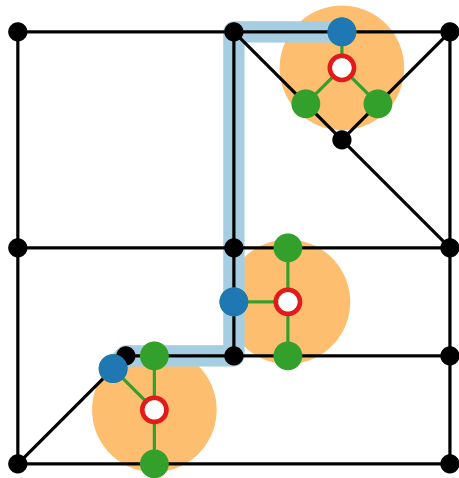
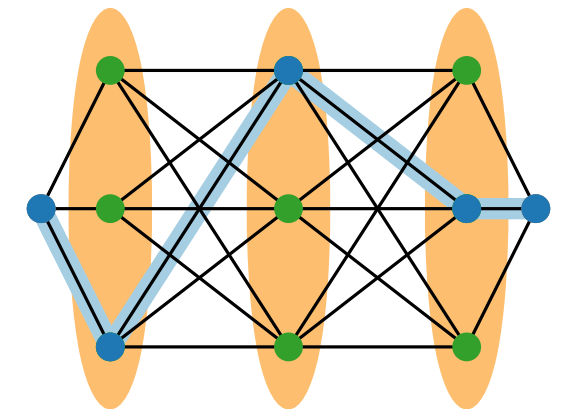


Algorithmen für geographische Informationssysteme

4. Vorlesung Map Matching Algorithmen

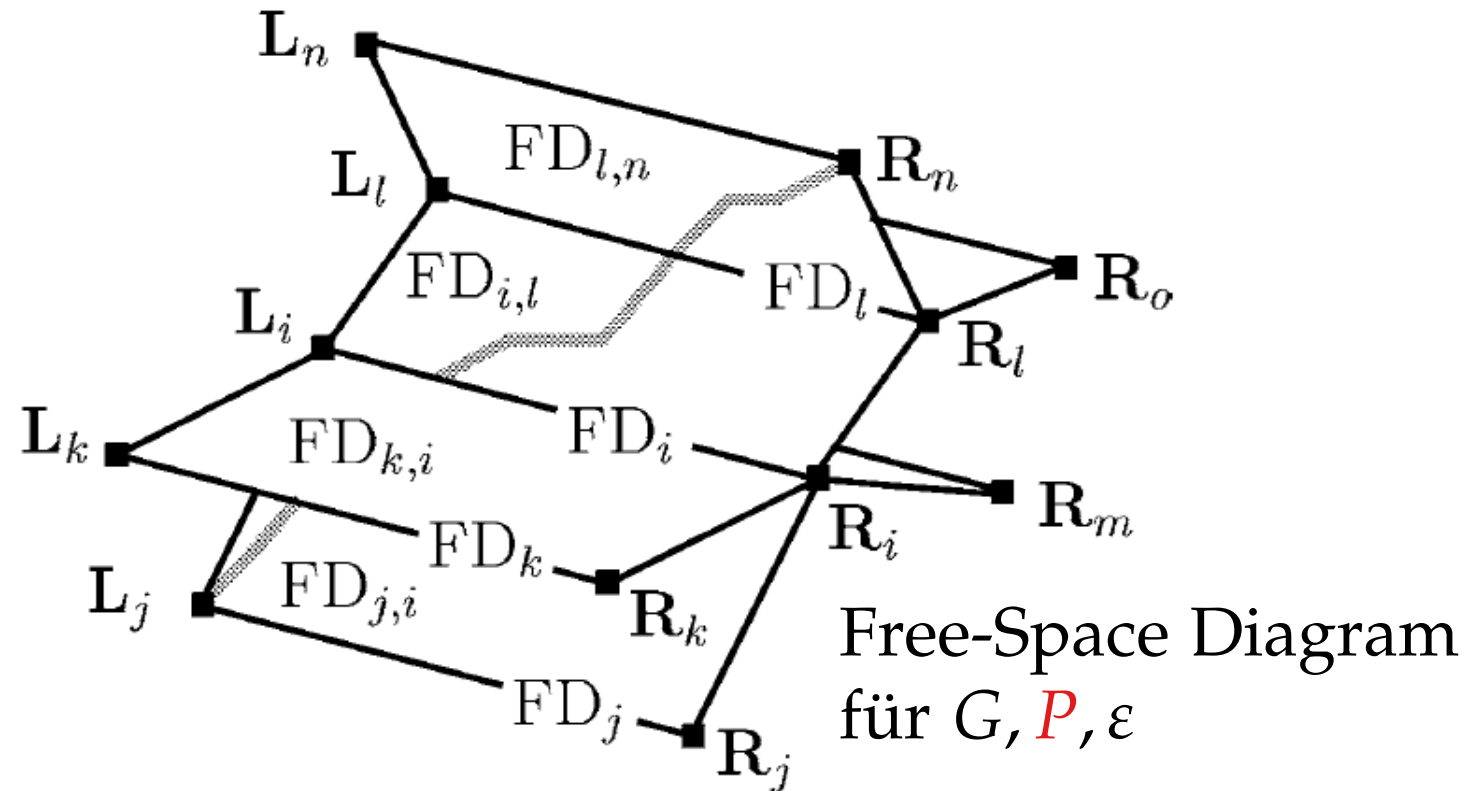


Teil I:
Map Matching mit
Fréchet Distanz

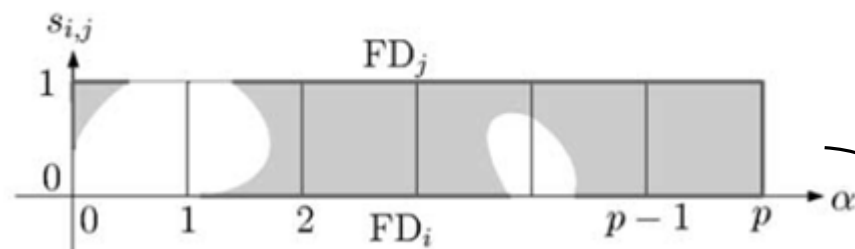


Optimierungsproblem:

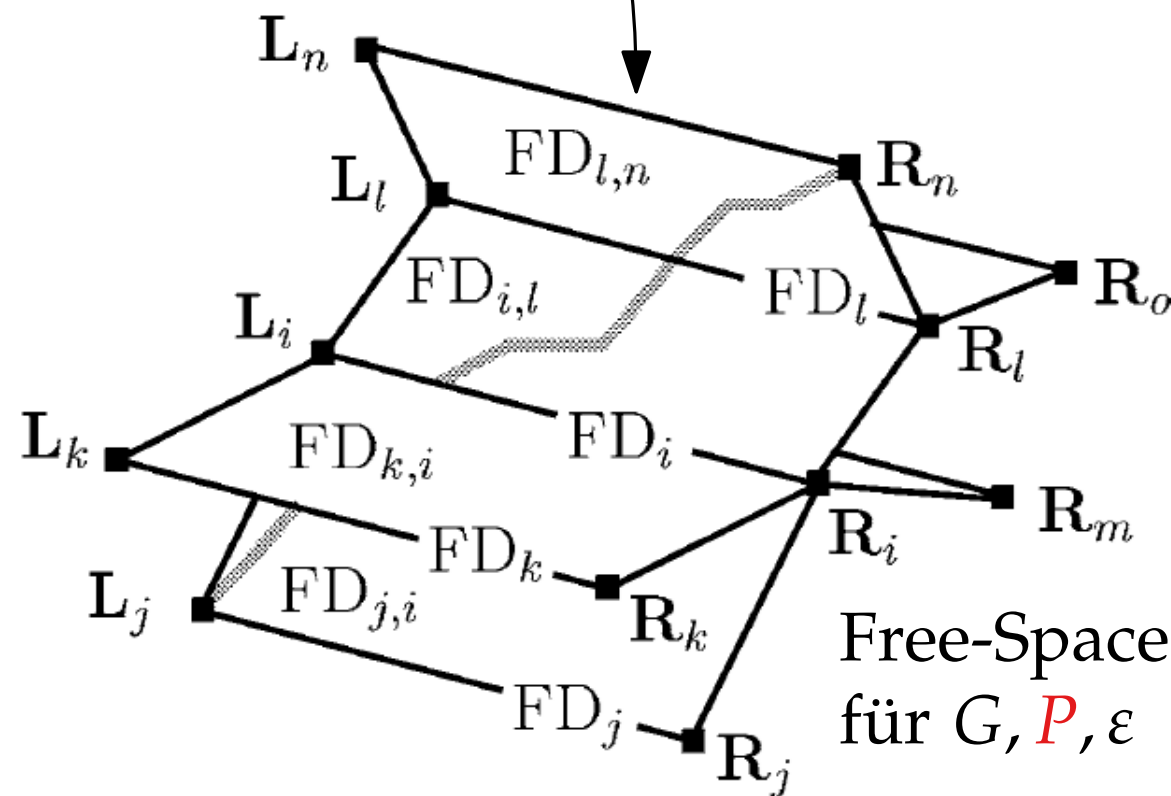
Entscheidungsproblem:



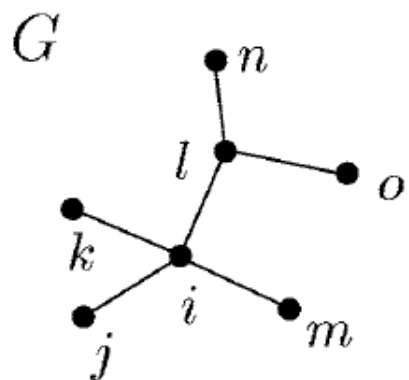
Map Matching mit Fréchet Distance



Free-Space Diagram für P, ε und eine Kante von G



Free-Space Diagram für G, P, ε





Map Matching mit Fréchet Distance

Gehe ähnlich wie DIJKSTRA vor:

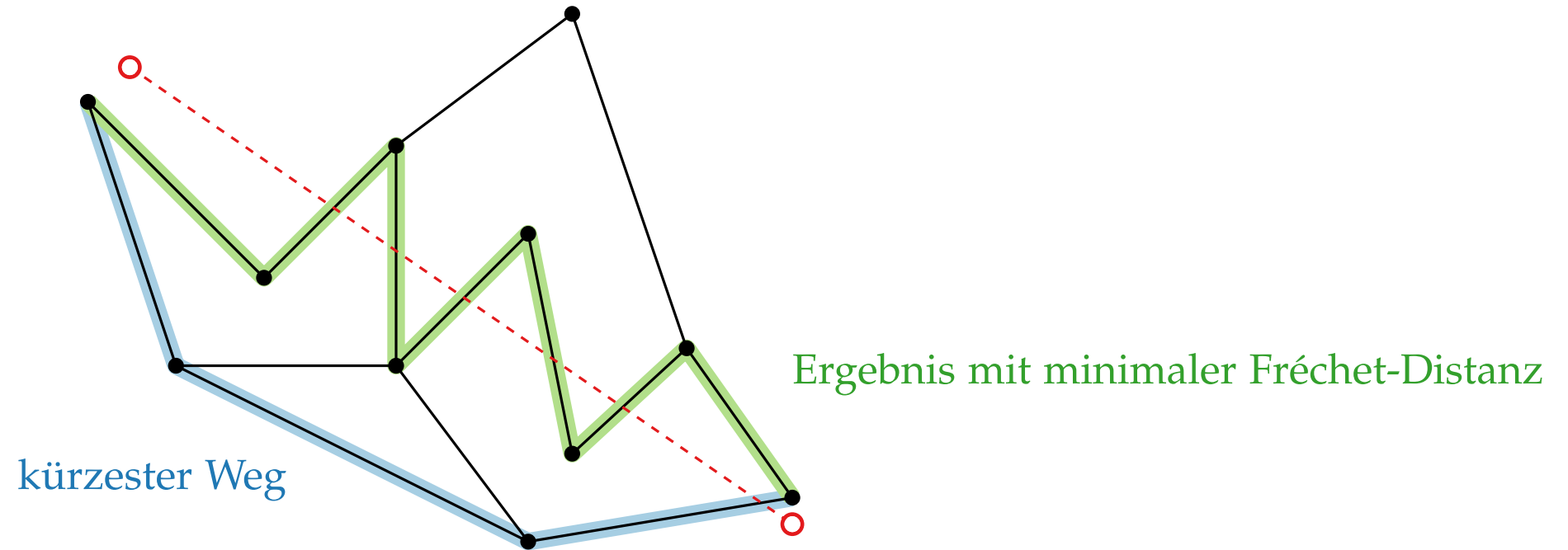
- Finde immer den ersten Punkt (von links nach rechts), der erreichbar ist.
- Dann betrachte anliegende Zellen und update ihren erreichbaren Punkte, wenn möglich.
- Laufzeit für Entscheidungsproblem: $\mathcal{O}(mp \log m)$
- Laufzeit für Optimiserungsproblem: $O(mp \log m \log mp)$

[Alt et al. 2003: Matching Planar Maps, J. Algorithm., 49:262–283]



Problem

GPS-Punkte der Trajektorie weisen häufig einen relativ großen Abstand zueinander auf.
z.B. [Lou et al.: Map-Matching for Low-Sampling-Rate GPS Trajectories, ACM GIS '09]:
Ein Punkt etwa alle zwei Minuten (z.B. zur Aufzeichnung & Analyse von Taxirouten)

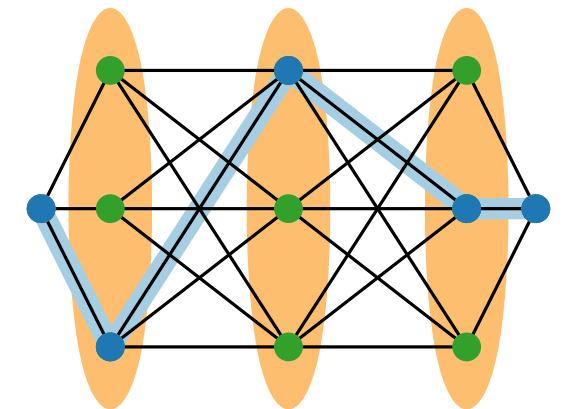
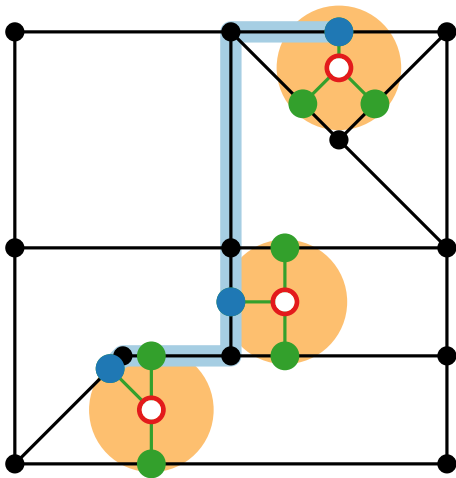


Fahrer wählen bevorzugt **kürzeste Wege** im Straßennetz.

Algorithmen für geographische Informationssysteme

4. Vorlesung Map Matching Algorithmen

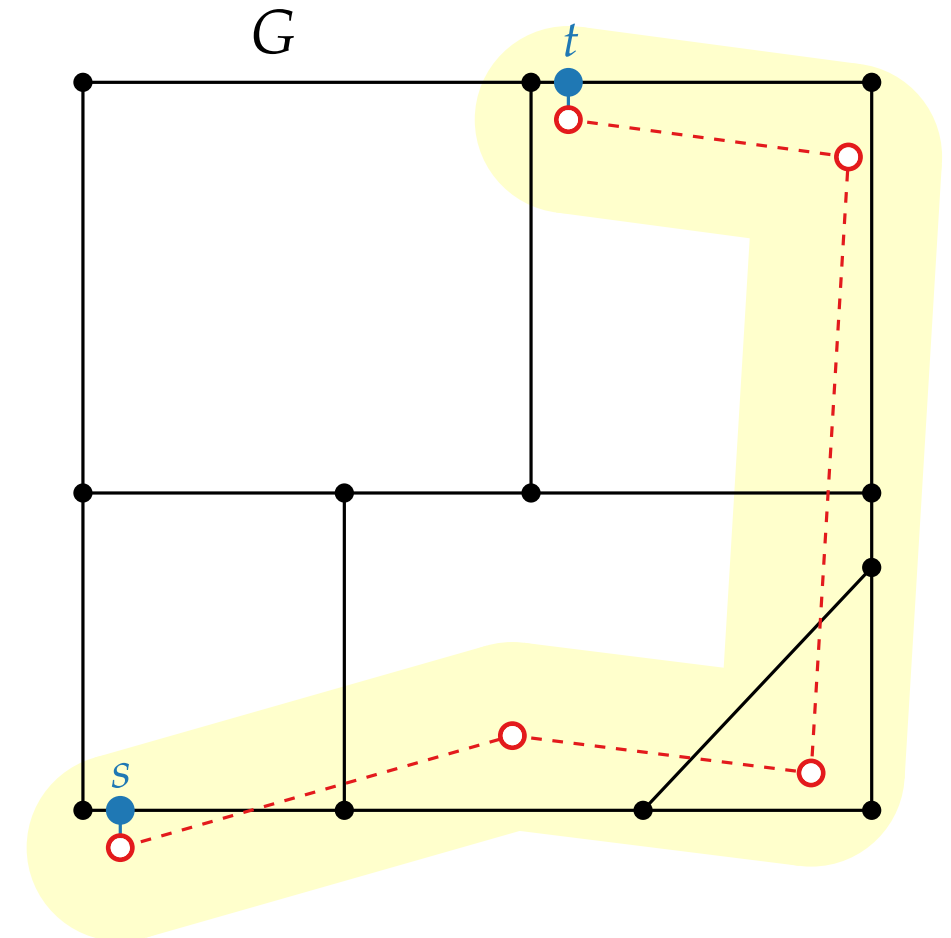
Teil II: Andere Ansätze



Dalumpines & Scott (2011)



- Erstelle Puffer der Größe ε (vom Benutzer festgelegt).
- Finde Punkte s und t in G mit min. Euklidischen Abstand zu Start und Ende der Trajektorie.
- Berechne Schnitt G' von G mit Puffer.

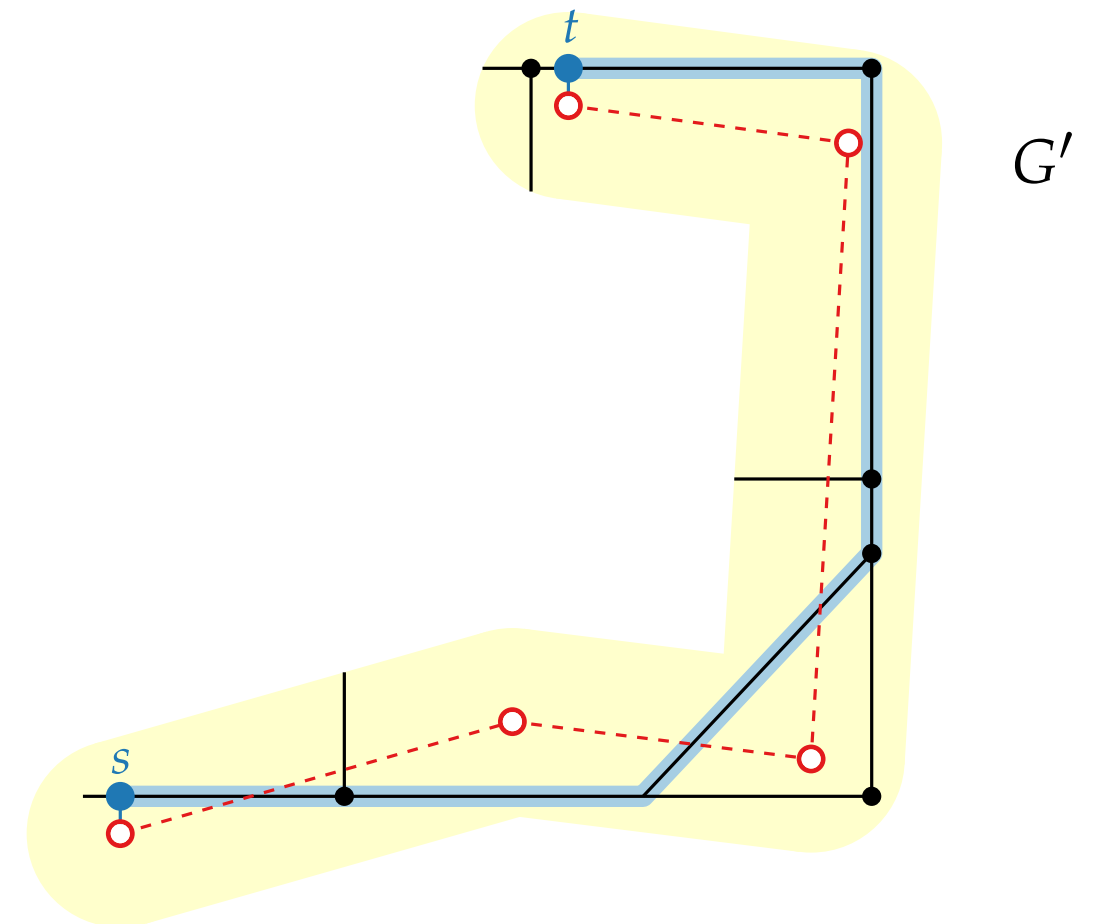




Dalumpines & Scott (2011)

- Erstelle Puffer der Größe ε (vom Benutzer festgelegt).
- Finde Punkte s und t in G mit min. Euklidischen Abstand zu Start und Ende der Trajektorie.
- Berechne Schnitt G' von G mit Puffer.
- Suche kürzesten s - t -Pfad in G' .

Problem: Algorithmus betrachtet nur Endpunkte!



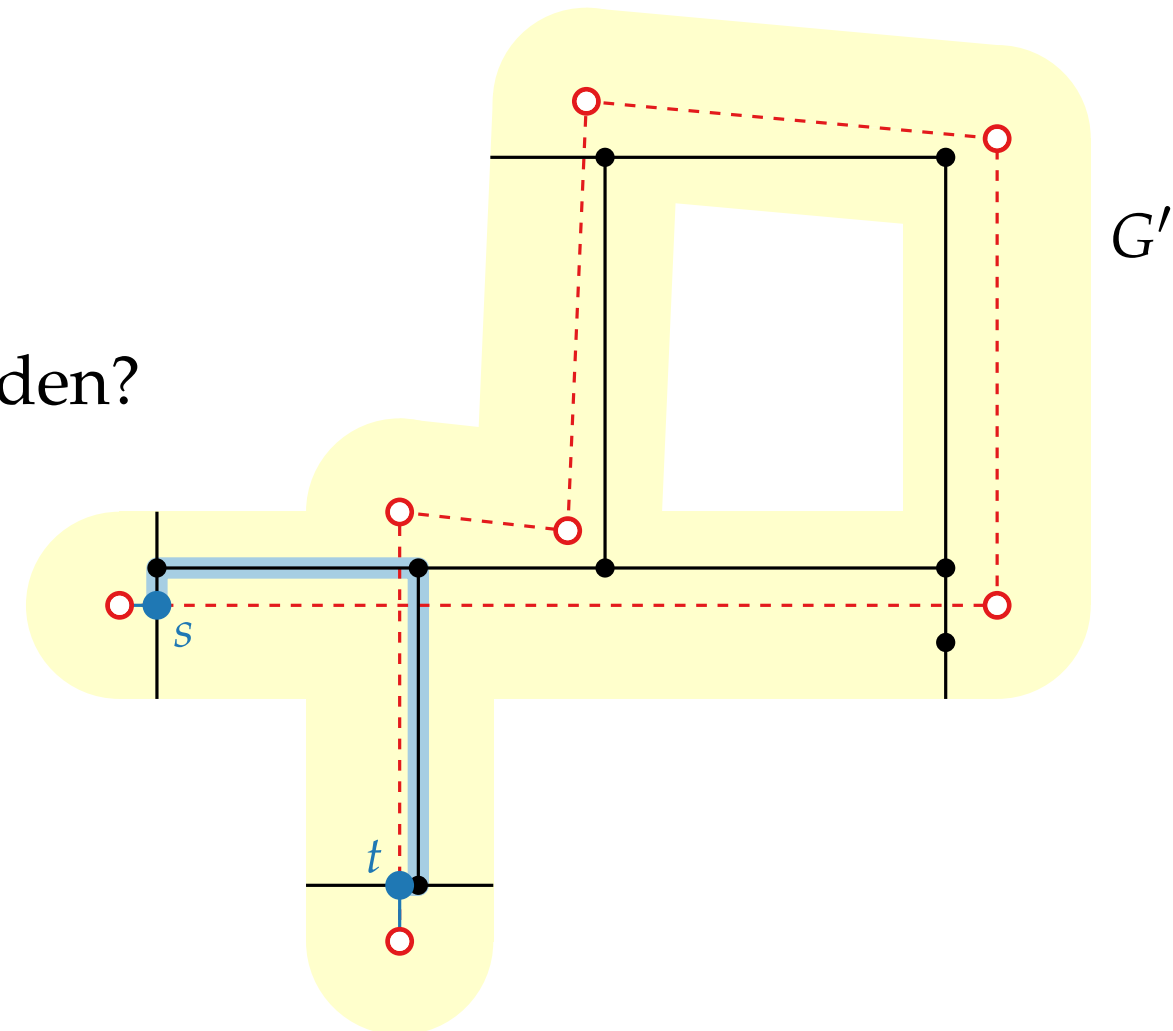
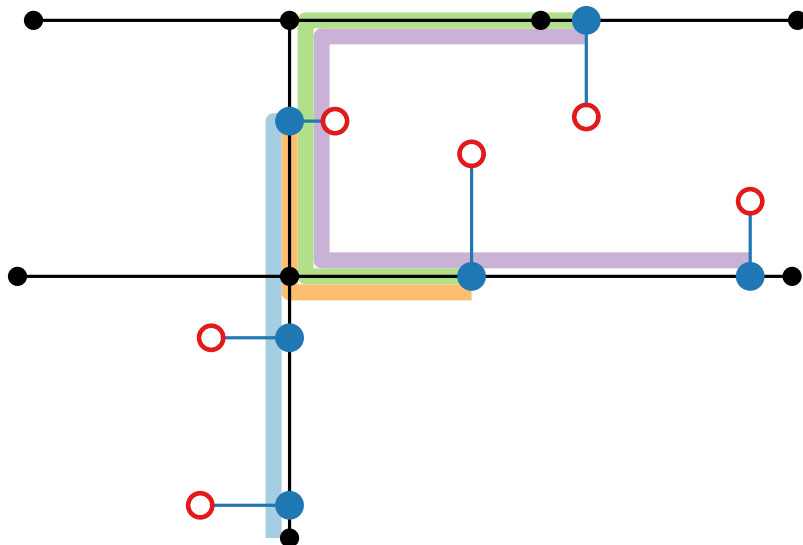


Dalumpines & Scott (2011)

- Erstelle Puffer der Größe ε (vom Benutzer festgelegt).
- Finde Punkte s und t in G mit min. Euklidischen Abstand zu Start und Ende der Trajektorie.
- Berechne Schnitt G' von G mit Puffer.
- Suche kürzesten s - t -Pfad in G' .

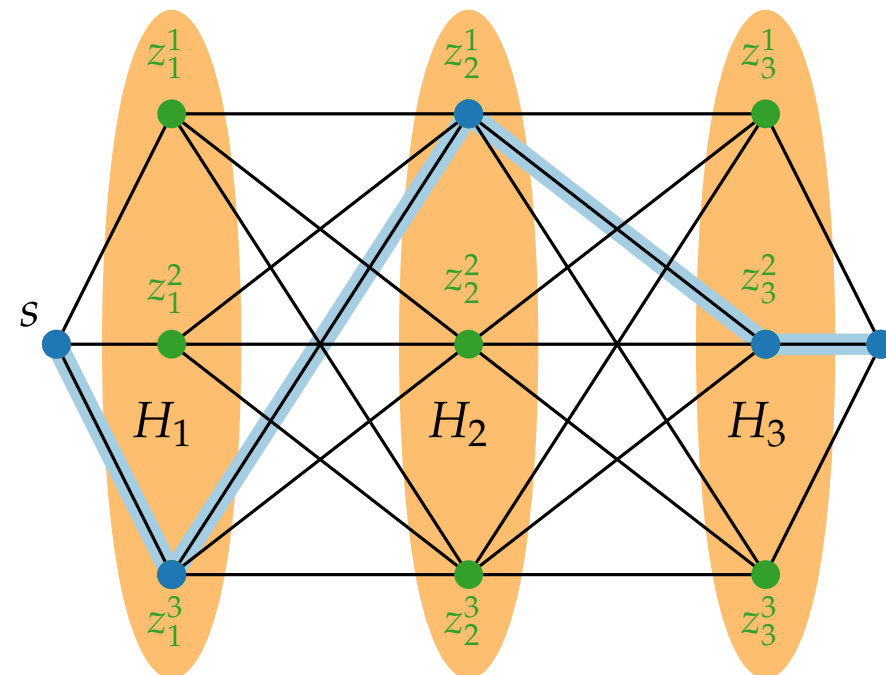
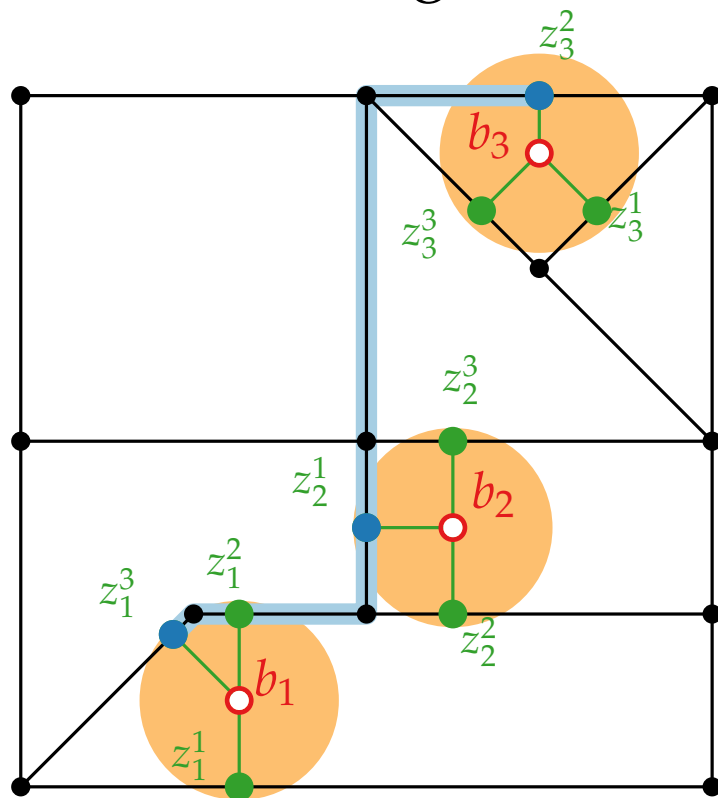
Problem: Algorithmus betrachtet nur Endpunkte!

Algorithmus zwischen jedem Punktepaar verwenden?



Newson & Krumm (2009)

Idee: Methode der größten Plausibilität + Markow-Kette + Verdecktes Markowmodell



- Vorgehen:**
- Berechne Kandidatenmenge für GPS Punkte
 - Berechne Trellis-Graph (Kanten entsprechen kürzesten Wegen in G , Kantengewichte basierend auf probabilistischem Modell)
 - Berechne s - t -Pfad mit größtem Gewicht
 - Rekonstruiere entsprechenden Graph in G

Newson & Krumm (2009)



Für jeden Punkt der Trajektorie gibt es zwei Zufallsvariablen:

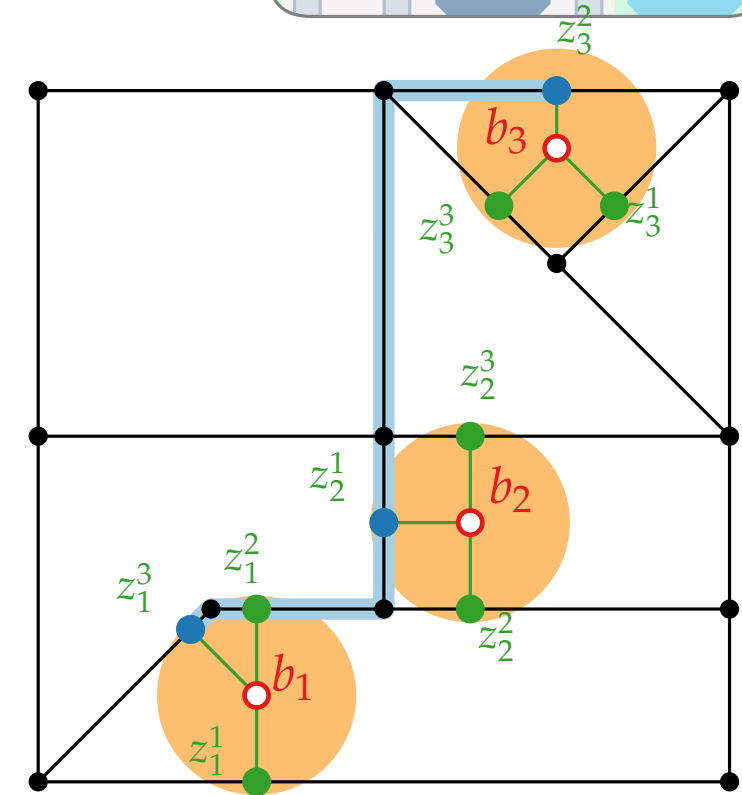
- B_i ist die beobachtete GPS Koordinate (Domäne $\mathbb{R}^2$)
- Z_i ist die Position im Straßennetz (diskrete Domäne $\{z_i^1, \dots, z_i^k\}$)
- Wahrscheinlichkeitsverteilung $P[B_i \mid Z_i]$ ist bekannt.

Gegeben Position im Straßennetz, was ist die Wahrscheinlichkeit der GPS Beobachtung?

Üblicherweise wird Normalverteilung um Position im Straßennetz angenommen.

- Wahrscheinlichkeitsverteilung $P[Z_i \mid Z_{i-1}]$ ist bekannt.

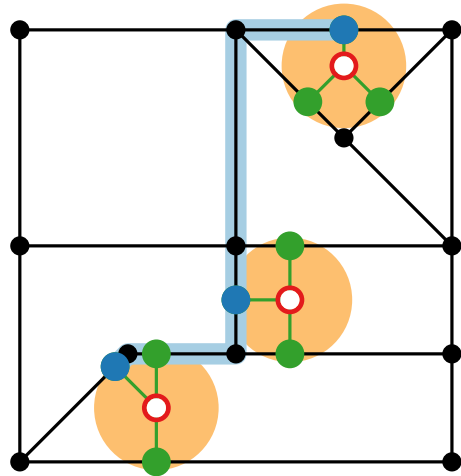
Gegeben Position im Straßennetz zu Zeitpunkt $i - 1$, was ist die Wahrscheinlichkeit der Position im Straßennetz zum Zeitpunkt i ?
Hängt insbesondere von Distanz zwischen den zwei Positionen im Straßennetz ab.



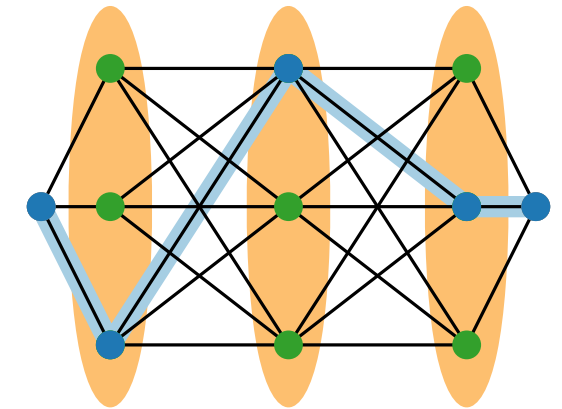
Frage: Welche Sequenz der Positionen im Straßennetzwerk ist am wahrscheinlichsten?

Algorithmen für geographische Informationssysteme

4. Vorlesung Map Matching Algorithmen



Teil III:
Methode der größten
Plausibilität



Methode der größten Plausibilität



Wahrscheinlichkeit vs. Plausibilität

- Wir werfen eine Münze.

- $P[\text{Kopf}] = p$
- $P[\text{Zahl}] = 1 - p$

Modell

- Wir werfen eine Münze 3 mal.

- Was ist die **Wahrscheinlichkeit**, das Ergebnis $x = (K, K, Z)$ zu erhalten? **Ergebnis?**

$$P[x(K, K, Z)] = p \cdot p \cdot (1 - p)$$

- Angenommen, die Münze ist fair ($p = \frac{1}{2}$)

Parameter

~~$$\rightarrow P[x(K, K, Z) | p = \frac{1}{2}] = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$~~

p ist keine Zufallsvariable!

$$\rightarrow P[x(K, K, Z); p = \frac{1}{2}] = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

Methode der größten Plausibilität



Wahrscheinlichkeit vs. **Plausibilität**

- Wir werfen eine Münze.
 - $P[\text{Kopf}] = p$
 - $P[\text{Zahl}] = 1 - p$
- Wir **haben** eine Münze 3 mal **geworfen**.
- Das Ergebnis war $x = (K, K, Z)$.
- Was ist die ~~Wahrscheinlichkeit~~ **Plausibilität**, dass die Münze fair ist ($p = \frac{1}{2}$)?

Modell

Ergebnis

Parameter?



Plausibilitätsfunktionen

- Gegeben ein Parameterwert θ , wie **wahrscheinlich** ist das Ergebnis x ?
- Gegeben ein Ergebnis x , wie **plausibel** ist ein bestimmter Parameterwert θ ?
- Definiere Plausibilitätsfunktion $L(\theta | x) = P[x; \theta]$
- Das Ergebnis war $x = (K, K, Z)$
- Wie **plausibel** ist es, dass die Münze fair ist ($p = \frac{1}{2}$)?

Keine Wahrscheinlichkeit!

$$\rightarrow L(p = \frac{1}{2} | x = (K, K, Z)) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = 0,125$$

$$\rightarrow L(p = 0.1 | x = (K, K, Z)) = 0,1 \cdot 0,1 \cdot 0,9 = 0,009$$

$$\rightarrow L(p = \frac{2}{3} | x = (K, K, Z)) = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} = 0,\overline{148}$$



Wahrscheinlichkeit vs. Plausibilität

Wahrscheinlichstes Ergebnis

- Gegeben Parameter $p = \frac{2}{3}$
- Das wahrscheinlichste Ergebnis ist (K, K, K)
 - ➔ $P[x = (K, K, K); p = \frac{2}{3}] = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} = 0,296$
 - ➔ $P[x = (K, K, Z); p = \frac{2}{3}] = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} = 0,148$

Größte Plausibilität

- Gegeben Ergebnis $x = (Z, K, Z)$
- Die größte Plausibilität ist $p = \frac{1}{3}$
 - ➔ Der Wert von p „explärt das Ergebnis am Besten“
 - ➔ Der Wert von p macht das Ergebnis „am wenigsten überraschend“

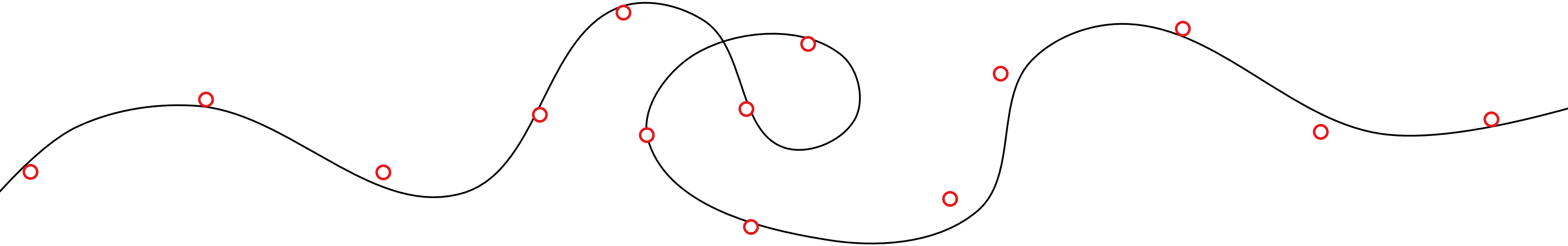
Zurück zu Karten

Modell?
Parameter?
Ergebnis?



Wahrscheinlichstes Ergebnis

- Gegeben ein Pfad im Straßennetz
- Welche GPS Trajektorie wird am wahrscheinlichsten beobachtet? (mit Ungenauigkeit?)

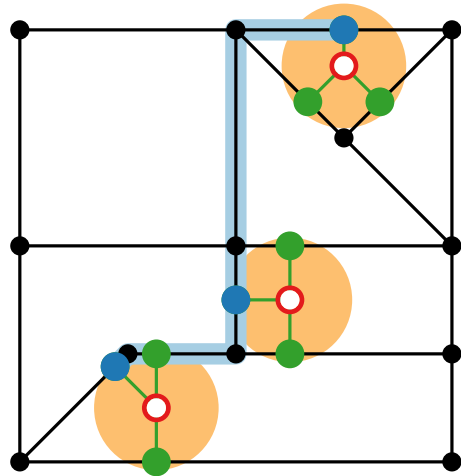


Größte Plausibilität

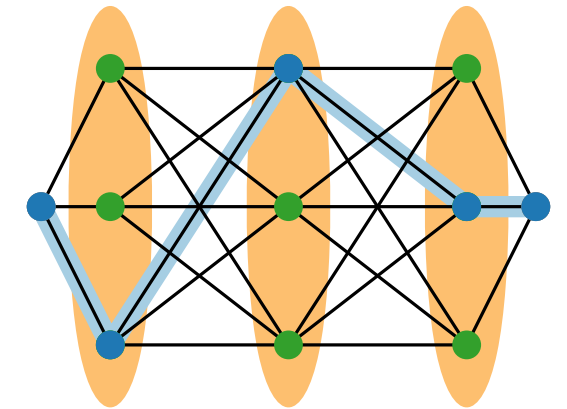
- Gegeben eine beobachtete GPS Trajektorie
- Welcher Pfad im Straßennetz ist am plausibelsten?
- „Welcher Pfad durchs Straßennetz erklärt am Besten die beobachtete Trajektorie?“

Algorithmen für geographische Informationssysteme

4. Vorlesung Map Matching Algorithmen



Teil IV:
Markow-Ketten



Markow-Ketten



Gegeben Sequenz von Zufallsvariablen $X_1, X_2, X_3, \dots$

Markow-Eigenschaft:

$$P[X_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_1 = x_1] \\ = P[X_{n+1} \mid X_n = x_n]$$

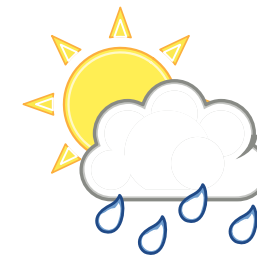
„Wenn man die Gegenwart kennt, ist die Zukunft unabhängig von der Vergangenheit.“

Wir schreiben $P[X_1]$, $P[X_2 \mid X_1]$, $P[X_3 \mid X_2]$, etc ...



Andrei Andrejewitsch Markow
*1856, Rjasan
†1922, Petrograd

Markow-Ketten – Beispiel



Zwei mögliche Zustände: $Z = \{\text{Sonne}, \text{Regen}\}$.

- $P[X_1 = \text{Regen}] = 0,3$

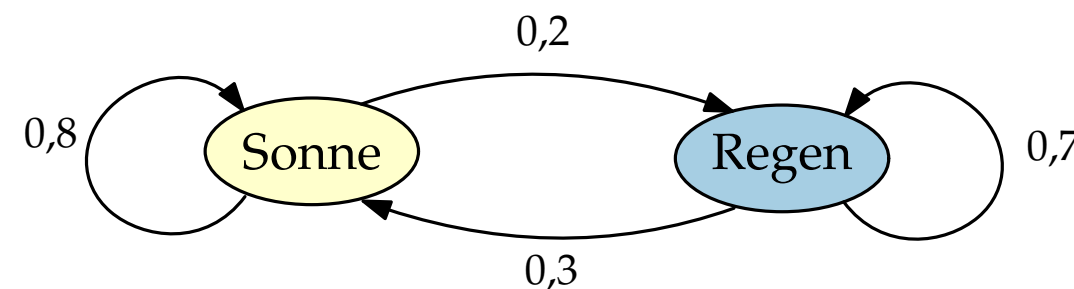
- $P[X_1 = \text{Sonne}] = 0,7$

- $P[X_{n+1} = \text{Regen} \mid X_n = \text{Regen}] = 0,7$

- $P[X_{n+1} = \text{Sonne} \mid X_n = \text{Regen}] = 0,3$

- $P[X_{n+1} = \text{Regen} \mid X_n = \text{Sonne}] = 0,2$

- $P[X_{n+1} = \text{Sonne} \mid X_n = \text{Sonne}] = 0,8$



- $P[X_2 = \text{Regen}] = 0,3 \cdot 0,7 + 0,7 \cdot 0,2 = 0,21 + 0,14 = 0,35$

- $P[X_2 = \text{Sonne}] = 0,65$

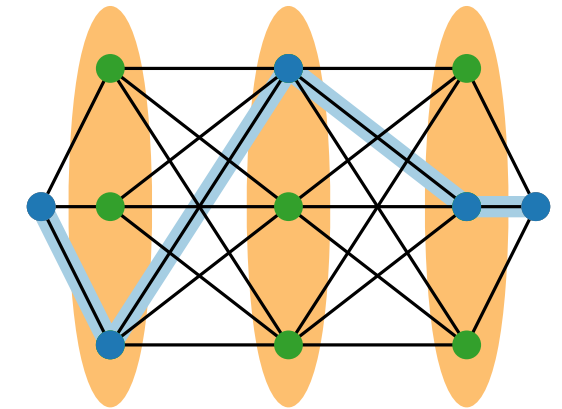
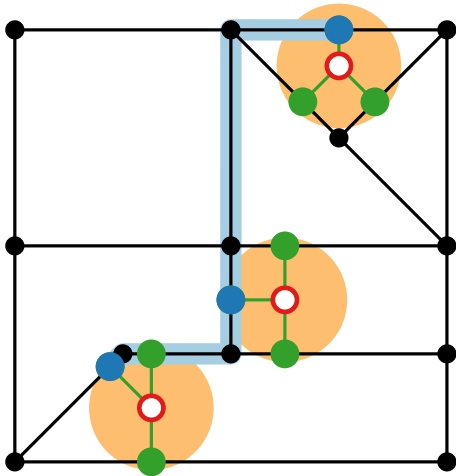
- $P[X_3 = \text{Regen}] = 0,375$

- $P[X_3 = \text{Sonne}] = 0,625$

Algorithmen für geographische Informationssysteme

4. Vorlesung Map Matching Algorithmen

Teil V: Verdecktes Markow-Modell





Verdecktes Markow-Modell

Modelliert ein System durch eine Markow-Kette

- Zu jedem Zeitpunkt befindet sich das System in einem bestimmten **Zustand**

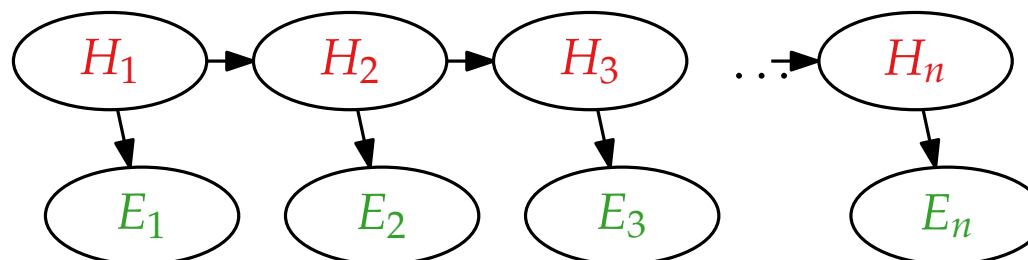
Aber: der Zustand ist **verdeckt**!

- Wir können den Zustand nicht beobachten.

Wir bekommen nur **Emissionen**

- Beobachtbare Ausgabesymbole
- Eine Emission pro Zeitpunkt
- Je nach Zustand unterschiedliche Wahrscheinlichkeiten
- Emissionen in jedem Schritt unabhängig

Ziel ist es, aus den Emissionen auf die plausibelsten verdeckten Zustände zu schließen.



Verdeckte Variablen (Zustände)

Beobachtete Variablen (Emissionen)



Verdecktes Markow-Modell

- Gegeben:**
- Beobachtungen $B = (b_1, \dots, b_n)$, je eine pro Zeitpunkt
 - Mögliche Systemzustände Z
 - Für jeden Systemzustand $z \in Z$ und jede Beobachtung o
 - die Wahrscheinlichkeit $P[b \mid z]$ oder Falls diskreter Zustandsraum
 - die Wahrscheinlichkeitsdichtefunktion $f[b \mid z]$, Falls kontinuierlicher Zustandsraum
 - dass b bei Zustand z beobachtet wird
 - Für jedes Paar von Zuständen $z_1, z_2 \in Z$ die Wahrscheinlichkeit $P[z_2 \mid z_1]$ (**Übergangswahrscheinlichkeit**)
 - Für jeden Zustand $z \in Z$ die (a-priori-)Wahrscheinlichkeit $P[z]$

Gesucht: Folge $Z = (z_1, \dots, z_n)$ von Zuständen, die sich am Besten mit den gegebenen Beobachtungen erklären lässt, also $P[Z \mid B]$ maximiert.

Satz von Bayes

konstant

$$P[Z \mid B] = P[B \mid Z] \cdot P[Z] / P[B]$$

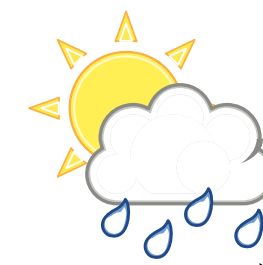
$$P[B \mid Z] = P[b_1 \mid z_1] \cdot \dots \cdot P[b_n \mid z_n]$$

$$P[Z] = P[z_1] \cdot P[z_2 \mid z_1] \cdot \dots \cdot P[z_n \mid z_{n-1}]$$



Thomas Bayes
*1701, London
†1761, Tunbridge Wells

Verdecktes Markow-Modell – Beispiel



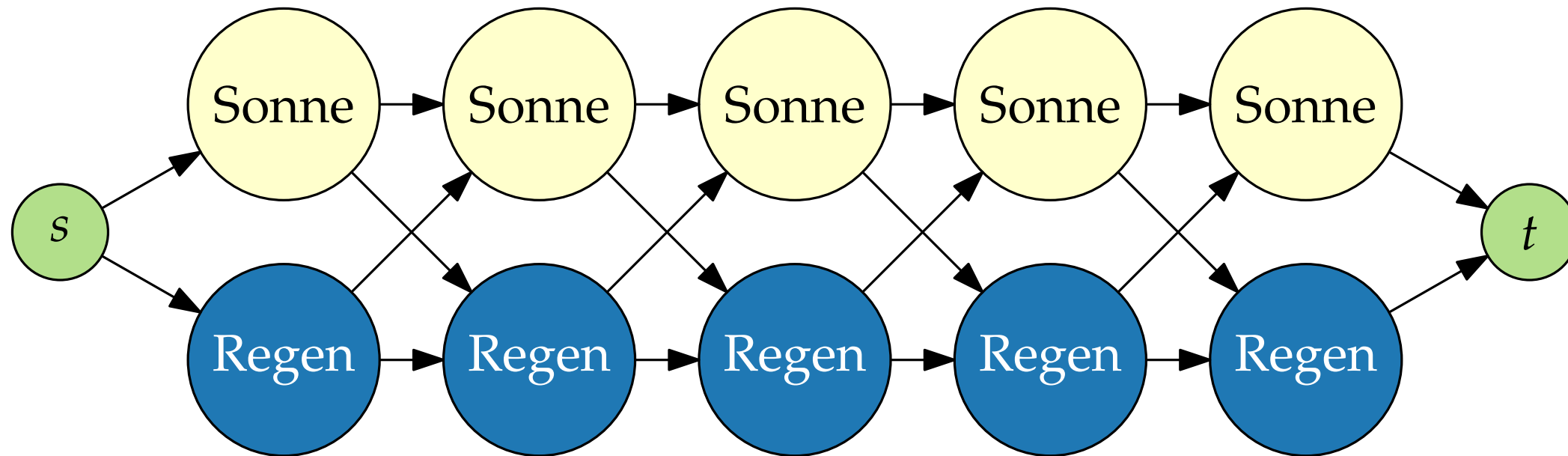
Gegeben:

- Beobachtungen $B = (\text{nass}, \text{nass}, \text{trocken}, \text{nass}, \text{trocken})$
- Mögliche Systemzustände $Z = \{\text{Sonne}, \text{Regen}\}$
- Beobachtungswahrscheinlichkeiten:
 - $P[\text{nass} \mid \text{Sonne}] = 0,1$
 - $P[\text{trocken} \mid \text{Sonne}] = 0,9$
 - $P[\text{nass} \mid \text{Regen}] = 0,95$
 - $P[\text{trocken} \mid \text{Regen}] = 0,05$
- Übergangswahrscheinlichkeiten:
 - $P[\text{Sonne} \mid \text{Sonne}] = 0,8$
 - $P[\text{Regen} \mid \text{Sonne}] = 0,2$
 - $P[\text{Sonne} \mid \text{Regen}] = 0,3$
 - $P[\text{Regen} \mid \text{Regen}] = 0,7$
- A-priori-Wahrscheinlichkeiten:
 - $P[\text{Sonne}] = 0,7$
 - $P[\text{Regen}] = 0,3$

Gesucht: Folge $Z = (z_1, \dots, z_n)$ von Zuständen, so dass $P[Z \mid B]$ maximal ist.

Verdecktes Markow-Modell – Beispiel

■ Beobachtungen $B = (\text{nass}, \text{nass}, \text{trocken}, \text{nass}, \text{trocken})$

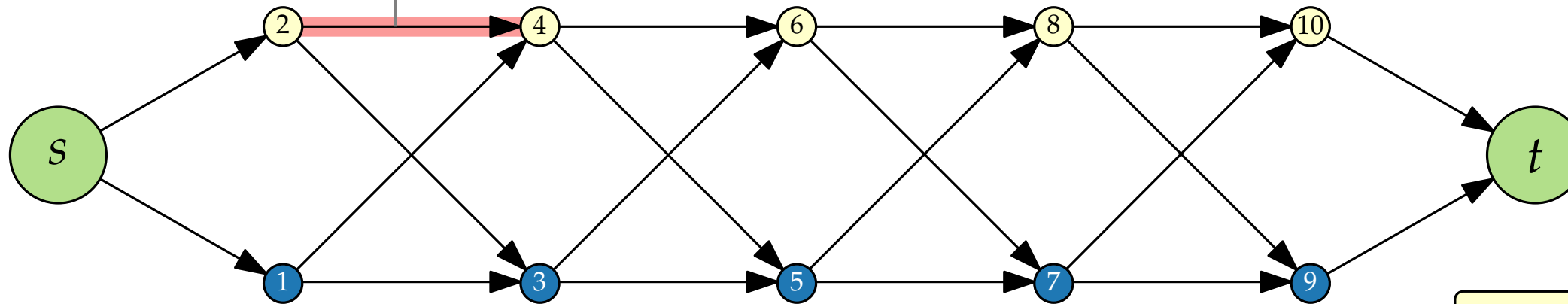




Verdecktes Markow-Modell – Beispiel

- Beobachtungen $B = (\text{nass}, \text{nass}, \text{trocken}, \text{nass}, \text{trocken})$

$$P[B_1 = \text{nass} \mid Z_1 = \text{Sonne}] \cdot P[Z_2 = \text{Sonne} \mid Z_1 = \text{Sonne}]$$

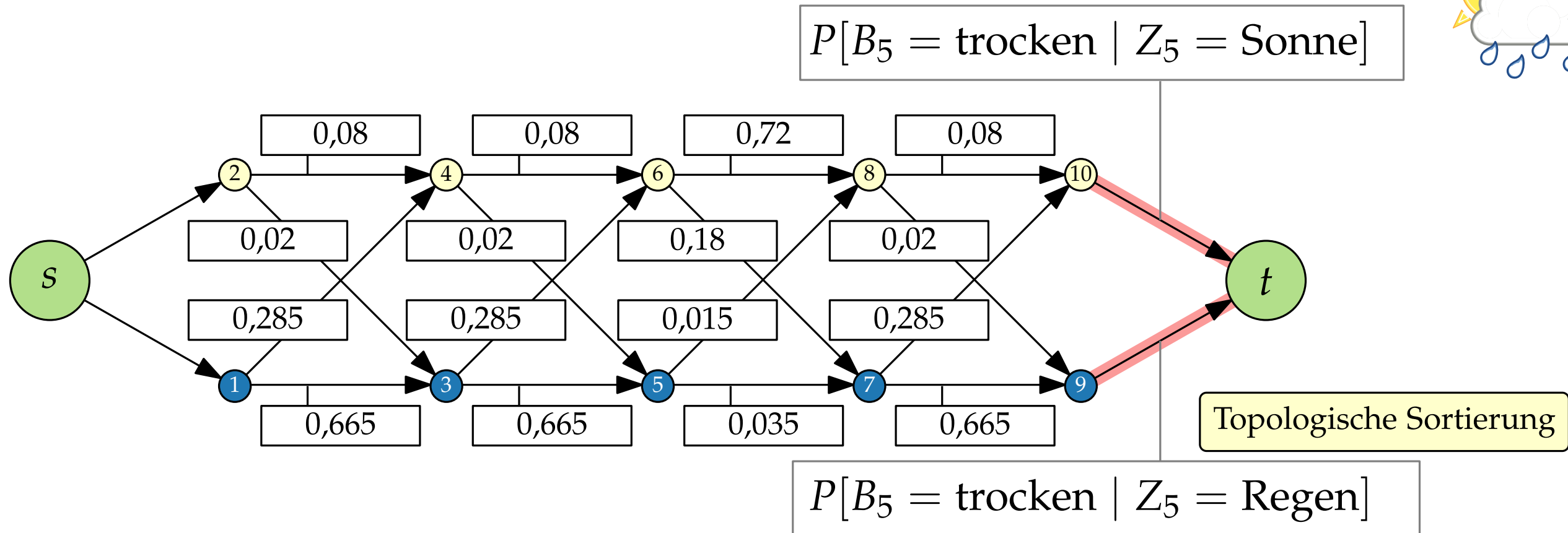


Topologische Sortierung

Verdecktes Markow-Modell – Beispiel

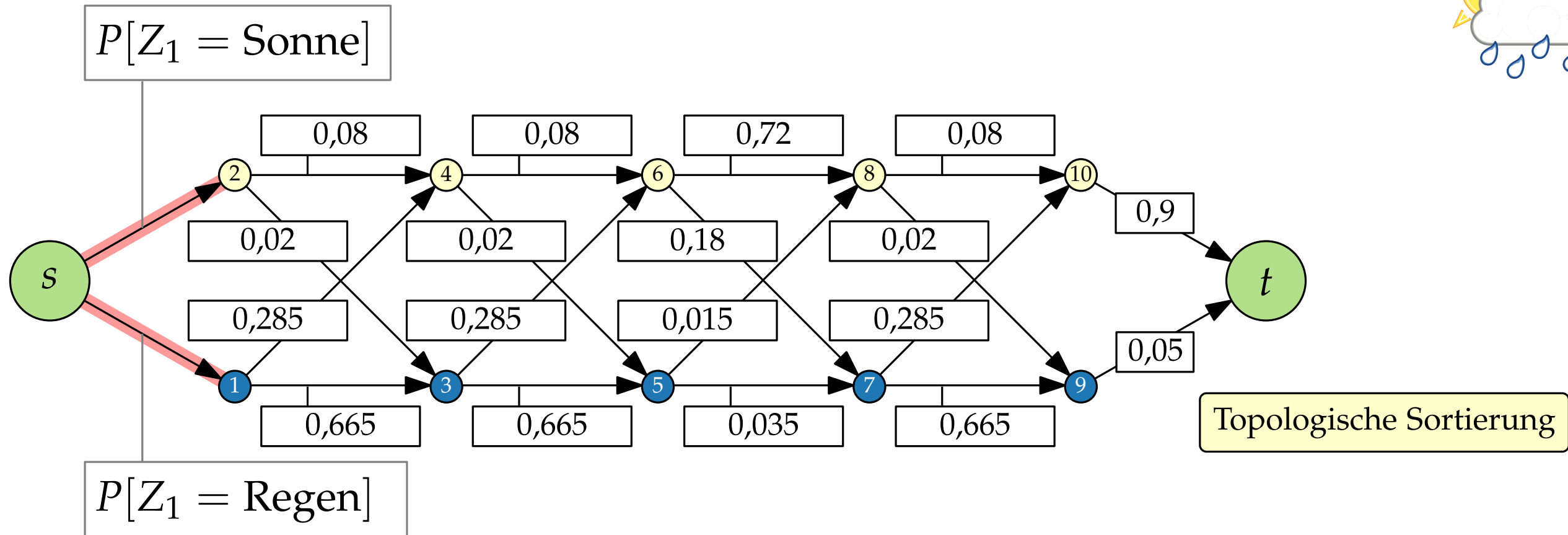


■ Beobachtungen $B = (\text{nass}, \text{nass}, \text{trocken}, \text{nass}, \text{trocken})$



Verdecktes Markow-Modell – Beispiel

■ Beobachtungen $B = (\text{nass}, \text{nass}, \text{trocken}, \text{nass}, \text{trocken})$



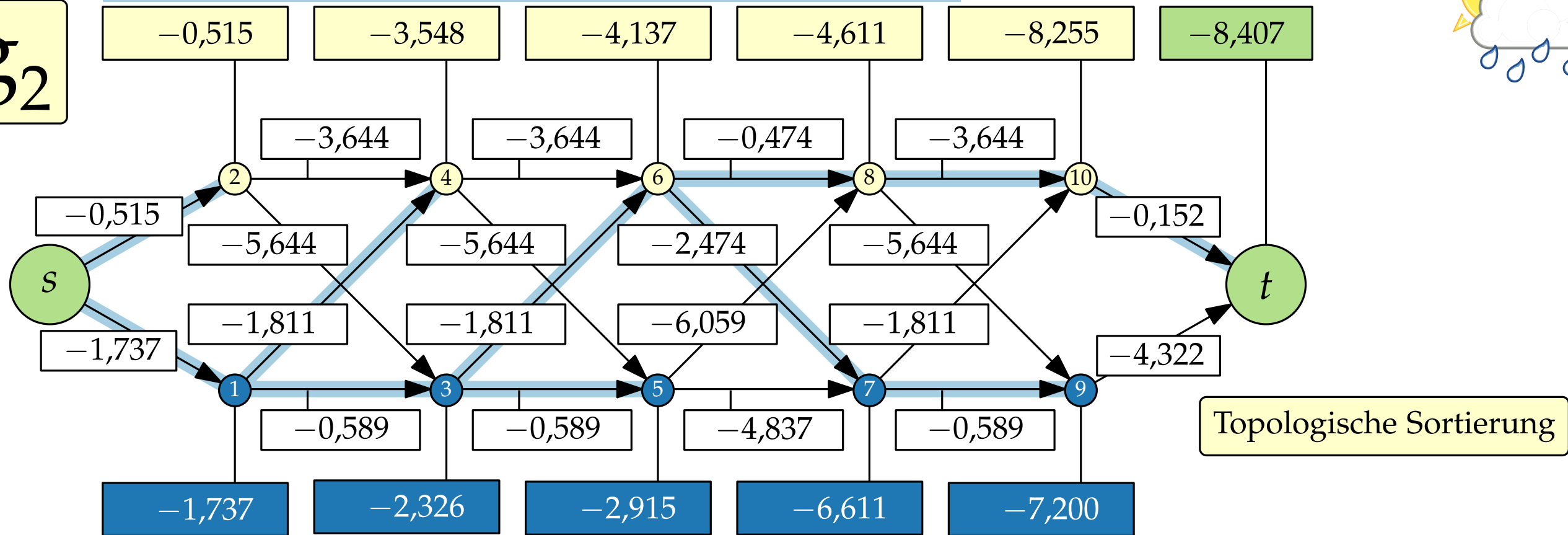


Verdecktes Markow-Modell – Beispiel

■ Beobachtungen $B = (\text{nass}, \text{nass}, \text{trocken}, \text{nass}, \text{trocken})$

➔ $Z = (\text{Regen}, \text{Regen}, \text{Sonne}, \text{Sonne}, \text{Sonne})$

$\log_2$



Produkt $x_1 \cdot x_2 \cdot \dots \cdot x_k$ ist maximal genau dann wenn $\log(x_1 \cdot x_2 \cdot \dots \cdot x_k) = \log x_1 + \log x_2 + \dots + \log x_k$ maximal ist.

LONGESTPATH()

for $i = 1$ **to** n **do**

└ Berechne schwersten $s-i$ -Pfad.



Algorithmus

LONGESTPATH(Gerichteter azyklischer Graph $G = (V, E)$)

Berechne topologische Sortierung von $V \Rightarrow v_1, \dots, v_n$

$v_1.d = 0$

$v_1.\pi = \text{nil}$

for $j = 2$ **to** n **do**

$v_j.d = -\infty$

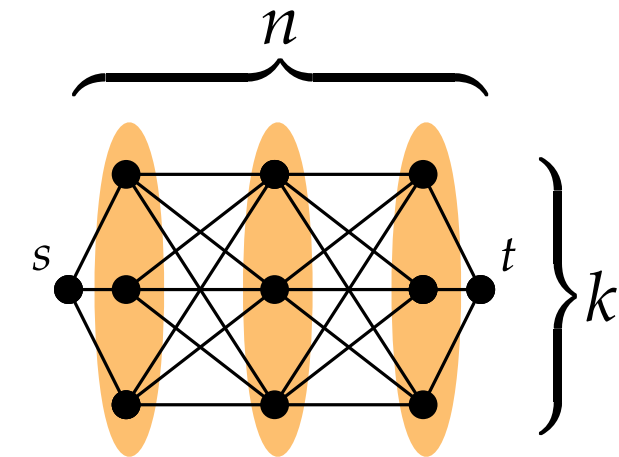
$v_j.\pi = \text{nil}$

for $v_i v_j \in E$ **do**

if $v_i.d + w(v_i v_j) > v_j.d$ **then**

$v_j.d = v_i.d + w(v_i v_j)$

$v_j.\pi = v_i$



Laufzeit $\mathcal{O}(n + m)$

DIJKSTRA: $\mathcal{O}(n \log n + m)$

Laufzeit bei n Messungen und k Zuständen? $nk + 2$ Knoten, $(n - 1)k^2 + 2k$ Kanten

➔ $\mathcal{O}(nk^2)$.

Newson & Krumm (2009) – letzte Details



Für jeden Punkt der Trajektorie gibt es zwei Zufallsvariablen:

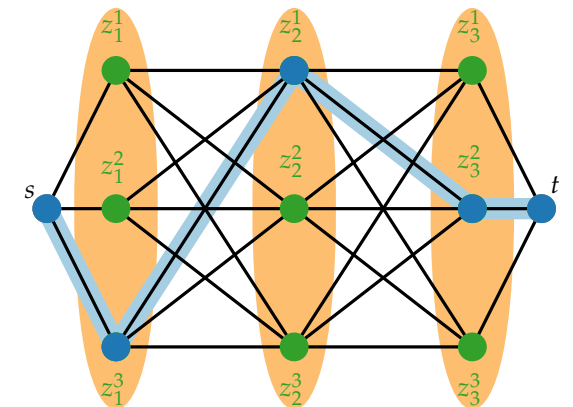
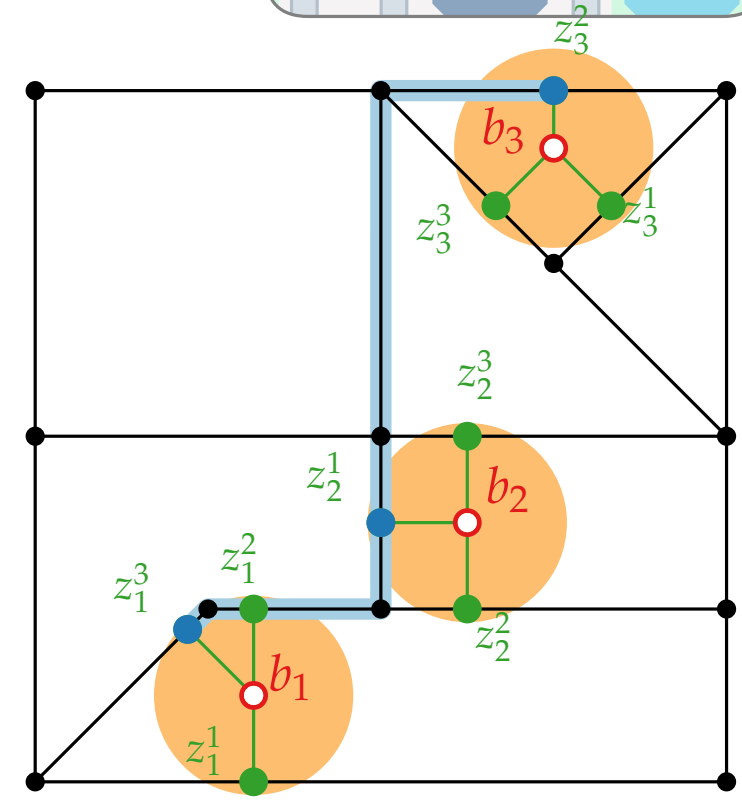
- B_i ist die beobachtete GPS Koordinate (Domäne $\mathbb{R}^2$)
- Z_i ist die Position im Straßennetz (diskrete Domäne $\{z_i^1, \dots, z_i^k\}$)
- Wahrscheinlichkeitsverteilung $P[B_i | Z_i]$ ist bekannt.

$$P[B_i = b_i | Z_i = z_i^j] = \frac{1}{\sqrt{2\pi\epsilon}} e^{-\frac{d_{\text{Eudlid}}(b_i, z_i^j)^2}{2\epsilon^2}}$$

(Normalverteilung um z_i^j)

- Wahrscheinlichkeitsverteilung $P[Z_i | Z_{i-1}]$ ist bekannt.

$$P[Z_i = z_i^r | Z_{i-1} = z_{i-1}^s] = \frac{d_{\text{Euklid}}(b_i, b_{i-1})}{d_{\text{Graph}}(z_i^r, z_{i-1}^s)}$$



Frage: Welche Sequenz der Positionen im Straßennetzwerk ist am wahrscheinlichsten?