Recommend Orienteering-based Tourist Trip Planning with Social Sensing

(work in progress)

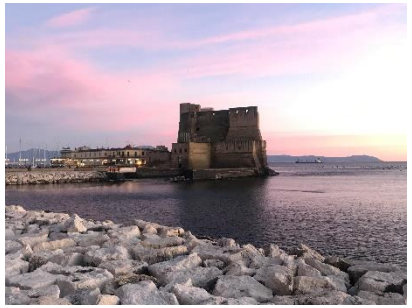
Mouzhi Ge

Masaryk University, Czech Republic

Shortest path is not enough

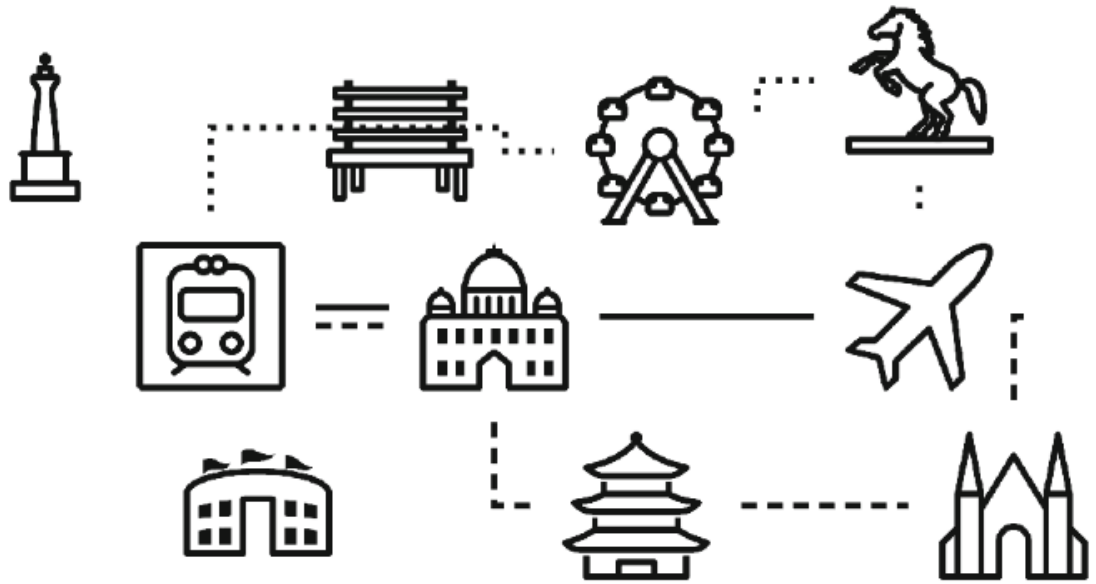
- In **tourist trip planning**, it is **not always useful** to determine **shortest paths** from source nodes to destination nodes.

- It is **more important** to discover routes covering the **most attractive points,** which involves solving an **orienteering problem.**



Orienteering Problem

- The **Orienteering Problem (OP)** in an **NP-hard problem** aimed at discovering a **path** from a starting node to an ending node in an **edge-weighted graph** with a **score** for each **node**
- **Maximizing** the **total score** while staying **within a certain time budget**



Recap some recent solution

- Paolo Bolzoni, Sven Helmer, Hybrid Best-First Greedy Search for Orienteering with Category Constraints. Advances in Spatial and Temporal Databases (SSTD) 2017: 24-42
- The authors defined an **efficient algorithm** for solving the **orienteering problem** with category constraints using a **probabilistic approach**
  - **decrease the size** of a problem instance by selecting nodes from a graph according to **probabilities** assigned to these nodes
  - run an **optimization algorithm** on the reduced graph

What we have and what we want to do

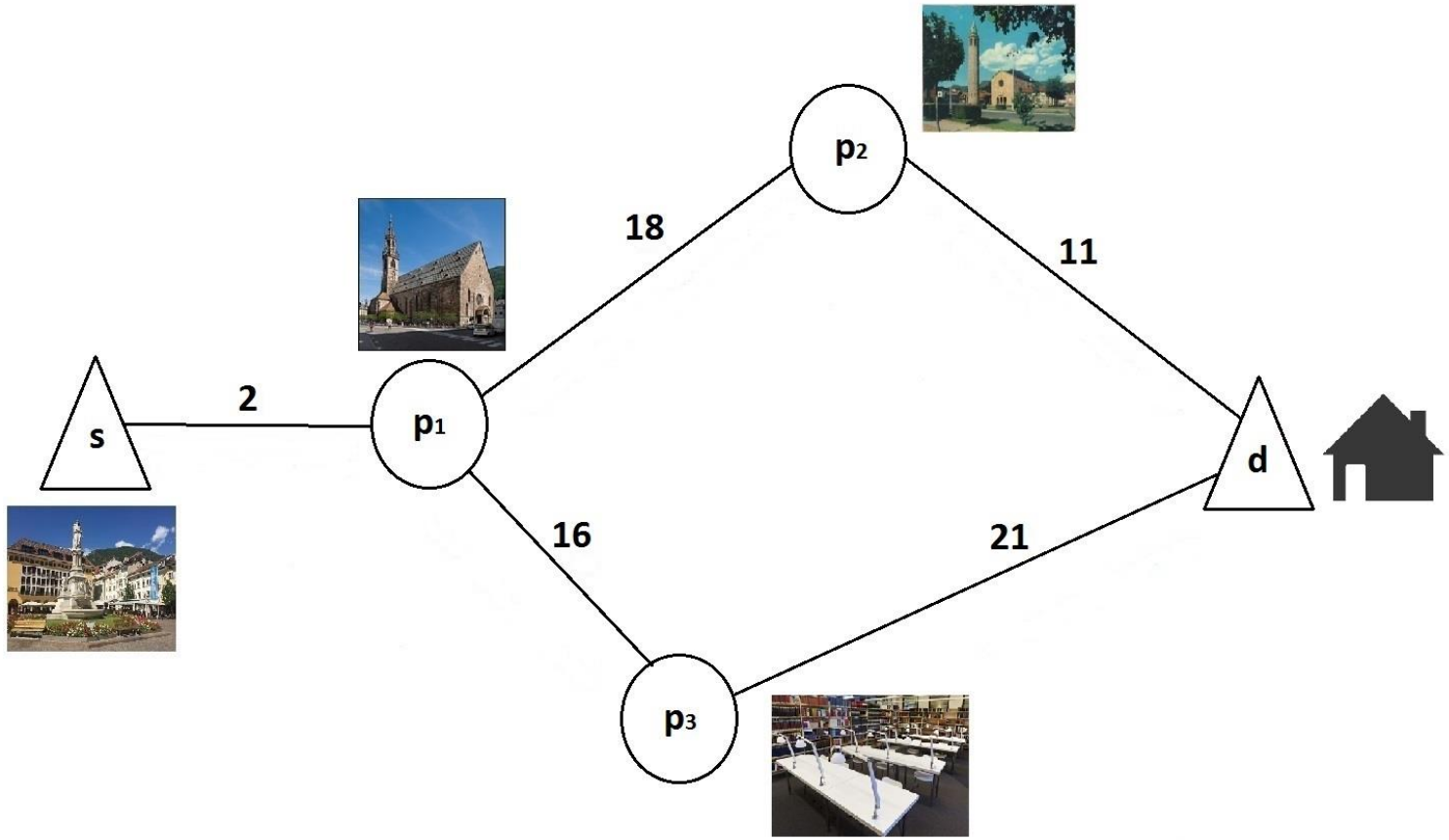
- We have algorithms to solve the **orienteering problem.**
- However, with the same starting point and ending point, the solution is the **same to every user.**
- We want to improve OP algorithms by **including personalization.**
- Personalization can be done by mining the user preference and score POIs only for this user.
  (POI = point of interest)

Aim of the research

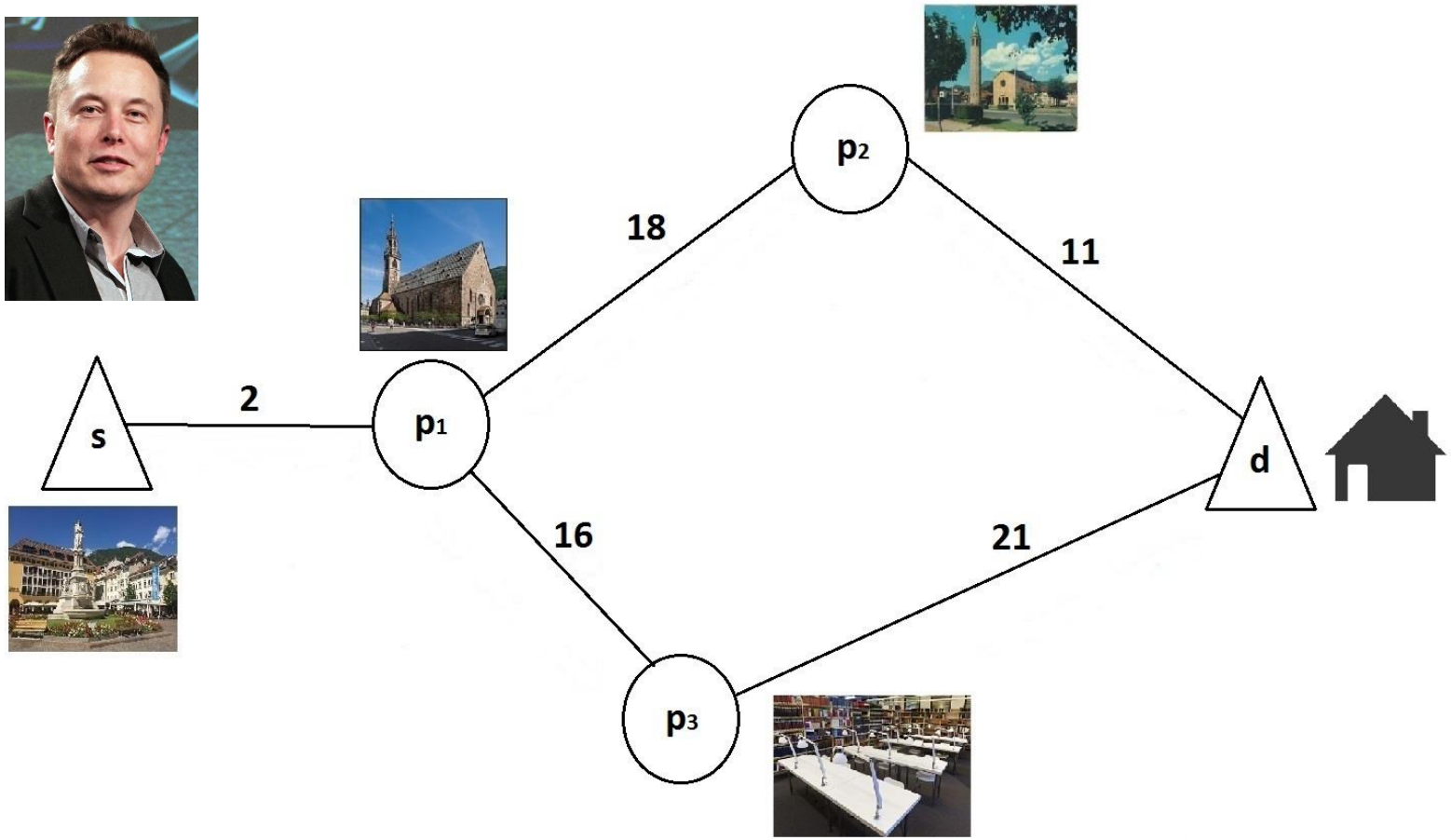
- We plan to propose a **new method** that **combines social sensing** with an OP algorithm to improve the orienteering problem solution**.**

- Social sensing is an approach to analyze user-generated content in social networks to make it usable for different applications.

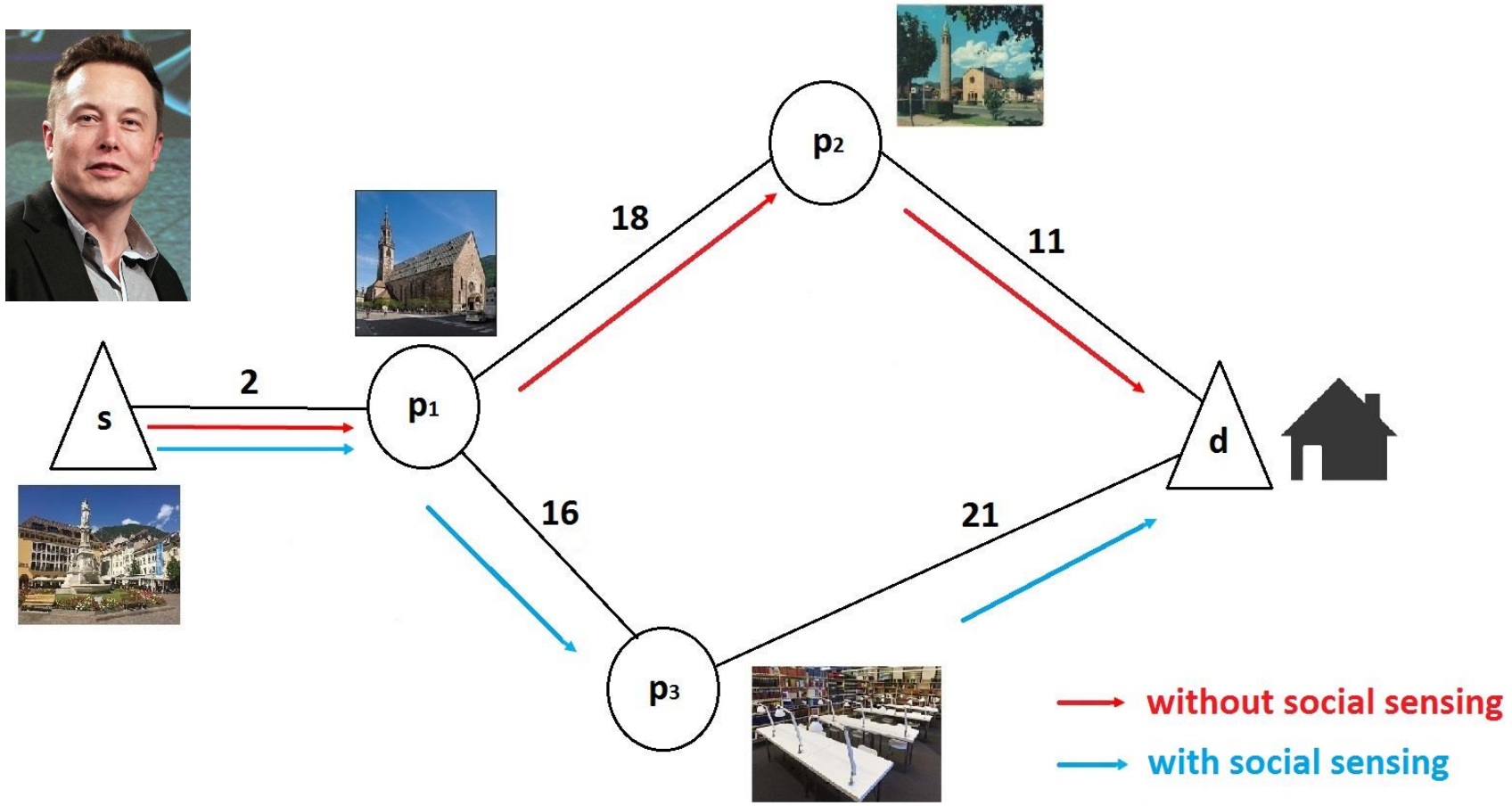

Visiting Brno



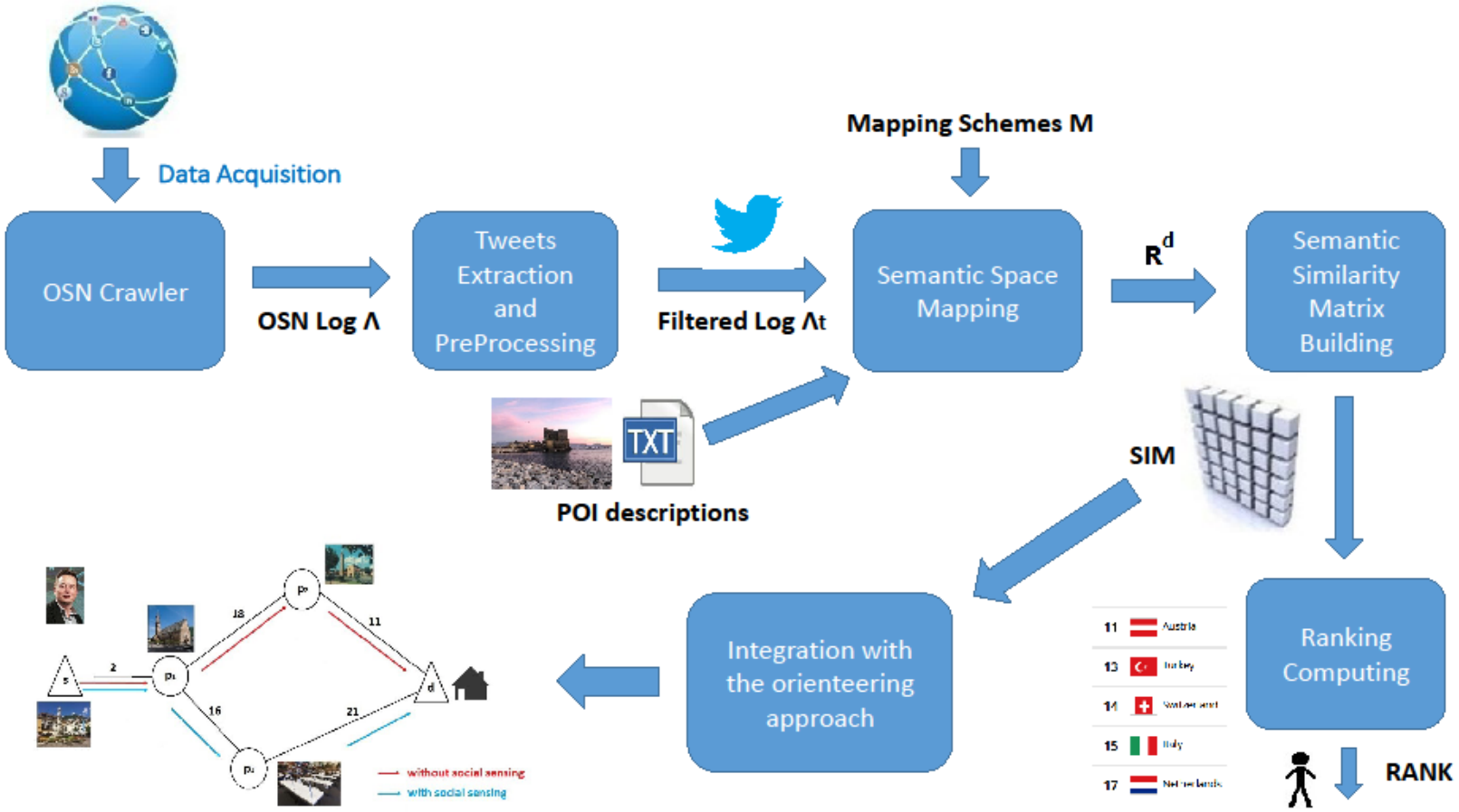
Elon Musk plans to visit Brno

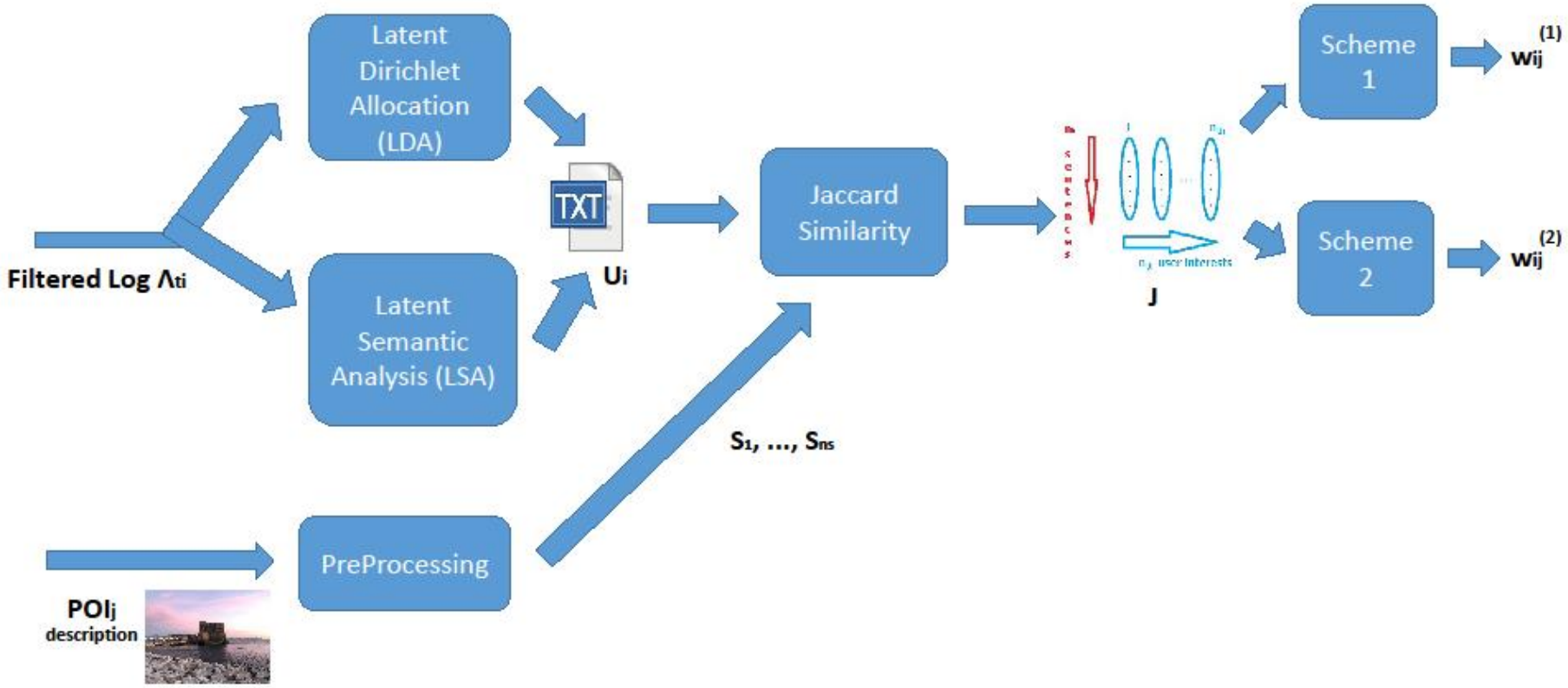


Social Sensing

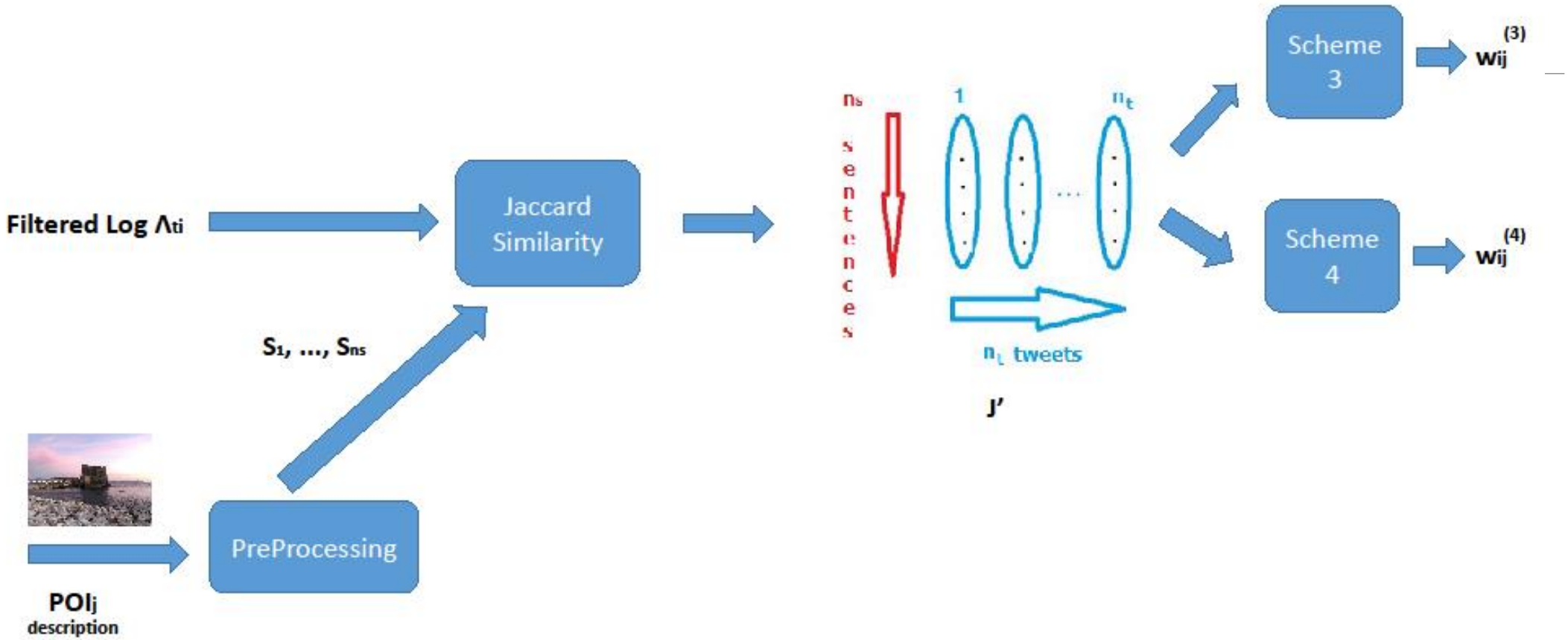


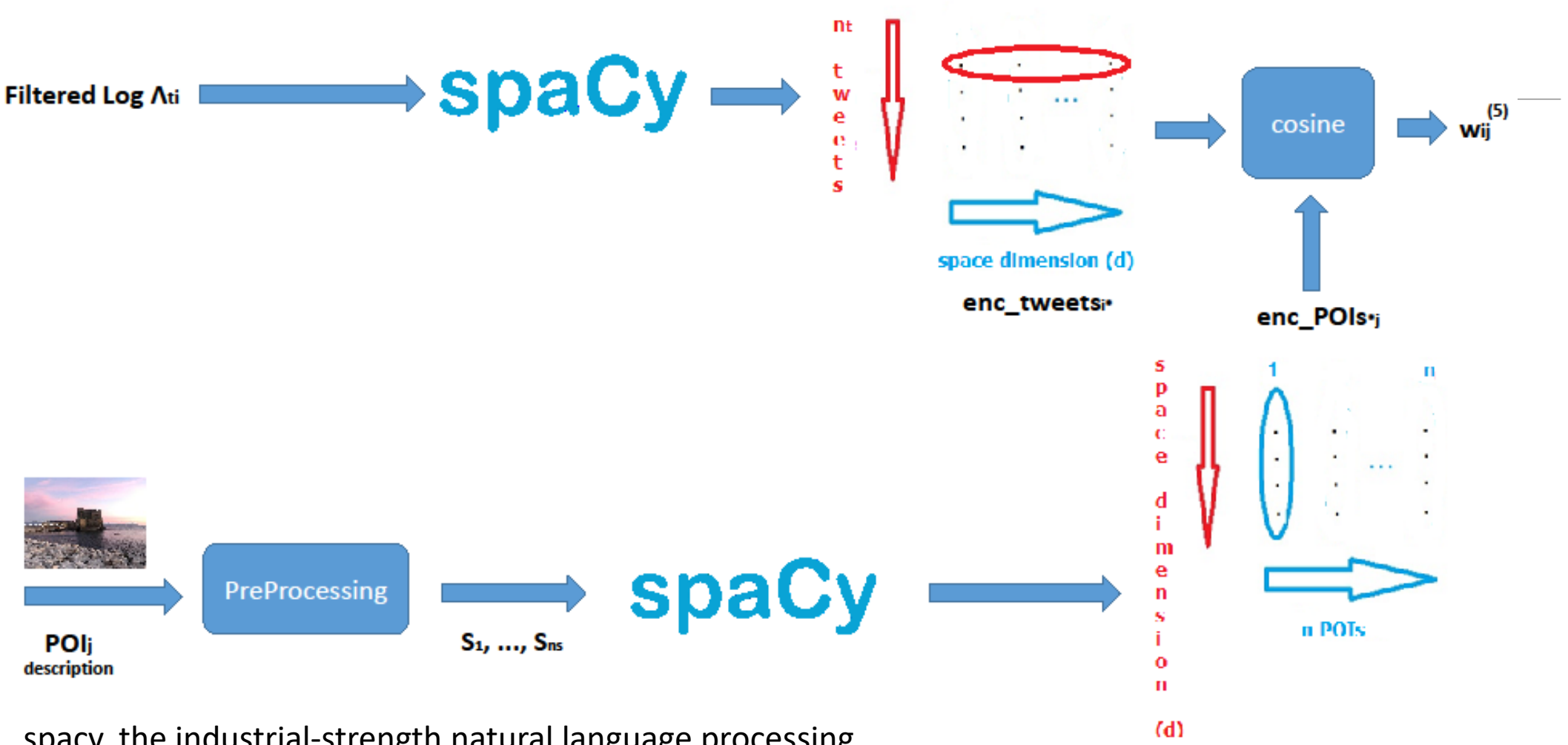
Overall process



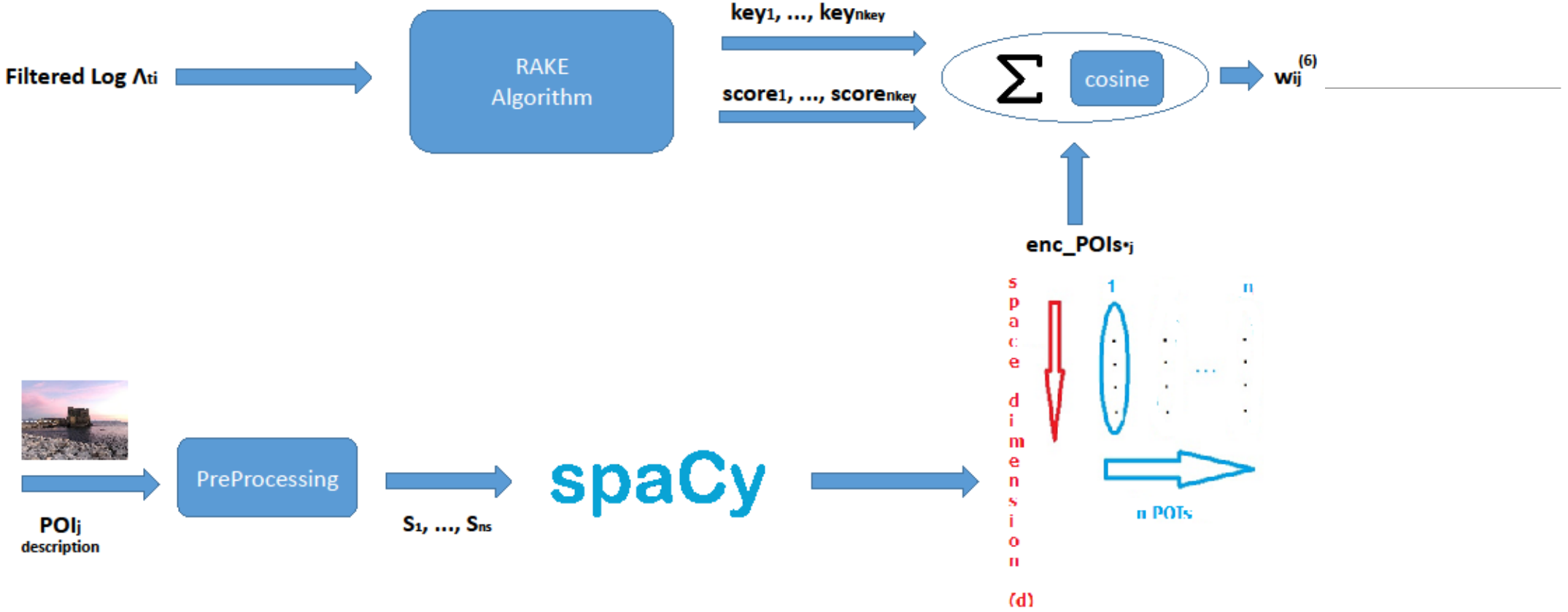


- Latent Dirichlet Allocation (LDA) is a methodology for computing a Bayesian probabilistic model of text corpora. Its aim is finding topics in documents.
- Latent Semantic Analysis (LSA) technique is a literature methodology useful for obtaining a vector encoding of words and their semantics.


Filtered Log Λti
Jaccard Similarity
S1, ..., Sns
PreProcessing
POIj
description
ns
sentences
1
nt
nL tweets
J'
Scheme 3
Scheme 4
(3)
wij
(4)
wij

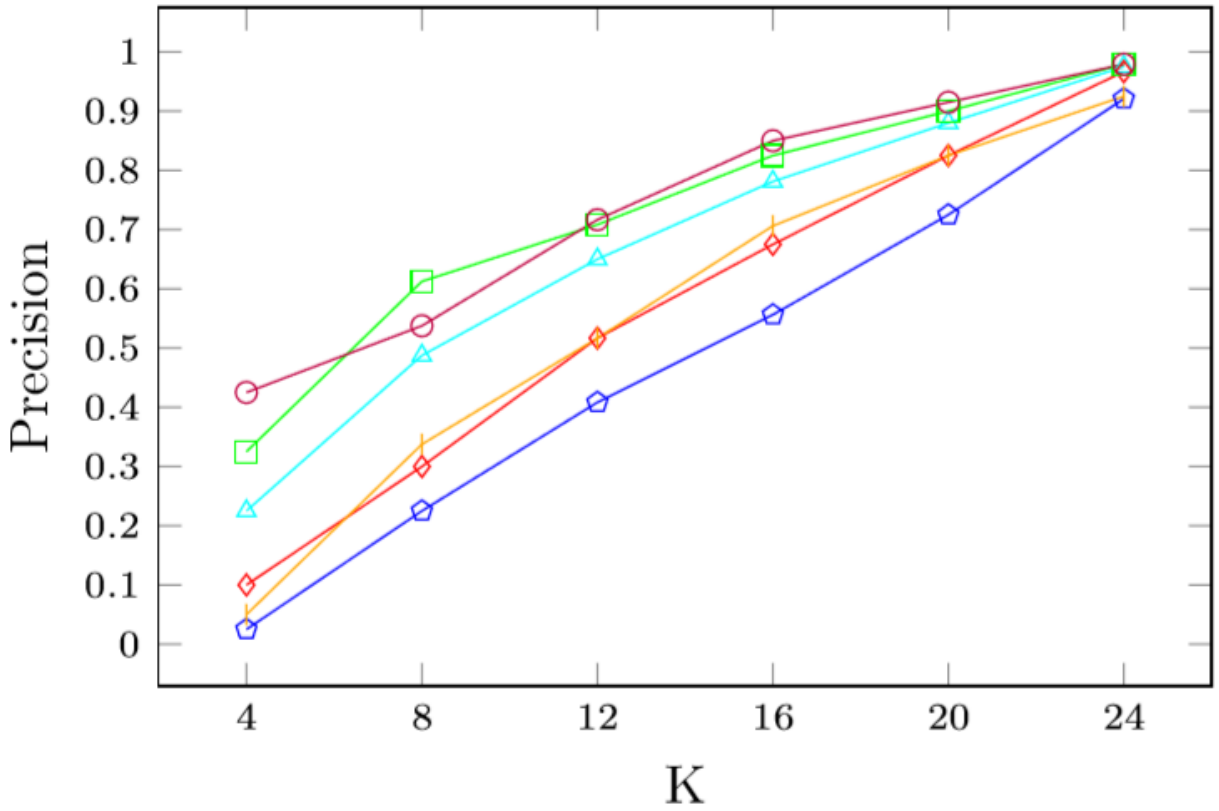


spacy, the industrial-strength natural language processing (2018). URL https://spacy.io/



(RAKE algorithm) S. Rose, D. Engel, N. Cramer, W. Cowley, Automatic keyword extraction from individual documents, in: M. W. Berry, J. Kogan (Eds.), Text Mining. Applications and Theory, John Wiley and Sons, Ltd, 2010, pp. 1–20.

Precision evaluation



(Scheme 1) Jaccard LSA+LDA max
(Scheme 2) Jaccard LSA+LDA average
(Scheme 3) Jaccard Tweet max
(Scheme 4) Jaccard Tweet average
(Scheme 5) Spacy unique doc
(Scheme 6) Rake user

$$A = \frac{|R_u^a \cap R_u^h|}{k}, \forall u \in U'$$

Scenario

Consider you would like to visit Brno in Czech Republic and every POI there is new to you.

Given the fixed visiting time as 3 hours, you are not able to visit all the POIs.

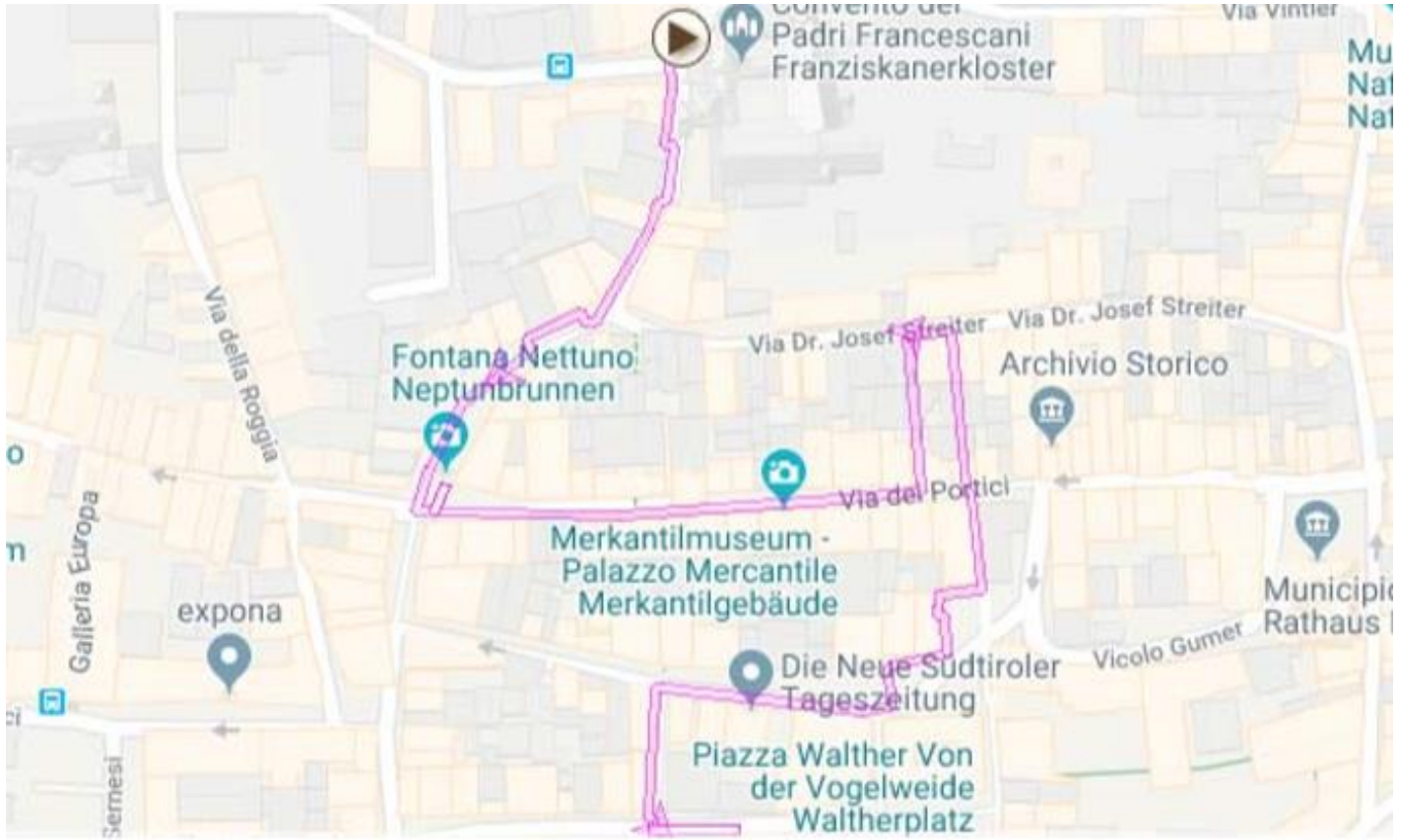
We have prepared a total of N proposed routes.

Each route contains a set of POIs with an orienteering route.

The transportation means is walking.

After reviewing all the proposed routes and related POIs, rank the N proposed routes based on their own preferences.

GUI in User evaluation



Research Roadmap

1. we plan to include user preference into the existing algorithm, to **make the existing solution personalized**.

2. personalization is done by social sensing, which is measuring **similarity between user SN posts and POI text corpus**, in this research, SN is Twitter and text corpus is POI descriptions.

3. we widen the research by using **different schemes to calculate the similarity**.

4. we can evaluate the **precision and performance**

5. we can evaluate the **user satisfaction**, we ask user to choose which recommended route he/she prefers. The **user decision can be used to choose which similarity scheme we will use**.

Thank you

Mouzhi Ge, Faculty of Informatics, Masaryk University, mouzhi.ge@muni.cz