

# Graphen und biologische Netze (WS 2016/17)

# Inhaltsverzeichnis

# 1 Vorlesung 14.10.2016

## 1.1 Grundlagen der Graphen und biologische Netze

Graph: Knoten, Kanten (binäre Relationen)

Transitivität: implizite Verbindung (abhängig vom Kontext)

Labeled Graphs:

- Graph:  $(V, E)$
- Labels:  $L_V$  (Knotenlabel),  $L_E$  (Kantenlabel)

$e \in E \Rightarrow \exists x, y \in V : x \text{ und } y \text{ sind die Endpunkte von } e$

Knoten-Labelfunktion  $\alpha: V \rightarrow L_V : v \mapsto \alpha(v)$

Kanten-Labelfunktion  $\beta: E \rightarrow L_E : e \mapsto \beta(e)$

### ungerichtete Graphen

- Kante ist eine Menge von 2 (verschiedenen) Knoten
- $e = \{x, y\} = \{y, x\} \rightarrow$  Reihenfolge egal
- $E \subseteq V^{(2)} \rightarrow$  Kante ist Teilmenge von 2 Knoten

### gerichtete Graphen

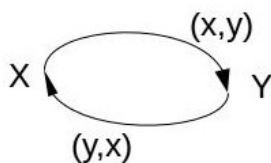
- Kante ist ein geordnetes Paar von 2 (verschiedenen) Knoten
- $e = (x, y)$  entspricht  $x \rightarrow y$ ,  $(y, x)$  entspricht  $y \rightarrow x$
- $E \subseteq V \times V$
- gerichtete Kante besteht aus head (in Pfeilrichtung) und tail

Funktionen gerichteter Graphen:

$h : E \rightarrow V : e \mapsto \text{head}(e)$

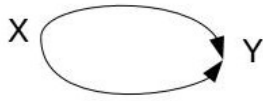
$t : E \rightarrow V : e \mapsto \text{tail}(e)$

Graphen in denen Kanten zwei verschiedenen Endpunkte haben **UND** zu jedem Paar von Kanten höchstens eine Kante gehört heißen EINFACH oder SIMPLE im gerichteten Fall:



trotzdem einfacher Graph!

erst:



ist Multigraph

**Loops:**



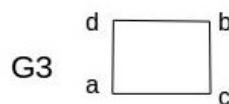
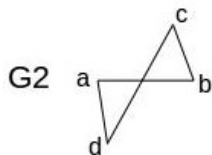
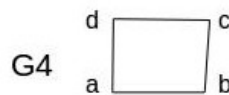
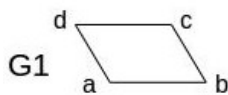
Abbildung 1: links: gerichtet; rechts: ungerichtet

⇒ einfacher Graph mit Loops

Durch Unterteilung der Kanten in Multigraphen kann eine Transformation in Graphen erzeugt werden:

- ungerichtet: zweifache Unterteilung mittels zweier Knoten
- gerichtet: einfache Unterteilung mittels Knoten

## 1.2 Gleichheit von Graphen

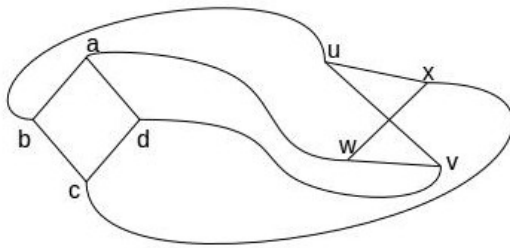


als labeled graphs:  $G_1 = G_2 = G_4 \neq G_3$

⇒ 2 Graphen  $G_1 = (V_1, E_1)$  und  $G_2 = (V_2, E_2)$  sind isomorph wenn es einen bijektive Abbildung<sup>1</sup>  $\pi : V_1 \rightarrow V_2$  gibt, sodass  $\{x, y\} \in E_1 \Leftrightarrow \{\pi(x), \pi(y)\} \in E_2$

---

<sup>1</sup>[https://de.wikipedia.org/wiki/Bijektive\\_Funktion](https://de.wikipedia.org/wiki/Bijektive_Funktion)



bijektive Abbildung: jedes Element von

1. wird zu genau einem Element von 2. zugeordnet

$$\pi(a) = w, \pi(b) = u, \pi(c) = x, \pi(d) = v$$

→ hier ergibt bijektive Abbildung keinen Isomorphismus, da Bild(d) und Bild(c) Kante haben, jedoch v und x keine Kante haben

Durch folgende bijektive Abbildung wird aber Isomorphie erreicht:

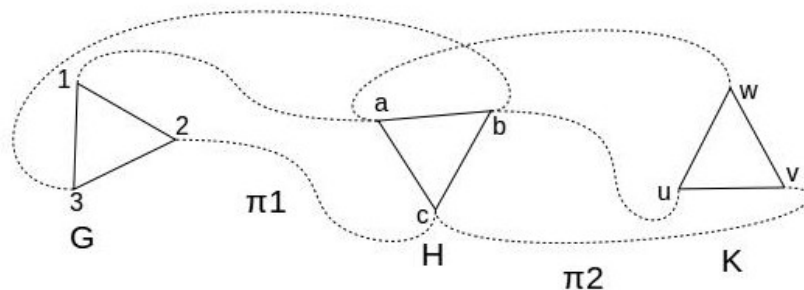
$$\pi(a) = w, \pi(b) = x, \pi(c) = u, \pi(d) = v$$

Bezogen auf die Labels kann es mehrere mögliche Isomorphismen geben.

Schreibweise:  $G \simeq H$  (G ist isomorph zu H) mit  $G \xrightarrow{\pi} H, G \xleftarrow{-\pi} H$  sodass  $\pi$  isomorph ist

Reflexivität: Ein Graph ist zu sich selbst immer isomorph:  $G \simeq G$

Symmetrie:  $G \simeq H \Leftrightarrow H \simeq G$  Transitivität:  $G \simeq H, H \simeq K \Rightarrow G \simeq K$



$\simeq$  ist eine Äquivalenzrelation → Isomorphie teilt Graphen in Klassen ein (Isomorphieklassen)

Nebenbemerkung: Labeled Graphen?

Zusätzliche Bedingung benötigt:  $\lambda(\pi(x)) = \lambda(x) \rightarrow$  Labels müssen erhalten bleiben!

### Testen auf Gleichheit

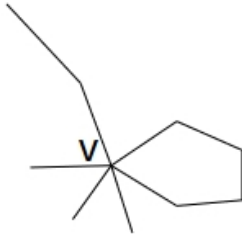
Gegeben:  $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$

Frage: Sind die Graphen isomorph?

Grundbedingungen:

1.  $|V_1| = |V_2| \rightarrow$  gleiche Anzahl von Knoten
2.  $|E_1| = |E_2| \rightarrow$  gleiche Anzahl von Kanten

### 1.3 Eigenschaften von Graphen



Nachbarknoten von  $v$ :  $N(v) := \{y \in V \mid \{v, y\} \in E\}$

$$\deg(v) := |N(v)|$$

$$\delta(G) := \min_{v \in V} \deg(v)$$

$$\Delta(G) := \max_{v \in V} \deg(v)$$

Def: Ein Graph heißt **REGULÄR** wenn  $\Delta(G) = \delta(G)$   
(wenn alle Knoten gleichen Grad haben)

Gradfolge von  $G$ :

$$\mathcal{F} = (n_0, n_1, n_2, \dots, n_{|V|-1}) \text{ mit } n_k := |\{x \in V \mid \deg(x) = k\}|$$

$$\delta(G) \geq 0$$

$$\Delta(G) \leq |V| - 1$$

Beispiel:



$$\mathbf{F} = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 4 & 0 & 0 & 1 \end{pmatrix}$$



$$\mathbf{F} = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

bei Isomorphie:  $\mathcal{F}_1 = \mathcal{F}_2 \rightarrow$  Isomorphismus  $\pi$  erhält Grad der Knoten!

## 1.4 Graph-Invarianten

Eigenschaften, die unter Isomorphie erhalten bleiben

$\mathcal{G}$ ... Menge aller Graphen

F... ist ein Graph invariant wenn

$$F : \mathcal{G} \rightarrow X \quad (1)$$

die Eigenschaft hat, dass

$$G \simeq H \Rightarrow F(G) = F(H) \quad (2)$$

Invarianten bis jetzt:  $|V|, |E|$ , Gradfolge  $\mathcal{F}$

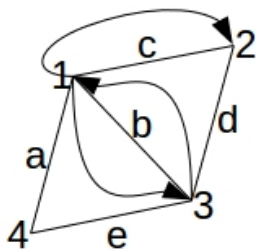
Wenn  $F(G) \neq F(H)$  für irgendeine Grapheninvariante  $\Rightarrow G \not\simeq H$

## 1.5 Pfade und Zusammenhänge

Kantenzug: Folge von Kanten in  $G$

$x_0, e_1, x_1, e_2, x_2, \dots, e_l, x_l$  sodass  $e_i := \{x_{i-1}, x_i\}$

Beispiel:



Weg: Kantenzug sodass  $e_i \neq e_j$  für  $i \neq j$  (keine Kante doppelt verwenden)

Pfad: Kantenzug sodass  $x_i \neq x_j$  für  $(i, j) \neq (0, l)$  mit  $0$ =Startknoten und  $l$ =Endknoten des Pfades (keinen Knoten mehrfach bis auf  $x_0, x_l$ )

- offen:  $x_0 \neq x_e$



- geschlossen:  $x_0 = x_e$  (nur hier 1 Knoten doppelt benutzt!)



Definition:  $G$  ist zusammenhängend wenn es zwischen je zwei Knoten  $x, y \in V$  einen Kantenzug gibt

Frage:

1. Ist Zusammenhang eine Grapheninvariante?
2. Kann man in der Definition Kantenzug durch Weg, Pfad oder Kreis ersetzt?

## **2 Vorlesung 21.10.2016**



### 3 Vorlesung 28.10.2016

#### 3.1 Inzidenzstrukturen

Struktur aus Punktmenge und Menge von Blöcken.

Tripel:  $(p, B, I)$

- $p \cap B = \emptyset$
- $I \subseteq p \times B$
- $p$  = Punkte - z.B. Vertices
- $B$  = Blöcke - z.B. Kanten
- $I$  = Inzidenzmatrix

Die Punkte  $p$  „inzidieren“ demnach mit den Blöcken  $B$ , „liegen auf“ einem Block. Dieser Block kann, wie in unserem Fall bei Graphen, eine Gerade sein.

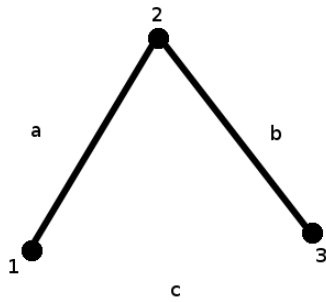
##### 3.1.1 Inzidenzmatrize - ungerichtete Graphen

- $n$  Knoten,  $m$  Kanten
- $n \times m$  Matrix  $B = b_{i,j}$
- $G = (V, E)$   $V = \{v_1, \dots, v_n\}$   
 $E = \{e_1, \dots, e_m\}$   
$$b_{i,j} \begin{cases} 1, & v_i \in e_j \\ 0, & \text{sonst} \end{cases}$$

	a	b	c
1	1	0	0
2	1	1	0
3	0	1	0

In dieser Inzidenzmatrize steht Vertex 1 nur auf Kante a, Vertex 2 auf Kante a und b (kann nicht auf c liegen) und Vertex 3 auf Kante b. In Form einer Bildes

könnte das so aussehen:

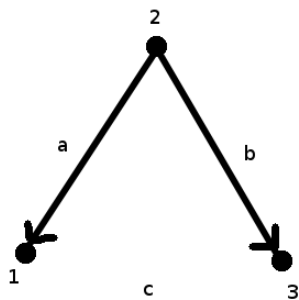


### 3.1.2 Inzidenzmatrize - gerichtete Graphen

$$\bullet \quad b_{i,j} = \begin{cases} 1, & e_j = (v_i, x) \\ 0, & v_i \ni e_j \\ -1, & e_j = (x, v_i) \end{cases}$$

	a	b	c
1	-1	0	0
2	1	1	0
3	0	-1	0

Hier sind die Kanten gerichtet. Im Gegensatz zur ungerichteten Inzidenzmatrize erhalten „ankommende“ Kanten hier ein negatives Vorzeichen. Siehe Kante a zu Vertex 1 und Kante b zu Vertex 3.



### 3.2 Laplace-Matrix

- $G=(V,E)$
- Gradmatrix  $D=d_{i,j}$
- Adjazenzmatrix  $A=a_{i,j}$
- Laplace-Matrix  $L = D-A = l_{i,j}$

$$d_{i,j} \begin{cases} \deg(v_i), & i = j \\ 0, & \text{sonst} \end{cases}$$

$$a_{i,j} \begin{cases} 1, & (i,j) \in E \\ 0, & \text{sonst} \end{cases}$$

- A ist symmetrisch für ungerade Graphen

$$L \begin{cases} \deg(v_i), & i = j \\ -1, & i \neq j, (i,j) \in E \\ 0, & \text{sonst} \end{cases}$$

- Zusammenhang zur Inzidenzmatrix:  $L=B \times B^T$

Beispiel:



$$\text{Gradmatrix: } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\text{Adjazenzmatrix: } \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

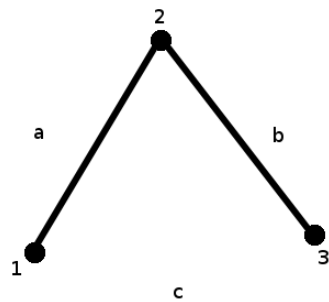
$$\text{Laplace-Matrix: } \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

Eigenschaften:

- symmetrisch
- die Zeilen- und Spaltensumme = 0
- Eigenwert  $\lambda_0=0$ ,  $v_0=(1,\dots,1) \Rightarrow 2 v_0=\lambda_0$
- Anzahl der 0 Eigenwerte  $\Rightarrow$  Anzahl der connected components
- special gap: kleinster Eigenwert  $\neq 0$
- algebraische Konnektivität (Fiedler-Wert)
  - zweit-kleinsten Eigenwert - positiv-semidefinit
  - $\lambda_i \geq 0$

### 3.2.1 Algebraische Konnektivität

- beschreibt wie gut verbunden der Graph, global gesehen, ist
- algebraische Konnektivität  $\leq$  Vertex-Konnektivität
- $|V|=n$ , min. Durchmesser von  $d$  (längster Pfad)
  - alg. Konn  $\geq 4/nd$



Beispiel:

- $|V|=3$
- $d=2$
- vert. conn = 1
- alg. conn =  $\lambda_z = 0,666$

### 3.2.2 Fiedler Vektor

- Eigenvektor zur Alg. Konn.
- eignet sich zur Graphpartitionierung



Beispiel:

- $F = \langle 0.4, 0.3, 0.1, -0.2, 0.2, -0.8 \rangle$
- $\{4,6\}, \{1,2,3,5\}$

### 3.2.3 Interlacing Theorem

- Sei  $A$  eine reelle, symmetrische Matrix
- mit Eigenvektoren  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$
- sei  $A'$  principal Submatrix von  $A$
- Beispiel: principal Submatrix  $\Leftrightarrow$  induzierter Subgraph, ein Vertex weniger ( $G - \{v_i\}$ )
  - Eigenwerte zu  $A'$ :  $\eta_1 \geq \eta_2 \geq \dots \geq \eta_{n-1}$
  - dann gilt:  $\lambda_i \geq \eta_i \geq \lambda_{i+1} \dots i = 1, 2, \dots, n-1$

### 3.2.4 Anzahl nicht-isomorpher Graphen

- $V = \{1, \dots, n\}$
- $E \subseteq \binom{V}{2} = [V]^2$
- wie viele versch. Graphen gibt es?  
 $\Rightarrow 2^{\binom{n}{2}}$ 
  - einige dieser Graphen sind isomorph zueinander
  - wie viele Äquivalenzklassen gibt es für  $(\cong)$  auf  $V = \{1, \dots, n\}$ ?

**Approximation:**

- wie viele isomorphe Graphen gibt es für 6?
- Isomorphe-Bijektion  $\pi: V \rightarrow V$
- Anzahl Permutationen:  $n!$
- Es gibt maximal  $n!$  isomorphe Graphen auf  $G$

$$\frac{2^{\frac{n}{2}}}{n!} = \frac{\text{Anzahl Graphen}}{\text{mgl. Bijektion für einen einzelnen Graphen}}$$

- wir haben mit  $\frac{2^{\frac{n}{2}}}{n!}$  paarweise nicht-isomorphe Graphen  
 $n! \leq 2^{\frac{n}{2}}$

$$\log_2(2^{\frac{n}{2}}) = \log_2\left(2^{\binom{n}{2}}\right) = \binom{n}{2} = \frac{n^2}{2} \left(1 - \frac{1}{n}\right)$$

$$\log_2\left(\frac{2^{\binom{n}{2}}}{n!}\right) = \binom{n}{2} - \log(n!) \geq \frac{1}{2}n^2 - \frac{1}{2}n - n\log_2(n)$$

$$= \frac{n^2}{2} \left(1 - \frac{1}{n} - \frac{2\log_2(n)}{n}\right)$$

Fehler geht gegen  $0(2^{O(n^2)})$

### 3.2.5 Isomorphismus auf Bäumen

- das „das ist einfach“  $\rightarrow$  Programm mit polynomieller Zeit bauen
- Baum codieren via 2n 0en und 1en  
 $\hookrightarrow$  das sei der Code von Baum  $T$
- zu beweisen: isomorph Bäume haben gleiche codes
- drei Klassen von Bäumen:
  - Bäume
  - gewurzelte Bäume
  - gewurzelte Bäume mit Geschwisterordnung

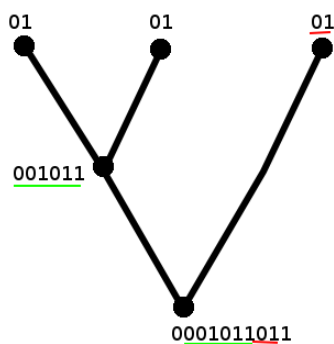
Beispiel:



Algorithmus (gewurzelt, geordnet):

- K1: Blätter werden kodiert als 01
- K2:  $v \in V$  mit Kiner  $c_1, \dots, c_n \in V$
- Sei  $A_i$  der Code von  $c_i$
- Dann codiert 0  $A_1 A_n$  1
  - Isomorphe Bäume werden mit gleichem Code codiert
  - Baum aus Code: Zeigt das nicht-isomorphe Bäume verschiedenen Code haben

Beispiel:



**Induktion:**

- einzelne Wurzel „01“
- Schritt: Code  $K$ , Länge  $2(n+1)$  mit Form  $0A1$ ,  $A=A_{1-n}$   
bestimme  $A_1$ : Kleinsten 0/1 kann mit gleicher Zahl von 0 und 1
- $A_i$  ist via Induktion des Codes für dazugehörigen gewurzelten, geordneten Baum  $T_i$

„Kleinste maximale Entfernung zwischen den Blättern als Wurzel wählen“



## **4 Vorlesung 11.11.2016**

## 5 Vorlesung 18. und 25.11.2016

### 5.1 Ohrenzerlegung

Es sei nur dann ein Graph  $G(V,E)$  mit  $|E| \geq 2$ , der 2-Vertex-verbunden ist, gegeben, wenn es eine offene Ohrenzerlegung gibt. Jede Ohrenzerlegung definiere eine Kreisbasis.

- **Kreisbasis:**

Ein Ohr Sei ein maximaler Pfad  $P$ ,  $|P| \geq 1$ , so dass  $P$  nur an Endpunkten Kanten aus  $E \setminus P$  berührt. Die Knoten in  $P$ , die keine Endknoten sind haben immer Grad  $\deg = 2$ .

- **offene Ohrenzerlegung**

Eine Folge von Ohren  $P_1, P_2, \dots, P_k$  ist eine offene Ohrenzerlegung, wenn  $P_1$  ein Kreis,  $P_k$  und alle anderen  $P_i$  Ohren in  $G_i = G_{i+1} \setminus P_{i+1}$

#### 5.1.1 Algorithmus der Ohrenzerlegung

1. Finde Spannbaum  $T$  für  $G$  und wähle eine Wurzel
2. Für jede Kante  $(u,v)$ , die nicht Teil des Spannbaums ist, finde den *common lowest ancestor* der Knoten  $u$  und  $v$ .
3. Für jede Kante  $(u,v)$  soll die Hauptkante  $(w,x)$  gefunden werden, wobei  $(u,v)$  und  $(w,x)$  Teil eines Kreises sind und  $(w,x)$  einen *lowest common ancestor* so nah wie möglich an der Wurzel haben und  $(w,x) \notin T$  ist.
4. Für alle  $(w,x)$  die nicht aus dem Spannbaum sind, trenne alle Kanten mit gleichem Wert ab. Diese Kanten bilden ein Ohr.
5. Ordne die Ohren nach ihrem Gewicht.

### 5.2 Unabhaengigkeitssysteme und Matroide

Viele *Greedy Probleme* lassen sich mittels Matroiden beschreiben (insbesondere Graphenprobleme).

- **Unabhaengigkeitssystem**

Ein Unabhaengigkeitssystem ist ein Paar  $M = (S, l)$  mit endlicher Menge  $S$  und  $l \subseteq P_S(S)$  (Powerset von  $S$ ). Es besitzt folgende Eigenschaften:

- 1  $\emptyset \in l \rightarrow$  Die leere Menge ist unabhängig
- 2  $x \subseteq Y \in l \rightarrow$  Erbllichkeitseigenschaft

Die Elemente  $x \in l$  und  $y \in P_S(S) \setminus l$  sind unabhängig.  
 $\rightarrow$  Kostenfunktion  $C: S \rightarrow \mathbb{R}$

*Austauscheigenschaft:*

Falls  $A \in l, B \in l, |A| < |B|$ , dann  $\exists x, x \in B \setminus A : A \cup \{x\} \in l$

*Matroid*

Ein Unabhängigkeitssystem sei ein Matroid falls die Eigenschaften eines Unabhängigkeitssystems und die Austauscheigenschaft erfüllt sind.

*grafischer Matroid*

Ein graphischer Matroid  $M_G = (S_G, l_G)$  mit  $S_G = E(G)$

→ für  $A \subseteq S_G : l_G \leftrightarrow A$  ist kein Kreis

→ Menge an Kanten ist nur dann unabhängig, wenn  $G' = (V(G), A)$  einen Wald bilden →  $G'$  bildet keinen Kreis

*Extension:*

Sei  $M = (S, l)$  und  $x \notin A$  gegeben, so sei  $l$  eine Erweiterung von  $A$ , falls  $A \cup \{x\}$  unabhängig ist. ( $A \cup \{x\} \in l$ )

→ Kante  $l$  ist eine Erweiterung, falls  $A \cup \{x\}$  keinen Kreis bilden.

*Maximalität:*

$A \in l$  ist maximal, falls es keine Erweiterung für  $A$  gibt.

### **Theorem**

Alle maximal unabhängigen Teilmengen in einem Matroid haben die selbe Größe.

### **Beweis**

Wenn das Theorem nicht gilt, so wären die maximal unabhängigen Elemente  $A$  und  $B$  mit  $|A| < |B|$ . Damit würde die Austauscheigenschaft zeigen, dass  $x$  abhängig von  $A$  ist:  $\exists x : B \setminus A : A \cup \{x\}$ . → Beweis durch Widerspruch

### **Beispiel**

Gegeben Graph  $G$  gilt, dass alle Spannbäume, die gleiche Anzahl an Kanten haben:  $|E| = |V| - 1$ .

- **gewichtetes Matroid**

Ein Matroid  $M=(S, l)$  ist gewichtet, falls es eine Gewichtsfunktion  $w(x) > 0 \forall x \in S$  gibt. →  $w(A), A \subseteq S$  mit  $w(A) = \sum_{x \in A} w(x)$ .

### 5.2.1 Greedy Algorithmen auf Matroiden

Gegeben sei  $M(S, l)$ . Finde  $A \in l$ , sodass  $w(A)$  maximal ist.

**Beispiel:** Minimallänge Spannbäume mit  $w'(x) = w_0 - w(x)$  mit  $w_0 = \max_x(w(x)) + \epsilon$

#### *Greedy-Algorithmus*

- 1  $A \leftarrow \{\}$
- 2 sortiere  $S[M]$  nach absteigendem Gewicht
- 3  $\forall x \in S[M] \{ \text{if } A \cup \{x\} \in l: \text{ then } A \leftarrow A \cup \{x\} \}$
- 4 return  $A$

#### **Theorem**

$A$  ist die optimale Lösung des Greedy-Algorithmus

#### **Beispiel**

Betrachte und bilde Kreisbasen  $L1$  oder  $L2$ .  $L1$  hat insgesamt weniger Knoten als  $L2$  und ist somit eine bessere Lösung.  $C(G)$  ist ein Matroid mit den Elementgewichten  $|C| \rightarrow$  diese definieren  $S \rightarrow$  die minimalen Kreisbasen werden in polynomieller Zeit in  $|C|$  berechnet.

### 5.2.2 Horton-Algorithmus (1972)

- 1 Konstruiere die kürzesten Pfadbäume mittels Matroide.
- 2 Extrahiere die fundamentalen/minimalen Kreise der Pfadbäume.

$\rightarrow$  Berechnung ist polynomiell abhängig von  $|V|$

- essentielle Kreise sind Kreise, die in allen minimalen Kreisbasen vorkommen.
- relevante Kreise sind Kreise, die in mindestens einer minimalen Kreisbasenlösung vorkommen.
- Falls  $C$  nicht in ein einfacher Kreis ist, dann ist  $C$  keine minimale Kreisbase.

## 5.3 Graphen in der Ebene und planare Graphen

### Topologie

- euklidische Ebene:  $\mathbb{R}^2$
- Liniensegmente  $\{ p+x(q-p) \}$  mit  $p,q \in \mathbb{R}^2$  und  $p \neq q$
- homöomorphisch (bijektive stetige Abbildung) zum Einheitskreis
- Polygonzug (zusammenhängende Folge von Kanten)

#### **Theorem: Jordan'scher Kurvensatz**

Für jedes Polygon  $P \subseteq \mathbb{R}^2$  hat  $\mathbb{R}^2 \setminus P$  zwei Regionen, wobei  $P$  die Grenze bildet  
→ definiertes außen und innen

#### **Lemma**

Seien  $P_1, P_2, P_3$  Polygonzüge mit den Endpunkten  $x$  und  $y$ , so hat  $\mathbb{R}^2 \setminus \{P_1 \cup P_2 \cup P_3\}$  drei Facetten mit den Grenzen  $P_1, P_2, P_3$  (ohne die Endpunkte  $x$  und  $y$ ):

- $b_1 := P_1 \cup P_2$
- $b_2 := P_1 \cup P_3$
- $b_3 := P_2 \cup P_3$

→ falls  $P_4$  ein Polygonzug zwischen  $\dot{P}_1$  und  $\dot{P}_3$  mit  $\dot{P}_4, \dot{P}_3 \subset \mathbb{R}^2 \setminus \{P_1 \cup P_3\}$  ist, so schneiden sich  $P_4$  und  $P_2$  ( $\dot{P}_4 \cap \dot{P}_2$ )

#### **Graph in Ebene:**

Ein Graph in der Ebene  $(V,E)$  hat folgende Eigenschaften

- 1  $V \subset \mathbb{R}^2$
- 2 Jede Kante sei ein Polygonzug zwischen zwei Knoten
- 3 unterschiedliche Kanten haben unterschiedliche Mengen von Knoten
- 4 das Innere einer Kante enthält keinen Knoten und keinen Punkt einer anderen Kante

somit ist ein Graph in der Ebene *kreuzungsfrei*






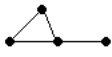


#### **Facette**

Falls  $G$  ein Graph in der Ebene ist, so sind die Regionen  $\mathbb{R}^2 \setminus G$  die Facetten. Die äußere Facette liegt außerhalb des Graphen und die anderen Facetten liegen per Definition innerhalb.

### Lemma

Sei  $G$  ein Graph in der Ebene  $f \in F(G)$  eine Facette und  $H \subseteq G$  ein Subgraph, so gilt:

- $H$  hat eine Facette  $f'$  mit  $f' \subseteq f$
- Falls der Rand von  $f \subset H$  ist, dann ist  $f' = f$

	$p=1$	$l=0$	$f=1$	$p-l+f=2$
	$p=2$	$l=1$	$f=1$	$p-l+f=2$
	$p=3$	$l=2$	$f=1$	$p-l+f=2$
	$p=3$	$l=3$	$f=2$	$p-l+f=2$
	$p=4$	$l=3$	$f=1$	$p-l+f=2$
	$p=4$	$l=4$	$f=2$	$p-l+f=2$
	$p=4$	$l=5$	$f=3$	$p-l+f=2$
	$p=4$	$l=6$	$f=4$	$p-l+f=2$

### Theorem: Eulersche Formel

Sei  $G$  ein verbundener Graph in der Ebene mit  $n$  Knoten,  $m$  Kanten und  $l$  Facetten, so gilt:  $n - m + l = 2$

### Beweis

Sei  $n$  fixiert  $\rightarrow$  induziere über  $m$ .

- $if m < n - 1$  : Graph ist nicht verbunden
- $if m = n - 1$  : Graph ist ein Baum
- $if m \geq n$  : Sei  $e \in E(G)$  Kante auf einem Kreis, dann ist  $G' = G - e \rightarrow e$  liegt auf der Grenze zweier Facetten  $f_1$  und  $f_2$  von  $G$  und es gibt eine Facette  $f_e$  von  $G'$ , die  $e$  enthält.

Zeige, dass  $F(G) \notin \{f_1, f_2\} = F(G') \notin \{f_e\}$ , womit  $G'$  eine Facette und eine Kante weniger hat als  $G$ .

$\rightarrow$  Das entfernen einer Kante kombiniert 2 Facetten

$\rightarrow$  Fügt man eine Kante hinzu so wird auch eine Kante hinzugefügt

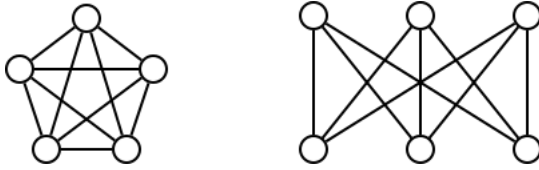
$\rightarrow |V'| = |V| + 1$  und  $|E'| = |E| + 1 \rightarrow$  q.w.e.d.

### Korollar 1

- 1) Ein Graph in der Ebene mit  $n \geq 3$  Knoten hat maximal  $3n - 6$  Kanten
- 2) Jede Triangulation mit  $n$  Knoten hat genau  $3n - 6$  Kanten

**Korollar 2(nach Kuratowski)**

Ein Graph in der Ebene hat weder  $K_5$  noch  $K_{3,3}$  als topologischen Minor ( $K_5$  und  $K_{3,3}$  sind nicht mehr kreuzungsfrei zeichenbar).

**Minor**

H ist ein Minor von G, falls G in H durch folgende Operationen transformiert werden kann:

- $G - e, e \in E(G)$
- $G - v, v \in V(G)$  (hierbei werden auch nichtverbundene Kanten gelöscht)
- Kontraktion von  $e \in E(G)$  mit  $e = u,v, u,v \in V(G)$ , wobei  $u,v$  zu einem Knoten vereinigt werden und dieser zu allen Nachbarn inzident wird.

**planarer Graph**

Ein Graph G sei planar, wenn er folgende Eigenschaften erfüllt:

- Eine endliche Anzahl von Facetten bildet eine Kreisbasis
- jede Kante, die zwei Kreisen zugehörig ist, heißt innere Kante
- jede Kante, die einem Kreis zugehörig ist, heißt äußere Kante

**2-Basis**

Jede Kante ist genau 2 mal in den Kreisbasen vertreten.

**Konstruktion 2-Basis**

- $F = \bigoplus_{c \in B^c} \text{Rand der äußeren Facette (äußerer Rand)}$
- $2\text{-Basis} = B \cup F$

### **Satz von Mc Lane**

$G$  ist planar, wenn  $G$  auch eine 2-Basis hat.

#### **Beweis Satz von Mc Lane**

1  $B$  sei 2-Basis von  $G$  und  $G$  sei nicht planar

nach Kuratowski ist Subdivision von  $K_5$  oder  $K_{3,3}$   $H \subset G$  möglich

Behauptung:  $\rightarrow H$  hat ebenfalls eine 2-Basis

2  $G \notin e, \forall e$  hat eine 2-Basis

es ist nur in einem Kreis  $C \in B$  vorhanden, womit  $B \notin C$  entsteht.

$e$  ist in zwei Kreisen vorkommend, womit  $B \notin \{C_1, C_2\} \cup \{C_1 \oplus C_2\}$  eine 2-Basis ist

$\rightarrow$  somit haben alle Teilgraphen eine 2-Basis. Da die Behauptung belegt ist, wird ein Widerspruch

*zyklometrische Zahl:* Die Anzahl der Basiselemente einer Kreisbasis nennt man zyklometrische Zahl.

#### **Beispiel: vollständiger Graph $K_5$**

- $|V| = 5; |E| = 10$
- 2-Basis mit 5 inneren und 5 äußeren Kanten
- $\mu(K_5) = -5 + 10 + 1 = 7$
- 2-Basis hat demnach  $(2n - 6) * 7$  Kanten = 21 Kanten  $\rightarrow$  **falsche Aussage**

#### **Beispiel: bipartiter Graph $K_{3,3}$**

- $|V| = 6; |E| = 9$
- $\mu(K_{3,3}) = -6 + 9 + 1 = 4$
- 2-Basis hat demnach  $(2n - 6) * 4$  Kanten = 20 Kanten  $\rightarrow$  **falsche Aussage**

#### **Planaritätstest**

- 1 zähle die Kanten
- 2 Tiefensuche  $\rightarrow$  Konstruiere einen Spannbaum
- 3 Teste für Kanten  $e \leftarrow G \notin T$ , ob  $K_5$  oder  $K_{3,3}$  entsteht.

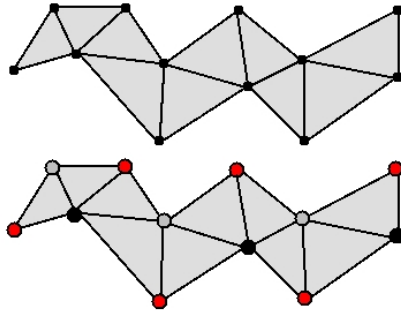


## 5.4 Färbung von Graphen

### Vertexfärbung

Zwei durch eine Kante verbundene Knoten haben unterschiedliche Farben.

Beispiel wäre eine Landkarte auf der mit so wenig wie möglich Farben die Länder ausgemalt werden, ohne zwei benachbarte Länder gleichfarbig zu haben. Hierbei entspricht jede Facette einen Knoten.



## 6 Vorlesung 02.12.2016

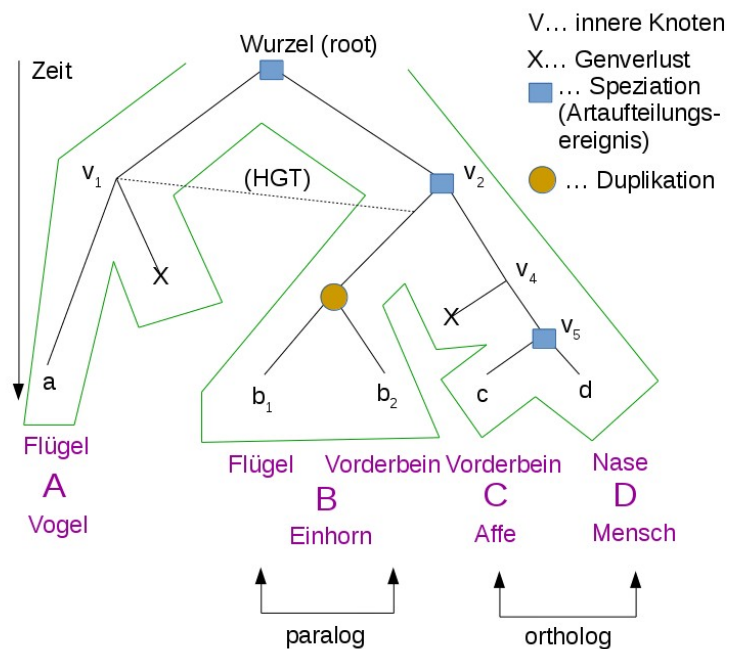
### 6.1 Cographen & Cotrees

#### Phylogenetik

- Erforschung von Abstammung
- Rekonstruktion von phylogenetischen Bäumen („Stammbäume“)
- Speziesbäume/Genbäume

Ergebnisse:

- Genverlust (loss)
- Aufspaltung zu einer neuen Spezies (speciation)
- Duplikation von Genen (duplication)
- horizontaler Gentransfer (HGT)



Def.: Baum (tree)

Ein Baum  $T=(V,E)$  ist ein zusammenhängender Graph, der keine Kreise enthält (azyklisch).

Def.: Zusammenhang

Ein Graph  $G=(V,E)$  ist zusammenhängend, wenn es zwischen jedem möglichen Paar von Knoten einen Pfad gibt.

Theorem:

$T=(V, E)$  ist ein Baum  $\Leftrightarrow \exists!$  Pfad zwischen zwei zufällig gewählten Knoten existiert. ( $\Leftrightarrow \dots$  aus dem folgt;  $\exists!$   $\dots$  genau einem)

Beweis:

$\Rightarrow$  : Da  $T$  zusammenhängend ist, gibt es einen Pfad zwischen  $v, u \in V(T)$ ,  $\forall v, u \in V(T)$ . Angenommen es gäbe noch einen 2. Pfad, dann gibt es einen Kreis; Widerspruch zur Definition.

$\Leftarrow$  : Wenn genau ein Pfad existiert, ist  $T$  zusammenhängend, Also gibt es auch keine Kreise;  $T$  ist ein azyklischer Graph = Baum.

Def.: Distanz

Die Distanz  $d(u,v)$  zwischen zwei Knoten  $u, v \in V$  ist gleich der Anzahl der Kanten im kürzesten Pfad zwischen  $u$  und  $v$ .

Def.: Lowest Common Ancestor (lca)

Seien  $x,y \in V(T)$  Blätter im Baum  $T$  mit Wurzel  $r$ . Sei  $P_x = \{x, x_1, x_2, \dots, r\}$  der Pfad von  $x$  nach  $r$  und  $P_y = \{y, y_1, y_2, \dots, r\}$  der Pfad von  $y$  nach  $r$ . Dann  $lca(x, y) = \min(d(d, v_i), d(y, v_i))$  mit  $v_i \in (P_x \cap P_y)$   
 $v_i \dots$  mehrere  $v$ 's (kann auch  $r$  sein)  
 $r \dots$  root (Wurzel)

- $P_{b_2r} = \{b_2, v_3, v_2, r\}$
- $P_{dr} = \{d, v_5, v_4, v_2, r\}$
- $P_{b_2r} \cap P_{dr} = \{v_2, r\}$ 
  - $d(b_2, v_2) = 2$
  - $d(b_2, r) = 3$

Def.:

- Homologie: 2 Gene sind homolog, wenn sie die selben Vorfahren haben
- Orthologie: 2 Gene sind ortholog, wenn ihr lca eine Speziation (Artaufteilungsereignis) ist
- Paralogie: 2 Gene sind paralog, wenn ihr lca eine Duplikation ist

Def.:  $\Theta$ -Relation (Orthologie-Relation)

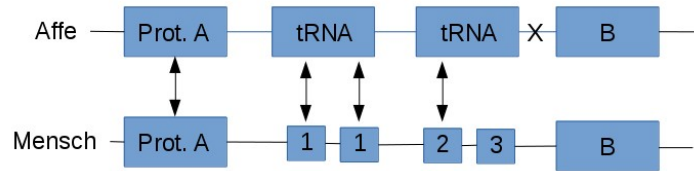
Seien  $x,y \in H$ ,  $H$  = Menge von Genen  
 $(x,y) \in \Theta \Leftrightarrow lca(x,y)$  ist eine Speziation.

Diese Relation ist reflexiv (rückbezüglich), symmetrisch, aber nicht transitiv (mit sich ziehend).

Bestimmung von Orthologie:

- Sequenzähnlichkeit

- Syntenie („Gemeinsamkeiten in der Reihenfolge von Genen oder Gensegmenten auf verschiedenen chromosomalen Abschnitten. [...] ist ein Maß für die genetische Verwandtschaft der beiden Arten.“[Wikipedia])



z.B. Tool: ProteinOrtho

Def.:  $\sim$ -Relation (fast-Orthologie)

$(x,y) \in \sim$ , wenn x,y als ortholog eingestuft werden.

Ziel: Korrigieren  $\sim$  sodass wir  $\Theta$  erhalten. Dazu stellen wir  $\sim$  und  $\Theta$  als Graphen dar.

$$G_{\Theta} = (V_{\Theta}, E_{\Theta})$$

$$G_{\sim} = (V_{\sim}, E_{\sim})$$

$$V_{\Theta} = V_{\sim} = Gene$$

$$E_{\Theta} = \{(x,y) \in \binom{V}{2} \mid x\Theta y\}$$

$$E_{\sim} = \{(x,y) \in \binom{V}{2} \mid x \sim y, y \sim x\}$$

$\binom{V}{2}$  ... alle möglichen Kombinationen von zwei Knoten

Def.: Komplementgraph (complement)

Sei  $G=(V,E)$  ein Graph. Das Komplement  $\overline{G}$  von  $G$  ist der Graph  $\overline{G} = (V, \overline{E})$  mit  $\overline{E} = \{(u,v) \in \binom{V}{2} \mid (u,v) \notin E\}$

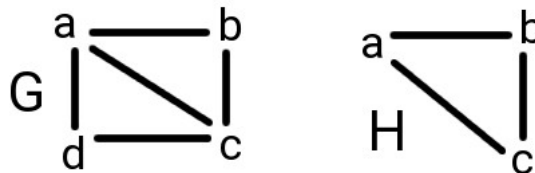


Def.: Teilgraph

Sei  $G=(V,E)$  ein Graph und  $H \subseteq G$ .  $H$  ist Teilgraph von  $G$ , wenn  $V(H) \subseteq V(G)$ ,  $E(H) \subseteq E(G)$ . Ein induzierter Teilgraph ist ein Teilgraph  $H$  von  $G$  bei dem alle Knoten die in  $G$  benachbart sind, auch in  $H$  benachbart sein müssen.

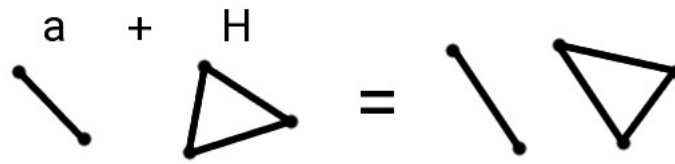
$$(v,u) \in E(G) \wedge u,v \in V(H) \Leftrightarrow (v,u) \in E(H)$$

$\wedge \dots$  Konjunktion



Def.: disjunkte Vereinigungen

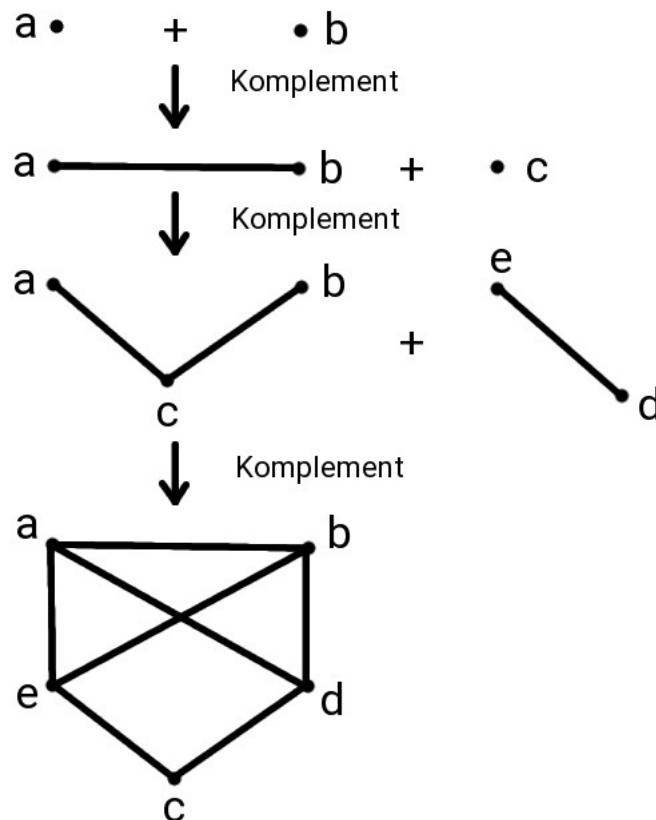
Graphen  $G,H$ :  $G+H$  ist ein Graph mit  $V(G) \cup V(H)$  und  $E(G) \cup E(H)$ .



Def.: Cograph

1.  $K_1$  ist ein Cograph  $\bullet_{K_1}$
2.  $G$  ist ein Cograph  $\Leftrightarrow \overline{G}$  ist ein Cograph
3.  $G, H$  sind Cographen  $\Leftrightarrow G+H$  ist ein Cograph

Erstellung von Cographen:

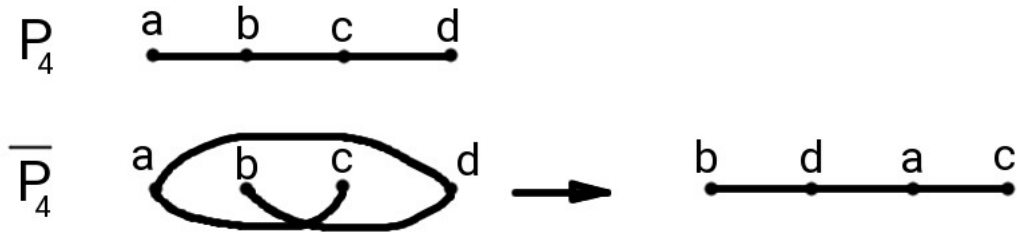


Eigenschaften von Cographen:

Sei  $G=(V,E)$  ein Cograph und  $H \subseteq G, H$  Cograph

- i)  $G$  enthält keine induzierten  $P_4$ 's
- ii)  $H$  ist zusammenhängend  $\Leftrightarrow \overline{H}$  ist nicht zusammenhängend

iii)  $G$  kann aus einzelnen Knoten ( $K_1$ ) zusammengesetzt werden



$$\Rightarrow P_4 = \overline{P_4}$$

Ein Cograph muss jedoch ein  $P_4$ -freier Graph sein.

Cograph =  $P_4$ -free graphs = complement reducible graphs

Test ob  $G=(V,E)$  ein Cograph ist:

Input: Graph  $G$

```

i ≤ Cograph (G) {
    if (| V(G) | < 4) {return true;}
    c = {Zusammenhangskomponenten von G}
    if(| c | = 1) {c'={Komponenten von Ḡ}}
    if (| c' | = 1) {return false;}
    else {
        foreach (c ∈ C)
            {isCograph (c) }
    }
}

```

Bei isCograph: je nachdem ob  $c$  oder  $c'$  rausgekommen ist, muss  $c$  oder  $c'$  geprüft werden.

Theorem:

$\sim = \Theta \Leftrightarrow G_\Theta = G_\sim$  und  $G_\Theta$  ist ein Cograph

Damit können wir testen, ob  $G_\sim$  ein Orthologiegraph ist.

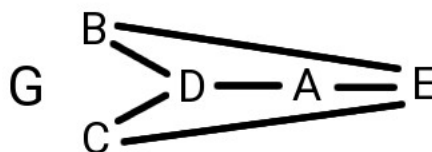
Was passiert wenn  $\sim \neq \Theta$  bzw.  $G_\sim$  kein Graph?

$\Rightarrow$ aktuelle Forschung  $\Rightarrow$ Es gibt Lösungen  $G_\sim$  zu editieren mit optimalen Kriterien, sodass der editierte  $G_\sim$  ein Cograph ist. Z.B. ILP (integer linear program), Cograph-editing. Alle Algorithmen, die exakte Möglichkeiten liefern, brauchen sehr lange und sind in der Praxis nicht nutzbar.

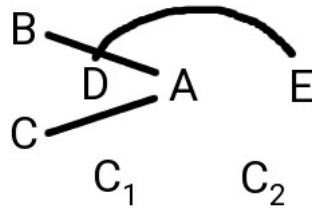
Weitere Literatur: Marc Hellmuth

Theorem:

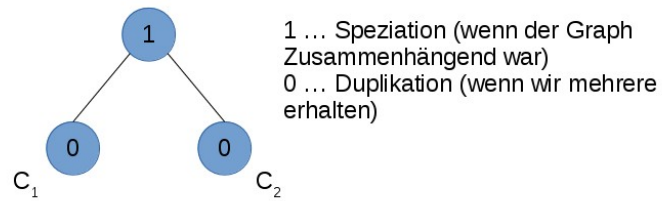
Für jeden Cographen gibt es einen eindeutigen Cotree (Cobaum)



1. Schritt: Komplement

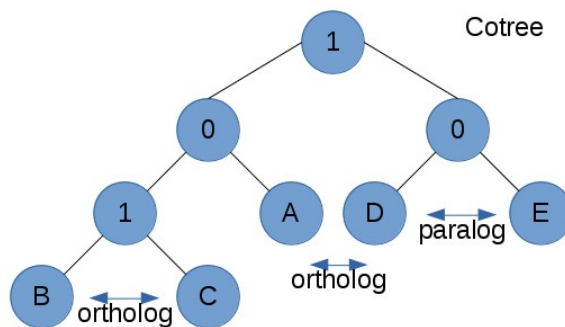
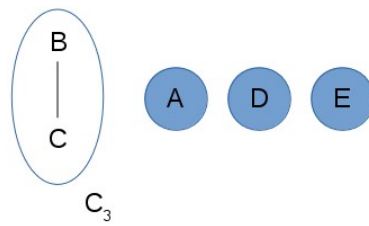


2. Schritt: umgekehrte disjunkte Vereinigung



1 ... Speziation (wenn der Graph  
Zusammenhängend war)  
0 ... Duplikation (wenn wir mehrere  
erhalten)

3. Schritt: Komplement



## **7 Vorlesung 09.12.2016**



## 8 Vorlesung 16.12.2016

Metrik:

1.  $d_{uu} = 0$
2.  $d_{uv} = 0 \Rightarrow u = v$
3.  $d_{uv} = d_{vu}$
4.  $d_{uv} + d_{vw} \geq d_{uw}$  (Dreiecksungleichung)

Pseudometrik: -,1,2,3

Metrik: 0,1,2,3

Distanzfunktion: 1,2

### 4-Punkte-Bedingung:

Eine Distanzfunktion  $d$  ist eine additive (Baum) Metrik wenn je vier Punkte so geordnet werden können, daß:

$$d_{xy} + d_{uv} \leq d_{xu} + d_{yv} = d_{xv} + d_{yu} \Leftrightarrow \forall x,y,u,v \text{ gilt:}$$

$$d_{xy} + d_{uv} \leq \max\{d_{xu} + d_{yv}, d_{xv} + d_{yu}\}$$

### Isolationsindex:

$$l(e) = \alpha(A|B) = \max(0, \min_{\substack{x,y \in A \\ u,v \in B}} \frac{1}{2} [\max\{d_{xu} + d_{yv}, d_{xv} + d_{yu}\} - (d_{xy} + d_{uv})])$$

=Länge der Baumkante, die A,B trennt oder  $\leq 0$  wenn  $A|B$  keine Teilbäume bestimmt.

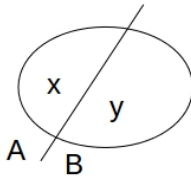
Wenn  $d$  eine additive Distanzfunktion:

- $\alpha(A|B) \geq 0$
- $A|B$  entspricht einer Kante im Baum  $\Leftrightarrow \alpha(A|B) > 0$

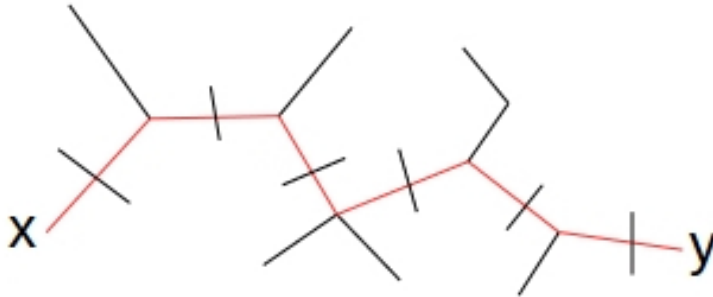
Splitpseudometrik:

$$\delta_{A|B}(x, y) = \begin{cases} 1 : x \in A, y \in B \\ 1 : x \in B, y \in A \\ 0 : x, y \in A \\ 0 : x, y \in B \end{cases} \quad (3)$$

$x, y$  durch  $A|B$  getrennt  $\Leftrightarrow \delta_{A|B}(x, y) = 1$



$$d_T(x, y) = \sum_{(A|B) \in \Sigma(T)} \alpha(A|B) \cdot \delta_{A|B}(x, y)$$



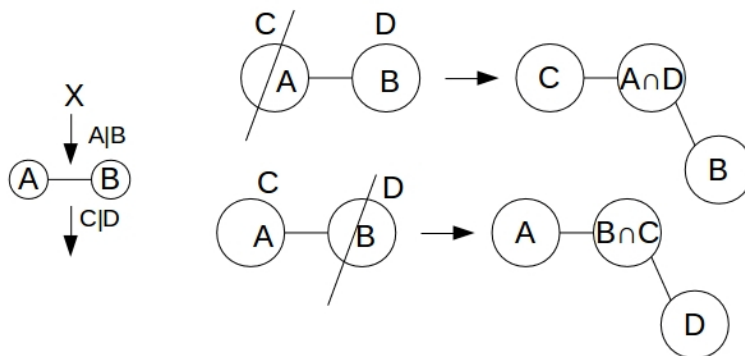
Genau die splits entlang des Pfades von x und y trennen x,y

### Splits $\Sigma(T) \rightarrow \text{Baum}$

wir wissen  $\Sigma(T)$  ist kompatibel

$A|B, C|D \in \Sigma(T)$  dann mindestens einer der vier Durchschnitte:

$A \cap C, A \cap D, B \cap C, B \cap D$  leer



jeder split-Teil GENAU eine der Mengen

**Frage:** Wie können Isolationsindizes, schnell und und alle Möglichkeiten durch-  
zuprobieren, erzeugt werden?

Lösung: effiziente Berechnung von  $\alpha(A|B) > 0$

Idee: erweitere X schrittweise

$|A|, |B| = 1$

$X' \leftarrow X \cup \{w\}$

$A \cup B = X$

in  $X'$ :

- $X \setminus \{w\}$
- $A \cup \{w\} | B$
- $B \cup \{w\} | A$

$$\beta_{xy|uv} := \frac{1}{2} \max\{d_{xu} + d_{yv}, d_{xv} + d_{yu}\} - (d_{xy} + d_{uv})$$

erster Fall:

$$\alpha(\{x\}|X) = \min_{u,v \in X} \beta_{wx|uv} = \min_{u,v \in X} \frac{1}{2}(d_{wu} + d_{wv} - d_{uv})$$

zweiter Fall:

$$\alpha(A|B) = \min_{\substack{x,y \in A \\ u,v \in B}} \beta_{xy|uv}$$

$$\alpha(A \cup \{w\}|B) = \min\left\{\min_{\substack{x,y \in A \\ u,v \in B}} \beta_{xy|uv}, \min_{\substack{y \in A \\ u,v \in B}} \beta_{yw|uv}, \min_{\substack{x \in A \\ u,v \in B}} \beta_{xw|uv}\right\}$$

$$\Rightarrow \alpha(A \cup \{w\}|B) \leq \alpha(A|B)$$

Also: wenn  $\alpha(A|B) \leq 0 \Rightarrow \alpha(A \cup \{w\}|B)$  auch  $\leq 0$

$\Rightarrow$  nur Splits auf X mit  $\alpha(A|B) > 0$  müssen erwartet werden

Wenn d additiv  $\Rightarrow$  Baum  $\Rightarrow$  splits  $\Sigma(T)$  kompatibel  $\Rightarrow$  es gibt nicht mehr als  $2|X|$  splits

$\Rightarrow$  Die Isolationsindizes aller Splits mit  $\alpha(A|B) > 0$  können in  $\mathcal{O}(|x|^5)$  berechnet werden:

$|x|$  Erweiterungsschritte für  $\mathcal{O}(|x|)$  splits mit Aufwand  $\mathcal{O}(|x|^3)$

Theorem:[Bandelt,Dress]

Sei d eine Pseudometrik auf X. Dann gibt es eine Pseudometrik  $d^0$  auf X sodaß

$$d(x, y) = \sum_{A|B} \underbrace{\alpha(A|B)}_* \cdot \delta_{A|B}(x, y) + d^0(x, y)$$

$$* \alpha(A|B) = 0 \text{ wenn } \min_{\substack{x,y \in A \\ u,v \in B}} \beta_{xy|uv} < 0$$

außerdem gilt:  $\Sigma(d) = \{(A|B)\}$

$\alpha(A|B) > 0$  hat höchstens  $\mathcal{O}(|x|^2)$  Elemente

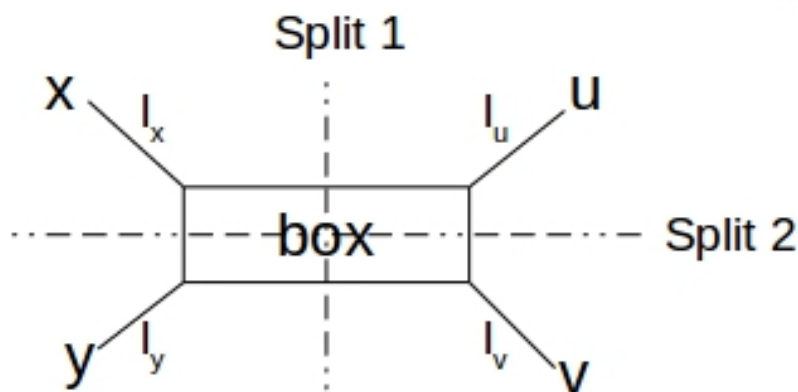
alle  $\alpha(A|B) > 0$  können in  $\mathcal{O}(|x|^6)$  Elemente berechnet werden.

- d additiv  $\Rightarrow d^0 = 0$
- $d^0$  heißt split-primer
- d heißt total zerlegbar wenn  $d^0 = 0$

### allgemeine Pseudometrik auf 4 Punkten

Anzahl unabhängigen Distanzen: 6





$$d_{xu} + d_{xy} - d_{uy}$$

$$(l_x + a + l_u) + (l_x + b + l_y) - l_u - a - b - l_y = 2l_x$$

$$l_x = \frac{1}{2} [ \underbrace{d_{xu} + d_{xy} - d_{uy}}_{\geq 0 (\text{Dreiecksungleichung})} ]$$

Split 1:

$$d_{xv} + d_{yu} - (d_{xy} + d_{uv}) =$$

$$l_x + a + b + l_v$$

$$+ l_y + a + b + l_u$$

$$- l_x - b - l_y$$

$$- l_u - b - l_v = 2a$$

Split 2:

$$d_{xu} + d_{yv} - (d_{xy} + d_{uv}) =$$

$$l_x + a + l_u$$

$$+ l_y + a + l_v$$

$$- l_x - b - l_y$$

$$- l_y - b - l_u = 2(a - b) \leq 2a$$

$$\alpha(\{xy\}|\{uv\}) = a$$

$$\alpha(\{xu\}|\{yv\}) = b$$

Baum  $\Rightarrow b=0$

**Messung der Baumartigkeit:**

$$B := \frac{1}{\binom{n}{4}} \sum_{\substack{i < j < k < l \\ i, j, k, l \in X}} \frac{b_{ijkl}}{a_{ijkl} + b_{ijkl}}$$

Mittelwerte von in der Box

$B \approx$  Baumartig

$B \approx \frac{1}{2}$  völlig verrauscht, netzwerk-artig

## Travelling sales person problem (TSP)

geschlossene Tour Voraussetzung

$|X| > 1$  (Anzahl der Städte größer 1)

Metrik  $d$  auf  $X$  gegeben

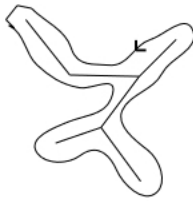
Tour: Permutation von  $X: \pi$

$$L(\pi) = \sum_{i=1}^{|X|} d_{\pi(i-1)\pi(i)} \quad (\text{lesen als indices modulo } |X|)$$

Definition Mastertour:

Einschränkung von  $\pi$  auf  $X' \subseteq X$  löst das TSP auf  $X$

Wenn  $d$  eine additive Metrik (Baum) ist dann existiert eine Mastertour (optimale Lösung) die genau ein Mal um den Baum herum führt.



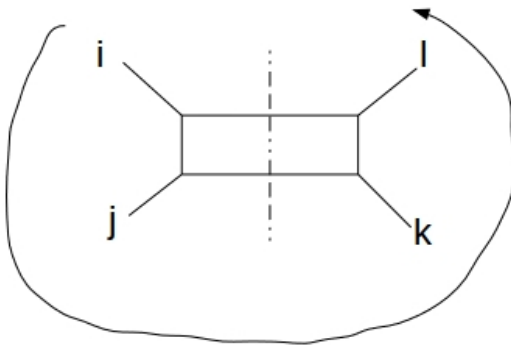
Eine Metrik hat die KALMANSON-Eigenschaft, wenn man  $X$  so ordnen kann, daß

$$d_{ij} + d_{kl} \leq d_{ik} + d_{jl} \quad \forall i < j < k < l$$

und

$$d_{il} + d_{jk} \leq d_{ik} + d_{jl} \quad \forall i < j < k < l$$

→ für jedes Quadrupel tauchen höchstens die Splits  $ij|kl$ ,  $il|jk$  auf  
 $d$  ist Kalmanson  $\Leftrightarrow$  das TSP mit Distanz  $d$  einen Mastertour hat



Wenn  $d$  Kalmanson ist (zirkulär zerlegbar)  $\Rightarrow d$  splitzerlegbar (planar darstellbar)

$\nLeftarrow$  (Umkehr falsch)

$$d = \underbrace{\sum_{A|B} \alpha(A|B) \cdot \delta_{A|B}}_{\text{fast immer Kalmanson}} + \underbrace{\delta^0}_{\substack{\text{Rauschen} \\ (\text{split } \text{Primaeranteil})}}$$

Anteil der Distanz ohne phylogenetische Information:

$$\frac{\sum_{x \neq y} \delta^0(x, y)}{\sum_{x \neq y} \delta(x, y)}$$

(Maß für die Größe des Rauschens  $\rightarrow$  keine phylogenetische Information)

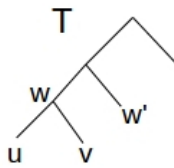
## 9 Vorlesung 21.12.2016

### 9.1 neighbor joining

geg: Distanzmatrix (d) auf Menge X von Taxa  $\rightarrow$  Baum (ungewurzelt)

Iteration:

1. suche  $\operatorname{argmin}_{x,y} \tilde{d}_{xy} = \{u, v\}$
2. ersetze  $\{u, v\} \rightarrow w$  (neuer Knoten)
3. berechne  $d_{wz}$  für  $z \neq u, v \rightarrow$  Schritt 1



$\mathbf{d} \rightarrow \mathbf{T}$

$\tilde{d}$  Transformation von d

$F: d \mapsto \tilde{d}$

$d_{wz} = \phi(d_{uz}, d_{vz}, d_{uv})$

Ein Baumrekonstruktionsalgorithmus  $\mathcal{A}: d \mapsto T$  ist konsistent wenn:

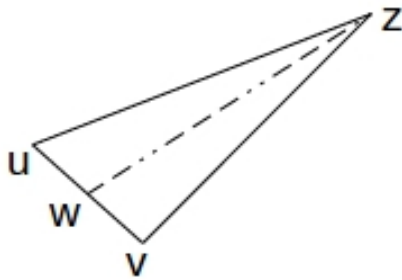
Falls d ein additive Baum-Metrik mit Baum  $\hat{T}$  ist, dann ist  $\mathcal{A}(d) = \hat{T}$

Beispiel:

$\tilde{d} = d$

$d_{wz} = \frac{1}{2} \cdot d_{uz} + \frac{1}{2} \cdot d_{vz}$  (WPGMA)

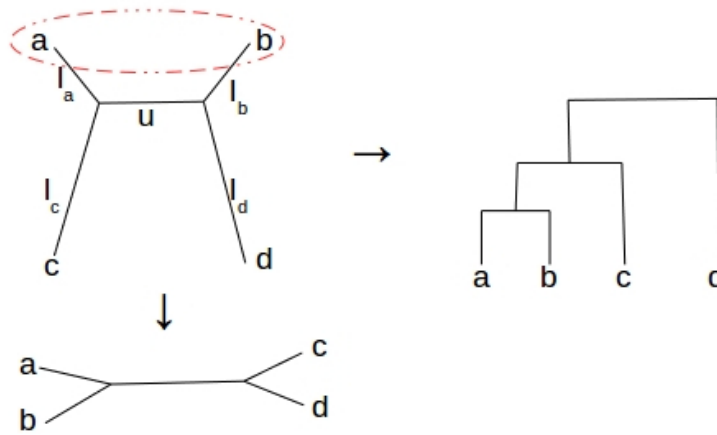
$d_{wz} = \frac{|u|}{|u|+|v|} \cdot d_{uz} + \frac{|v|}{|u|+|v|} \cdot d_{vz}$  (UPGMA)



Ist der zugehörige Algorithmus konsistent?

Gegenbeispiel:

$l_a, l_b, q \ll l_c, l_d \Rightarrow \operatorname{argmin}_{x,y} \tilde{d}_{xy} = \{a, b\}$



(Problem: LBA - long branch attraction)

Lösung:

Abstand eines Punktes von allen anderen Punkten berechnen:  $r(u) = \sum_{x \neq u} d(x, u)$

$$\tilde{d}_{xy} = d_{xy} - \alpha \cdot r(x) - \beta \cdot r(y)$$

$$\alpha = \beta = \frac{1}{n-2} \text{ mit } n = \text{Zahl der Taxa}$$

**Lemma:** Wenn  $d$  eine additive Baum-Metrik ist und  $\{u, v\} = \operatorname{argmin}_{x, y} \tilde{d}_{xy} = \{u, v\} \Rightarrow u, v$  wird cherry genannt.

$\{u, v\} \mapsto w$  (u und v mittels Vaterknoten w vereinigen)

$$d(u, w) = \frac{1}{2} \cdot d(u, v) + \frac{1}{2} \cdot \frac{1}{n-2} [r(u) - r(v)]$$

$$\text{durch Symmetrie: } d(v, w) = \frac{1}{2} \cdot d(u, v) + \frac{1}{2} \cdot \frac{1}{n-2} [r(v) - r(u)]$$

$$d(w, z) = \frac{1}{2} \cdot [d(u, z) - d(u, w)] + \frac{1}{2} \cdot [d(v, z) - d(v, w)]$$

$$= \frac{1}{2} \cdot [d(u, z) + d(v, z)] - d(u, w)$$

[Paper: Gascuel + Steel, Mol Biol Evol, 23 Seite 1997-2000 (2006)<sup>2</sup>]

## 9.2 Neighbor Net

Kalmanson Metrik

→ zirkuläre Ordnung der Taxa

- Auswahl der Nachbarn
- Update der Distanzen

Initialisierung: Jeder Punkt ist in einem separaten Cluster  $C_i$ , mit Punkten  $x, y, \dots$

$$d(C_i, C_j) := \frac{1}{|C_i||C_j|} \sum_{\substack{x \in C_i \\ y \in C_j}} d(x, y)$$

<sup>2</sup><http://mbe.oxfordjournals.org/content/23/11/1997.long>



$$Q(C_i, C_j) := (m-2) \cdot d(C_i, C_j) - \underbrace{\sum_{k \neq i} d(C_i, C_k)}_{(m-2) \cdot r(C_i)} - \underbrace{\sum_{k \neq j} d(C_j, C_k)}_{(m-2) \cdot r(C_j)}$$

mit  $m$  = Anzahl Cluster

(NI-Formale für Cluster)

Bestimme  $i^*, j^* = \operatorname{argmin}_{i,j} Q(C_i, C_j)$

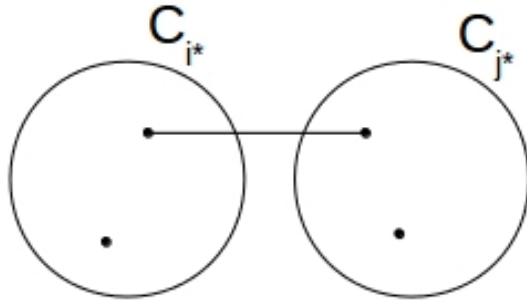
$C_i, C_j$  enthält jeweils entweder 1 oder 2 Knoten

für Punkte in  $x_i \in C_{i^*}$  und  $x_j \in C_{j^*}$

$$\hat{Q}(x_i, x_j) = (\hat{m} - 2) \cdot d(x_i, x_j) - \sum_k d(x_i, C_k) - \sum_k d(x_j, C_k)$$

$$\hat{m} = m - \underbrace{2}_{i^*, j^*} + |C_{i^*}| + |C_{j^*}|$$

Erkläre  $x^*, y^*$  mit  $x^* \in C_{i^*}, y^* \in C_{j^*}$  (mit jedem Schritt eine Kante mehr)



$y$  hat 2 (verschiedene) Nachbarn  $x, z$

$a \neq x, y, z, u, v$

$$d(u, a) = \alpha \cdot d(x, a) + \beta \cdot d(y, a)$$

$$d(v, a) = \beta \cdot d(y, a) + \gamma \cdot d(z, a)$$

$$d(u, v) = \alpha \cdot d(x, y) + \beta \cdot d(x, z) + \gamma \cdot d(y, z)$$

mit  $\alpha + \beta + \gamma = 1$ ;  $\alpha, \beta, \gamma \geq 0$ ;  $\alpha = \beta = \gamma = \frac{1}{3}$

**Theorem:** Wenn d Kalmanson Eigenschaften hat  
 $\Rightarrow$  Neighbor Net erzeugt die zugehörige zirkuläre Ordnung und identifiziert damit  
 alle Splits mit nichtnegativen  $\beta_{A|B}$

letzter Schritt im Neighbor Net Algorithmus:

$$\min_{\substack{\beta_{A|B} \forall A|B \\ \text{cirkuläre Splits}}} \left( \sum_{x,y} (d(x, y) - \sum_{\text{splits}} \beta_{A|B} \cdot \delta_{A|B}(x, y))^2 \right) \text{ mit } \beta_{A|B} \geq 0$$