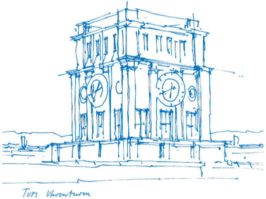
Building a Geo-Referenced Microsimulation Model with Discrete Optimization

Graphen und diskrete Optimierung



Kendra Reiter

10.05.2023

Joint work with Ulf Friedrich and Ralf Münnich

Outline

Motivation

Motivation

Census (2011): large-scale and comprehensive survey of German citizens

Motivation

Census (2011): large-scale and comprehensive survey of German citizens

- number of inhabitants

Motivation

Census (2011): large-scale and comprehensive survey of German citizens

- number of inhabitants
- housing conditions

Motivation

Census (2011): large-scale and comprehensive survey of German citizens

- number of inhabitants
- housing conditions
- working conditions

Motivation

Census (2011): large-scale and comprehensive survey of German citizens

- number of inhabitants
- housing conditions
- working conditions
- ...

Motivation

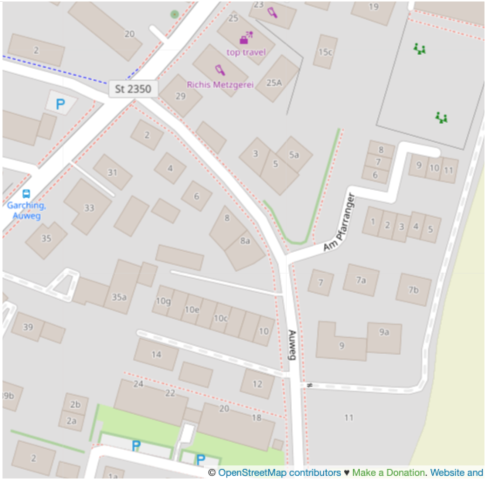
Census (2011): large-scale and comprehensive survey of German citizens

- number of inhabitants
- housing conditions
- working conditions
- ...

MikroSim: "Multi-sectoral Regional Microsimulation Model" (Trier University)

Problem Background

Address Selection



Address Selection

Address Selection

Address Selection: Goal

Goal

Create an assignment between households and dwellings which reflects the statistical properties of the real-life population.

Goal

Create an assignment between households and dwellings which reflects the statistical properties of the real-life population.

- Primary focus: *feasibility* of the assignment

Goal

Create an assignment between households and dwellings which reflects the statistical properties of the real-life population.

- Primary focus: *feasibility* of the assignment
- Secondary focus: statistical *quality* of the assignment

Challenge

Combine several data sources with different resolutions.

Data

Assumption (for this talk)

All necessary data are available, complete, cleaned, and pre-processed.

Data Structure

We have two datasets: Household data and Address data.

Data Structure

We have two datasets: **Household** data and Address data.

Unique ID	Household Size [number of persons]
	$= 1, \ldots, 6$

Data Structure

We have two datasets: Household data and **Address** data.

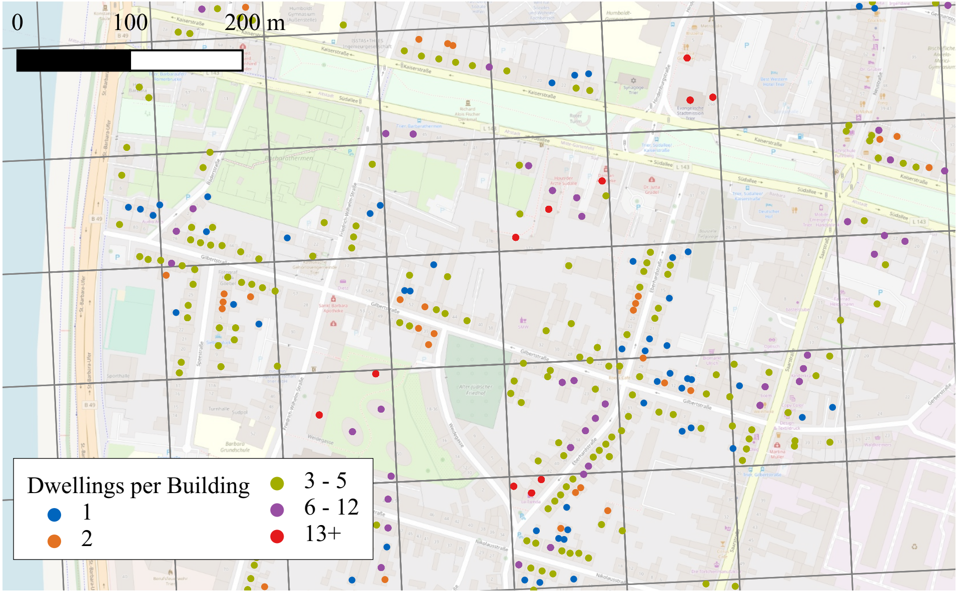
Unique ID	X-coord.	Y-coord.	Grid Cell	Dwellings [amount per building]
				$= 1, \ldots, 5$

Data Structure

We have two datasets: Household data and **Address** data.

Unique ID	X-coord.	Y-coord.	Grid Cell	Dwellings [amount per building]
				$= 1, \ldots, 5$

Each Address is split into individual dwellings according to a discrete Gaussian distribution with a peak at the real-life average.


0
100
200 m
Dwellings per Building
1
2
3 - 5
6 - 12
13+

Assumption

The more dwellings per address, the smaller the household size per dwelling.

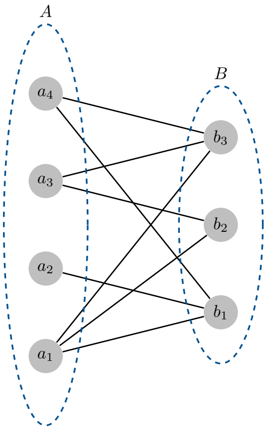
Graph Theory

Definitions

Definition (Bipartite Graph (Bipartiter Graph))
Consider a graph $G = (V, E)$ where the vertices V can be partitioned into two disjoint subsets A and B and each edge is of the form $e = \{a, b\}$ with $a \in A$ and $b \in B$. Then G is called a bipartite graph and is denoted by $G = (A, B, E)$.

Definitions

Definition (Bipartite Graph (Bipartiter Graph))
Consider a graph $G = (V, E)$ where the vertices V can be partitioned into two disjoint subsets A and B and each edge is of the form $e = \{a, b\}$ with $a \in A$ and $b \in B$. Then G is called a bipartite graph and is denoted by $G = (A, B, E)$.

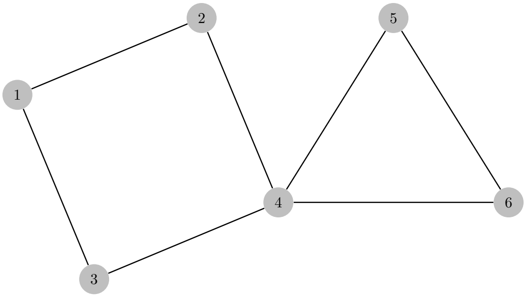


Definitions

Definition (Matching (Matching))
A matching M in a graph $G = (V, E)$ is a subset of the edges such that no two edges are incident to the same node.

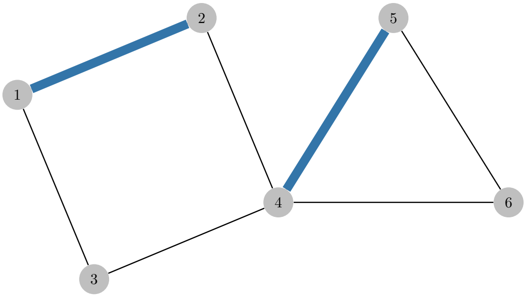
Definitions

Definition (Matching (Matching))
A matching M in a graph $G = (V, E)$ is a subset of the edges such that no two edges are incident to the same node.



Definitions

Definition (Matching (Matching))
A matching M in a graph $G = (V, E)$ is a subset of the edges such that no two edges are incident to the same node.



Definitions

Definition (Maximal Matching (nicht erweiterbar))
A matching M of a graph G which is not subset of any other matching is called a maximal matching.

Definitions

Definition (Maximal Matching (nicht erweiterbar))
A matching M of a graph G which is not subset of any other matching is called a maximal matching.

Definition (Maximum Matching (größte Paarung))
A matching M of a graph G which contains the largest possible number of edges in G is called a maximum matching.

Definitions

Definition (Maximal Matching (nicht erweiterbar))
A matching M of a graph G which is not subset of any other matching is called a maximal matching.

Definition (Maximum Matching (größte Paarung))
A matching M of a graph G which contains the largest possible number of edges in G is called a maximum matching.

Every maximum matching is maximal, but not every maximal matching is maximum.

Matching Problems

Maximum Matching
Given: Graph $G = (V, E)$

Want: Find Matching M in G with the maximum number of edges.

Matching Problems

Maximum Matching
Given: Graph $G = (V, E)$

Want: Find Matching M in G with the maximum number of edges.

Maximum Weight Matching
Given: Graph $G = (V, E)$ and weights $w : E \to \mathbb{R}$

Want: Find Matching M in G of maximum weight, i.e. $\sum_{e \in M} w_e$ maximal.

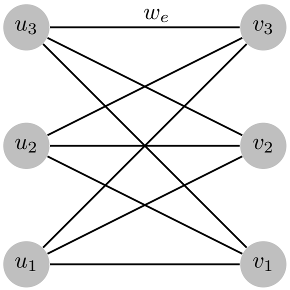
Matching Problems with Weights

Assignment Problem (Zuordnungsproblem)
Given: Bipartite Graph $G = (U, V, E)$ with $|U| = |V|$ and weights $w : E \to \mathbb{R}$
Want: Find maximum Matching M in G of minimum weight, i.e. $\sum_{e \in M} w_e$ is minimal.

Matching Problems with Weights

Assignment Problem (Zuordnungsproblem)
Given: Bipartite Graph $G = (U, V, E)$ with $|U| = |V|$ and weights $w : E \to \mathbb{R}$
Want: Find maximum Matching M in G of minimum weight, i.e. $\sum_{e \in M} w_e$ is minimal.

Example
Given three delivery drivers and three parcels, each driver has a different delivery speed for each parcel. The goal is to deliver all parcels in the shortest time possible. Which driver should deliver which parcel?

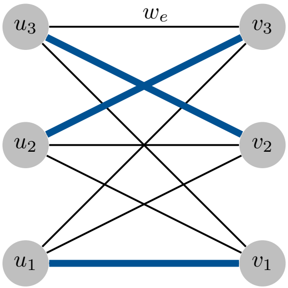


Matching Problems with Weights

Assignment Problem (Zuordnungsproblem)
Given: Bipartite Graph $G = (U, V, E)$ with $|U| = |V|$ and weights $w : E \to \mathbb{R}$
Want: Find maximum Matching M in G of minimum weight, i.e. $\sum_{e \in M} w_e$ is minimal.

Example

Given three delivery drivers and three parcels, each driver has a different delivery speed for each parcel. The goal is to deliver all parcels in the shortest time possible. Which driver should deliver which parcel?



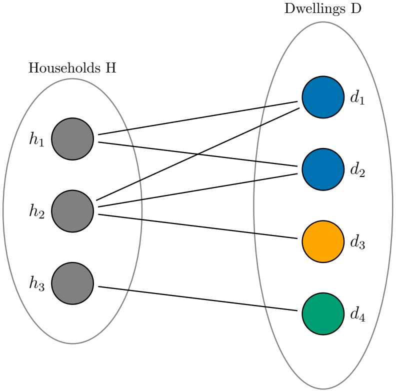
Mathematical Formulation

Matching

- Natural structure: bipartite graph $G = (H, D, E)$

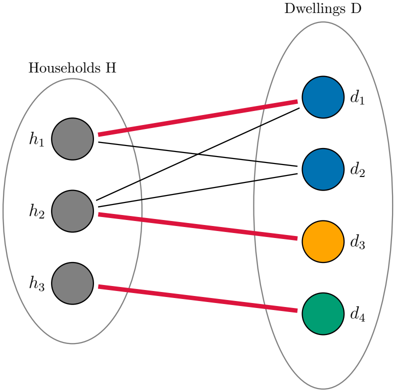
Matching

- Natural structure: bipartite graph $G = (H, D, E)$
- Draw edge $\{h, d\}$ if size of h is at most the capacity of dwelling d



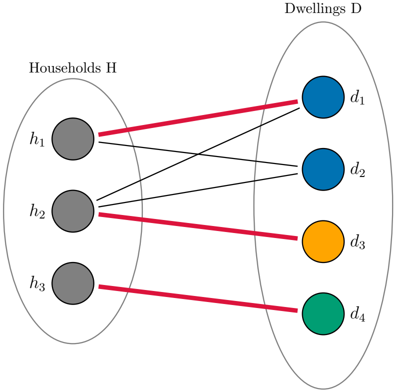
Matching

- Natural structure: bipartite graph $G = (H, D, E)$
- Draw edge $\{h, d\}$ if size of h is at most the capacity of dwelling d



Matching

- Natural structure: bipartite graph $G = (H, D, E)$
- Draw edge $\{h, d\}$ if size of h is at most the capacity of dwelling d
- Goal: match all households (if possible) or find a maximum matching

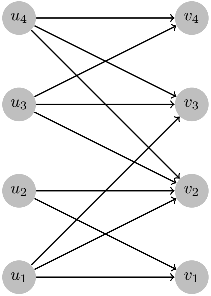


Solution Strategies

Finding Matchings

Given: Bipartite Graph $G = (U, V, E)$

Want: Create a flow network $\tilde{G} = (\tilde{W}, \tilde{E})$

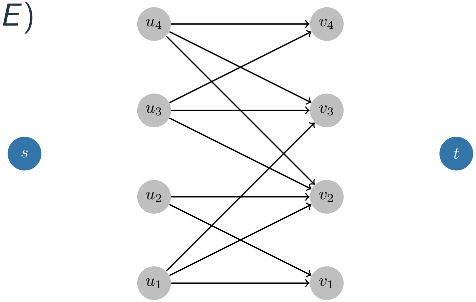


Finding Matchings

Given: Bipartite Graph $G = (U, V, E)$

Want: Create a flow network $\tilde{G} = (\tilde{W}, \tilde{E})$

- Introduce two new nodes: s (source) and t (sink)

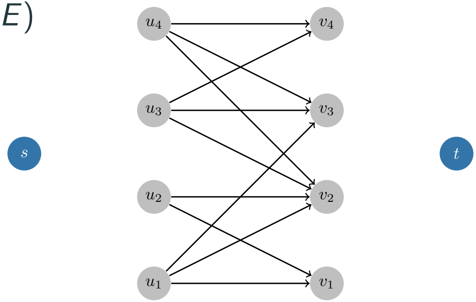


Finding Matchings

Given: Bipartite Graph $G = (U, V, E)$

Want: Create a flow network $\tilde{G} = (\tilde{W}, \tilde{E})$

- Introduce two new nodes: s (source) and t (sink)
- Set $\tilde{W} = U \cup V \cup \{s, t\}$

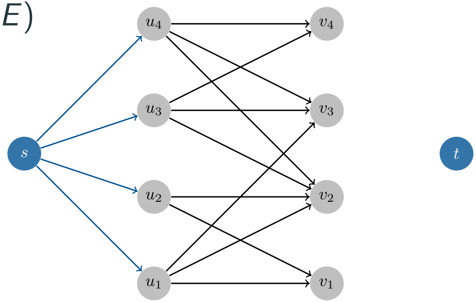


Finding Matchings

Given: Bipartite Graph $G = (U, V, E)$

Want: Create a flow network $\tilde{G} = (\tilde{W}, \tilde{E})$

- Introduce two new nodes: s (source) and t (sink)
- Set $\tilde{W} = U \cup V \cup \{s, t\}$
- Construct $\tilde{E}$: add edges $(s, u)\ \forall u \in U$

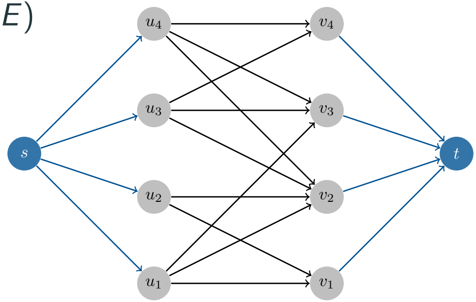


Finding Matchings

Given: Bipartite Graph $G = (U, V, E)$

Want: Create a flow network $\tilde{G} = (\tilde{W}, \tilde{E})$

- Introduce two new nodes: s (source) and t (sink)
- Set $\tilde{W} = U \cup V \cup \{s, t\}$
- Construct $\tilde{E}$: add edges $(s, u)\ \forall u \in U$ and $(v, t)\ \forall v \in V$ with capacity 1

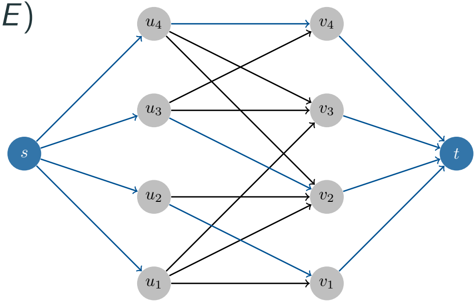


Finding Matchings

Given: Bipartite Graph $G = (U, V, E)$

Want: Create a flow network $\tilde{G} = (\tilde{W}, \tilde{E})$

- Introduce two new nodes: s (source) and t (sink)
- Set $\tilde{W} = U \cup V \cup \{s, t\}$
- Construct $\tilde{E}$: add edges $(s, u)\ \forall u \in U$ and $(v, t)\ \forall v \in V$ with capacity 1
- Find maximum flow in $\tilde{G}$



Theorem

The cardinality of the maximum matching in a bipartite Graph G is equal to the value of the maximum $s - t$-flow in $\tilde{G}$.

Theorem

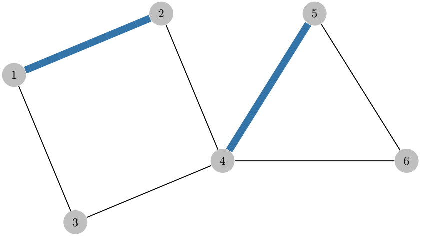
The cardinality of the maximum matching in a bipartite Graph G is equal to the value of the maximum $s - t$-flow in $\tilde{G}$.

Proof: Übungszettel 4.

Augmenting Path

Definition (Augmenting Path)
Let M be a matching on a graph $G = (V, E)$. A path $p = (v_1, \dots, v_k)$ is called M-augmenting if all of the following hold:

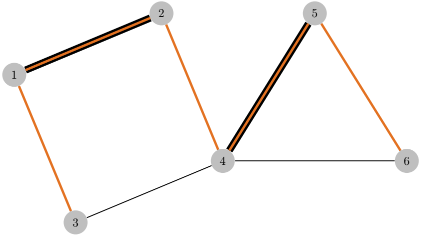
1. p has odd length, i.e. $k - 1$ is odd,
2. v_1 and v_k are not incident to a matched edge in M,
3. the edges of p are alternating in and out of M.



Augmenting Path

Definition (Augmenting Path)
Let M be a matching on a graph $G = (V, E)$. A path $p = (v_1, \ldots, v_k)$ is called M-augmenting if all of the following hold:

1. p has odd length, i.e. $k - 1$ is odd,
2. v_1 and v_k are not incident to a matched edge in M,
3. the edges of p are alternating in and out of M.



Augmenting Paths

Lemma

Consider a graph $G = (V, E)$ with a matching M. Let P be the set of edges of an augmenting path $p = (v_1, \dots, v_k)$. Then

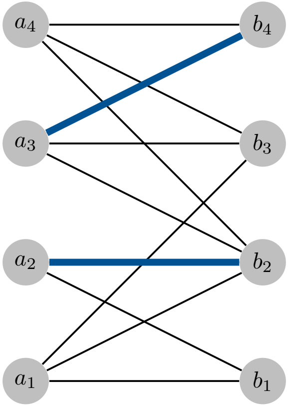
$$M' := (M \setminus P) \cup (P \setminus M) = M \triangle P$$

is a matching of cardinality $|M| + 1$ in G.

Augmenting Paths

Given: Bipartite Graph $G = (U, V, E)$

Want: Find a maximum matching in G.

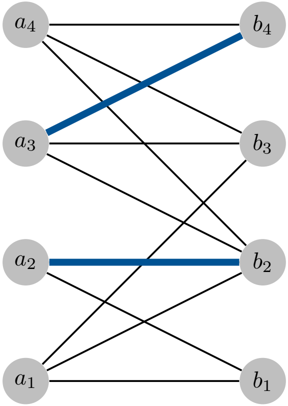


Augmenting Paths

Given: Bipartite Graph $G = (U, V, E)$

Want: Find a maximum matching in G.

- Start with an existing matching M

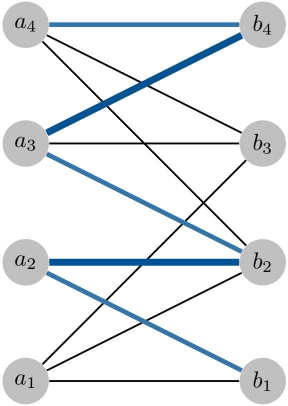


Augmenting Paths

Given: Bipartite Graph $G = (U, V, E)$

Want: Find a maximum matching in G.

- Start with an existing matching M
- Find an augmenting path p

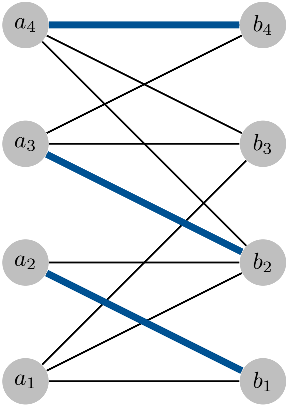


Augmenting Paths

Given: Bipartite Graph $G = (U, V, E)$

Want: Find a maximum matching in G.

- Start with an existing matching M
- Find an augmenting path p
- Construct a new matching $M' = M \triangle P$

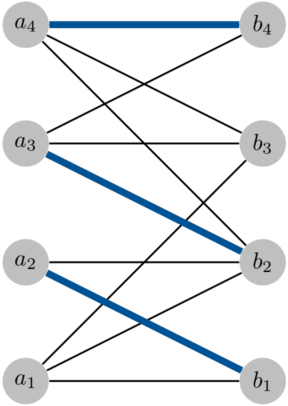


Augmenting Paths

Given: Bipartite Graph $G = (U, V, E)$

Want: Find a maximum matching in G.

- Start with an existing matching M
- Find an augmenting path p
- Construct a new matching $M' = M \triangle P$
- Repeat until no further augmenting paths can be found.



Berge's Lemma

Theorem

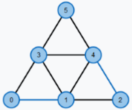
Let $G = (V, E)$ be a graph and let M be a matching in G. Then M is a maximum matching if and only if there does not exist an augmenting path in G with respect to M.

Main Algorithms

- **Using max. flow**: maximal matching in a bipartite graph ($\mathcal{O}(\sqrt{|V|} \cdot |E|)$)
- **Hopkraft-Karp-Algorithm**: maximal matching in a bipartite graph ($\mathcal{O}(\sqrt{|V|} \cdot |E|)$)
- **Blossom Algorithm (Edmonds)**: maximal matching in an arbitrary graph ($\mathcal{O}(|V|^2 \cdot |E|)$)
- **Hungarian Method**: maximal weighted matching in bipartite graph ($\mathcal{O}(|V|^3)$)

Main Algorithms

Click here to try.

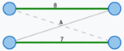


Blossom Algorithm

The Blossom Algorithm computes a maximal matching in an arbitrary graph (in contrast to the Hopcroft-Karp algorithm which requires a bipartite graph).

Hopcroft-Karp Algorithm

The Hopcroft-Karp Algorithm computes for two given sets with possible assignments a maximal relation.



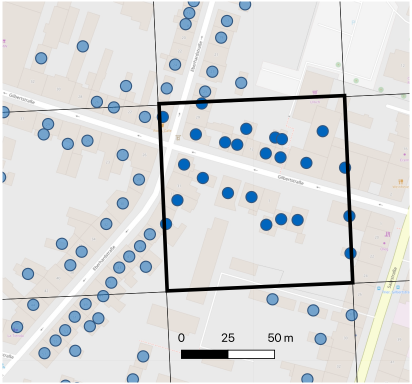
Hungarian Method

The Hungarian Method determines for two given sets with weighted assignments a maximal (with respect to the weights) relation.

Mathematical Formulation II

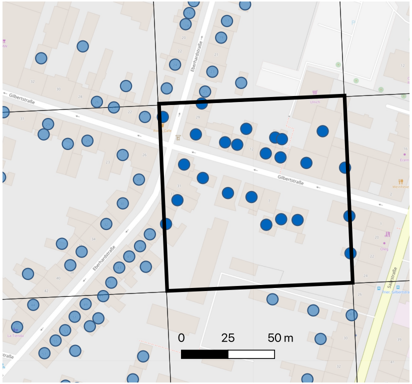
Constraints

- Per Grid Cell



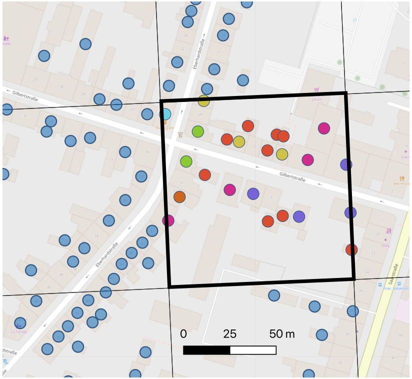
Constraints

- Per Grid Cell
- C_1: Limit the amount of households per grid cell



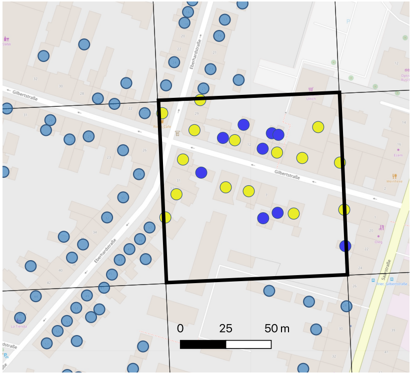
Constraints

- Per Grid Cell
- C_1: Limit the amount of households per grid cell
- C_2: Limit the amount of persons per grid cell



Constraints

- Per Grid Cell
- C_1: Limit the amount of households per grid cell
- C_2: Limit the amount of persons per grid cell
- C_3: Ensure that the household size 'fits' into the dwelling capacity



Constraints

- Per Grid Cell
- C_1: Limit the amount of households per grid cell
- C_2: Limit the amount of persons per grid cell
- C_3: Ensure that the household size 'fits' into the dwelling capacity

Want: Maximum weight Matching which respects *all* constraints.

Constraints

- Per Grid Cell
- C_1: Limit the amount of households per grid cell
- C_2: Limit the amount of persons per grid cell
- C_3: Ensure that the household size 'fits' into the dwelling capacity

Want: Maximum weight Matching which respects *all* constraints.

How?

Matching Formulation: Objective Function

Want: Maximum weight Matching which respects *all* constraints.

Idea: Introduce weights $\lambda_1, \lambda_2, \lambda_3 > 0$

Matching Formulation: Objective Function

Want: Maximum weight Matching which respects *all* constraints.

Idea: Introduce weights $\lambda_1, \lambda_2, \lambda_3 > 0$

$$\text{Maximum Weights} \tag{1}$$

$$+\lambda_1 \cdot C_1 : \text{ Limit the amount of households per grid cell} \tag{2}$$

$$+\lambda_2 \cdot C_2 : \text{ Limit the amount of persons per grid cell} \tag{3}$$

$$+\lambda_3 \cdot C_3 : \text{ Ensure that the household size `fits' into the dwelling capacity} \tag{4}$$

Solving *this* Optimization...

... is not easy.

Solving *this* Optimization...

... is not easy.

Cannot use the objective function as weights on the graph ⇒ solved as Mixed-Integer Linear Program (MILP).

More details in Part 2 of this course!

Outlook

Use the implementation on larger datasets:

Assign a population of 105k in 56k households to 59k dwellings at 20k addresses

References

[1] Ulf Friedrich, Ralf Münnich, and Kendra M. Reiter. **"Building a geo-referenced microsimulation model with discrete optimization".** Online Conference. Conference on New Techniques and Technologies for Statistics. 2021. URL: https://coms.events/NTTS2021/data/abstracts/en/abstract_0002.html.

[2] Ralf Münnich et al. **"A Population Based Regional Dynamic Microsimulation of Germany: The MikroSim Model".** In: *methods, data, analyses* (2021), p. 23.

[3] Christos H. Papadimitriou and Kenneth Steiglitz. ***Combinatorial Optimization: Algorithms and Complexity.*** Dover Publications, 1998.

[4] Kendra M. Reiter. **"A Weighted Matching Model for Georeferenced Microsimulations".** MA thesis. Technical University Munich, 2021.

[5] Alexander Schrijver. ***Combinatorial Optimization.*** Springer, 2002.

[6] Statistische Ämter des Bundes und der Länder. ***Overview of the register-based census.*** 2011. URL: https://www.zensus2011.de/EN/2011Census/Methodology/Methodology_node.html.