

Graphen und biologische Netze (WS 2016/17)

Inhaltsverzeichnis

1	Vorlesung 14.10.2016	1
1.1	Grundlagen der Graphen und biologische Netze	1
1.2	Gleichheit von Graphen	2
1.3	Eigenschaften von Graphen	4
1.4	Graph-Invarianten	5
1.5	Pfade und Zusammenhänge	5
2	Vorlesung 21.10.2016	6
2.1	Zusammenhang	6
2.2	Cut-Vertex (Schnittknoten)	7
2.3	Spezielle Graphen	10
2.4	algebraische Graphentheorie	11
2.4.1	Matrizendarstellung von Graphen	11
2.4.2	Permutationsmatrix	11
2.5	Spektrum einer Matrix	13
3	Vorlesung 28.10.2016	14
3.1	Inzidenzstrukturen	14
3.1.1	Inzidenzmatrize - ungerichtete Graphen	14
3.1.2	Inzidenzmatrize - gerichtete Graphen	15
3.2	Laplace-Matrix	16
3.2.1	Algebraische Konnektivität	17
3.2.2	Fiedler Vektor	18
3.2.3	Interlacing Theorem	18
3.2.4	Anzahl nicht-isomorpher Graphen	18
3.2.5	Isomorphismus auf Bäumen	19
4	Vorlesung 04.11.2016	22
4.1	Kreise	22
4.2	Kreisbasen	25
5	Vorlesung 11.11.2016	29
5.1	Ohrenzerlegung	29
5.1.1	Algorithmus der Ohrenzerlegung	30
5.2	Unabhängigkeitssysteme und Matroide	31
5.2.1	Matroid	31
5.2.2	Greedy Algorithmen auf Matroiden	32
5.2.3	Horton-Algorithmus (1972)	33
6	Vorlesung 18.11.2016	34
6.1	Graphen in der Ebene und planare Graphen	34

7	Vorlesung 25.11.2016	38
7.1	Färbung von Graphen	38
7.2	Zufallsgraphen	41
7.3	Eigenschaften fast aller Graphen	41
8	Vorlesung 02.12.2016	42
8.1	Cographen & Cotrees	42
9	Vorlesung 09.12.2016 (Spezialvorlesung 1)	48
9.1	Phylogenetische Kombinatorik	48
9.1.1	Phylogenetische Distanzen	48
9.1.2	Modell für Sequenzevolution	50
9.1.3	Evolutionäre Distanzen → Baum	51
9.1.4	Splitsystem eines Baums	54
10	Vorlesung 16.12.2016 (Spezialvorlesung 2)	55
10.1	Metriken	55
10.1.1	Messung der Baumartigkeit:	58
10.1.2	Travelling sales person problem (TSP)	59
11	Vorlesung 21.12.2016 (Spezialvorlesung 3)	61
11.1	neighbor joining	61
11.2	Neighbor Net	63

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 $\alpha: \alpha : V \rightarrow L_V : v \mapsto \alpha(v)$

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

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:

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

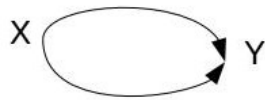
$t : E \rightarrow V : e \mapsto tail(e)$

Graphen in denen Kanten zwei verschiedenen Endpunkte haben **UND** zu jedem Paar von Knoten 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

vollständiger Graph: jeder Knoten ist mit allen anderen Knoten verbunden →
Anzahl Kanten = $\binom{n}{2}$

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 Abbildung¹ $\pi : V_1 \rightarrow V_2$ gibt, sodass $\{x, y\} \in E_1 \Leftrightarrow \{\pi(x), \pi(y)\} \in E_2$

¹https://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:

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

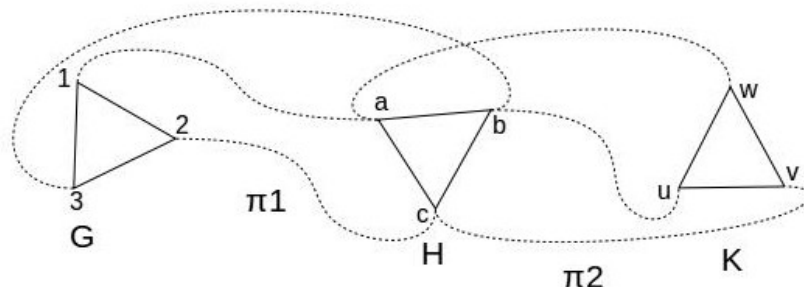
Bezogen auf die Labels kann es mehrere mögliche Isomorphismen geben.

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

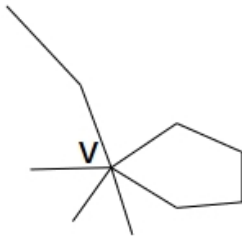
Gegeben: $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$

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

$\deg(v) := |N(v)| \dots$ Anzahl der Nachbarknoten

$\delta(G) := \min_{v \in V} \deg(v) \dots$ Knoten aus G mit wenigstens Nachbarn

$\Delta(G) := \max_{v \in V} \deg(v) \dots$ Knoten aus G mit meisten Nachbarn

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

Gradfolge von G : aufsteigende Folge der Knotengrade aller Knoten eines Graphen² $\mathcal{F} = (n_0, n_1, n_2, \dots, n_{|V|-1})$ mit $n_k := |\{x \in V \mid \deg(x) = k\}|$

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

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

Beispiel:



$$\begin{array}{cccccc} & 0 & 1 & 2 & 3 & 4 \\ \mathbf{F} = & (0 & 4 & 0 & 0 & 1) \end{array}$$



$$\begin{array}{cccc} & 0 & 1 & 2 & 3 \\ \mathbf{F} = & (0 & 0 & 4 & 0) \end{array}$$

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

²<https://de.wikipedia.org/wiki/Gradfolge>

1.4 Graph-Invarianten

Eigenschaften, die unter Isomorphie erhalten bleiben³

\mathcal{G} ...Menge aller Graphen, F ...ist ein Graphinvariant und X eine Eigenschaft
wenn $F: \mathcal{G} \rightarrow X$ die Eigenschaft hat, dass $G \simeq H \Rightarrow F(G) = F(H)$

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: (4,a,1,b,3,b,1,c,2)



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

Beispiel: (1,c,2,d,3,b,1,a,4,e,3))

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

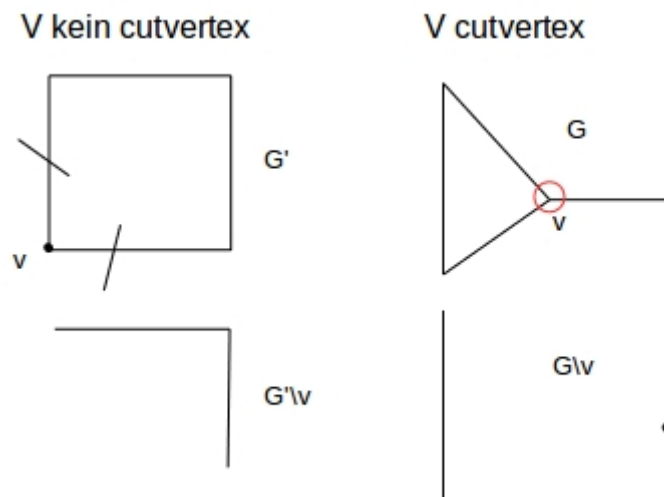
Frage:

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

³https://en.wikipedia.org/wiki/Category:Graph_invariants

2.2 Cut-Vertex (Schnittknoten)

Sei G ein zusammenhängender Graph. Dann ist v ein Cut vertex wenn $G \setminus \{v\}$ [= Graph der entsteht wenn v und alle inzidenten⁴ Kanten entfernt werden] in wenigstens 2 Zusammenhangskomponenten zerfällt.



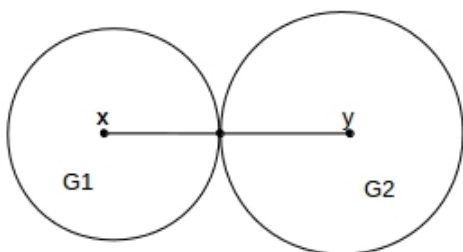
Theorem: Sei G ein zusammenhängender Graph mit mehr als 2 Knoten, dann seien folgende Aussagen äquivalent:

- G ist Kreiszusammenhängend $\Rightarrow G$ hat keinen cut-vertex

Beweis/Idee: cut-vertex $\Rightarrow \neg$ Kreiszusammenhang



$$|V(G)| \geq 3$$



$x \in G_1, y \in G_2 \Rightarrow \nexists$ Kreis durch x und y
 $\Rightarrow \neg$ Kreiszusammenhang

⁴Ein Knoten v und eine Kante e heißen inzident, wenn e den Knoten v mit einem anderen Knoten verbindet

zu 2.:

Kreiszusammenhang $\Rightarrow \nexists$ cut-vertex ($|V(G)| \geq 3$)

Kreiszusammenhang $\Rightarrow \exists$ mindestens 2 disjunkte Wege von x nach y, d.h. es gibt keinen Knoten durch den alle Wege von x nach y gehen \Rightarrow es kann keinen cut-vertex geben

Spezialfälle:

1. $|V(G)| = 1$ (nur ein Knoten) \Rightarrow zusammenhängend, \neg Kreiszusammenhängen, kein cut-vertex
2. $|V(G)| = 2$ (zwei Knoten, eine Kante) \Rightarrow zusammenhängend, \neg Kreiszusammenhängen, kein cut-vertex

Definition.: Ein Graph G ist 2-zusammenhängend wenn $G \setminus v$ für alle $v \in V(G)$ zusammenhängend und nicht ein einzelner Knoten oder leer ist $\Leftrightarrow G \setminus v$ zusammenhängend und enthält wenigstens 1 Kante

Spezialfälle 1 und 2 sind zusammenhängend aber nicht 2-zusammenhängend

G ist 2-zusammenhängend $\Leftrightarrow G$ enthält keinen cut-vertex



Sei $W \subseteq V(G)$ und $G \setminus W$ der Graph der aus G entsteht, wenn alle Knoten in W und deren inzidenten Kanten entfernt werden.

G ist k -zusammenhängend, wenn $G \setminus W$ für alle W mit $|W| = k - 1$ zusammenhängend und weder K_1 noch leer ist.

	zshgd	2-zshgd	3-zshgd	4-zshgd
	$\kappa(G)=1$ ✓	—	—	—
	$\kappa(G)=2$ ✓	✓	—	—
	$\kappa(G)=3$ ✓	✓	✓	—

Definition: $\kappa(G)$ ist die größte Zahl k sodass G k -zusammenhängend $\Leftrightarrow G$ ist k -zusammenhängend aber nicht $(k+1)$ -zusammenhängend

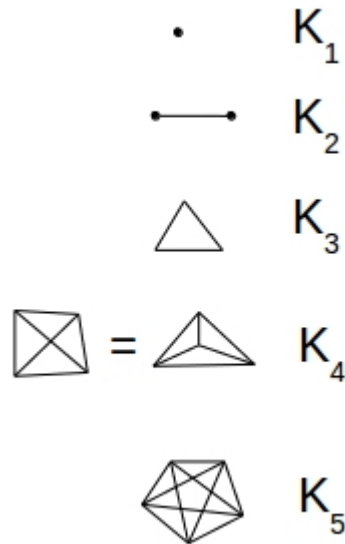
$\kappa(G)$ heißt die Konnektivität von G



Bemerkung: G ist k -zusammenhängend $\Rightarrow G \setminus v$ ist $(k-1)$ -zusammenhängend für alle $v \in V(G)$

2.3 Spezielle Graphen

vollständige Graphen:



$K_n \dots n = V(K_n)$ Knoten, $\binom{n}{2}$ Kanten

Kreise:



Baum: zusammenhängender Graph der keine Kreise enthält

Wald: disjunkte Vereinigung von Bäumen \Leftrightarrow kreisfreier Graph

Teilgraphen: $H(W,F)$ ist Teilgraph von $G(V,E)$ wenn

1. $W \subseteq V$
2. $F \subseteq E$
3. $\{x,y\} \in F \Rightarrow x,y \in W$

Induzierter Teilgraph: H ist induzierter Teilgraph von G wenn

1. H Teilgraph von G
2. $x,y \in W$ und $\{x,y\} \in E \Rightarrow \{x,y\} \in F$

spannende Teilgraphen (spanning subgraphs): Teilgraphen $W=V$

Spannbäume (spanning trees): Teilbaum von G , der alle Knoten enthält
jeder zusammenhängende Graph hat einen Spannb Baum

2.4 algebraische Graphentheorie

Idee: Graphen \Rightarrow Matrizen \Rightarrow Eigensch. der Matrizen \Rightarrow Eigensch. der Graphen

2.4.1 Matrizendarstellung von Graphen

Adjazenz-Matrix $A \rightarrow$ fixe Reihenfolge der Knoten ($x < y$)

$$A_{xy} = \begin{cases} 1 & \text{wenn } x,y \text{ benachbart } \{x, y\} \in E \\ 0 & \text{sonst} \end{cases}$$

Beispiel:



	1	2	3	4
1	0	1	1	0
2	1	0	1	0
3	1	1	0	1
4	0	0	1	0

Eigenschaften von A:

- symmetrisch
- für simple Graphen $A_{xx}=0$ für alle A
- für unterschiedlich Nummerierungen unterschiedliche A-Matrizen

2.4.2 Permutationsmatrix

quadratische Matrix P sodass in jeder Zeile und jeder Spalte genau eine 1 steht und sonst 0 ($P=(p_{ij})$)

Satz: zwei Graphen mit Adjazenz-Matrizen A und B sind isomorph genau dann wenn es eine P gibt sodass $A \cdot P = P \cdot B$

$$PP^T = P^T P = I \quad (\hat{=} \text{Einheitsmatrix}) \quad [\dots \text{ mit } P_{ij}^T = P_{ji}]$$

$$A \cdot P = P \cdot B$$

$$A \cdot \underbrace{PP^T}_I = P \cdot B \cdot P^T$$

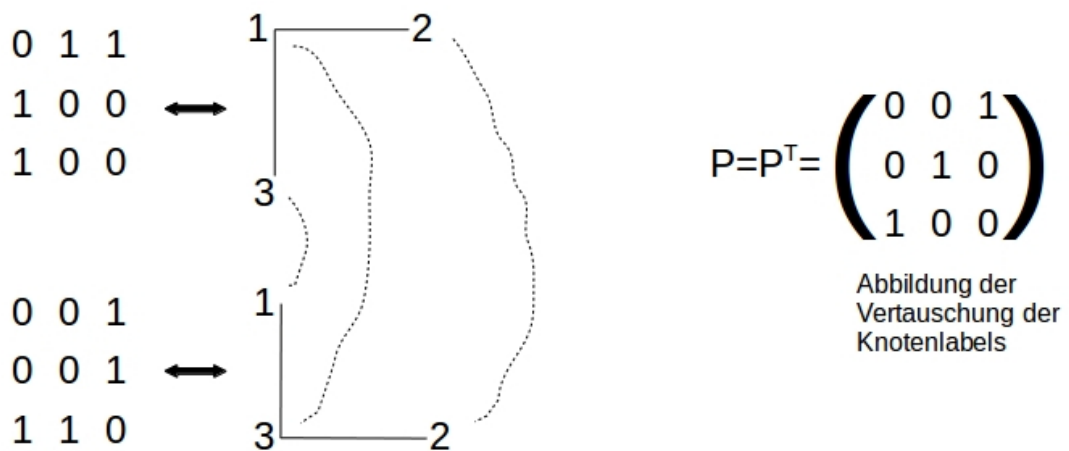
Zwei quadratische Matrizen A, B heißen ähnlich wenn es einer invertierbare Matrix Q gibt, sodass $A \cdot Q = Q \cdot B$

Satz: ähnliche Matrizen haben das gleiche Spektrum



$$\begin{aligned}
 j &= \pi(i) \\
 P_{ij} &= 1 \\
 P_{ij'} &= 0 \text{ für } j' \neq j \\
 P_{i'j} &= 0 \text{ für } i' \neq i
 \end{aligned}$$

Beispiel:



2.5 Spektrum einer Matrix

- für unsere Fälle quadratisch
- Eigenvektor x und Eigenwerte λ von A erfüllen $A \cdot x = \lambda \cdot x$

Eine $n \times n$ -Matrix hat höchstens n verschiedene Eigenwerte

- wenn A symmetrisch ($A=A^T$) sind alle Eigenwerte reell
- es gibt $\det(A - \lambda \cdot I) = 0 \rightarrow$ liefert Gleichung n -ten Grades für λ

Bemerkung: $\det(A) = \sum_{\pi \in S_n} a_{1\pi(1)} \cdot a_{2\pi(2)} \cdot \dots \cdot a_{n\pi(n)} \cdot (-1)^{\text{sgn}(\pi)}$

Definition: Das Spektrum einer Matrix ist die Menge der Eigenwerte
Graph \rightarrow Adjazenz-Matrix \rightarrow Spektrum der Adjazenz-Matrix

Satz: das Spektrum einer A -Matrix ist eine Graphinvariante einer Graphenklasse

3 Vorlesung 28.10.2016

3.1 Inzidenzstrukturen

Struktur aus Punktmenge und Menge von Blöcken.

Tripel: (p, B, I)

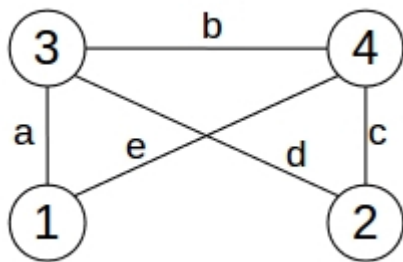
- $p \cap B = \emptyset$
- $I \subseteq p \times B$
- p = Punkte - z.B. Vertices
- B = Blöcke - z.B. Kanten
- I = Inzidenzmatrix

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

Beispiel:



		a	b	c	d	e
Inzidenzmatrix:	1	1	0	0	0	1
	2	0	0	1	1	0
	3	1	1	0	1	0
	4	0	1	1	0	1

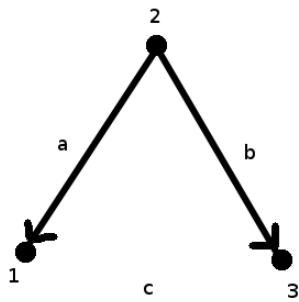
Spaltensumme: immer 2 (Start- und Endknoten)
 Zeilensumme entspricht Gradzahl des entsprechenden Knoten

3.1.2 Inzidenzmatrize - gerichtete Graphen

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

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

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 Graphen mit ungerade Knotenanzahl

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

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

Beispiel:



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

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

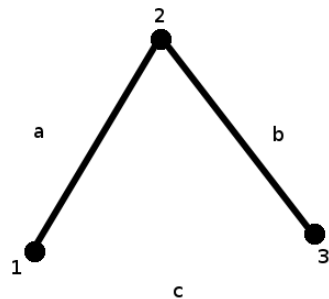
$$\text{Laplace-Matrix: } \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

Eigenschaften:

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

3.2.1 Algebraische Konnektivität (Fiedler-Wert)

- beschreibt wie gut verbunden der Graph, global gesehen, ist
- $\frac{4}{n \cdot d} \leq \text{algebraische Konnektivität} \leq \text{Vertex-Konnektivität}$
- $|V|=n$, min. Durchmesser von d (längster Pfad) (d = längster minimaler Pfad)



Beispiel:

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

weiteres Beispiel: https://en.wikipedia.org/wiki/Algebraic_connectivity

3.2.2 Fiedler Vektor

- Eigenvektor zur algebraischen Konnektivität
- eignet sich zur Graphpartitionierung



Beispiel:

- $F = \langle 0.4, 0.3, 0.1, -0.2, 0.2, -0.8 \rangle$ entspricht der Knotenreihenfolge: (1,2,3,4,5,6)
 \rightarrow umso kleiner umso schlechter mit Graphen verbunden
- \Rightarrow Partitionierung: $\{4,6\}, \{1,2,3,5\}$

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

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

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

$$- n! \leq n^n$$

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

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

$$- = \frac{n^2}{2} \left(1 - \underbrace{\frac{1}{n}}_{n \rightarrow \infty = 0} - \underbrace{\frac{2 \log_2(n)}{n}}_{n \rightarrow \infty = 0}\right) \Rightarrow \text{Fehler gegen 0} \rightarrow \text{Isomorphie selten!}$$

$$- \text{Komplexitätsklasse: } (2^{O(n^2)})$$

3.2.5 Isomorphismus auf Bäumen

- das „das ist einfach“ \rightarrow Programm mit polynomieller Zeit bauen
- Baum codieren via Werten: 0 und 1
 \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:

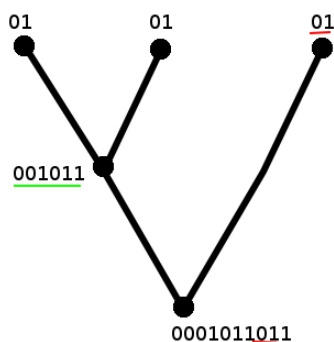


Letzes Beispiel gilt als isomorph wenn Geschwisterordnung nicht interessiert!

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 $0 A_1 \dots A_n 1$
 - 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

Bei nicht gewurzelten Bäumen:

„Kleinste maximale Entfernung zwischen den Blättern als Wurzel wählen“

4 Vorlesung 04.11.2016

4.1 Kreise

Sei $G(V,E)$ ein Graph. Teilgraphen H aus G beschreiben wir dann durch deren Kantenmenge $F \subseteq E \Leftrightarrow$ isolierte Knoten in H werden ignoriert
isoliert: $\deg_H(x)=0$

Kreis C - zusammenhängend und $\deg_C(x)=2$

verallgemeinerte Kreise: kantendisjunkte Vereinigung von Kreisen

eulerscher Kreis: zusammenhängende verallgemeinerte Kreise

Bemerkung: verallgemeinerte Kreise haben überall geraden Knotengrad



Theorem (Euler):

1. H ist verallgemeinerter Kreis \Leftrightarrow jeder Knoten geraden Grad
2. Wenn H zudem zusammenhängend, dann gibt es einen geschlossenen Weg, der Kanten von H (genau 1 mal) enthält.

Beweisskizze des Euler-Theorems

zu 1:

\Rightarrow erledigt in der oberen Bemerkung

\Leftarrow zeige: Sei H zusammenhängend und $\deg(x)$ ist gerade für alle Knoten $\Rightarrow H$ ist ein verallgemeinerter Kreis

\rightarrow sei u ein beliebiger Startpunkt. Konstruiere einen Weg w in H ausgehend von u



w erreicht nach endlich vielen Schritten einen bereits benutzten Knoten y zum zweiten Mal.

Teilweg von der 1. zur 2. Berührung von y ist ein Kreis, weil dieses Wegstück ein Pfad ist. Nennen diesen Kreis C_y .

Konstruiere den Graphen $H \setminus C_y =: H'$

Für H' gilt:

- $\deg_{H'}(x) = \deg_H(x)$ wenn x nicht auf C_y liegt
- $\deg_{H'}(x) = \deg_H(x) - 2$ wenn x auf C_y liegt

Also: In H gibt es einen Kreis C_y

- $H \setminus C_y$ hat wieder nur geraden Knotengrad
- $H \setminus C_y$ muss aber nicht mehr zusammenhängend sein
- außerdem hat $H \setminus C_y$ strikt weniger Knoten als H

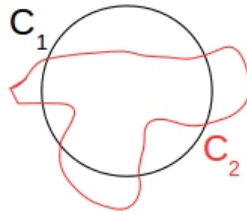
Wiederhole die Zerlegung $H \rightarrow H \setminus C_y$ für jede Zusammenhangskomponente von H'



\Rightarrow Alle entstehenden Kreise sind Kantendisjunkt. Die Vereinigung der Kreise ist ganz H .

zu 2:

Seien C_1 und C_2 zwei Kreise, die wenigstens einen Knoten gemeinsam haben



- Suche einen Startpunkt auf C_1 (beliebig!)
- Folge C_1 bis zum 1. Schnittpunkt mit C_2 ($=y$)
- folge $C_2 \rightarrow$ das führt zurück nach y
- folge C_1 zurück zum Startpunkt

\Rightarrow diese Vorschrift erzeugt einen geschlossenen Weg (nicht Pfad!) auf $C_1 \cup C_2$ der alle Kanten von C_1 und C_2 verwendet
wenn mehrere Kreise beteiligt:

- Suche einen Startpunkt auf C_1 (beliebig!)
- Folge C_1 bis zum 1. Schnittpunkt mit Q ($=y$)
- folge $Q \rightarrow$ das führt zurück nach y
- folge C_1 zurück zum Startpunkt

\Rightarrow diese Vorschrift erzeugt einen geschlossenen Weg (nicht Pfad!) auf $C_1 \cup Q$ der alle Kanten von C_1 und Q verwendet, d.h. $C_1 \cup Q$ enthalten einen Eulerschen Pfad

Kreis C_1 , kantendisjunkte Vereinigung Q von k Kreisen, zusammenhängend die C_1 in wenigstens einen Punkt schneidet

Angenommen auf Q gibt es einen Eulerschen Weg (d.h. einen geschlossenen Weg, der jede Kanten von Q enthält) \rightarrow siehe oben *wenn mehrere Kreise beteiligt*

Vollständige Induktion nach der Zahl der Kreise in Q .

Beweis Fall $k=1$ (mit k =Anzahl der Kreise), Q ist Kreis, der auch Eulerpfad in Q ist.

Iteration: $k-1=k$

Wenn Q mit $k-1$ Kreisen einen Eulerschen Pfad enthält (=Induktionsannahme) dann enthält $C_1 \cup Q$ (k -Kreise) auch einen Eulerpfad \rightarrow Aussage 2 damit korrekt

4.2 Kreisbasen

Seien H_1 und H_2 zwei Teilgraphen von G . Dann ist $H_1 \oplus H_2$ der Teilgraph von G mit Kantenmenge $(H_1 \cup H_2 \setminus (H_1 \cap H_2))$

\oplus heißt auch symmetrische Differenz (XOR)



Lemma: Seien C_1 und C_2 verallgemeinerte Kreise dann ist $C_1 \oplus C_2$ wieder verallgemeinerter Kreis

Beweis:

$$\deg_{C_1}(x) \bmod 2 = 0$$

$$\deg_{C_2}(x) \bmod 2 = 0$$

$$\deg_{C_1 \oplus C_2}(x) \bmod 2 = 0$$

$C_i(x)$ = Menge der Kanten in C_i ($i=1,2,1\cup 2$) die x enthalten

$$C_{1\cup 2}(x) = (C_1(x) \cup C_2(x)) \setminus (C_1(x) \cap C_2(x))$$

$$|C_{1\cup 2}(x)| = |C_1(x)| + |C_2(x)| - 2|C_1(x) \cap C_2(x)|$$

gerade \Leftarrow gerade + gerade - gerade

Charakteristischer Vektor eines Teilgraphen H

$$\chi_H(e) = \begin{cases} 1 & \text{wenn } e \in H \\ 0 & \text{sonst} \end{cases}$$

Lemma: Sei \mathcal{H} die (ungerichtete) Inzidenzmatrix von G

Dann gilt $\mathcal{H} \cdot \chi_H = 0 \Leftrightarrow H$ ein verallgemeinerter Kreis in G ist

$$\sum_{x \in G} \bigoplus_{e \in G} \mathcal{H}_{x,e} \cdot \chi_H(e) = 0$$

mit x alle Knoten in G und e alle Kanten in G (mit $\mathcal{H}_{x,e}$ 1 wenn x Endpunkt von e , 0 sonst)

Beweis: nachrechnen (Achtung Rechenregeln GF(2)⁵)
 $\text{GF}(2) = \{1, 0, \oplus, \}$

- $1 \oplus 1 = 0$
- $1 \oplus 0 = 0 \oplus 1 = 1$
- $0 \oplus 0 = 0$

Weil $\mathcal{H}(\chi_{H_1} \oplus \chi_{H_2}) = \mathcal{H} \cdot \chi_{H_1} \oplus \mathcal{H} \cdot \chi_{H_2} = 0$ (wieder verallgemeinerte Kreise) mit $\mathcal{H} \cdot \chi_{H_1} = 0$ und $\mathcal{H} \cdot \chi_{H_2} = 0$ weil H_1 und H_2 v.K.

\Rightarrow verallgemeinerte Kreise bilden einen Vektorraum

Deswegen gibt es eine Basis, d.h. eine Menge von Kreisen $B = \{C_1, C_2, \dots, C_\mu\}$ sodass jeder verallgemeinerte Kreis C dargestellt werden kann als

$$C = \bigoplus_{i=1}^{\mu} \lambda_i C_i \text{ mit } \lambda \in \{0, 1\}: \begin{cases} 1 & C_i \in C \\ 0 & C_i \notin C \end{cases}$$

\Leftrightarrow jeder v.K. C ist die Summe einer eindeutig bestimmten Teilmenge $B_C \subseteq B$

Basis: minimale Menge von erzeugenden Kreisen

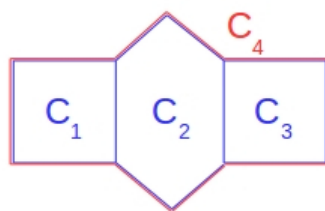
Vektorraum \Rightarrow alle minimal erzeugende Kreismenge sind gleichgroß

Anzahl Basiskreis $\hat{=}$ Dimension des Vektorraums

Menge $M = \{C_1, \dots, C_\mu\}$ von Kreisen (im allgemeinen Vektorraum) heißt linear unabhängig wenn gilt:

$$\bigoplus_{i=1}^{|M|} \lambda_i C_i \text{ mit } \lambda \in \{0, 1\} \Rightarrow \lambda_i = 0 \quad \forall i = 1 \dots |M|$$

Beispiel:



$$\underbrace{C_1 \oplus C_2 \oplus C_3}_{C_4 \oplus C_4 = 0} \oplus C_4 = 0$$

$M = \{C_1, C_2, C_3, C_4\}$ ist abhängig weil $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1$ die Gleichung

$$\bigoplus_{i=1}^4 \lambda_i C_i = 0 \text{ löst}$$

$M = \{C_1, C_2, C_3\}$ ist unabhängiges Argument. Jeder Kreis enthält eine Kante, die

⁵<https://en.wikipedia.org/wiki/GF%282%29>

in keinem anderen Kreis aus M vorkommt und daher nicht in einer Summe anderer Kreise auftreten kann.

daher nochmal Basis:

- maximale unabhängige Menge
- minimale erzeugende Menge

Konstruktion von Kreisbasen: Kirchhoff Basen (strikt fundamental Basen)

Graph G mit zusammenhängendem Spannbaum T von G

Für jede Kante $e \notin T$ enthält $T \cup \{e\}$ genau einen Kreis $\rightarrow C_{e,T}$

WARUM? Seien x, y die Endpunkte von e dann gibt es genau einen Pfad von x nach y in T . Der wird mit e zu genau einem Kreis.

$B_T = \{C_{e,T} \mid e \in T\}$ ist eine Kreisbasis

Beweis: zu zeigen

1. Jeder Kreis kann als Summe der $C_{e,T}$ erzeugt werden
2. Die Kreismenge B ist unabhängig

zu 1.: Sei Z ein Kreis, $Z \notin B$ (Z keine Kreisbasis)

$\Rightarrow Z$ Summe von Basiskreisen. Wenn $e \in Z \setminus T$ dann muss $C_{e,T}$ zu Z beitragen.

Wenn $e \notin Z \setminus T$ dann $e \in T$ gibt es keinen zugehörigen Basiskreis

$e \notin T$, $e \notin Z$ dann kann $C_{e,T}$ nicht zu Z beitragen

Also $\bigoplus_{e \in Z \setminus T} C_{e,T}$

zu 2.: $C_{e,T}$ enthält nicht e für $e, e' \in T$, $e \neq e'$

\Rightarrow jeder Basiskreis enthält eine Kante, die in keinem anderen Basiskreis enthalten ist \Rightarrow Unabhängigkeit

Wie viele Kreise enthält B (Kirchhoff-Basen)?

$$|B| = |E| - |T|$$

Kanten in Spannbaum $|T| = |V| - 1 \Rightarrow |B| = |E| - |V| + 1$

am Allgemeinen: $|B| = |E| - |V| + C(G)$ mit $C(G)$ = Anzahl der Zusammenhangskomponenten

- nicht notwendig zusammenhängend

Bemerkung: Die Kreise aus der Eulerzerlegung eines Graphen mit geraden Knotengrad bilden

- unabhängige Menge, da kantendisjunkt
- aber normalerweise keine Basis

Beispiel:



$$|E| = 12 \text{ mit } \epsilon = \text{Eulerkreise}$$

$$|B| = |E| - |V| + 1 = 12 - 8 + 1 = 5$$

Basis für G: 4 Dreiecke + 1 Viereck = 5

linear unabhängig?

für Dreiecke ja, da kantendisjunkt (wäre eine alternative Eulerzerlegung)

bei Viereck?



Viereck kann nicht durch Summe der Dreiecke dargestellt werden \Rightarrow linear unabhängig

Gibt es einen Spannbaum, der für G die Kirchhoffbasis (4 Dreiecke + Viereck) erzeugt?

\Rightarrow nein, da jeder Spannbaum, der das Viereck erzeugt, eines der 3-Ecke nicht erzeugen kann \Rightarrow keine Kreisbasis

5 Vorlesung 11.11.2016

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 definiert eine Ohrenbasis.



- **Kreisbasis:**

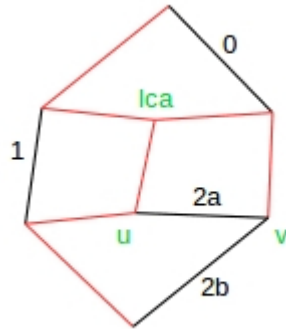
Ein Ohr sei ein maximaler Pfad P , $|P| \geq 1$, so dass P nur an Endpunkten Kanten aus $E \setminus P$ berührt. Die Knoten in P , die keine Endknoten sind haben immer $\text{Grad } \deg = 2$.

- **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 $(0,1,2a,2b)$, finde den *lowest common ancestor (lca)* 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 Unabhängigkeitssysteme und Matroide

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

- **Unabhängigkeitssystem:**

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

M1: $\emptyset = \{\} \in l \rightarrow$ Die leere Menge ist unabhängig

M2: $x \subseteq Y \in l \rightarrow x \in l$ (Erblichkeitseigenschaft)

Die Elemente $x \in l$ heißen unabhängig, die Elemente $z \in P_S(S) \setminus l$ heißen abhängig.

\rightarrow Kostenfunktion $C: S \rightarrow \mathbb{R}$

- **Austauscheigenschaft:**

M3: Falls $A \in l, B \in l, |A| < |B|$, dann $\exists x, x \in B \setminus A : A \cup \{x\} \in l$

5.2.1 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 $S_G = E(G)$

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

\rightarrow Menge an Kanten ist nur dann unabhängig, wenn $G' = (V(G), A)$ einen Wald bilden $\rightarrow G'$ 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. ($A \cup \{x\} \in l$)

\rightarrow 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 M3 zeigen, dass x abhängig von A ist: $\exists x : B \setminus A : A \cup \{x\}$. \rightarrow 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 Gewichtsfunction $w(x) > 0 \forall x \in S$ gibt. $\rightarrow w(A), A \subseteq S$ mit $w(A) = \sum_{x \in A} w(x)$.

5.2.2 Greedy Algorithmen auf Matroiden

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

Beispiel: Minimallänge Spannbäume mit $w'(x)0w_0 - w(x)$ 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

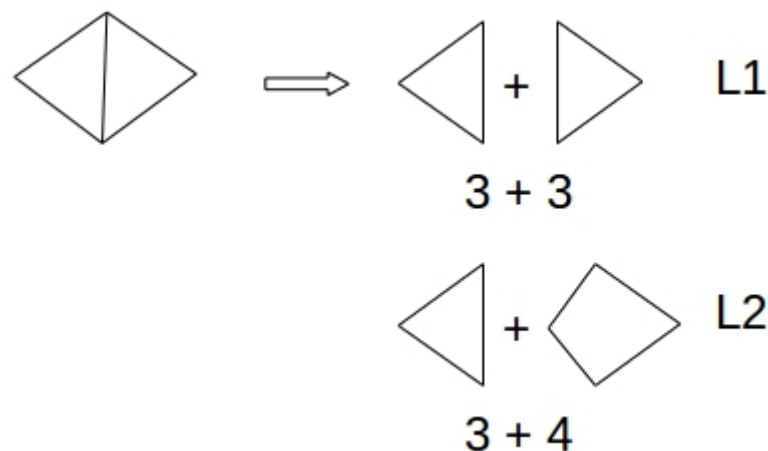
Theorem

A ist die optimale Lösung des Greedy-Algorithmus

Beispiel: minimal lange Kreisbasen

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.

5.2.3 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.

6 Vorlesung 18.11.2016

6.1 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 $\dot{P}_4, \dot{P}_2 \subset \mathbb{R}^2 \setminus \{P_1 \cup P_3\}$ ist, so schneiden sich P_4 und P_2 ($\dot{P}_4 \cap \dot{P}_2 \neq \emptyset$)

Graph in Ebene:

Ein verbundener Graph in der Ebene 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* (auch siehe oben \Rightarrow Kreuzungszug)

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 $f' = f$

Theorem: Eulersche Formel

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

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

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

\rightarrow Das entfernen einer Kante kombiniert 2 Facetten

\rightarrow Fügt man einen Knoten hinzu so wird auch eine Kante hinzugefügt

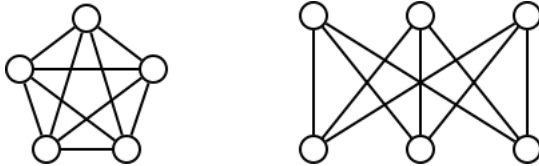
$\rightarrow |V'| = |V| + 1$ und $|E'| = |E| + 1 \rightarrow$ q.w.e.d.

Korollar 1

- 1) Ein kreuzungsfreier verbundener 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

- $F = \bigoplus_{c \in B} c$ Rand der äußeren Facette (äußerer Rand) mit B =Menge von Kreisen
- 2-Basis = $B \cup F$

Satz von Mc Lane

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

Beweis:

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

\Rightarrow **Widerspruch**

zyklometrische Zahl: Die Anzahl der Basiselemente einer Kreisbasis

$\mu(G) = |E(G)| - |V(G)| + k(G)$ mit $k(G) = \# \text{Zusammenhangskomponenten}$

Beispiel: vollständiger Graph K_5

- $|V| = 5; |E| = 10$
- 2-Basis: 5 inneren und 5 äußeren Kanten $\cdot 2 = 20$
- $\mu(K_5) = -5 + 10 + 1 = 6 + 1$ (Rand der unbeschränkten Facette) $= 7$
- 2-Basis hat demnach 20 Kanten $= 7 \cdot 3^*$ Kanten \rightarrow **falsche Aussage**
zu *: 3 von Graph mit drei Knoten um den K_5 einschließen zu können

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

- $|V| = 6; |E| = 9$
- 2-Basis: 9 Kanten $\cdot 2 = 18$
- $\mu(K_{3,3}) = -6 + 9 + 1 = 4 + 1$ (Rand der unbeschränkten Facette) $= 5$
- 2-Basis hat demnach 18 Kanten $= 5 \cdot 4^*$ Kanten \rightarrow **falsche Aussage**
zu *: 4 von Graph mit vier Knoten um den $K_{3,3}$ einschließen zu können

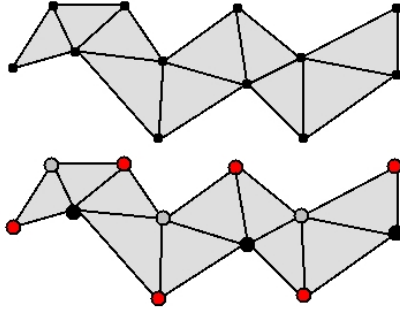
Planaritätstest

- 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.

7 Vorlesung 25.11.2016

7.1 Färbung von Graphen

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.



Vertexfärbung Eine Vertexfärbung eines Graphen $G = (V, E)$ ist eine Abbildung $\mathcal{C}: V \mapsto \mathcal{S}$, mit \mathcal{S} = Menge der Farben. Es gilt, dass $\mathcal{C}(v) \neq \mathcal{C}(w)$, mit $w, v \in \mathcal{S}$, wenn v und w adjazent ($\{v, w\} \in E$) sind. Die Elemente von \mathcal{S} heißen *Farben*.

k-Färbung Ein Graph G ist k -färbbar, wenn es für eine Abbildung \mathcal{C} eine Menge $\mathcal{S} = \{1, \dots, k\}$ gibt.

Chromatische Zahl Eine chromatische Zahl $\chi(G)$ ist die kleinste natürliche Zahl k , sodass G k -färbbar ist. $\chi(G) \leq \Delta(G) + 1$, mit $\Delta(G)$ = maximaler Grad von G

Proof (greedy) Färbe v_i der Vertices $v_1 \dots v_n$ mit der kleinsten Farbe, die nicht von einem Nachbarn von v_i benutzt wird. Da wir max. $\Delta(G)$ viele Nachbarn für v_i haben, gibt es immer eine freie Farbe.

$$\chi(G) \geq \text{Größe der größten Clique}$$

Lemma Für jeden einfachen planaren Graphen G ist der Durchschnittsgrad $d(G) < 6$

Proof $d(G) = 2 \cdot \frac{|E|}{|V|}$ mit $|V| \leq 3$, $|E| \leq 3 \cdot |V| - 6$,
dann $d(G) \leq \frac{2(3 \cdot |V| - 6)}{|V|} = 6 - \frac{12}{|V|}$

Theorem Jeder simple planare Graph G hat $\chi(G) \leq 6$

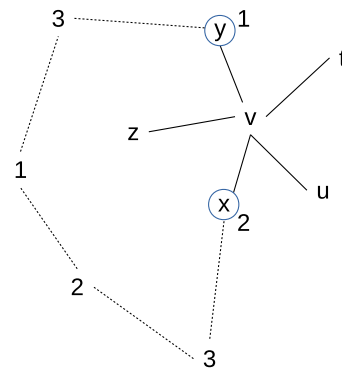
Proof Annahme: Jeder simple planare Graph mit $|V| = n$ ist 6-färbbar.

- Sei G hiermit ein simpler planarer Graph mit $|V| = n + 1$
- Vom Lemma wissen wir, dass $w \in V$ mit $\deg(w) \leq 5$ existiert
- Sei $G' = G \setminus \{w\}$. Via Induktionshypothese ist G' 6-färbbar. Das tun wir dann.
- Färbe w mit der (min.) freien Farbe, um G zu färben

Theorem Für jeden simplen planaren Graphen G gilt⁶, dass $\chi(G) \leq 5$

Proof Sei $G = (V, E)$ planar

1. Falls $|V| \leq 5 \rightarrow$ trivial
2. Für alle $v \in V(G)$ mit $\deg(v) < 5$, färbe v und arbeite mit $G \setminus \{v\}$



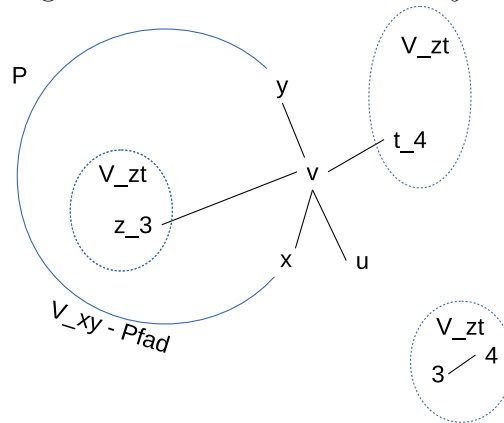
3. G hat Vertex v mit $\deg(v) = 5$
 - betrachte 5-Färbung $\mathcal{C}: V(G \setminus \{v\}) \mapsto \{1, \dots, 5\}$, die es nach Rekursion gibt
 - betrachte $x, y \in V_{xy}$ und sei $V_{xy} \subset V(G)$ die Menge der Knoten mit $\mathcal{C}(x)$ - oder $\mathcal{C}(y)$ -Färbung
 - (a) es gibt keinen Weg von $x \rightsquigarrow y$, der nur Knoten aus V_{xy} nutzt
 - Seien V'_{xy} die $s \in V(G \setminus \{v\})$, die von x nur via V_{xy} erreicht werden, $y \notin V'_{xy}$

$$\text{– Färbe um: } \mathcal{C}'(s) = \begin{cases} \mathcal{C}(s) & s \notin V'_{xy} \\ \mathcal{C}(y) & s \in V'_{xy}, \mathcal{C}(s) = \mathcal{C}(x) \\ \mathcal{C}(x) & s \in V'_{xy}, \mathcal{C}(s) = \mathcal{C}(y) \end{cases}$$

“Tausche Farben auf V'_{xy} ” $\Rightarrow \mathcal{C}'(x) = \mathcal{C}'(y) = \mathcal{C}(y)$

⁶siehe 5-Farben-Satz <https://de.wikipedia.org/wiki/F%C3%BCnf-Farben-Satz>

(b) es gibt einen solchen Pfad $x \rightsquigarrow y$ mit allen Knoten in V_{xy}



- V_{zt} : alle Knoten in $V(G \setminus \{v\})$, die $\mathcal{C}(t)$ oder $\mathcal{C}(z)$ gefärbt sind
- $V_{xy} \cap V_{zt} = \emptyset!$
- V'_{zt} kann nur via eines $s \in P$ einen Knoten in $V_{zt} \setminus V'_{zt}$ erreichen

Damit lassen sich z, t analog zu Fall (a) färben.

Theorem Jeder planare Graph ist 4-färbbar

Proof Es gibt eine Menge von 1936 4-färbbaren Karten, jede nicht Teil eines kleinsten Gegenbeispiels... (Appel, Haken, 1976). \Rightarrow Es folgt, dass es kein kleinstes Gegenbeispiel gibt.

7.2 Zufallsgraphen

Sei $G = (V, E)$ mit $V = \{1, \dots, n\}$ fixiert. Wir wollen nun Kanten zufällig auf dieser fixierten Knotenmenge auswählen, um zufällige Graphen zu generieren. Die Menge dieser Zufallsgraphen nennen wir \mathcal{G} . Jede Kante wird mit Wahrscheinlichkeit $p \in [0, \dots, 1]$ gewählt. Sei G_0 ein bestimmter Graph. Das Ereignis $\{G_0\}$ mit G_0 und m Kanten hat die Wahrscheinlichkeit $p^m \cdot (1 - p)^{\binom{n}{2} - m}$.

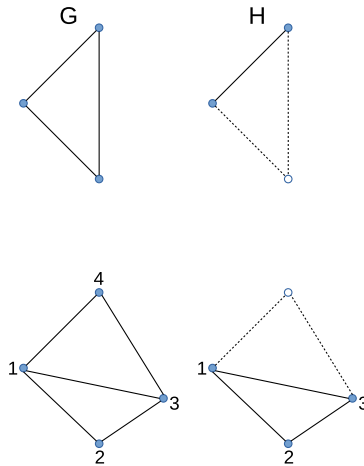
Wahrscheinlichkeitsmaß auf $\mathcal{G} \forall e \in [v]^2$

$\Omega_e = \{0_e, 1_e\}, \mathbb{P}(\{1_e\}) = p, \mathbb{P}(\{0_e\}) = 1 - p$

$\Omega_G = \prod_{e \in [v]^2} \Omega_e$

Beispiele Fixiere Graph $H, V(H) = V(G)$, ist $H \leq G$?

Mit $p^l, |E(H)| = l, |V(H)| = k$, aber falls H induzierter Teilgraph von G sein soll? \Rightarrow Nur $p^l(1 - p)^{\binom{k}{2} - l}$.



Und was ist mit Subgraph-Isomorphismus?

- Ereignismengen überlappen
- kompliziert

7.3 Eigenschaften fast aller Graphen

Falls die Wahrscheinlichkeit, dass $P(G \in \mathcal{G}) \rightarrow 1$ für $n \rightarrow \infty$, dann geschieht G fast sicher.

Pro: Gegeben jedes H als Isomorphie-Klasse, $n \rightarrow \infty$ und $p \in]0, 1[= (0, 1)$, induzierte Kopie von H . Dann haben fast alle G in $\mathcal{G}(n, p)$ mindestens eine induzierte Kopie von H .

Prf: Sei H gegeben, $K = V(H)$. Sei $K \leq n$. H ist (subgraph-) isomorph zu G mit Wahrscheinlichkeit $r > 0$ (G ist zufällig!). Teile G in $\lfloor \frac{n}{k} \rfloor$ Teilgraphen, um genau so viele “Versuche” (für $r > 0$) zu haben. $P(H \not\leq G \text{ induziert}) \leq (1 - r)^{\lfloor \frac{n}{k} \rfloor} \xrightarrow{n \rightarrow \infty} 0$

8 Vorlesung 02.12.2016

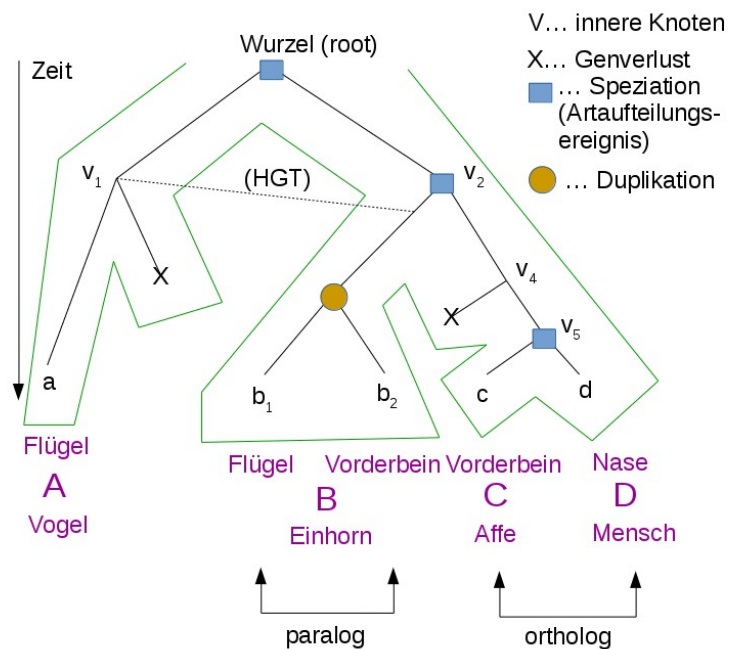
8.1 Cographen & Cotrees

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.: Θ -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 Θ erhalten. Dazu stellen wir \sim und Θ 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 $\overline{E} = \{(u,v) \in \binom{V}{2} \mid (u,v) \notin E\}$

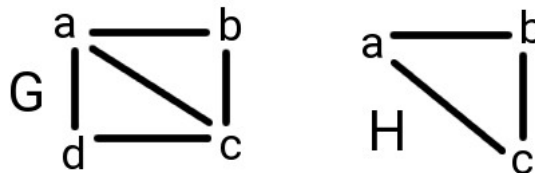


Def.: Teilgraph

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

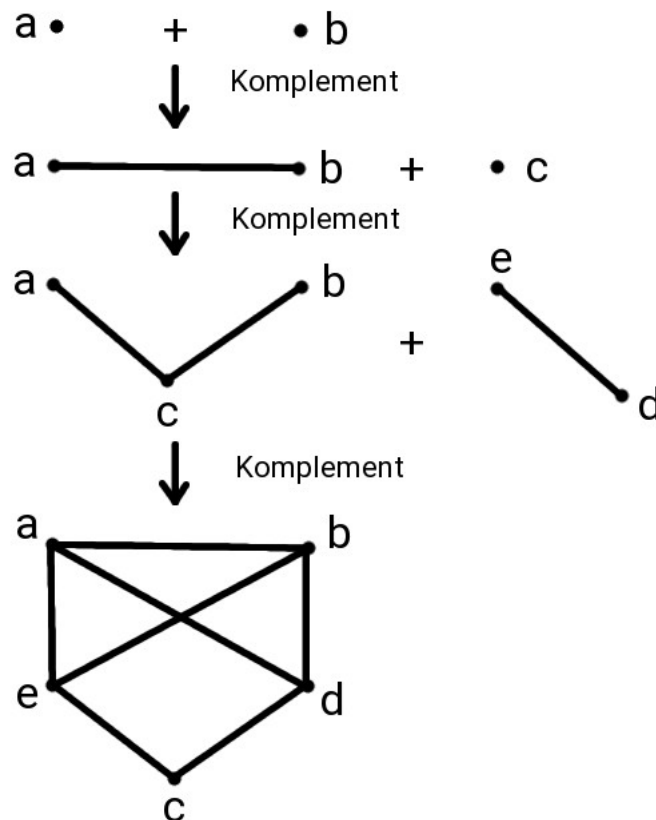
Graphen G,H : $G+H$ ist ein Graph mit $V(G) \cup V(H)$ und $E(G) \cup E(H)$.



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:

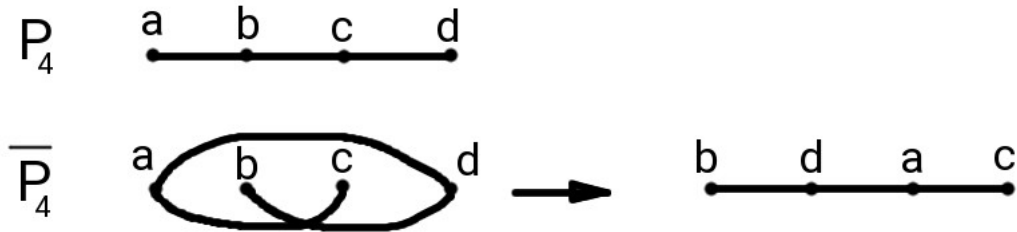


Eigenschaften von Cographen:

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

- i) G enthält keine induzierten P_4 's (Pfad mit 4 Knoten)
- 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

```
isCograph (G) {
    if (| V(G) | < 4) {return true;}
    c = {Zusammenhangskomponenten von G}
    if(| c | = 1) {c'={Komponenten von G-bar}}
    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_Θ ist ein Cograph

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

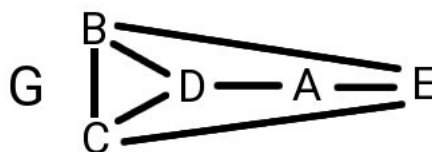
Was passiert wenn $\sim \neq \Theta$ bzw. G_\sim kein Graph?

\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.

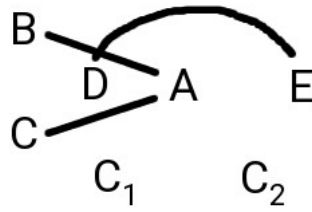
Weitere Literatur: Marc Hellmuth

Theorem:

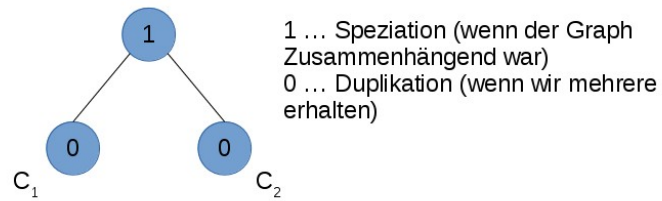
Für jeden Cographen gibt es einen eindeutigen Cotree (Cobaum)



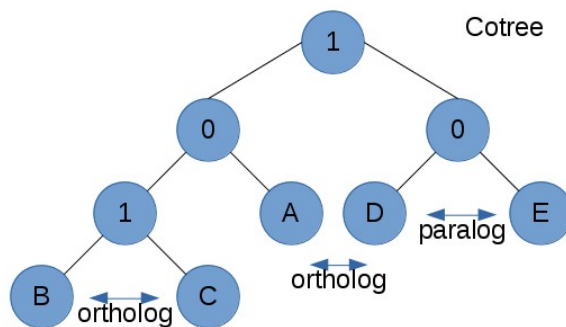
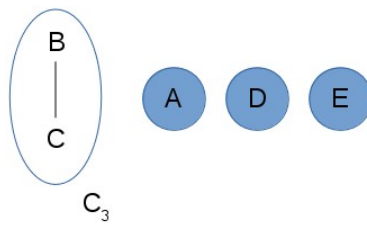
1. Schritt: Komplement



2. Schritt: umgekehrte disjunkte Vereinigung



3. Schritt: Komplement



9 Vorlesung 09.12.2016 (Spezialvorlesung 1)

9.1 Phylogenetische Kombinatorik

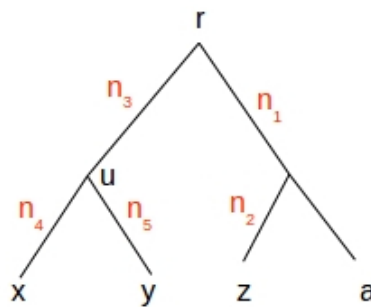
- Distanz Methoden / Clustering \leftrightarrow Netzwerke statt Bäume
- Supertree / Tripel-Methoden

9.1.1 Phylogenetische Distanzen

x = 0 1 1 1 0 0 1

y = 1 1 0 1 1 0 1

→ Hamming Distanz: $d_h(x, y) = 3$



Evolutionäre Distanz = Evol. Ereignisse; die x,y trennen

$$d_E(z, x) = n_2 + n_1 + n_3 + n_4$$

$$d_E(x, y) = n_4 + n_5$$

$$d_E(z, y) = n_2 + n_1 + n_3 + n_5$$

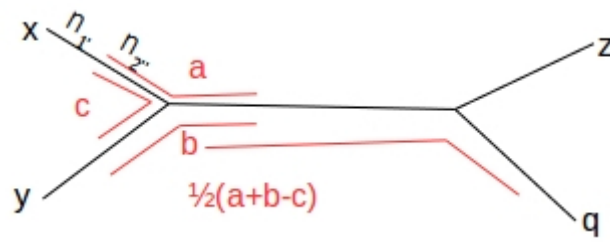
$$d_E(z, x) = d_E(z, u) + n_4$$

$$d_E(z, y) = d_E(z, u) + n_5$$

$$d_E(z, x) + d_E(z, y) = 2d_E(z, u) + n_4 + n_5$$

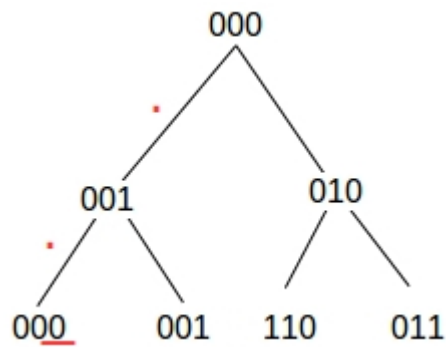
$$d_E(z, u) = \frac{1}{2} \{d_E(z, x) + d_E(z, y) - d_E(x, y)\}$$

- messbare Daten: Abstände zwischen Taxa \equiv Blätter des Baumes
 - Position der Wurzel:
 - * jede messbare Distanz enthält entweder $(n_1 + n_3)$ oder weder n_1 noch n_3
 - * genaue Position der Wurzel NICHT bestimmbar
- ungewurzelte Bäume



Entspricht Hamming-Distanz der evolutionären Distanz: $d_H(x, y) \Leftrightarrow d_E(x, y)$?

Nein wegen möglichen Rückmutationen, daher: $\underbrace{d_H(x, y)}_{\text{messbar}} \leq \underbrace{d_E(x, y)}_{\text{nicht messbar}}$



9.1.2 Modell für Sequenzevolution

Markov:

$$P_{a,b}(\tau) := P(a|b, \tau) = \left[e^{\mu\tau R} \right]_{a,b}$$

$P(a|b, \tau)$: die Wahrscheinlichkeit a zu sehen wenn ich b gesehen habe nach einer Zeit τ

μ = Mutationsrate

τ = Zeit

R = Matrix Mutationsmodell

$$R_1 = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$R_2 = \begin{pmatrix} -3 & 1 & 1 & 1 \\ 1 & -3 & 1 & 1 \\ 1 & 1 & -3 & 1 \\ 1 & 1 & 1 & -3 \end{pmatrix}$$

$$\text{Taylorreihe : } e^{\mu\tau R} := I + \tau * \mu * R + \frac{1}{2!} * \tau^2 * \mu^2 * R^2 + \frac{1}{3!} * \tau^3 * \mu^3 * R^3 + \dots$$

mit I =Einheitsmatrix

$$P_{a,b}(\tau) = \delta_{a,b} + \tau * \mu * R_{a,b} + \dots$$

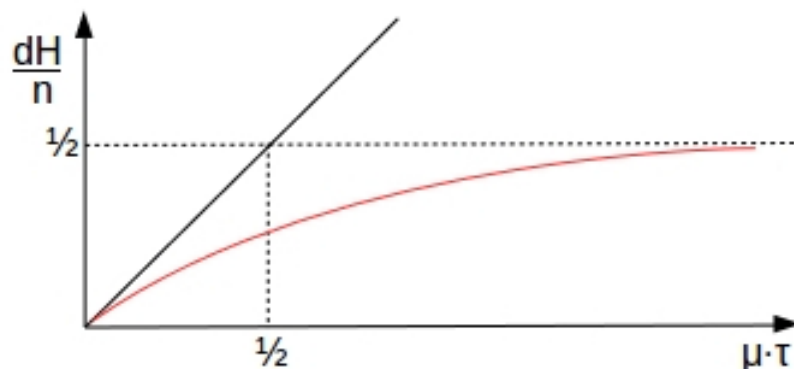
mit Kronecker-Delta: $\delta_{a,b} \begin{cases} 0 & \text{wenn } a \neq b \\ 1 & \text{wenn } a = b \end{cases}$

mit R_1 -Matrix:

$$\mathbf{a} = \mathbf{b}: P_{a,a}(\tau) = 1 + \tau * \mu * (-1) + \sigma(\tau^2) = 1 - \tau * \mu + \sigma(\tau^2)$$

$$\mathbf{a} \neq \mathbf{b}: P_{a,b}(\tau) = 0 + \tau * \mu * 1 + \sigma(\tau^2) = \tau * \mu + \sigma(\tau^2)$$

$\sigma(\tau^2)$ = Rückmutation, τ, μ können nicht getrennt werden \rightarrow Anzahl Ereignisse



$$\frac{dH}{n} = P_{a,b}(\tau) |\{a \neq b\}| = \tau * \mu + \sigma(\tau^2)$$

$$\lim_{\tau \rightarrow \infty} P_{a,b}(\tau) = \frac{1}{2}$$

$$\frac{dH}{n} = \frac{1}{2}(1 - e^{-\tau * \mu})$$

$\frac{1}{2}$ = weil 2 Buchstaben
 $\tau * \mu = dE$

$$dH(\tau) = \sum_{b \in A} \sum_{a \neq b} f_b^{Anfang} * P_{a,b}(\tau)$$

$$2 \frac{dH}{n} = 1 - e^{-\tau * \mu}$$

$$e^{-\tau * \mu} = 1 - 2 \frac{dH}{n}$$

$$\tau * \mu = -\ln\left(1 - 2 \frac{dH}{n}\right) = dE$$

Gilt nur für einfachen binären Fall (siehe R_1)

9.1.3 Evolutionäre Distanzen \rightarrow Baum

geg.: Baum T , Kantenlängen l : $E(T) \rightarrow \mathbb{R}_0^+$

$$\Rightarrow \text{Distanzen } d(x, y) = \sum_{e \in Pf_{ad_T}(x, y)} l(e) \quad (*)$$

Definition:

Eine metrische Distanz $d : X \times X \rightarrow \mathbb{R}_0^+$

heißt ADDITIV, wenn es einen Baum T

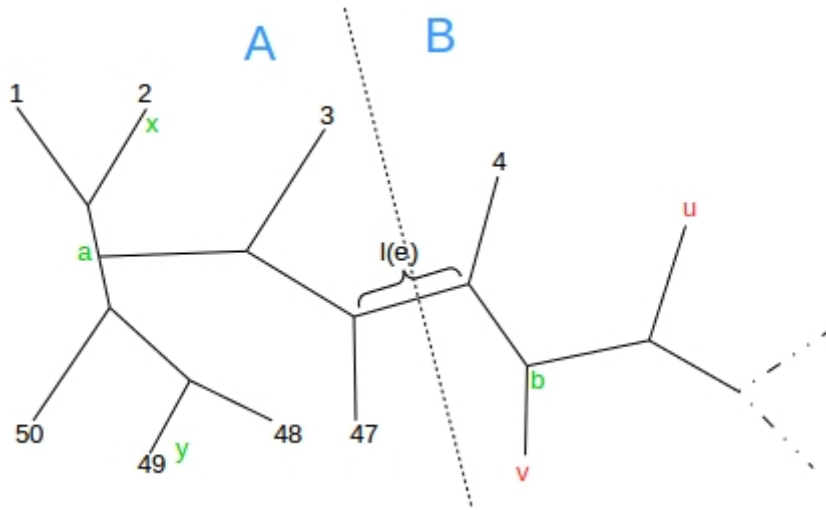
mit Blättern X und Kantenlängen l gibt sodass $(*)$ stimmt

Distanz heißt metrisch wenn:

- (i) $d(x, x) = 0$
- (ii) $d(x, y) = 0 \rightarrow x = y$
- (iii) $d(x, y) = d(y, x)$
- (iv) $d(x, y) + d(y, z) \geq d(x, z) \quad \forall x, y, z \in X$ (Dreiecksungleichung)

Wann ist eine Distanz additiv?

Bem.: wenn $l(e) > 0$ für alle Kanten $e \Rightarrow$ additive Distanz ist metrisch



Kante e induziert eine Partition von T und damit X in genau 2 Teilmengen (A, B) , nicht leer \Rightarrow **SPLIT**

$$A \cap B = \emptyset$$

$$A \neq \emptyset$$

$$B \neq \emptyset$$

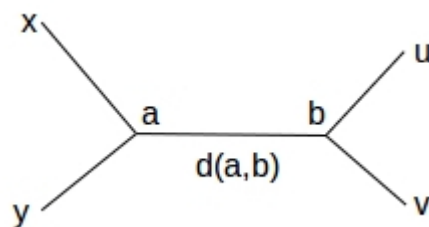
$$A \cup B = X$$

Was zeichnet Splits, die einer Kante im Baum entsprechen, aus?

Sei $A \mid B$ der Split, der zu e gehört

$x, y \in A, u, v \in B$, a innerer Knoten in A , b innerer Knoten in B :

$l(e) = \min_{a,b} d(a, b)$ für a, b als Knoten für Kante e gewählt



$$\underbrace{d(x, u) + d(y, v)}_{=Q} - d(x, y) - d(u, v) = 2d(a, b)$$

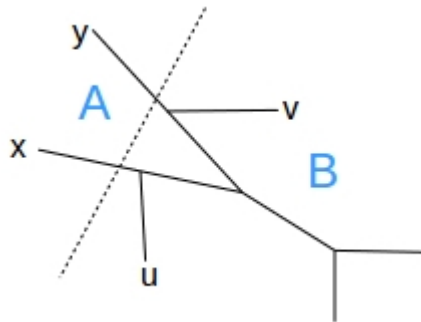
Da es ein Baum ist auch alternativ: $\underbrace{d(x, \underline{v}) + d(y, \underline{u})}_{=Q} - d(x, y) - d(u, v) = 2d(a, b)$

→ Q muss in Baum das Gleiche sein.

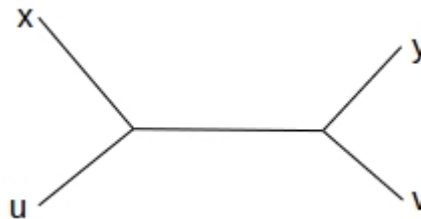
$$\Rightarrow l(e) = \frac{1}{2} \min_{\substack{x, y \in A \\ u, v \in B}} d(x, u) + d(y, v) - d(x, y) - d(u, v)$$

$$\text{oder } l(e) = \frac{1}{2} \min_{\substack{x, y \in A \\ u, v \in B}} d(x, \underline{v}) + d(y, \underline{u}) - d(x, y) - d(u, v)$$

Was wenn A, B nicht zu einer Kante gehören?



⇒ $l(e) < 0$ da:



$$\Rightarrow \underbrace{d(x, u) + d(y, v)}_{< d(x, y) + d(u, v)} - \underbrace{(d(x, y) + d(u, v))}_{> d(x, u) + d(y, v)}$$

wenn $A \mid B$ kein Split von $T \Rightarrow l(A|B) < 0$

wenn $A \mid B$ Split von $T \Rightarrow l(A|B) = l(e)$

$$l(A|B) = \frac{1}{2} \min_{\substack{x, y \in A \\ u, v \in B}} d(x, u) + d(y, v) - d(x, y) - d(u, v)$$

für je 4 Blätter betrachte:

$$d(x, u) + d(y, v) - d(x, y) - d(u, v)$$

$$d(x, v) + d(y, u) - d(x, y) - d(u, v)$$

$$d(x, y) + d(u, v) - d(x, u) - d(y, v)$$

2 dieser 3 Terme sind gleich, die 3te ist nicht größer. (da auch 0 zulässig)

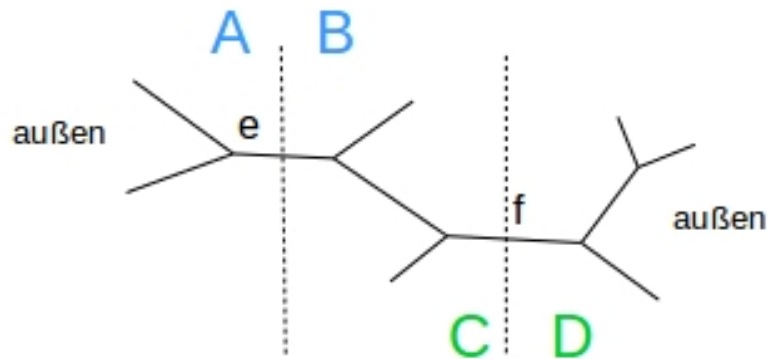
⇒ wenn das für alle $(x, y), (u, v)$ gilt, dann ist Metrik aus Baum abgeleitet

9.1.4 Splitsystem eines Baums

$\sum(T) = \{ \text{splits in } T, \text{ also die, die zu Kanten gehören.} \}$

1) $\{x\} \mid X \setminus \{x\} \in \sum(T)$ für alle $x \in X \rightarrow$ triviale Splits

zwei verschiedene Splitvarianten:



- e und f werden entfernt, es gibt genau eine Pfad zwischen den beiden Splitvarianten
- Weg von A nach D geht genau durch "zwischen"-Pfad $\Rightarrow A \cap B = \emptyset$

$\sum(T)$ heißt KOMPATIBEL, wenn für zwei Splits $A \mid B$ und $C \mid D$ gilt: mindestens ein Durchschnitt $A \cap C$, $A \cap D$, $B \cap C$, $B \cap D$ ist leer. (Trivialsplits gehören auch dazu)

Satz: Wenn \sum kompatibel, gibt es einen Baum T , sodass $\sum = \sum(T)$

10 Vorlesung 16.12.2016 (Spezialvorlesung 2)

10.1 Metriken

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 ≤ 0 wenn $A|B$ keine Teilbäume bestimmt.

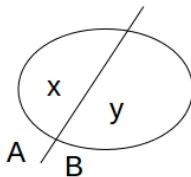
Wenn d eine additive Distanzfunktion:

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

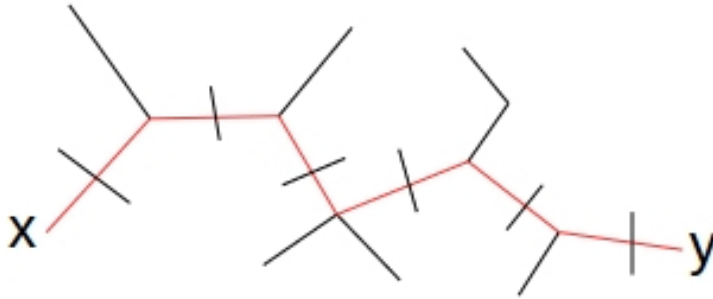
Splitpseudometrik:

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

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



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



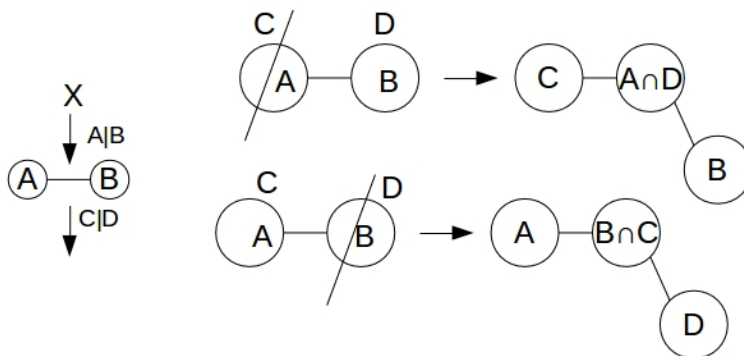
Genau die Splits entlang des Pfades von x und y trennen x,y

Splits $\Sigma(T) \rightarrow$ Baum

wir wissen $\Sigma(T)$ ist kompatible

$A|B, C|D \in \Sigma(T)$ 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 ohne alle Möglichkeiten durchzuprobieren, erzeugt werden? (Split-Decomposition⁷)

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' :

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

⁷Bandelt, H.-J. & Dress, A. W. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. Molecular phylogenetics and evolution 1, 242-252 (1992)

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

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

erster Fall:

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

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 ≤ 0

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

Wenn d additiv \Rightarrow Baum $\Rightarrow \text{splits}\Sigma(T)$ 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

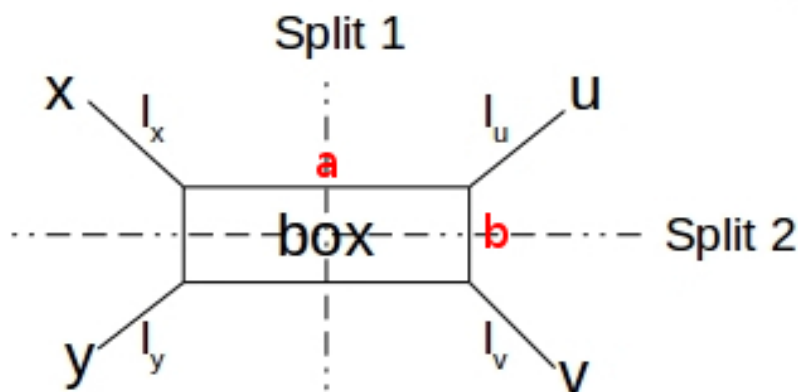
alle $\alpha(A|B) > 0$ können in $\mathcal{O}(|x|^6)$ Elemente berechnet werden.

- 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





$$\begin{aligned}
 d_{xu} + d_{xy} - d_{uy} &= 2l_x \\
 (l_x + a + l_u) + (l_x + b + l_y) - l_u - a - b - l_y &= 2l_x \\
 \Rightarrow l_x &= \frac{1}{2} [\underbrace{d_{xu} + d_{xy} - d_{uy}}_{\geq 0 \text{ (Dreiecksungleichung)}}]
 \end{aligned}$$

Split 1:

$$\begin{aligned}
 d_{xv} + d_{yu} - (d_{xy} + d_{uv}) &= \\
 l_x + a + b + l_v &+ \\
 + l_y + a + b + l_u &- \\
 - l_x - b - l_y &- \\
 - l_u - b - l_v &= 2a
 \end{aligned}$$

Split 2:

$$\begin{aligned}
 d_{xu} + d_{yv} - (d_{xy} + d_{uv}) &= \\
 l_x + a + l_u &+ \\
 + l_y + a + l_v &- \\
 - l_x - b - l_y &- \\
 - l_y - b - l_u &= 2(a - b) \leq 2a
 \end{aligned}$$

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

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

Baum $\Rightarrow b=0$

10.1.1 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

$B \approx 0 \rightarrow$ Baumartig

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

10.1.2 Travelling sales person problem (TSP)

geschlossene Tour Voraussetzung

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

Metrik d auf X gegeben

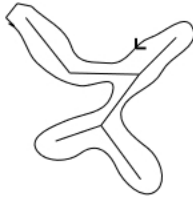
Tour: Permutation von $X: \pi$

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

Definition Mastertour:

Einschränkung von π 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.



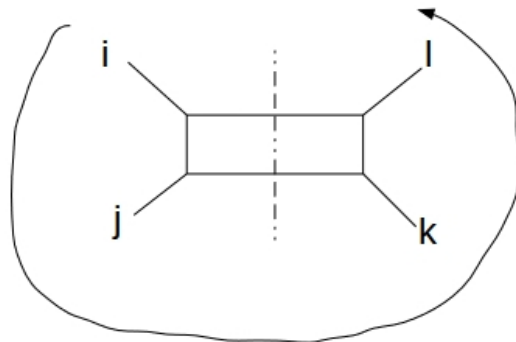
Eine Metrik hat die KALMANSON-Eigenschaft, wenn man X so ordnen kann, daß

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

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)

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

Anteil der Distanz ohne phylogenetische Information:

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

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

$$d_{\text{trivial}}(x, y) = \sum_{u \in X} \alpha(\{u\} | X \setminus \{u\}) \cdot \delta_{\{u\} | X \setminus \{u\}}(x, y)$$

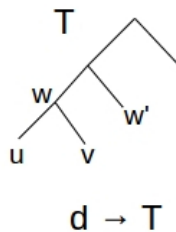
11 Vorlesung 21.12.2016 (Spezialvorlesung 3)

11.1 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 $\{u, v\} \rightarrow w^R$ (neuer Knoten)
3. berechne d_{wz} für $z \neq u, v \rightarrow$ Schritt 1



\tilde{d} Transformation von d

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

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

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

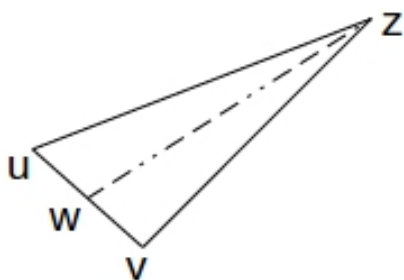
Falls d ein additive Baum-Metrik mit Baum \hat{T} ist, dann ist $\mathcal{A}(d) = \hat{T}$

Beispiel:

$\tilde{d} = d$ (keine Transformation)

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

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



Ist der zugehörige Alogrithmus 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 $\{u, v\} = \underset{x, y}{\operatorname{argmin}} \tilde{d}_{xy} = \{u, v\} \Rightarrow u, v$ 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)⁸]

⁸<http://mbe.oxfordjournals.org/content/23/11/1997.long>

11.2 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

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

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

mit m = Anzahl Cluster (NI-Formale für Cluster)

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

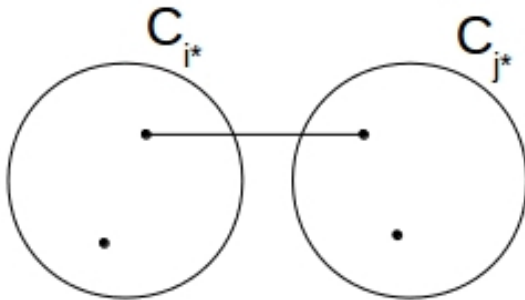
C_i, C_j enthält jeweils entweder 1 oder 2 Knoten

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

letzter Schritt im Neighbor Net Algorithmus:

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

Theorem: Wenn d Kalmanson Eigenschaften hat

⇒ Neighbor Net erzeugt die zugehörige zirkuläre Ordnung und identifiziert damit alle Splits mit nichtnegativen $\beta_{A|B}$