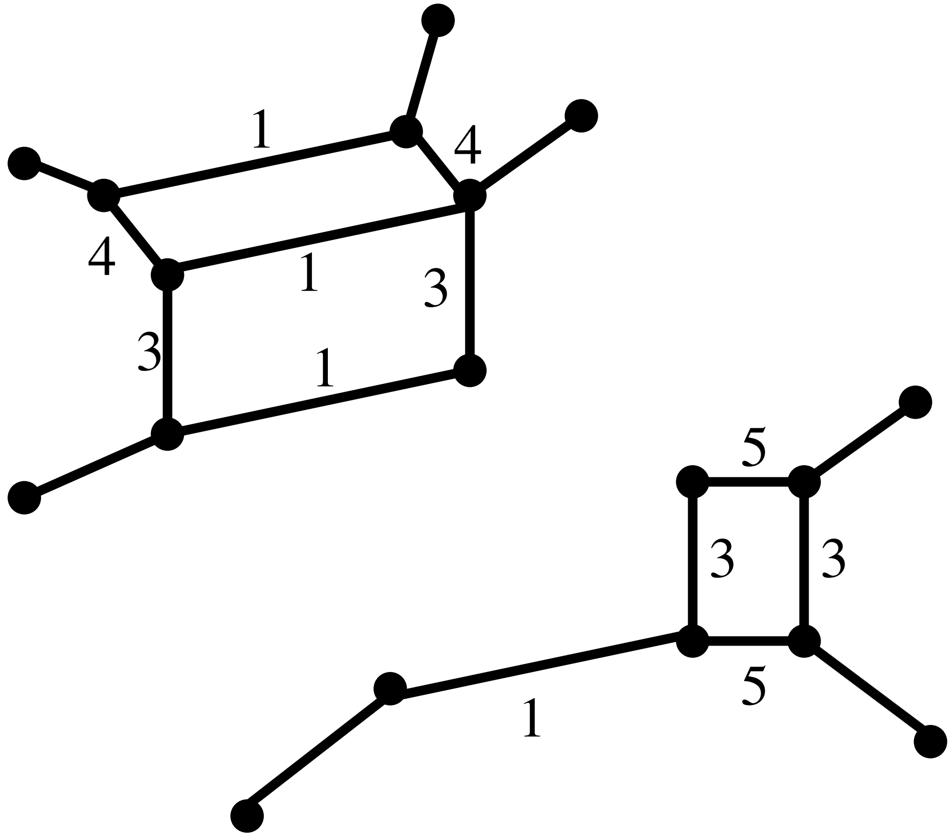
Solche Splitssysteme werden z.B. mit Hilfe der “Splitszerlegungsanalyse” von Bandelt und Dress aus Sequenz- oder Distanzdaten berechnet.

Splitgraphen



Splitgraph: zusammenhängend, bipartit, isometrisch

Beobachtung

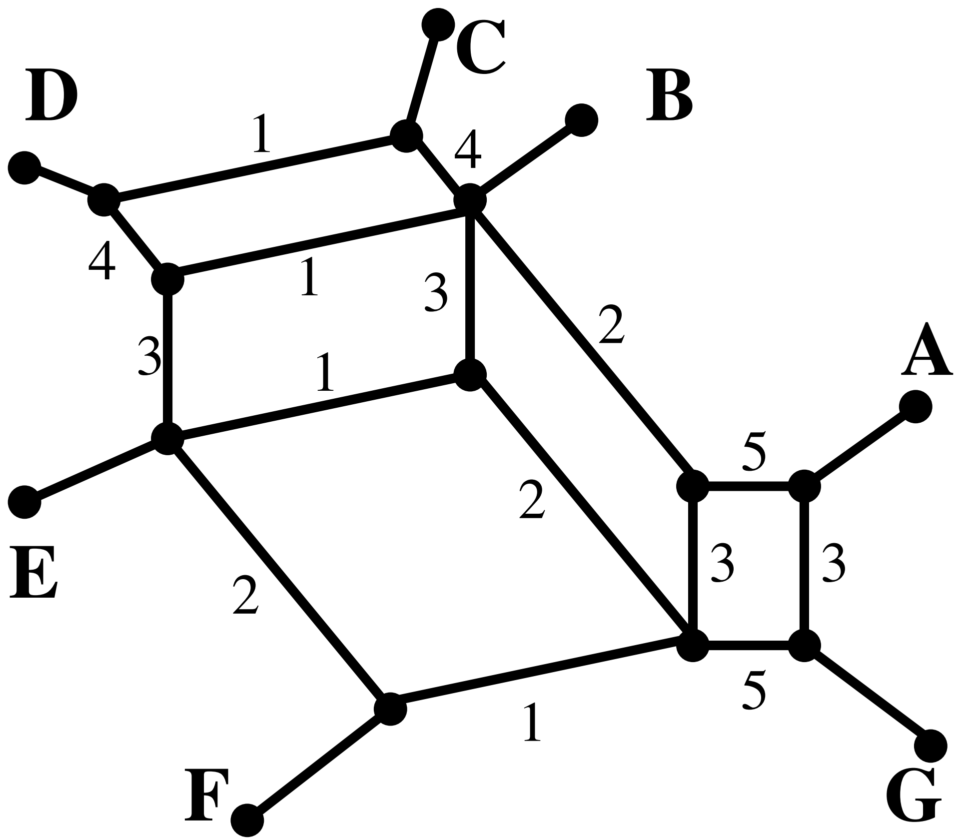


Entfernt man alle Kanten einer Farbe, so erhält man genau zwei Zusammenhangskomponenten.

Repräsentation von Splitsystemen

Gegeben Σ:

$\{D, E, F\}$,
$\{B, C, D, E\}$,
$\{E, F, G\}$,
$\{C, D\}$,
$\{B, C, D, E, F\}$.



Split S_i wird durch die Kanten der Farbe i repräsentiert.

Fragen

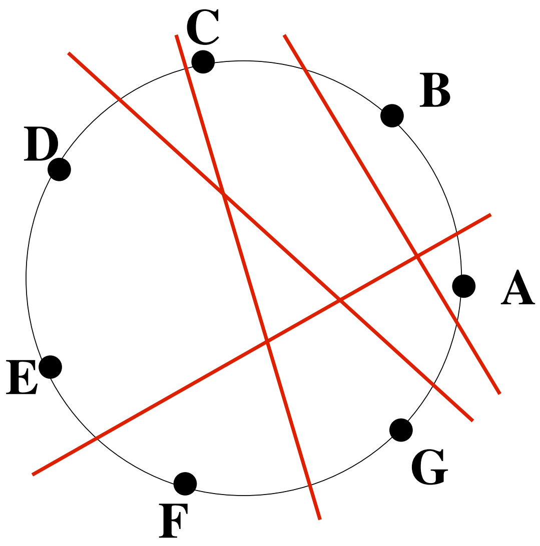
- Existenz?

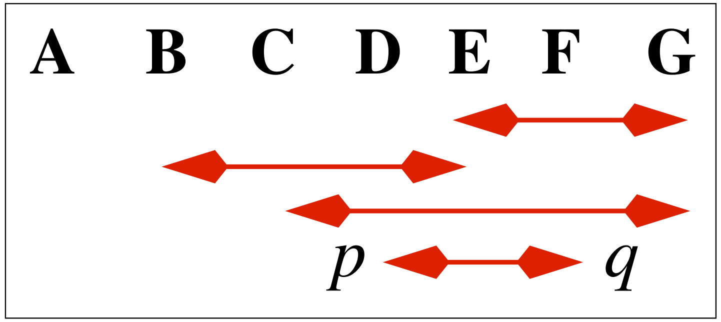
- Effiziente Berechnung?

- Eigenschaften, z.B. Planarität, Anzahl von Ecken, Kanten etc?

Zirkuläre Splitsysteme

Ein Splitsystem Σ heißt *zirkulär*, wenn es sich so darstellen läßt:





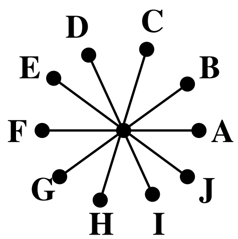
Intervalgraph

(Erkennung in $O(n \cdot k)$:
S. Booth & S. Luecker 1971)

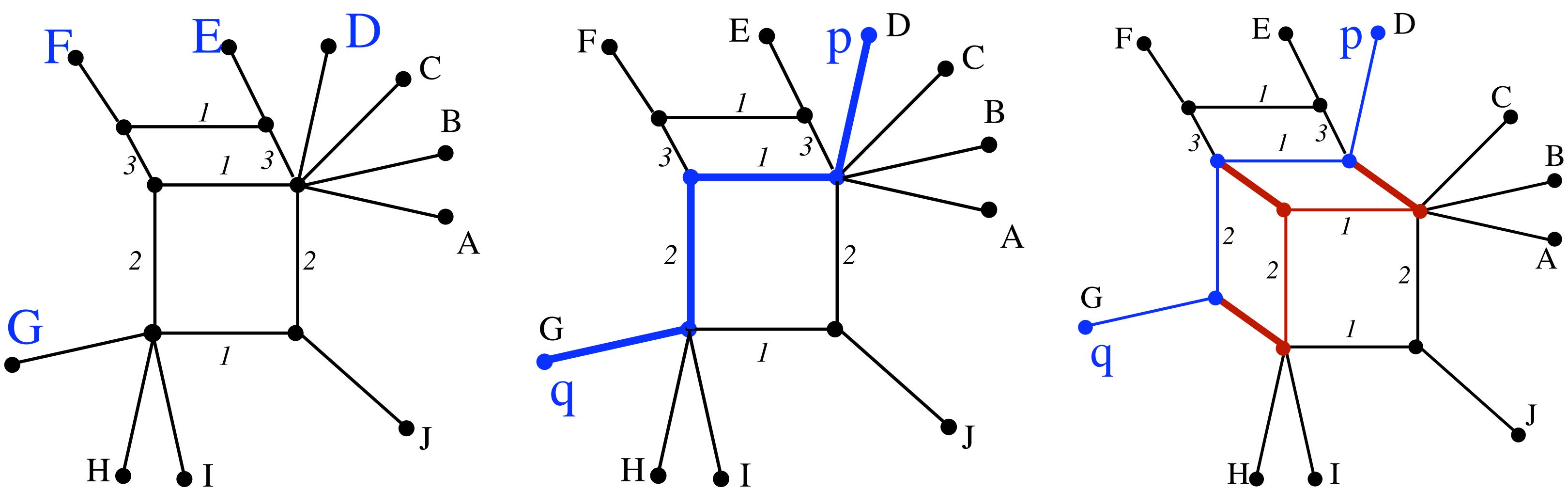
Biologische Relevanz: Die Splitzerlegungsanalyse (Bandelt & Dress) liefert für Sequenzdaten meistens ein zirkuläres System

Konstruktion des Splitgraphen

Gegeben ein Splitsystem Σ auf X, zirkulär (mit Anordnung $A, B, \ldots, J$).

G_0: 

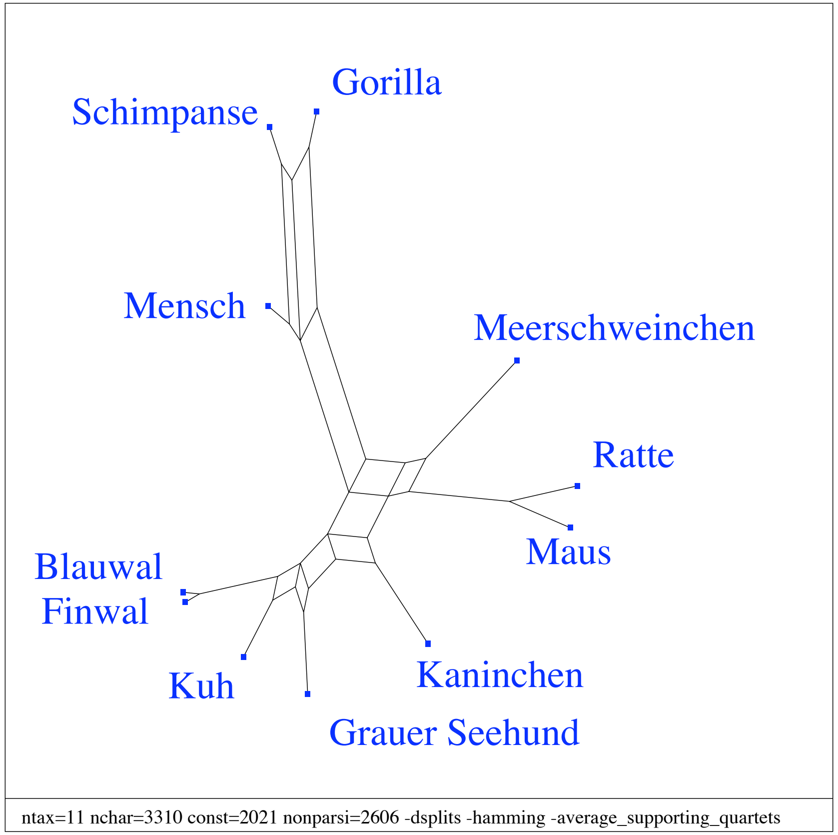
$G_t \to G_{t+1}$: Einbau des Splits $\{D, E, F, G\}$ vs $\{A, B, C, H, I, J\}$:



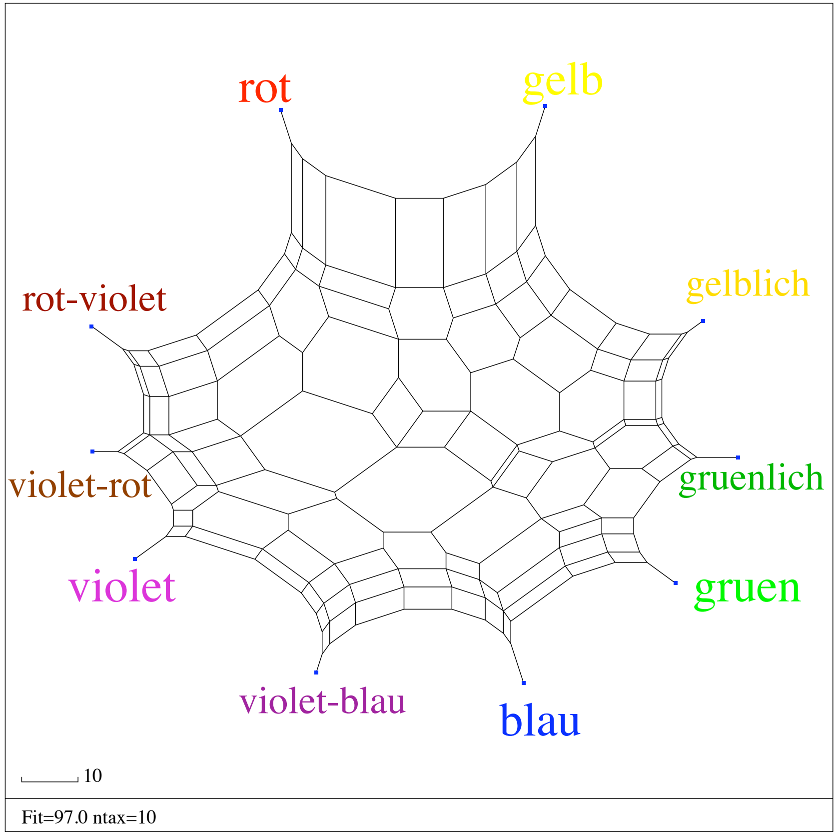
Eigenschaften des Graphen

- Anzahl der Außenkanten $= 2n + 2k$
- Anzahl der beschränkten Flächen = Anzahl der paare inkompatibler Splits in Σ
- Durchschnittlicher Eckengrad ist < 4
- Durchschnittlicher Eckengrad der unmarkierten Ecken $< \frac{n}{k} + 4$
- Zu erwartende Laufzeit des Algorithmus: $O(k^2)$

Beispiele

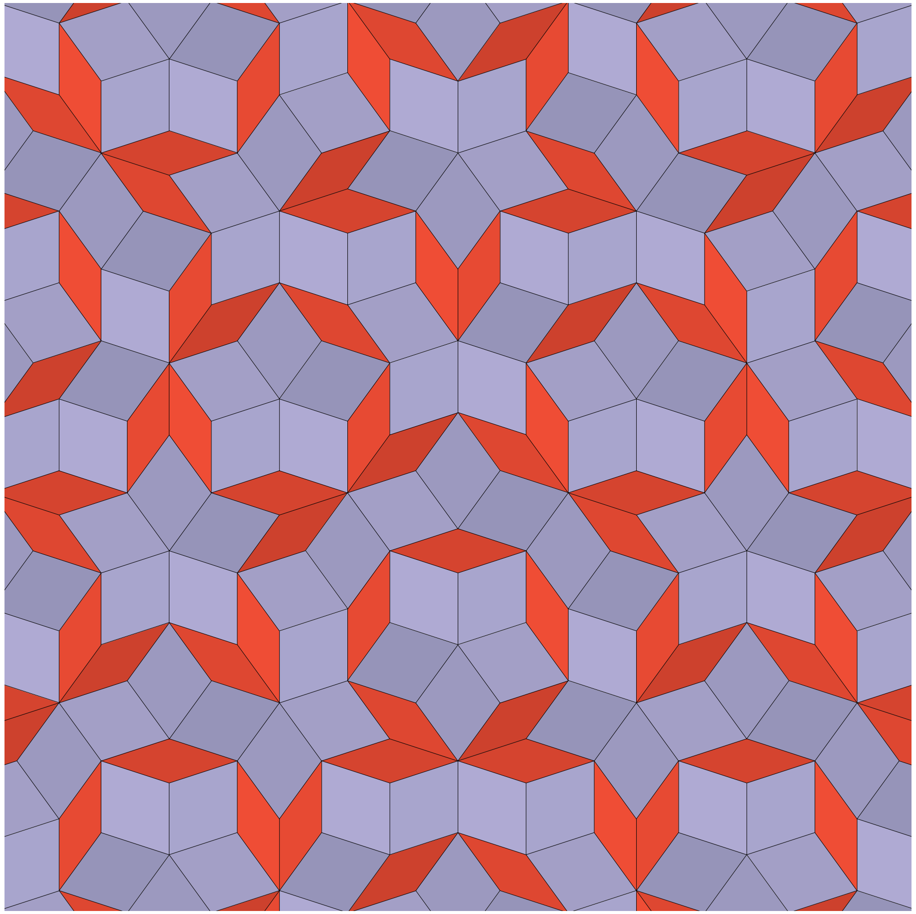


Säugetier mtDNA



Subjektive Farbdistanzen

Penrose Pflasterungen



Quasiperiodische Pflasterungen

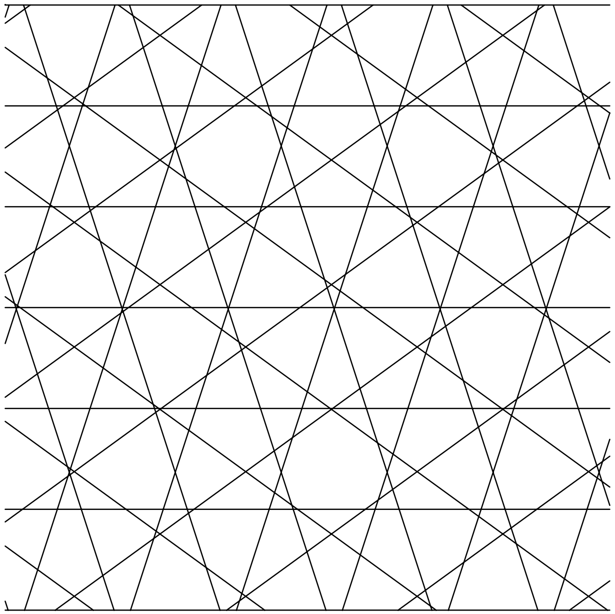
Roger Penrose 1973:

- Matching rules

N.G. de Bruijn 1981:

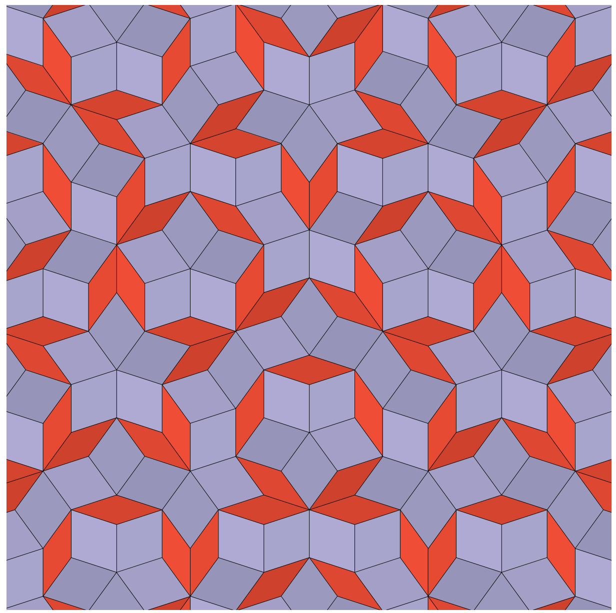
- Projektionsmethode
- Multigridmethode

de Bruijn'sche Multigrid Dualisierung



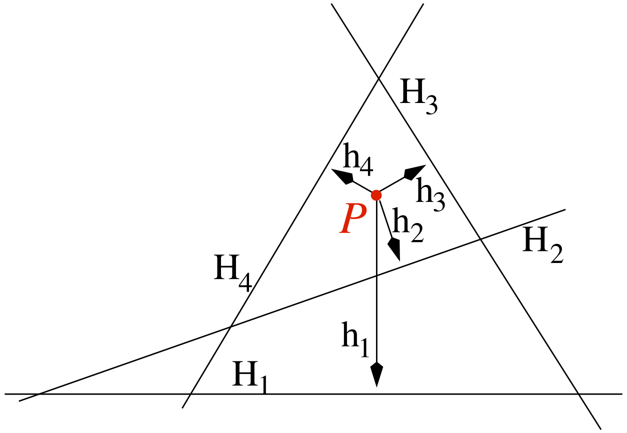
Multigrid

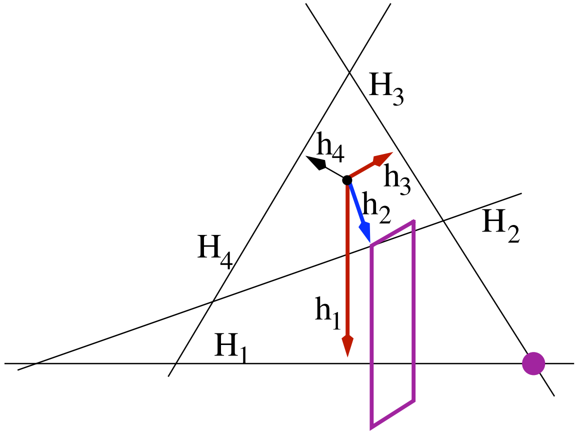
$\longrightarrow$

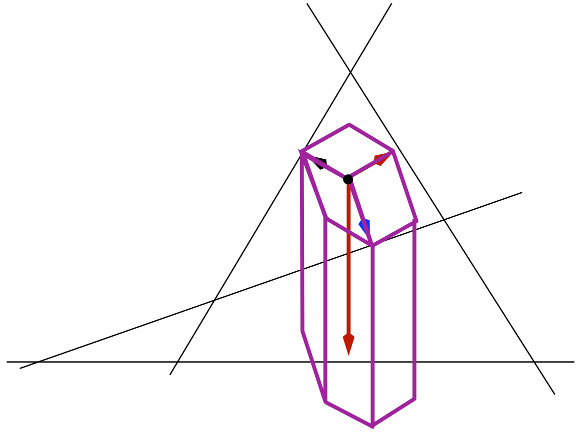


Dual

Berechnung des Duales



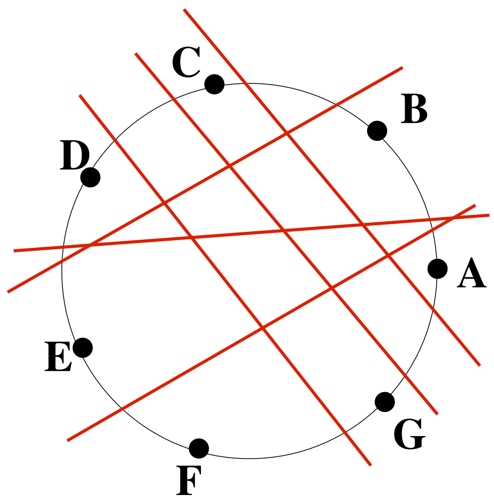




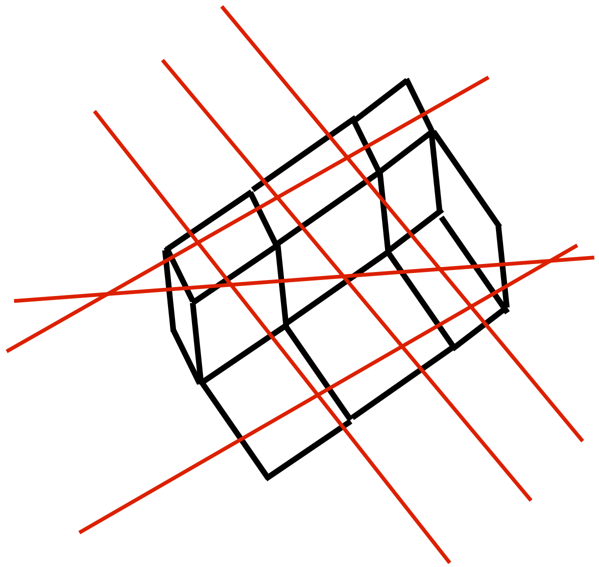
Nach Wahl eines Basispunktes $P \in T(\mathbb{R}^d)$ berechnet man das *Zonotop* $Z(Q)$ eines beliebigen Punktes $Q \in T(\mathbb{R}^d)$ wie folgt:

$$Z(Q) := \sum_i \lambda_i \cdot h_i \quad \text{mit} \quad \lambda_i = \left\{ \begin{array}{c} 1 \\ 0 \\ a \in [0,1] \end{array} \right\} \text{ falls } Q \left\{ \begin{array}{c} \textit{über} \\ \textit{unter} \\ \textit{auf} \end{array} \right\} H_i$$

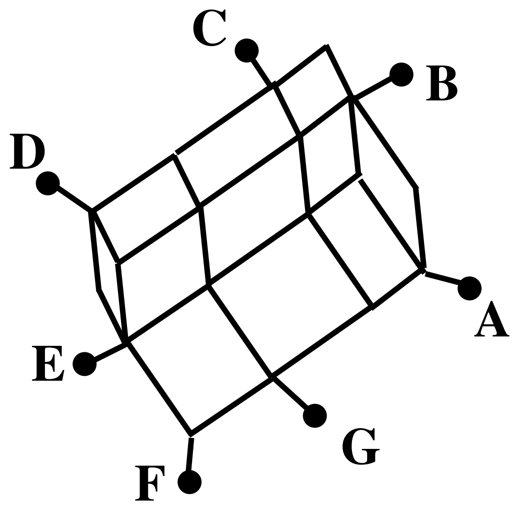
Zirkuläre Splitgraphen durch Dualisierung



Zirkuläres System



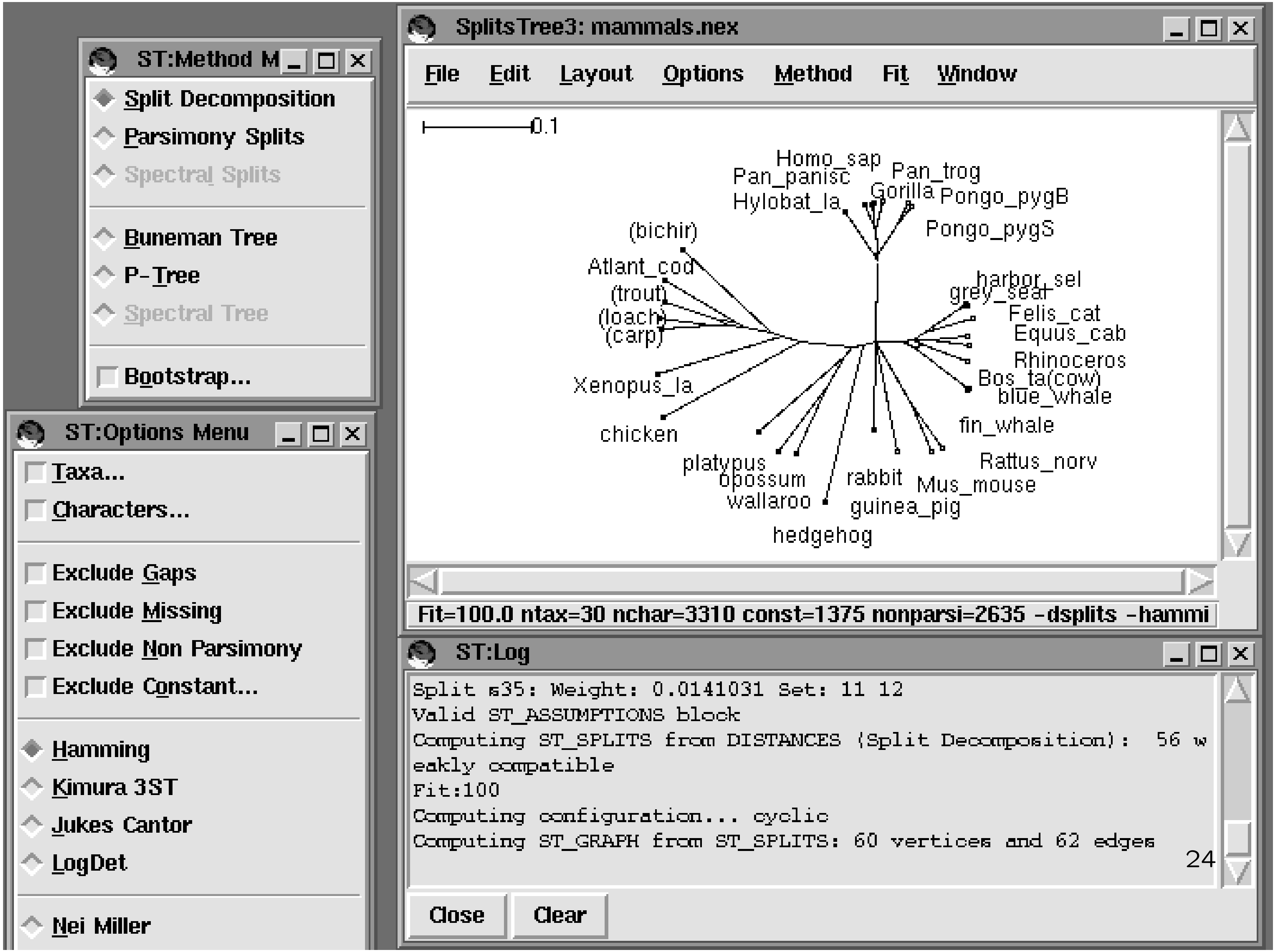
Dualisierung



Splitgraphen

Satz

Ein Splitsystem Σ auf X, welches alle trivialen Splits enthält, ist genau dann zirkulär, wenn es sich durch einen aussenmarkierten ebenen Graphen repräsentieren läßt.


ST:Method M
Split Decomposition
Parsimony Splits
Spectral Splits
Buneman Tree
P-Tree
Spectral Tree
Bootstrap...
ST:Options Menu
Taxa...
Characters...
Exclude Gaps
Exclude Missing
Exclude Non Parsimony
Exclude Constant...
Hamming
Kimura 3ST
Jukes Cantor
LogDet
Nei Miller
SplitsTree3: mammals.nex
File Edit Layout Options Method Fit Window
0.1
Homo_sap
Pan_panisc
Pan_trog
Gorilla
Pongo_pygB
Hylobat_la
Pongo_pygS
(bichir)
Atlant_cod
(trout)
(loach)
(carp)
harbor_sel
grey_seal
Felis_cat
Equus_cab
Rhinoceros
Bos_ta(cow)
blue_whale
Xenopus_la
chicken
fin_whale
Rattus_norv
platypus
opossum
wallaroo
rabbit
Mus_mouse
guinea_pig
hedgehog
Fit=100.0 ntax=30 nchar=3310 const=1375 nonparsi=2635 -dsplits -hammi
ST:Log
Split s35: Weight: 0.0141031 Set: 11 12
Valid ST_ASSUMPTIONS block
Computing ST_SPLITS from DISTANCES (Split Decomposition): 56 w
eakly compatible
Fit:100
Computing configuration... cyclic
Computing ST_GRAPH from ST_SPLITS: 60 vertices and 62 edges
Close
Clear