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 und y sind die Endpunkte von e

Knoten-Labelfunktion α: $\alpha : V \to L_V : v \mapsto \alpha(v)$
Kanten-Labelfunktion β: $\beta : E \to L_E : e \mapsto \beta(e)$

ungerichtete Graphen

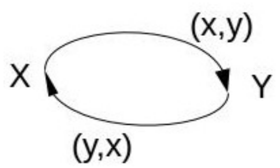
- Kante ist eine Menge von 2 (verschiedenen) Knoten
- $e = \{x, y\} = \{y, x\} \to$ Reihenfolge egal
- $E \subseteq V^{(2)} \to$ Kante ist Teilmenge von 2 Knoten

gerichtete Graphen

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

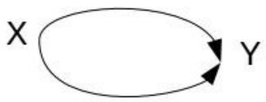
Funktionen gerichteter Graphen:
$h : E \to V : e \mapsto head(e)$
$t : E \to V : e \mapsto tail(e)$

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



trotzdem einfacher Graph!

erst:



ist Multigraph

Loops:

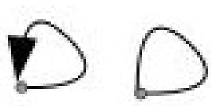


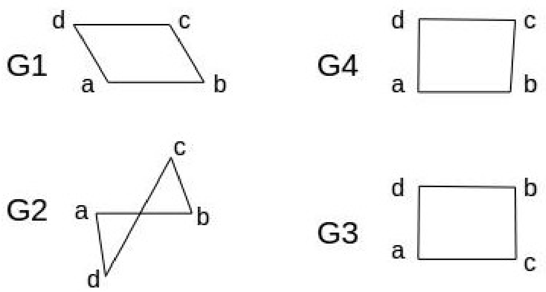
Abbildung 1: links: gerichtet; rechts: ungerichtet

⇒ 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

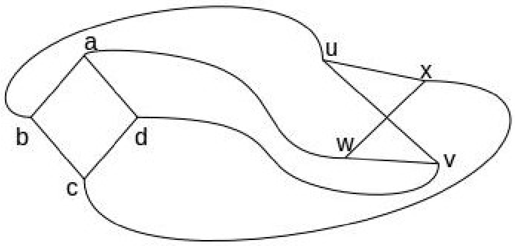
1.2 Gleichheit von Graphen



als labeled graphs: G_1=G_2=G_4 ≠G_3

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

[1] `https://de.wikipedia.org/wiki/Bijektive_Funktion`



bijektive Abbildung: jedes Element von 1. wird zu genau einem Element von 2. zugeordnet

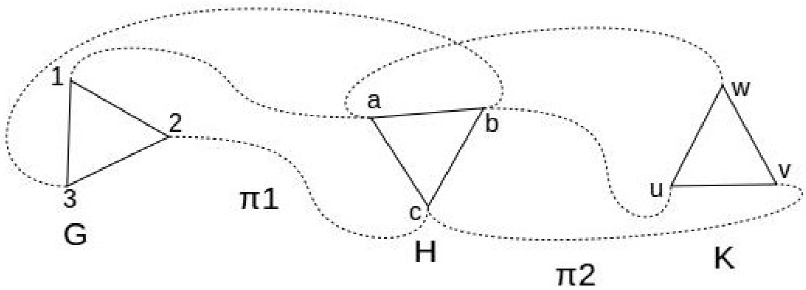
$\pi(a) = w, \pi(b) = u, \pi(c) = x, \pi(d) = v$
$\rightarrow$ hier ergibt bijektive Abbildung keinen Isomorpismus, da Bild(d) und Bild(c) Kante haben, jedoch v und x keine Kante haben

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 Isomorphien geben.

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

<u>Reflexivität:</u> Ein Graph ist zu sich selbst immer isomorph: $G \simeq G$
<u>Symmetrie:</u> $G \simeq H \Leftrightarrow H \simeq G$ <u>Transitivität:</u> $G \simeq H, H \simeq K \Rightarrow G \simeq K$



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

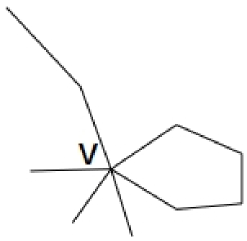
Nebenbemerkung: Labeled Graphen?
Zusätliche Bedingung benötigt: $\lambda(\pi(x)) = \lambda(x) \rightarrow$ Labels müssen erhalten bleiben!

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: $N(v) := \{y \in V | \{v, y\} \in E\}$
$deg(v) := |N(v)|$
$\delta(G) := \min_{v \in V} deg(v)$
$\Delta(G) := \max_{v \in V} deg(v)$

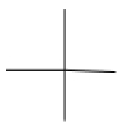
Def: Ein Graph heißt **REGULÄR** wenn $\Delta(G) = \delta(G)$
(wenn alle Knoten gleichen Grad haben)

Gradfolge von G:

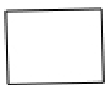
$\mathcal{F} = (n_0, n_1, n_2, \ldots, n_{|V|-1})$ mit $n_k := |\{x \in V | deg(x) = k\}|$

$\delta(G) \geq 0$
$\Delta(G) \leq |V| - 1$

Beispiel:



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



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

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

1.4 Graph-Invarianten

Eigenschaften, die unter Isomorphie erhalten bleiben
$\mathcal{G}\ldots$ Menge aller Graphen
F... ist ein Graph invariant wenn

$$F : \mathcal{G} \to X \tag{1}$$

die Eigenschaft hat, dass

$$G \simeq H \Rightarrow F(G) = F(H) \tag{2}$$

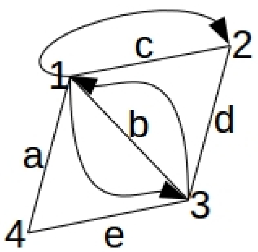
Invarianten bis jetzt: $|V|, |E|$, Gradfolge $\mathcal{F}$

Wenn F(G) $\neq$ F(H) für irgendeine Grapheninvariante $\Rightarrow$ G $\neg \simeq$ H

1.5 Pfade und Zusammenhänge

Kantenzug: Folge von Kanten in G"
$x_o, e_1, x_1, e_2, x_2, \ldots, e_l, x_l$ sodass $e_i := \{x_{i-1}, x_i\}$

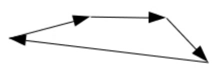
Beispiel:



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: $x_o \neq x_e$ 

- 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

4 Vorlesung 11.11.2016

5 Vorlesung 18.11.2016

6 Vorlesung 25.11.2016

7 Vorlesung 02.12.2016

8 Vorlesung 09.12.2016

9 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:
$d_{xy} + d_{uv} \leq d_{xu} + d_{yv} = d_{xv} + d_{yu} \Leftrightarrow$ ∀x,y,u,v 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.

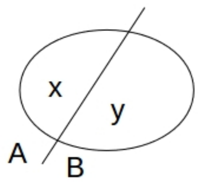
Wenn d eine additive Distanzfunktion:

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

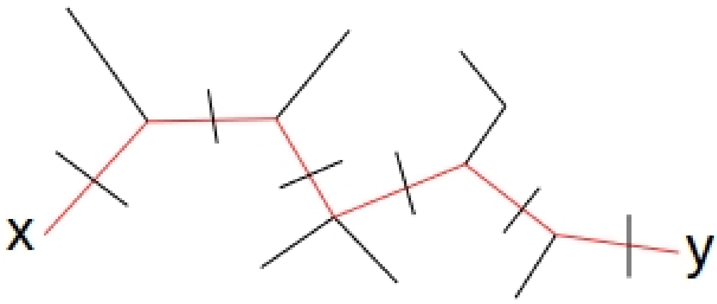
Splitpseudometrik:

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

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

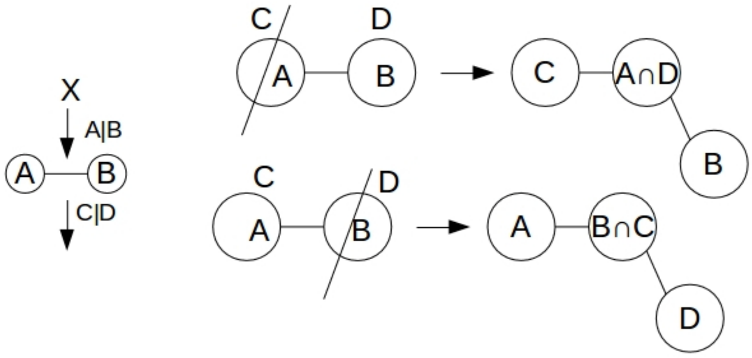


$$\boxed{d_T(x,y) = \sum_{(A|B)\in\Sigma(T)} \alpha(A|B)\cdot\delta_{A|B}(x,y)}$$



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 <u>GENAU</u> eine der Mengen

Frage: Wie können Isolationsindizes, schnell und und alle Möglichkeiten durchzuprobieren, 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' :
- $X|\{w\}$
- $A\cup\{w\}|B$
- $B\cup\{w\}|A$

$\beta_{xy|uv} := \frac{1}{2} max\{d_{xu} + d_{yv}, d_{xv} + d_{yu}\} - (d_{xy} + d_{uv})$
erster Fall:
$\alpha(\{x\}|X) = \min\limits_{u,v \in X} \beta_{ww|uv} = \min\limits_{u,v \in X} \frac{1}{2}(d_{wu} + d_{wv} - d_{uv})$
zweiter Fall:
$\alpha(A|B) = \min\limits_{\substack{x,y \in A \\ u,v \in B}} \beta_{xy|uv}$

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

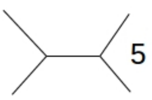
$\Rightarrow \alpha(A \cup \{w\}|B) \leq \alpha(A|B)$
Also: wenn $\alpha(A|B) \leq 0 \Rightarrow \alpha(A \cup \{w\}|B)$ auch ≤ 0
$\Rightarrow$ nur Splits auf X mit $\alpha(A|B) > 0$ müssen erwartet werden
Wenn d additiv $\Rightarrow$ Baum $\Rightarrow$ splits$\Sigma(T)$ kompatibel $\Rightarrow$ es gibt 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 $\mathcal{O}(|x|^3)$

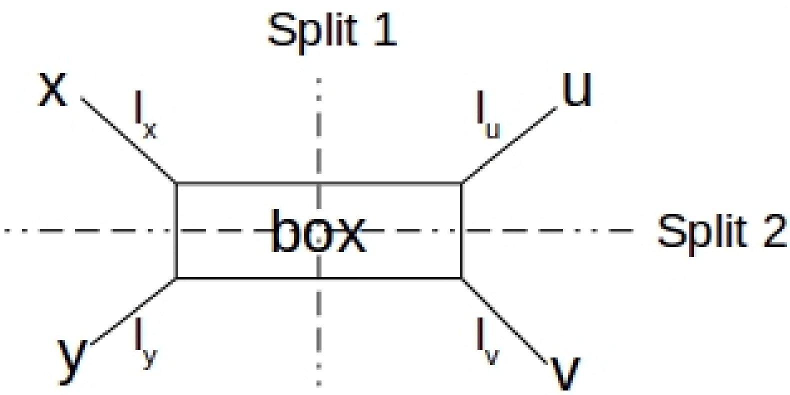
Theorem:[Bandelt,Dress]
Sei d eine Peusometrik auf X. Dann gibt es eine Pseudometrik d^0 auf X sodaß
$d(x,y) = \sum\limits_{A|B} \underbrace{\alpha(A|B)}_{*} \cdot \delta_{A|B}(x,y) + d^0(x,y)$
* $\alpha(A|B) = 0$ wenn $\min\limits_{\substack{x,y \in A \\ u,v \in B}} \beta_{xy|uv} < 0$

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

Baum mit 4 Blättern: 5 



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

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

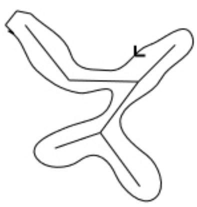
Messung der Baumartigkeit:
$B := \frac{1}{\binom{n}{4}} \sum_{\substack{i<j<k<l \\ i,j,k,l \in X}} \frac{b_{ijkl}}{a_{ijkl} + b_{ijkl}}$
Mittelwerte von in der Box
B $\approx$ Baumartig
B $\approx \frac{1}{2}$ völlig verrauscht, netzwerk-artig

Travelling sales person problem (TSP)
geschlossene Tour Voraussetzung
$|X| > 1$ (Anzahl der Städte größer 1)
Metrik d auf X gegeben
Tour: Permutation von X:π

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

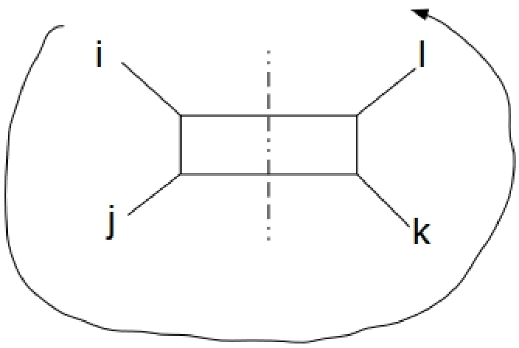
Definition Mastertour:
Einschränkung von π auf $X' \subseteq X$ löst das TSP auf X

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

$\rightarrow$ 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)
$\not\Leftarrow$ (Umkehr falsch)

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

Anteil der Distanz ohne phylogenetische Information:

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

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

10 Vorlesung 21.12.2016