Wildfire prediction using Spatio-Temporal Knowledge Graphs

D2R2'2023 28.05.2023

UNIVERSITÄT MANNHEIM

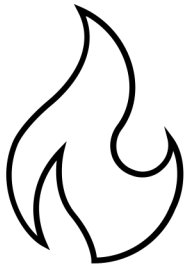
Agenda

UNIVERSITÄT MANNHEIM

- Introduction
- Use Case overview
  - Data format
  - Data overview
- Knowledge Graph creation
- Data Modeling
- Conclusion

Introduction

UNIVERSITÄT MANNHEIM

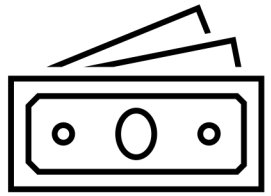


US (2021)

58.985 wildfires

California

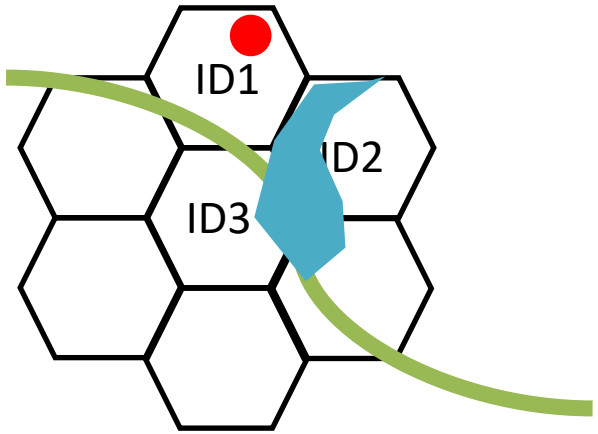
2M house at risk



California (2018)

148B $ economic damage

Current data preparation techniques



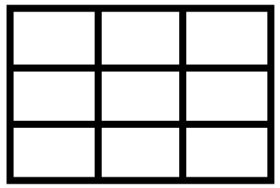
Grid Cell ID	hasRiver	hasCampfire	Wildfire
ID1	True	True	False
ID2	False	False	True
ID3	True	False	True

- Use Grid cells as single elements within datasets
- No inclusion of surrounding elements
- Limited data base

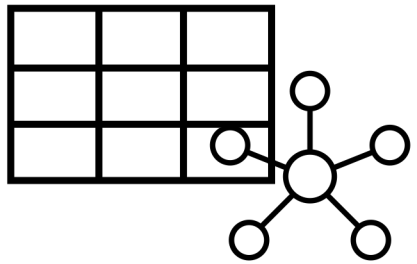
Can surrounding elements have a positive influence on wildfire prediction?

Use Case Overview

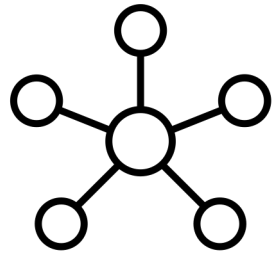
UNIVERSITÄT MANNHEIM



Base Case



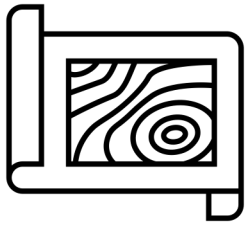
Hybrid Case



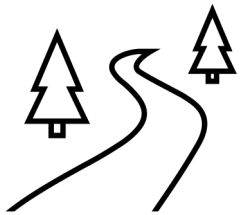
Graph Case

Input data

UNIVERSITÄT MANNHEIM



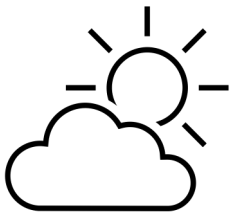
Elevation data



Openstreetmap data

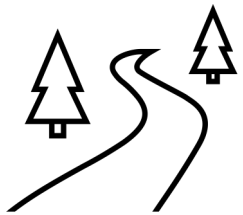
Landcover data

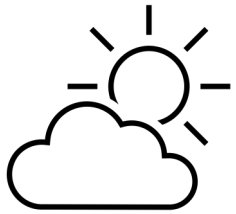
Wildfire area data

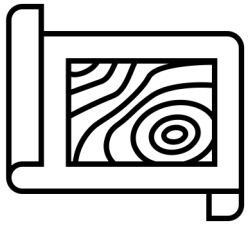


Weather data

Incorporating spatial relationships







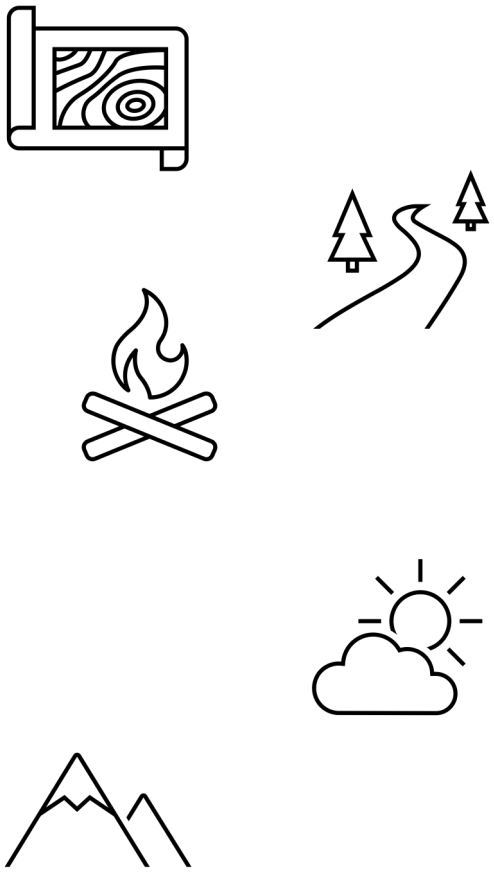
Interconnection of different data types

Transform data to spatial knowledge graph

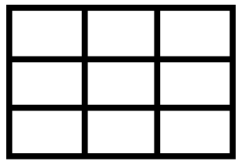
Starting point

- Divide area into regular spaces (called grid)
- Each area is a grid cell
- Possible geometric objects:
  - Triangle
  - Square
  - Hexagons

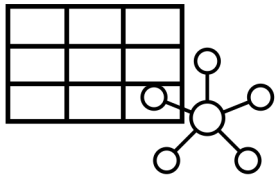
Data Preparation



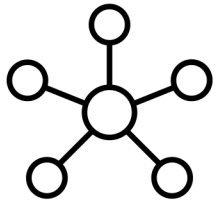
Still data preparation steps needed



Base Cases



Hybrid Cases



Graph Cases

Data Preparation - Weather

- Pointwise measurements of weather variables
- Need to interpolate data over created spatial grid
- Used interpolation technique:

  Kriging:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i * Z(s_i)$$

  Weight λ_i is determined by a semivariogram

  Semivariogram determines spatial autocorrelation and fits function to data

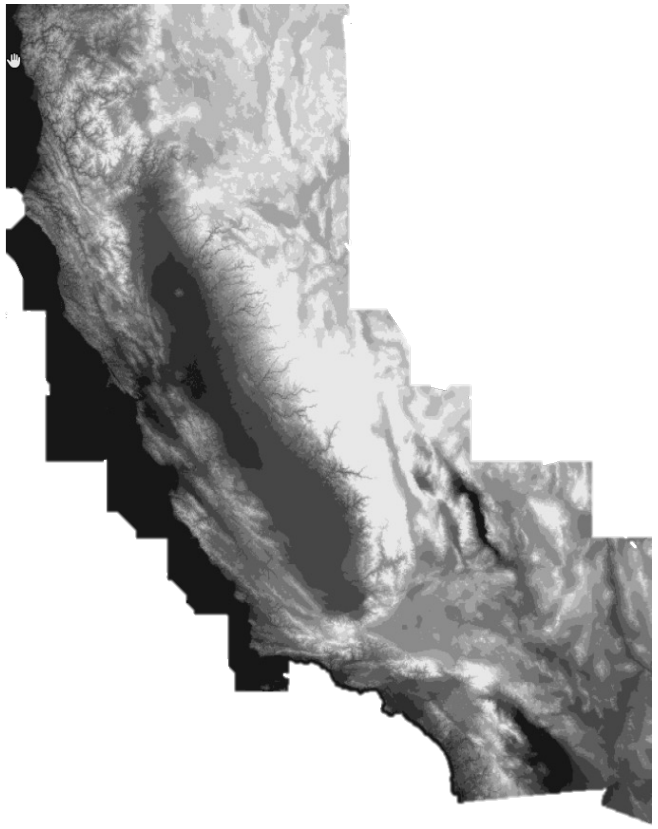
- Each constructed grid cell has now interpolated values for weather variables

- λ_i: Weight at i
- $Z(s_i)$: Value at point s_i
- $\hat{Z}(s_0)$: Prediction at point s_0

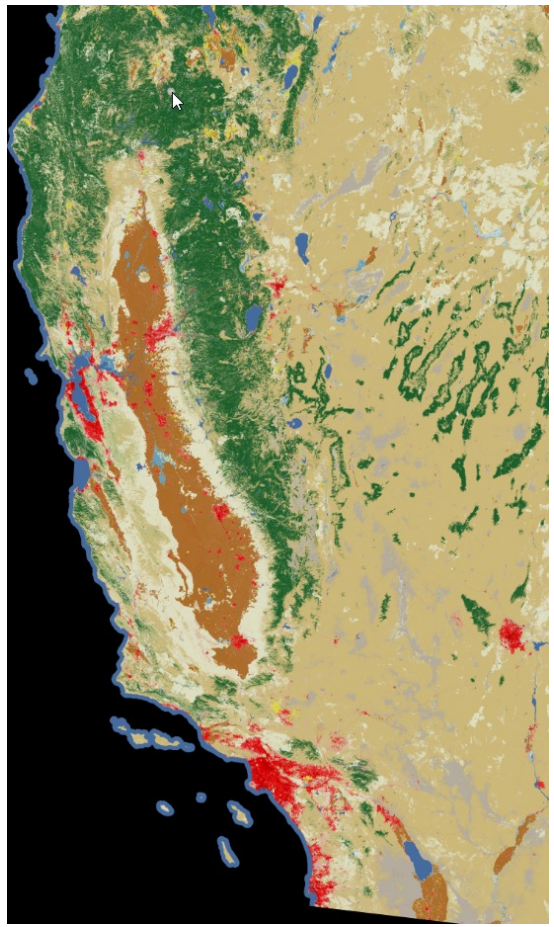
Data Preparation – Elevation & Landcover data

UNIVERSITÄT MANNHEIM

Elevation



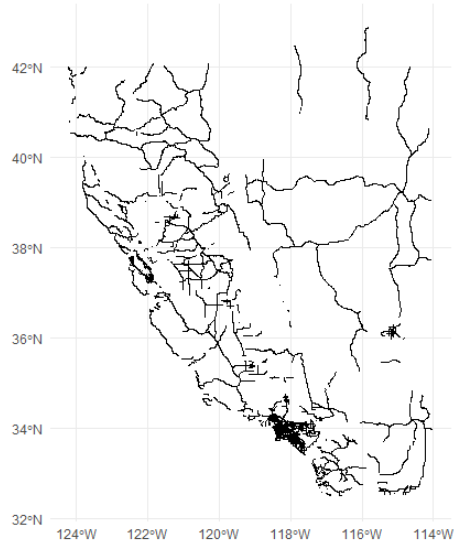
Land cover

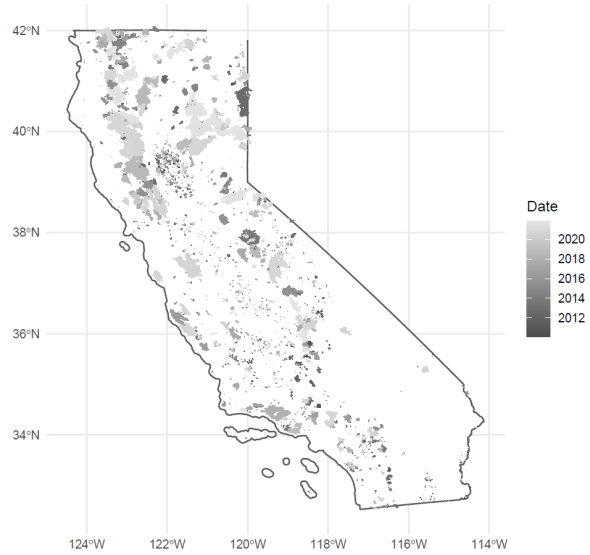


Data Preparation – Elevation & Landcover data

- Both datasets fine granular
  - Elevation 60m*60m tiles
  - Landcover 90m*90m tiles
- Elevation numeric dataset
- Landcover categorical dataset
- Elevation dataset gets aggregated with weighted mean to single grid cell
- Landcover dataset gets aggregated with weighted majority vote to single grid cell

Data Preparation – Openstreetmap & Wildfire data





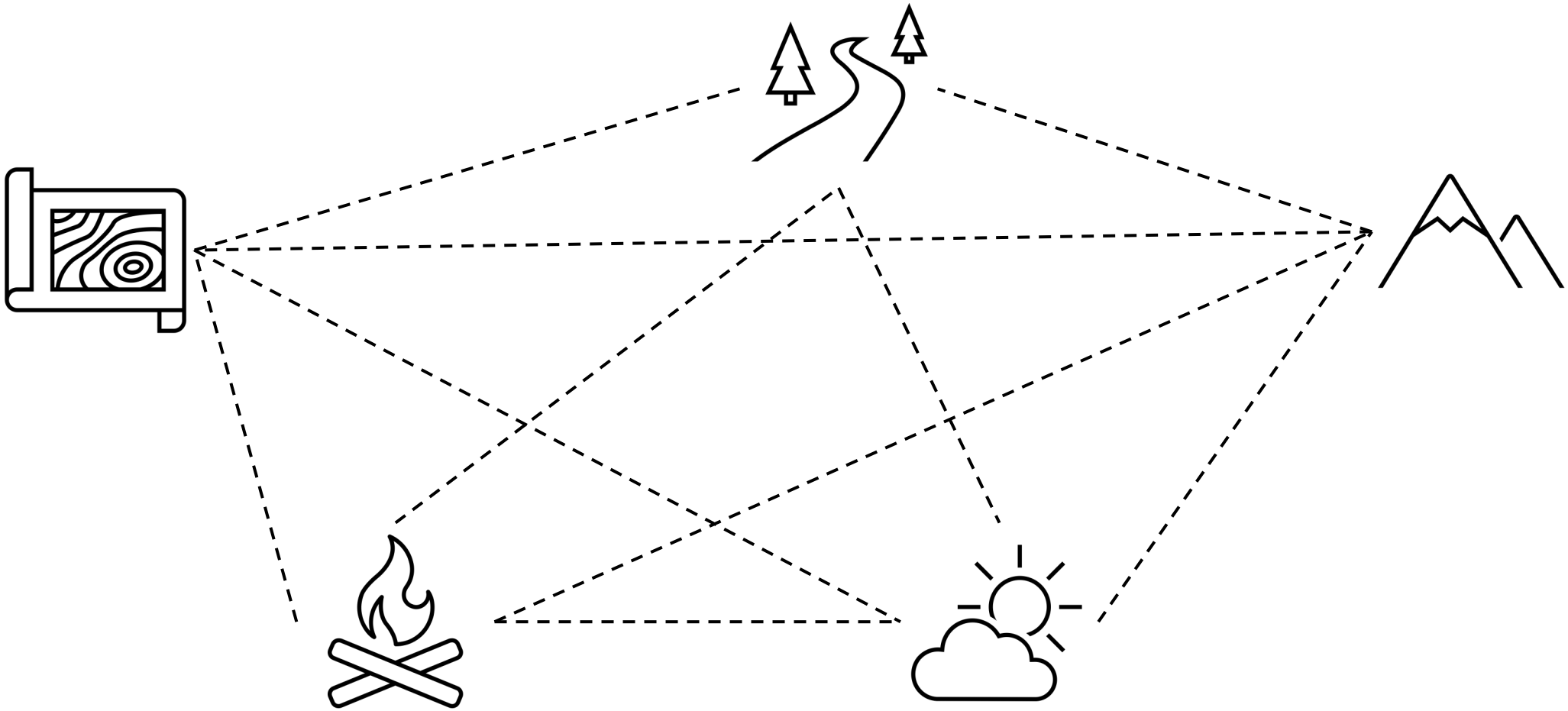
Openstreetmap

- Extract columns related to potential wildfires
- Extract necessary geometry types
- Join Openstreetmap to Grid Cell based on overlap

Wildfire

- Transform year and days to date
- Join wildfire to Grid Cell based on relation overlap

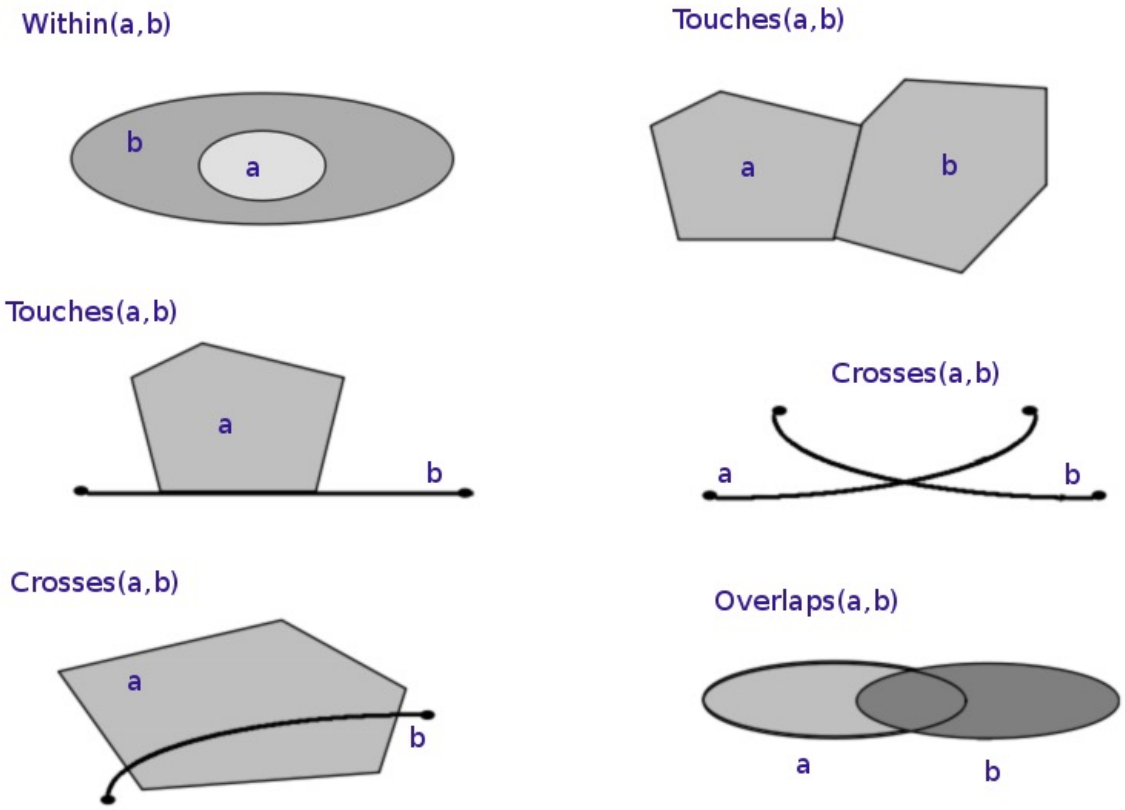
Build up spatial knowledge graph



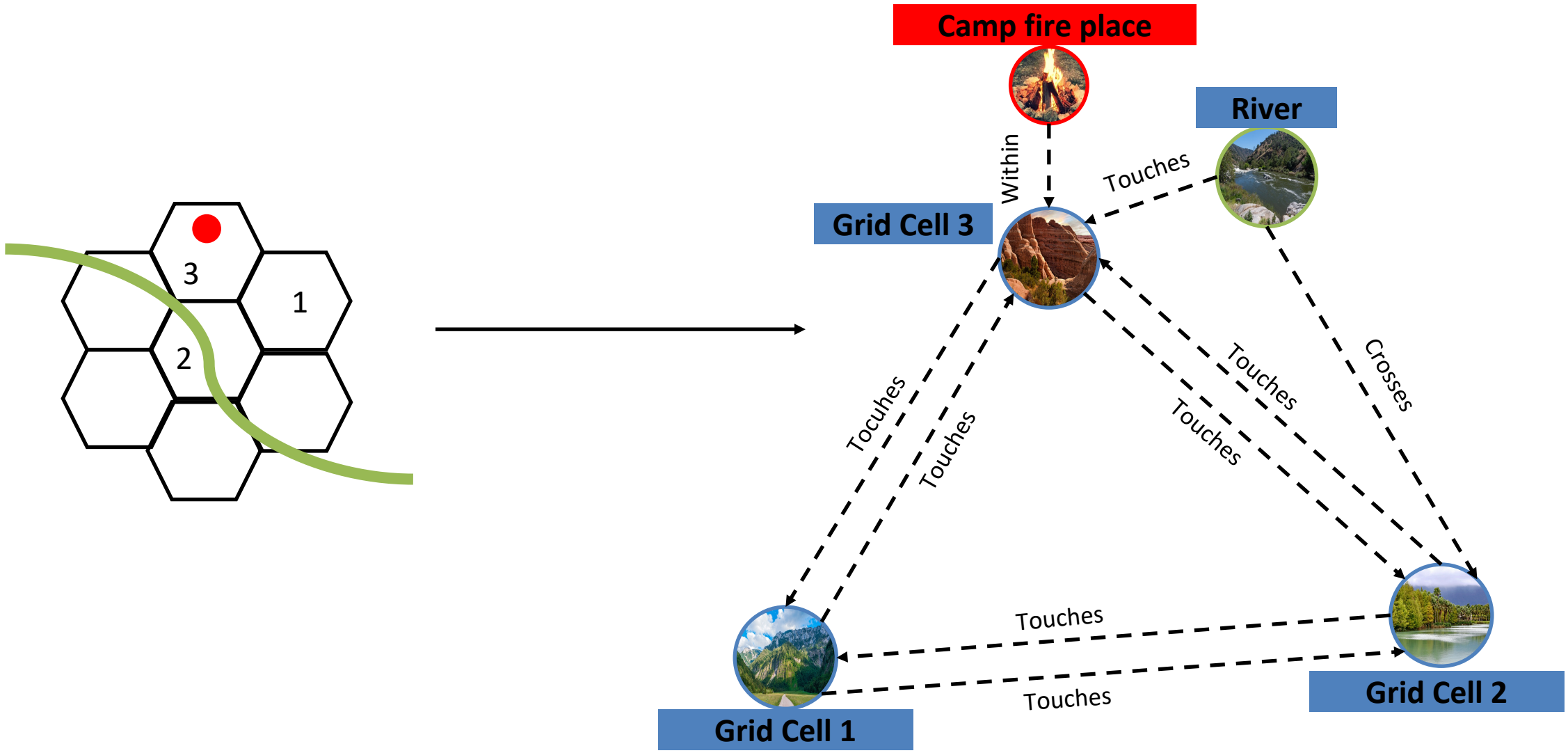
Build up spatial knowledge graph – DE-9IM

UNIVERSITÄT MANNHEIM

- DE-9IM is topological model to build relationships between geometric objects



Transform data to spatial knowledge graph

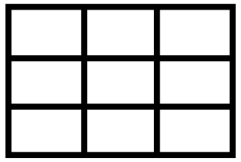


Wildfire detection – Create Vector representation (RDF2Vec)

UNIVERSITÄT MANNHEIM

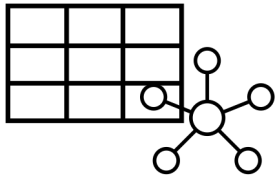
- RDF2Vec method transforms graphs to vector representations
- RDF2Vec is separated in two phases
  - Graph traversal phase with Breadth-First Search algorithm
  - Training of Word2Vec model
- Each extracted walk consists of Nodes and the Edge description transformed to sentence
- Resulting vector representation can be combined to dataset

Construct Base Case Dataset



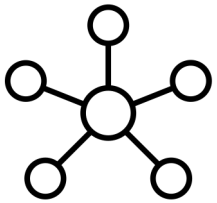
Base Cases

- Consists of tabular data where Grid Cell ID and Month build one row
- No relationship between neighboring grid cells



Hybrid Cases

- Consists of tabular data where Grid Cell ID and Month build one row
- Embeddings from OpenStreetMap knowledge graph are joined to dataset
- Neighbor semantics are modeled for single grid cell ID



Graph Cases

- Consists of embedding data where Grid Cell ID and Month build one row
- Embeddings from knowledge graph with all are joined

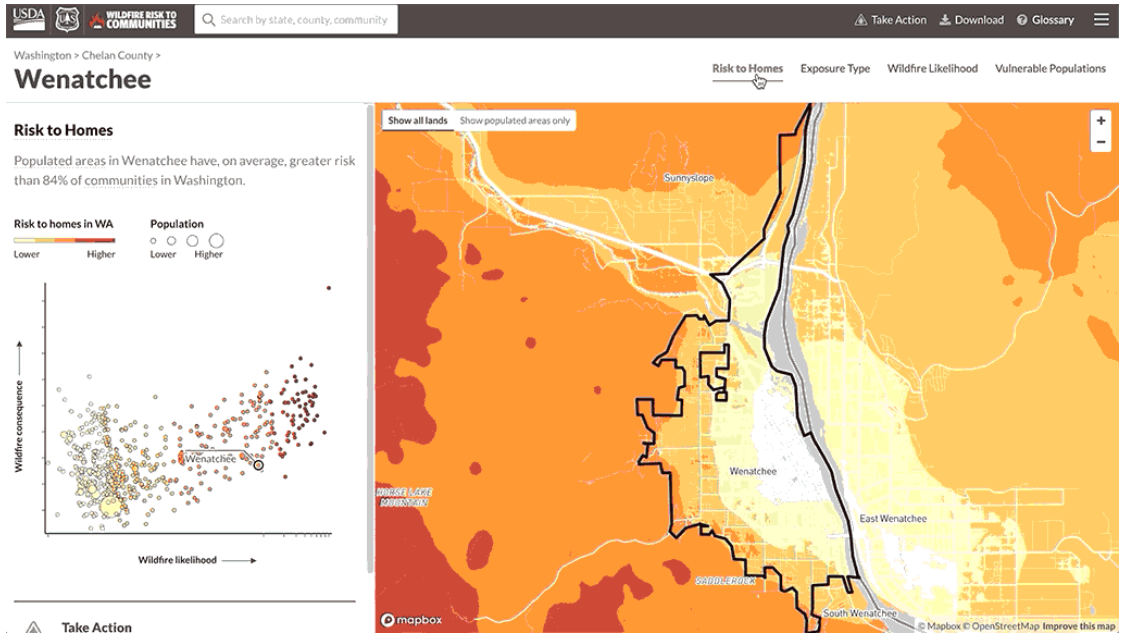
Result overview

UNIVERSITÄT MANNHEIM

Dataset	F1	AUC
BaseCase	0.3478	0.6816
HybridCase	0.3803	0.8748
NetworkCase	0.0107	0.5341

Conclusion and outlook

- Graph based inclusion in dataset improves results for Hybrid Case dataset
- Scenarios can be modelled more accurate due to surrounding factors and semantic relations
- Outlook:
  - Create more benchmark datasets related to geography
  - Compare constructed KG with other spatial Knowledge Graph on benchmark dataset
  - Embed Knowledge Graphs with different embedding methodologies



UNIVERSITÄT MANNHEIM

Thank you