Similarity Search on Uncertain Spatio-Temporal Data

Johannes Niedermayer, Andreas Züfle, Tobias Emrich, Matthias Renz, Nikos Mamoulis, Lei Chen, Hans-Peter Kriegel

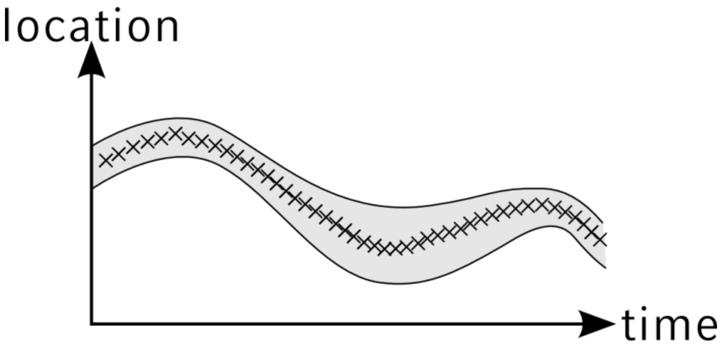
Ludwig-Maximilians-Universität München
Munich, Germany
www.dbs.ifi.lmu.de

Outline

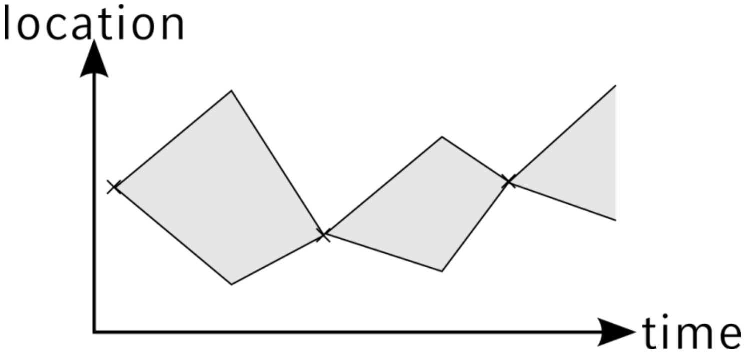
- Problem Definition
  - The Data
  - The Model
  - The Uncertain LCSS
- Algorithm
- Experimental Evaluation
- Conclusion and Future Work

Problem Definition: The Data

- Spatio-temporal data generally involves uncertainty
  - Position uncertainty arises from erroneous sensor measurements, e.g. in GPS or RFID systems.



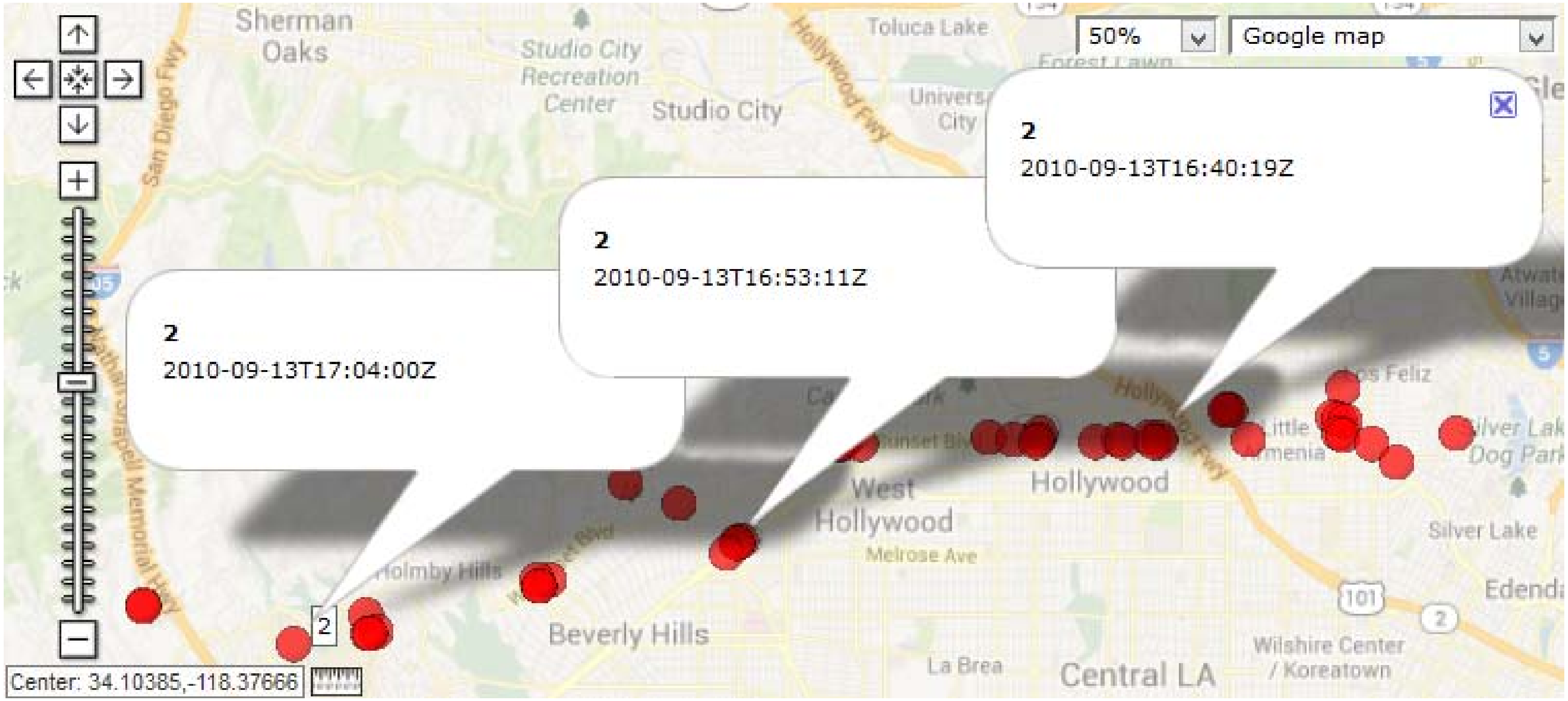
  - Motion uncertainty arises from time-discrete measurements.



  - In this talk, we address motion uncertainty.

Problem Definition: The Data - GoWalla (Suspended)

- Motion uncertainty arises for example in geo-social networks
  Example: GoWalla (Suspended)



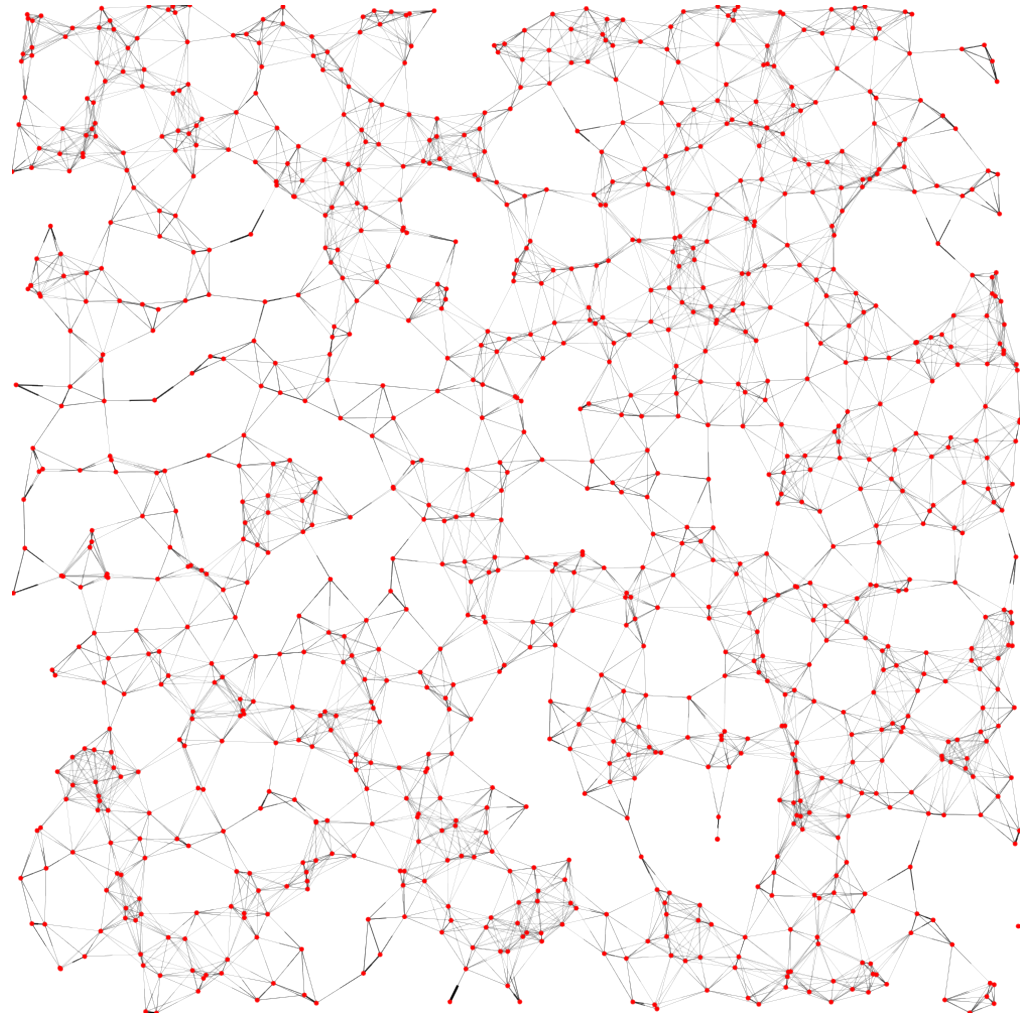
Problem Definition: The Model

- We model uncertain objects by 1st order Markov chains
- The spatial domain is modeled as a graph
- Nodes (red dots): possible positions of uncertain objects

  $\mathcal{S} = \{s_1, ..., s_{|\mathcal{S}|}\} \subset \mathbb{R}^d$

- Edges (black lines): transition probabilities

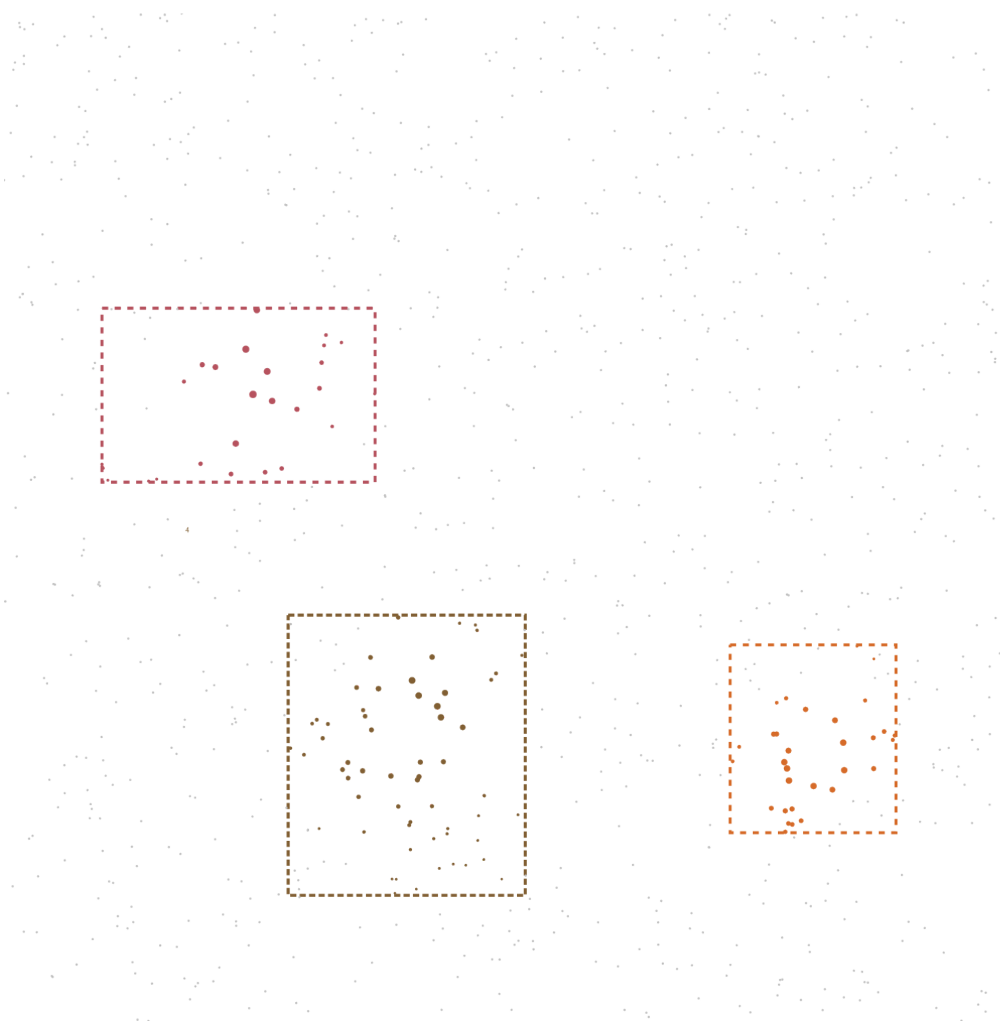
  $T^o_{ij}(t) := P(o(t+1) = s_j | o(t) = s_i)$



Problem Definition: The Model

- For a given point in time, the position of an uncertain object can be modeled by a probability distribution over all states.

- Mathematical representation as vectors:
  $\vec{s}_i^{\,o}(t) = P(o(t) = s_i)$

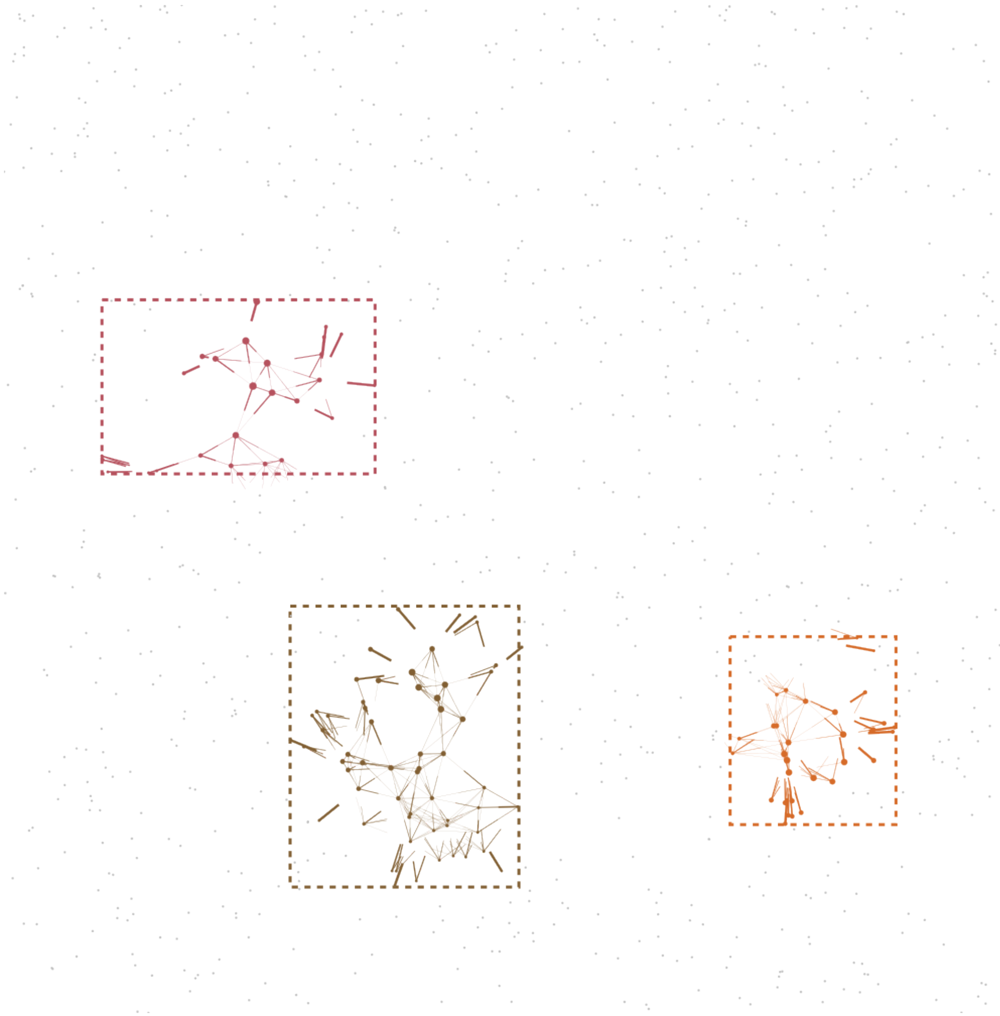


Problem Definition: The Model

- The actual motion can be described as transitioning between two states between consecutive points in time:

$$\vec{s}^o(t+1) = T^o(t)^T \cdot \vec{s}^o(t)$$

- This transition leads to temporal correlation of transitions
- The correlation has to be considered during query evaluation.



Problem Definition: The Uncertain LCSS

- In this talk, we address similarity search on uncertain spatio-temporal data exemplarily on the LCSS: Let $A = (a_1, \dots, a_n)$ and $B = (b_1, \dots, b_m)$, then:

$$LCSS_{\delta,\epsilon}(A,B) := \begin{cases} 0 \text{ if } A = \emptyset \text{ or } B = \emptyset, \\ 1 + LCSS_{\delta,\epsilon}(Head(A), Head(B)) \quad \text{if } dist(a_n - b_m) < \epsilon \text{ and } |n - m| \leq \delta \\ max(LCSS_{\delta,\epsilon}(Head(A), B), LCSS_{\delta,\epsilon}(A, Head(B))) \quad otherwise \end{cases}$$

- Based on this, we can define the Uncertain LCSS (ULCSS):

$$ULCSS_{\delta,\epsilon}(o_1, o_2) : \mathcal{D} \times \mathcal{D} \rightarrow (\mathbb{N} \rightarrow [0,1] \in \mathbb{R})$$

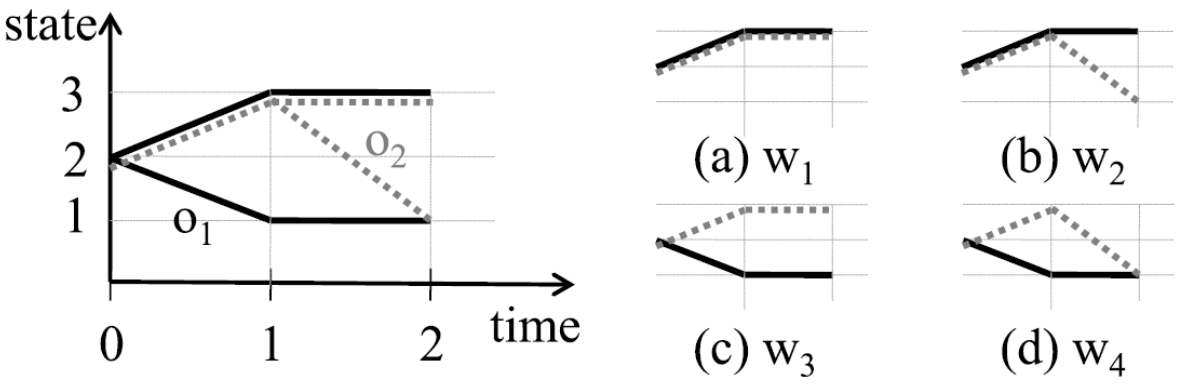
$$ULCSS_{\delta,\epsilon}(o_1, o_2) := pdf(x \in \mathbb{N}) = P(LCSS_{\delta,\epsilon}(o_1, o_2) = x)$$

- And the Uncertain Aligned LCSS (UALCSS):

$$UALCSS_{\epsilon}(o_1, o_2) = ULCSS_{0,\epsilon}(o_1, o_2)$$

Problem Definition: The Uncertain LCSS (Example)

- Example
  - Two uncertain objects o_1 and o_2
  - Each object consists of two possible worlds
  - We would like to compute the UALCSS in the time interval [0,2]
  - Uniform transition probabilities

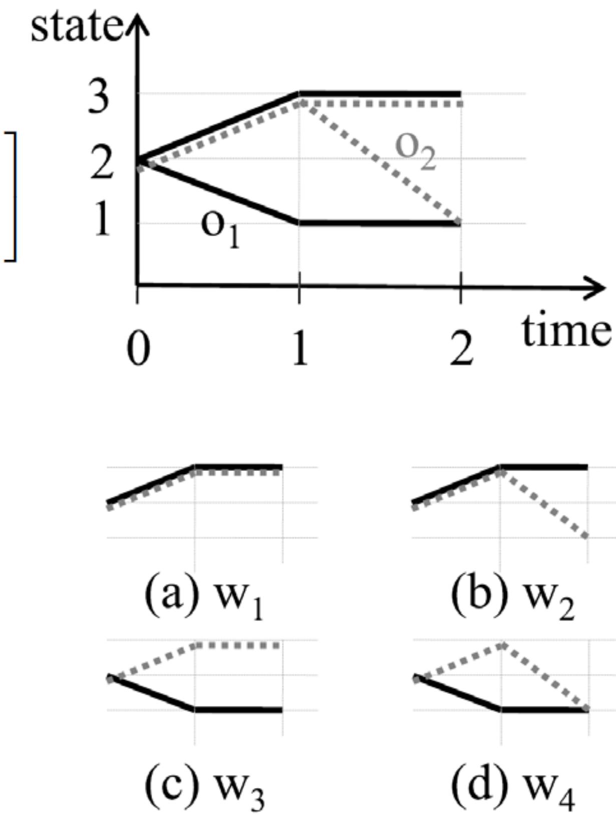


  - Result: UALCSS = [0, 0.25, 0.5, 0.25]
  - But number of possible worlds increases exponential with time!

→ Develop efficient algorithms

Algorithm

- Solution: build equivalence classes
  - An equivalence class contains the set of possible worlds where $o_1(t) = s_i$, $o_2(t) = s_j$, $UALCSS(t_0,t) = k$
  - $UALCSS(t_0,t_1)$ is computed iteratively:



$$\text{Induction Start } (t=0): \left[\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}, \begin{pmatrix}0&0&0\\0&1&0\\0&0&0\end{pmatrix}, \begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}, \begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}\right]$$

$$\text{Transition to } t=1: \left[\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}, \begin{pmatrix}\frac{1}{2}&0&0\\0&0&0\\\frac{1}{2}&0&0\end{pmatrix}, \begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}, \begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}\right]$$

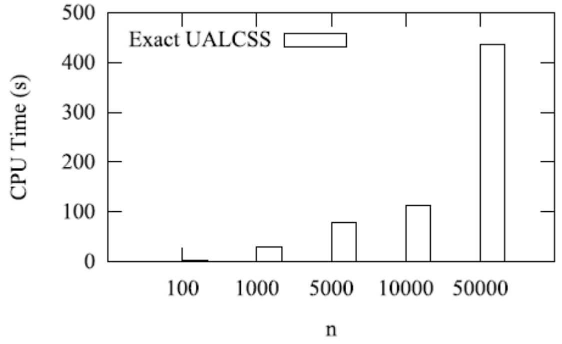
$$\mathit{Shift}: \left[\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}, \begin{pmatrix}0&0&0\\0&0&0\\\frac{1}{2}&0&0\end{pmatrix}, \begin{pmatrix}\frac{1}{2}&0&0\\0&0&0\\0&0&0\end{pmatrix}, \begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}\right]$$

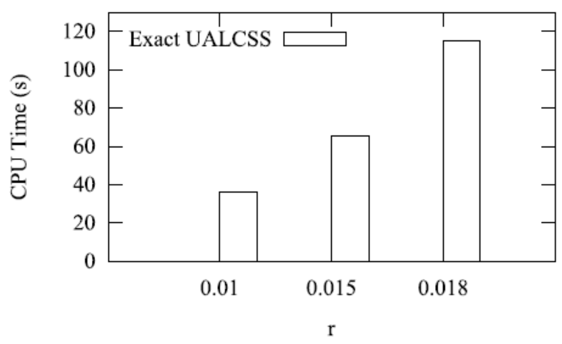
$$\text{Transition to } t=2: \left[\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}, \begin{pmatrix}0&0&0\\0&0&0\\\frac{1}{4}&0&\frac{1}{4}\end{pmatrix}, \begin{pmatrix}\frac{1}{4}&0&\frac{1}{4}\\0&0&0\\0&0&0\end{pmatrix}, \begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}\right]$$

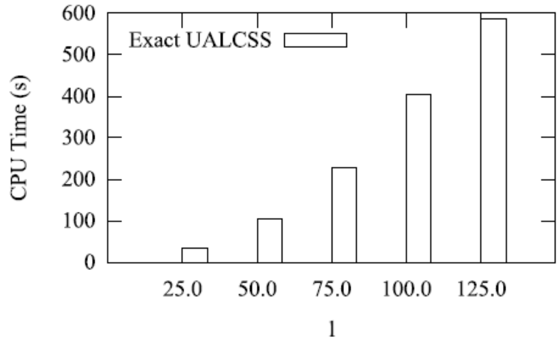
$$\mathit{Shift}: \left[\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}, \begin{pmatrix}0&0&0\\0&0&0\\\frac{1}{4}&0&0\end{pmatrix}, \begin{pmatrix}0&0&\frac{1}{4}\\0&0&0\\0&0&\frac{1}{4}\end{pmatrix}, \begin{pmatrix}\frac{1}{4}&0&0\\0&0&0\\0&0&0\end{pmatrix}\right]$$

Experimental Evaluation

- Experiments were conducted on synthetic data
  - Number of states: 100-50000 states in the unit space $[0,1]^2$
  - Range of connectivity: 0.01-0.018
  - Length of the query interval: 25-125 timesteps







Conclusion and Future Work

- We employed the UALCSS to describe the similarity of uncertain spatio-temporal objects modeled by Markov chains
- Computing the UALCSS is polynomial
- But computing the general ULCSS is hard
- Future work: Use sampling to compute the general ULCSS

DATABASE
SYSTEMS
GROUP

That‘s it!

Thank you!