

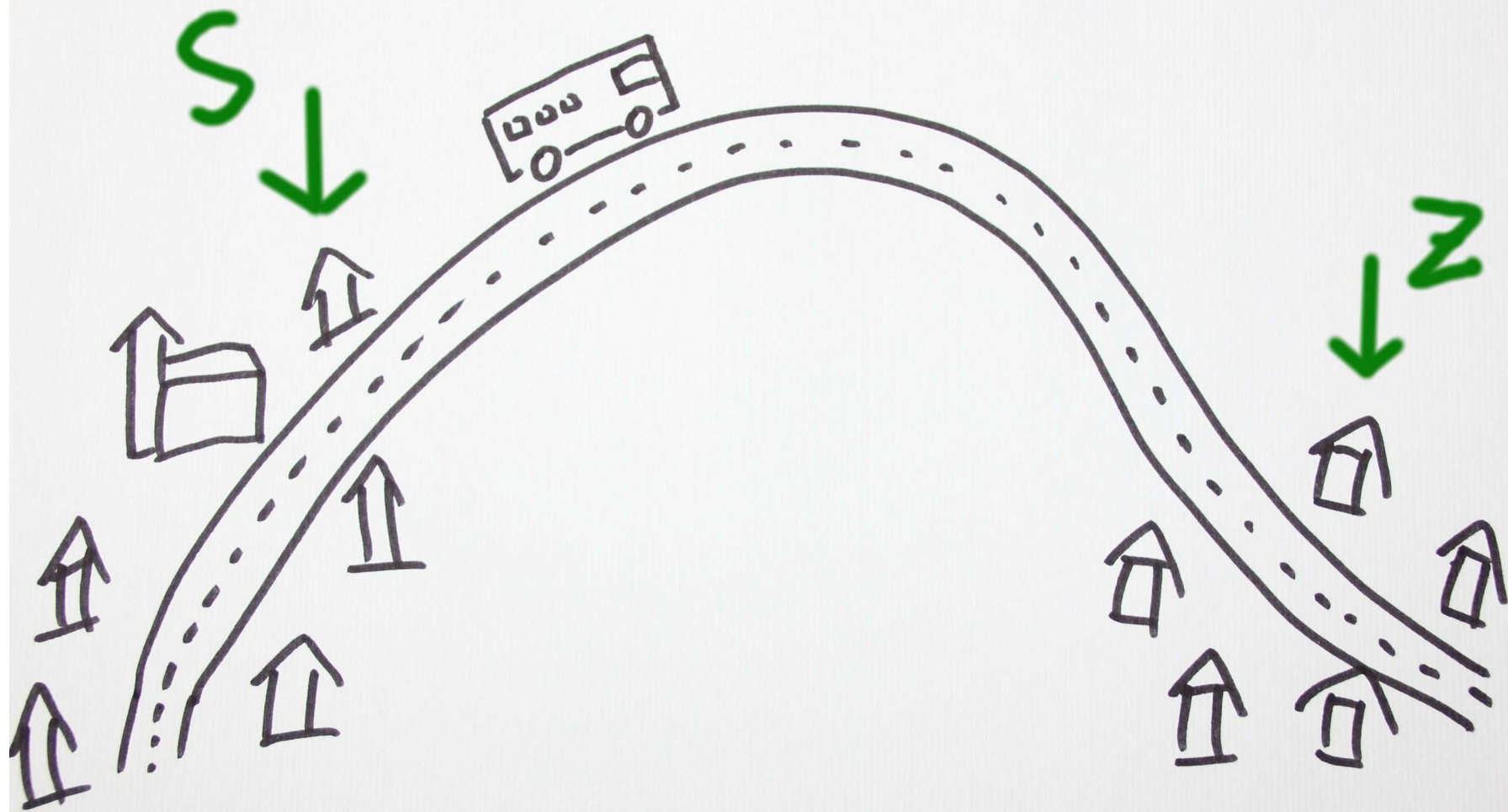
Clusterbildung von Geoobjekten im Straßennetz

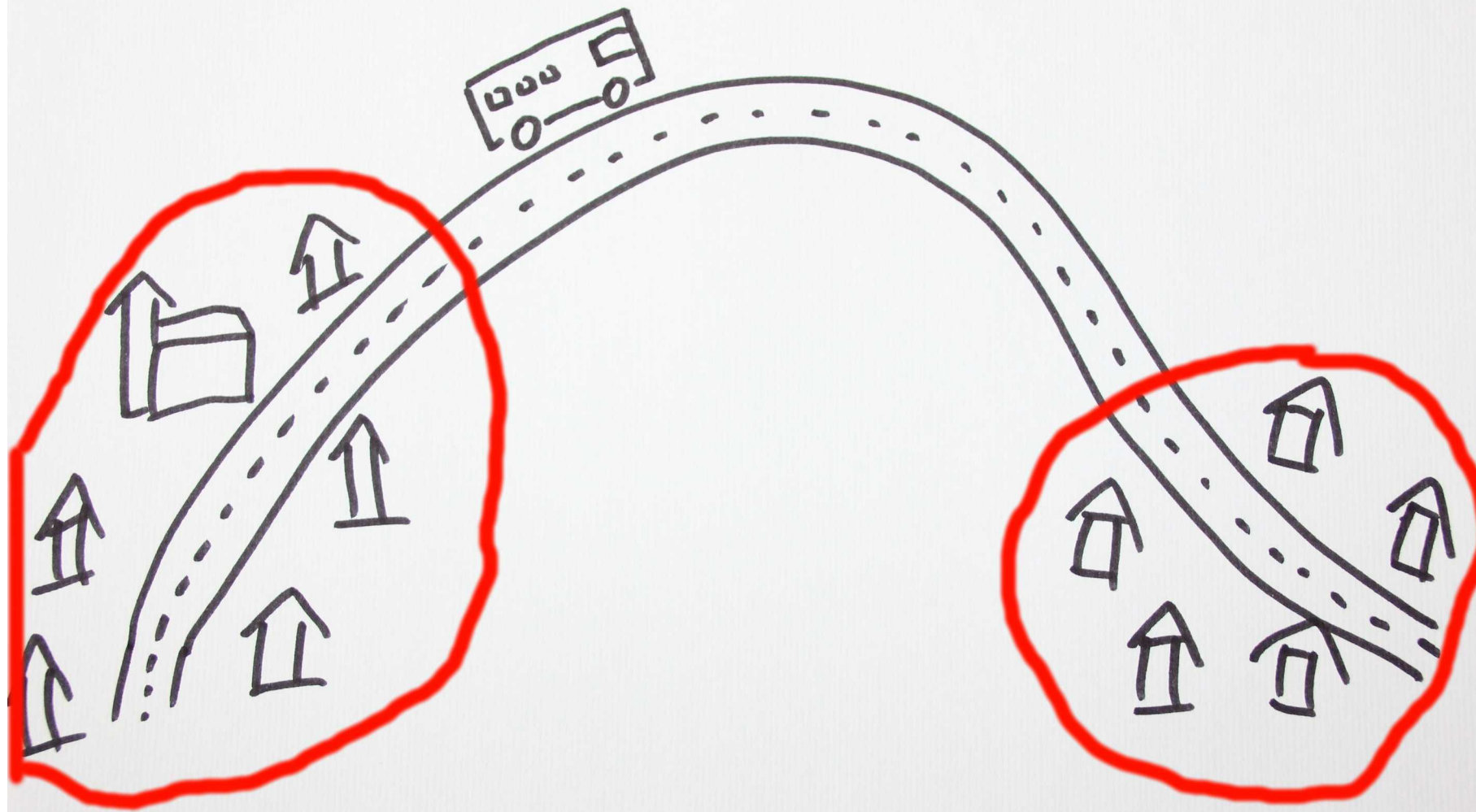
Georg Ottinger

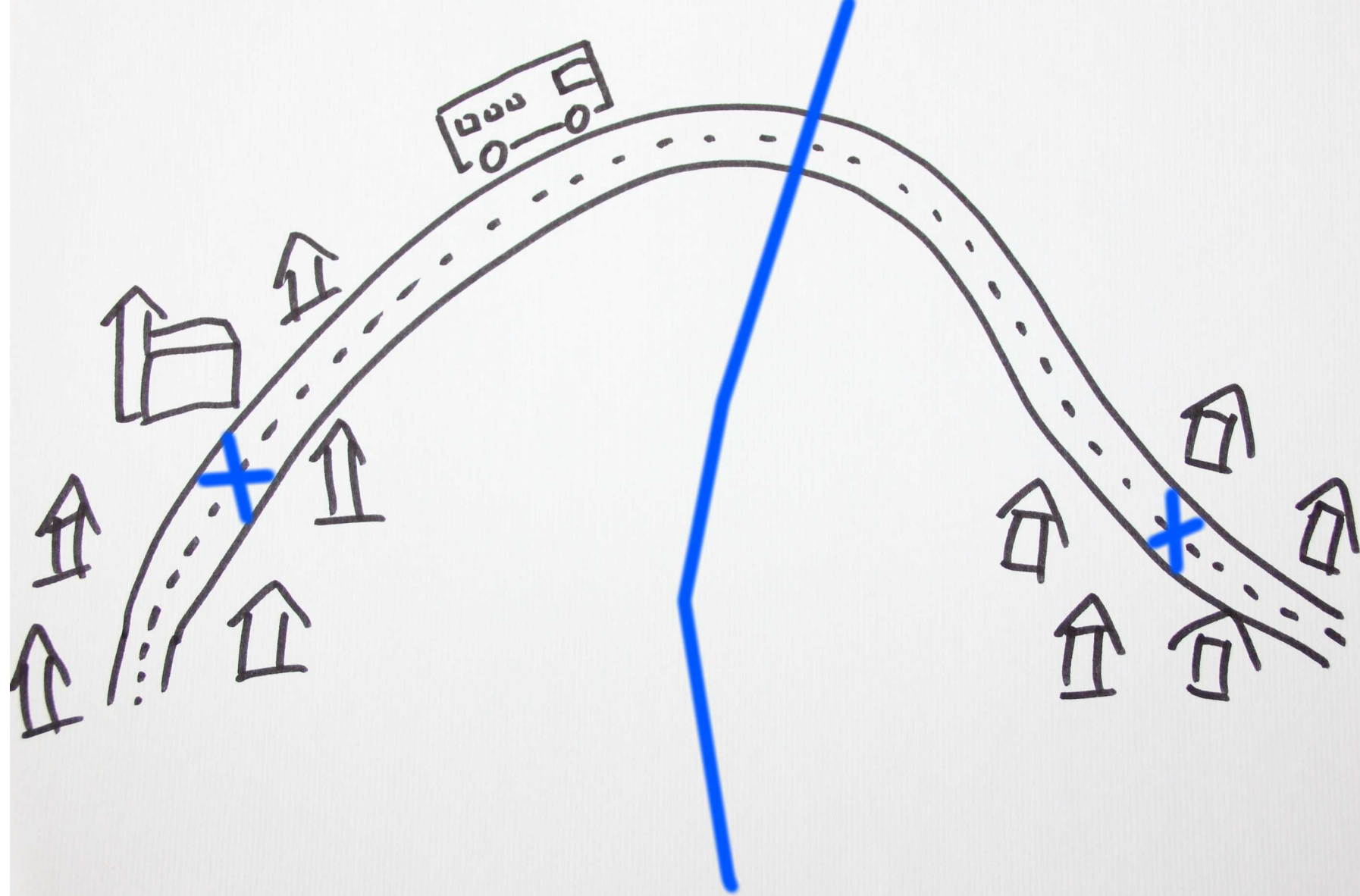
Überblick

- Problemstellung
- Beispiel: Clustern in der Ebene
- Ideen / Vorüberlegung
- Datenquellen / Tools
- Geoobjekte Aggregation (kurz GOA)

Problemstellung







Gesetzliche Rahmenbedingungen

„Um mehr Rechtssicherheit zu erreichen, sollten Auflösungsschwellen, ab denen, bei Personenbeziehbarkeit eines Datums, eine persönlichkeitsrechtliche Relevanz gegeben ist, für die gängigen Georeferenzierungen abgestimmt werden.

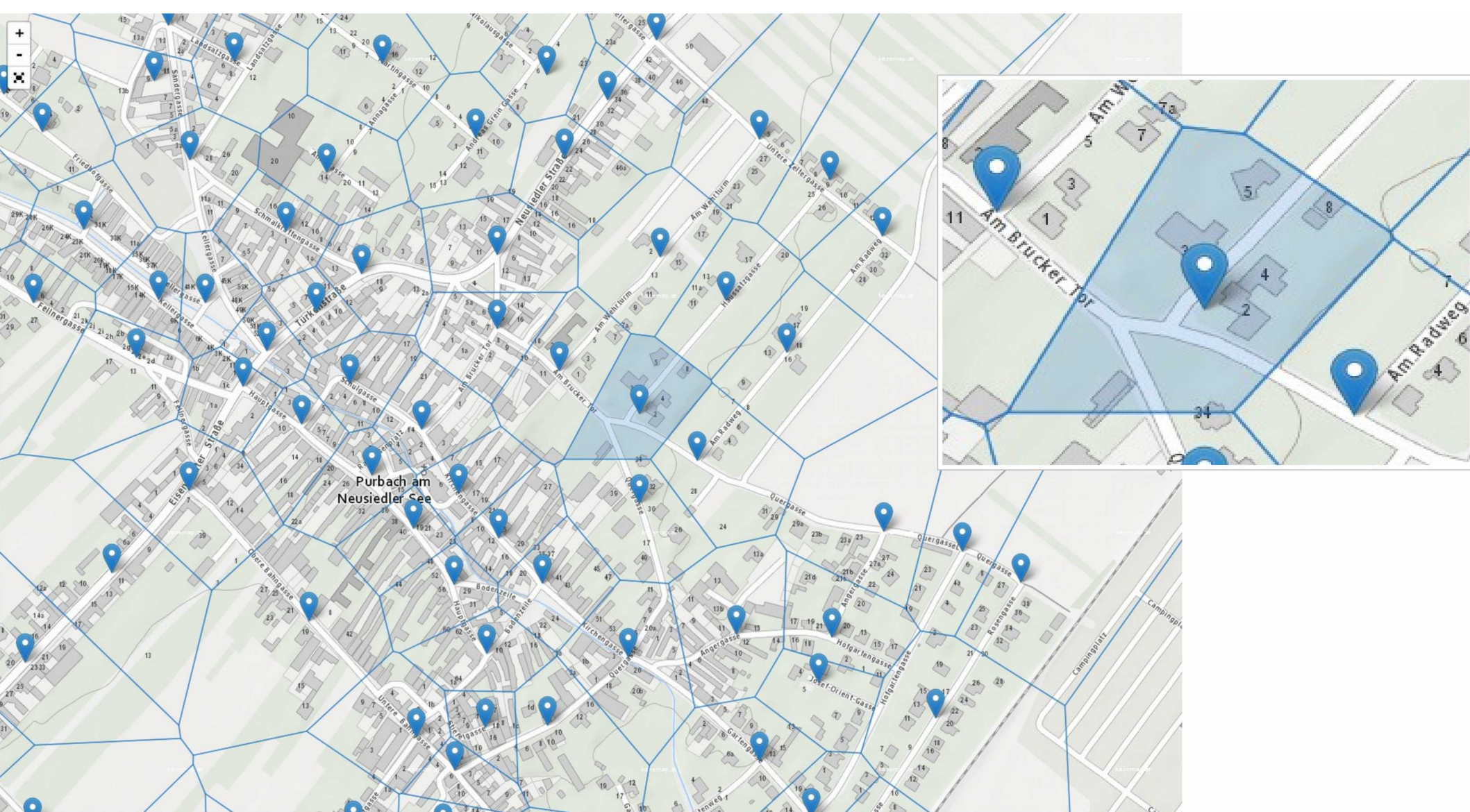
Keine Schutzbedürftigkeit könnte aus datenschutzrechtlicher Sicht bestehen **bei [..] mindestens auf vier Haushalte aggregierten Informationen“**

D. B. des Inneren. Dritter GeoFortschrittsbericht der Bundesregierung, 2012

Problemstellung

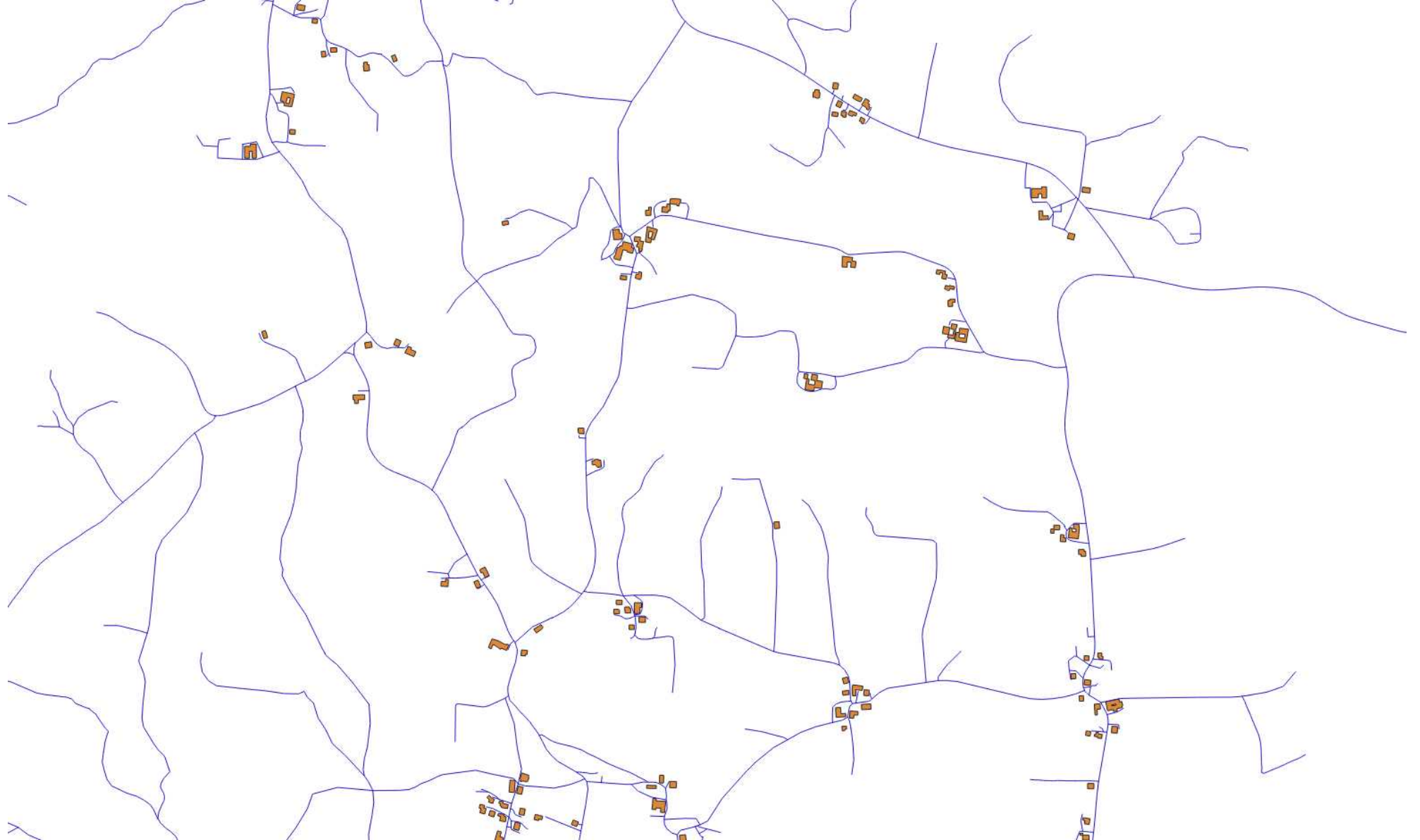
- Einteilung der Ebene in Territorien
- Jedes Territorium wird durch ein Zentrum repräsentiert
- Jedes Territorium muss eine Mindestanzahl an Geoobjekten beinhalten (α -Anonymität als untere Schranke)
- NB: Distanzen der Geoobjekte zu den entsprechenden Zentren sollen minimiert werden (“minsum” - Kompaktheit)

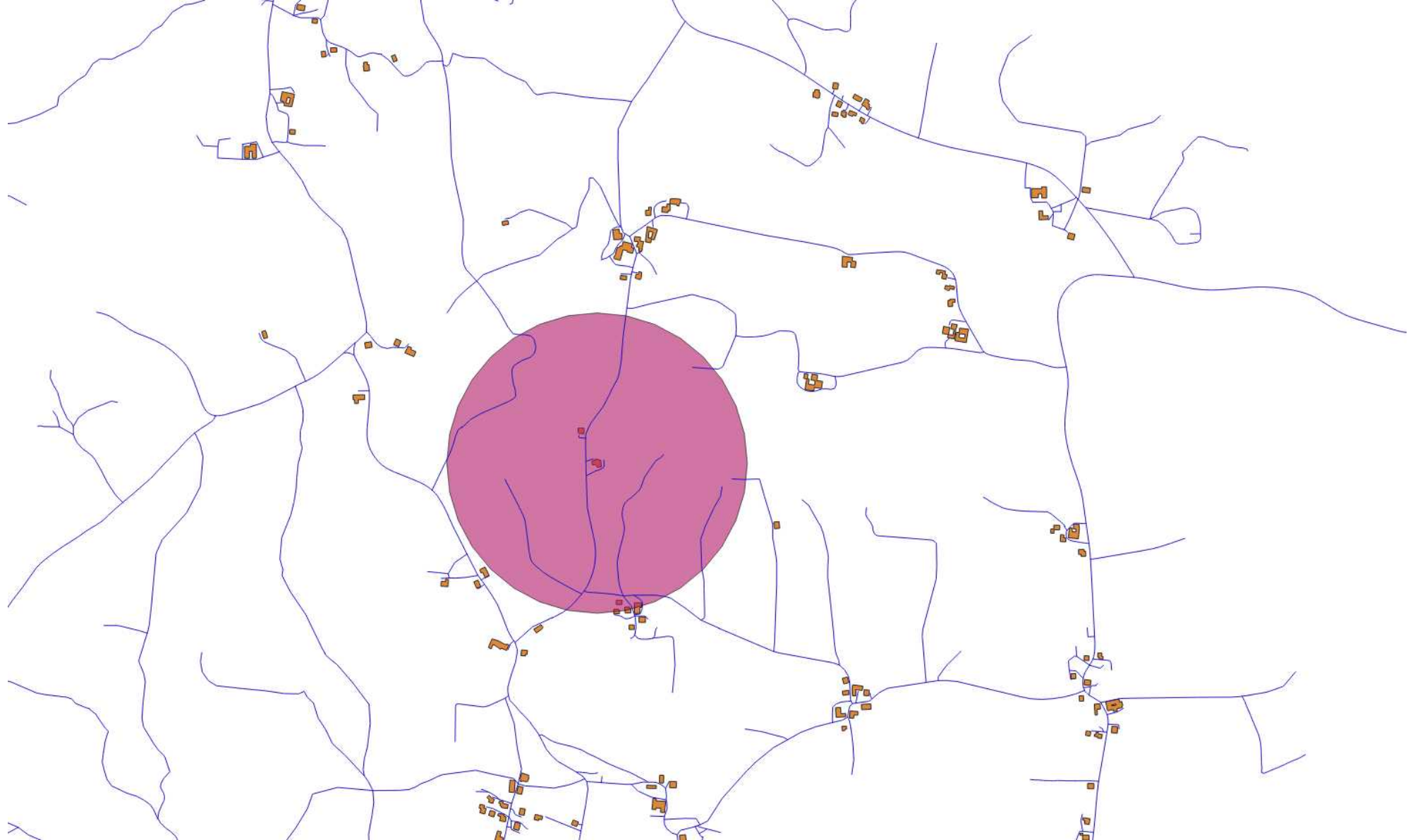
=> Problem der Gebietseinteilung

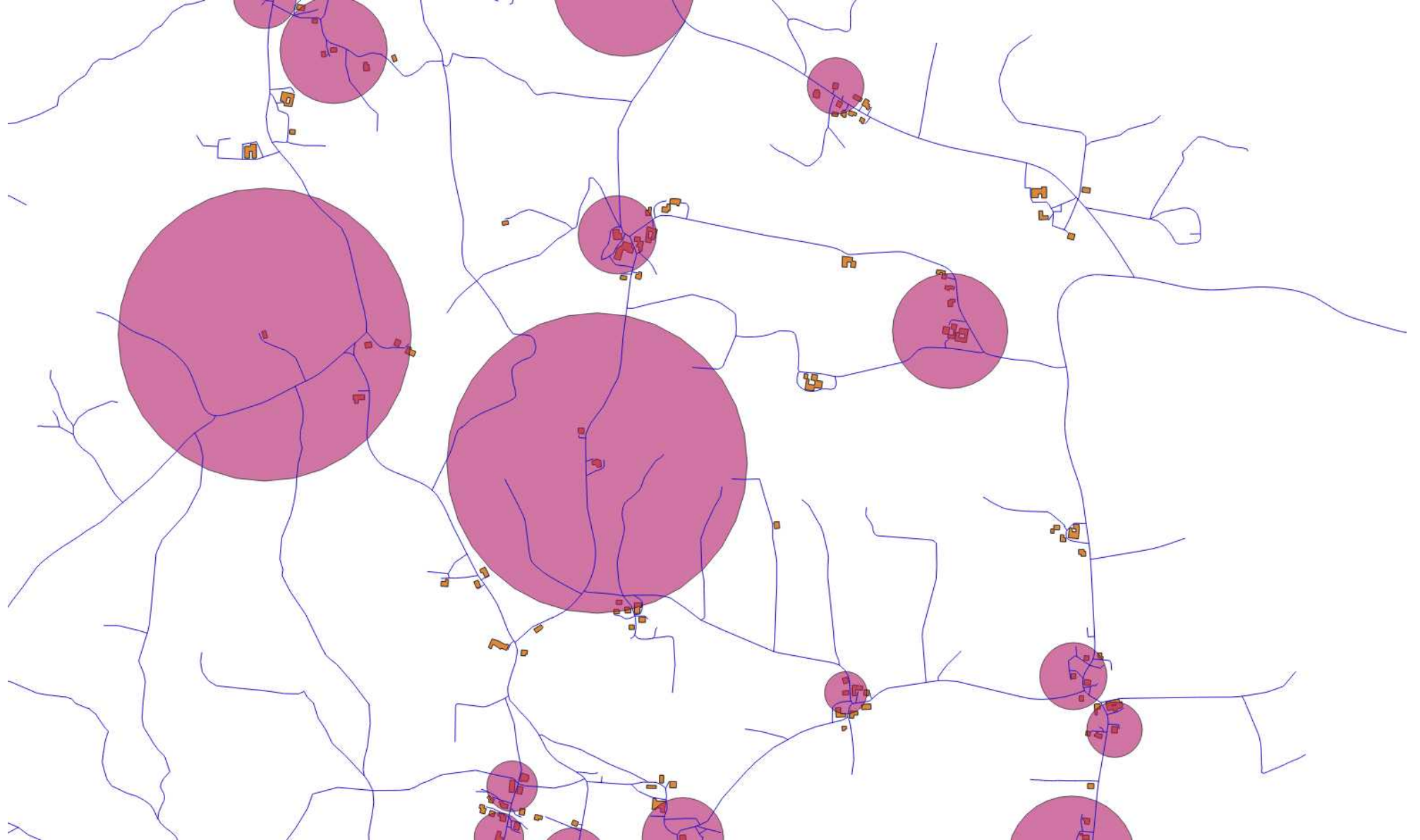


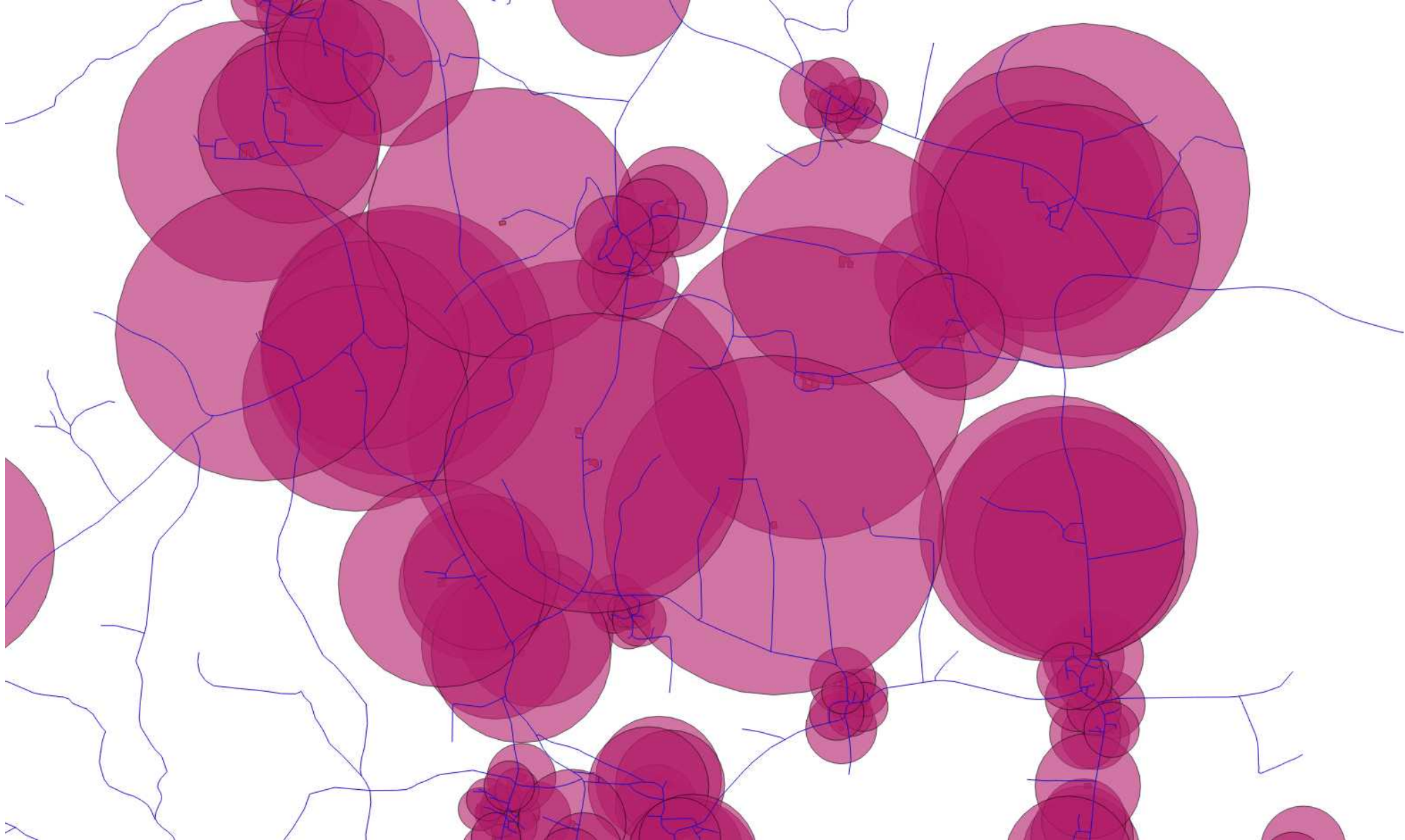
Clustern in der Ebene

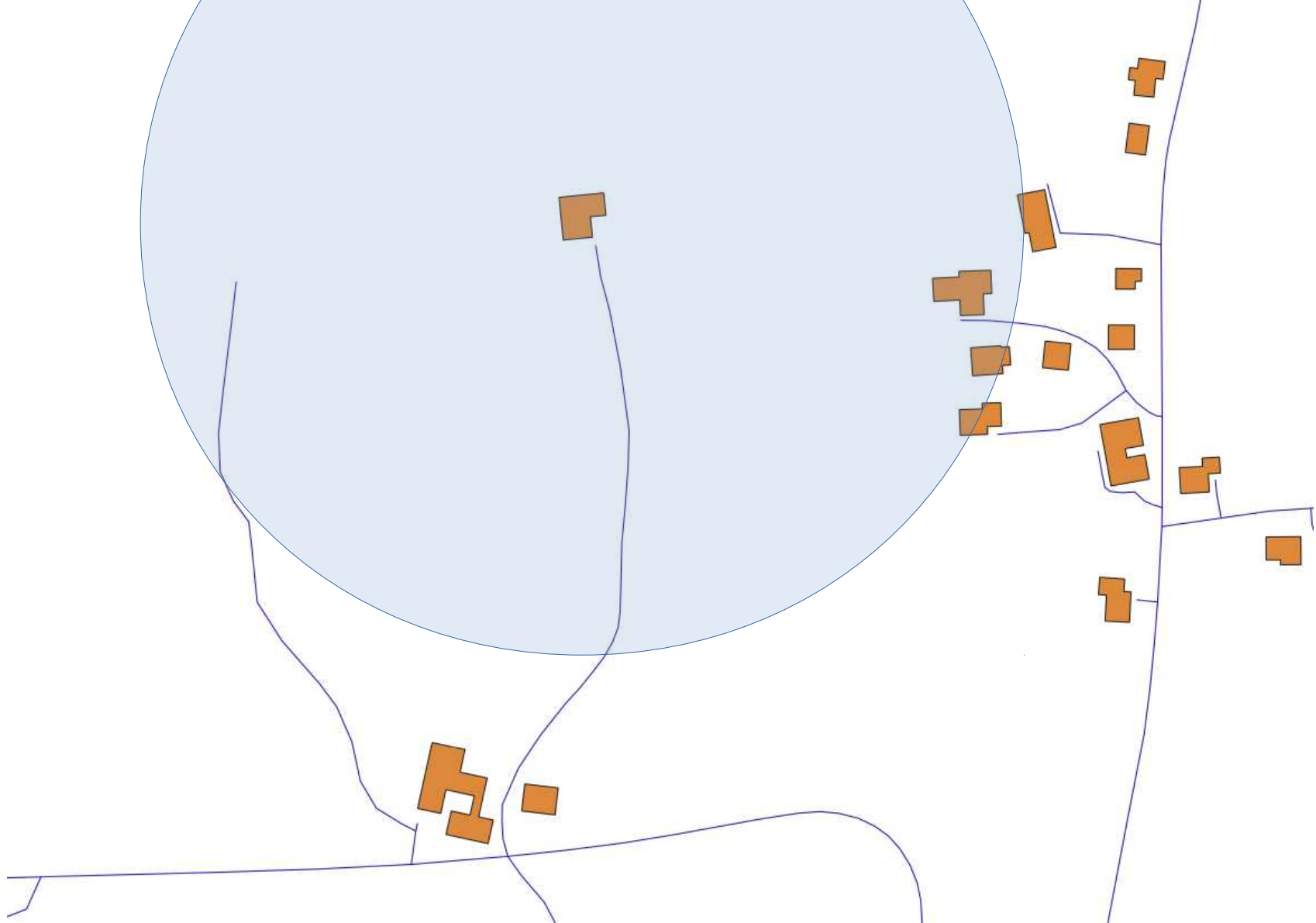
Similarity Partitioning nach Güting





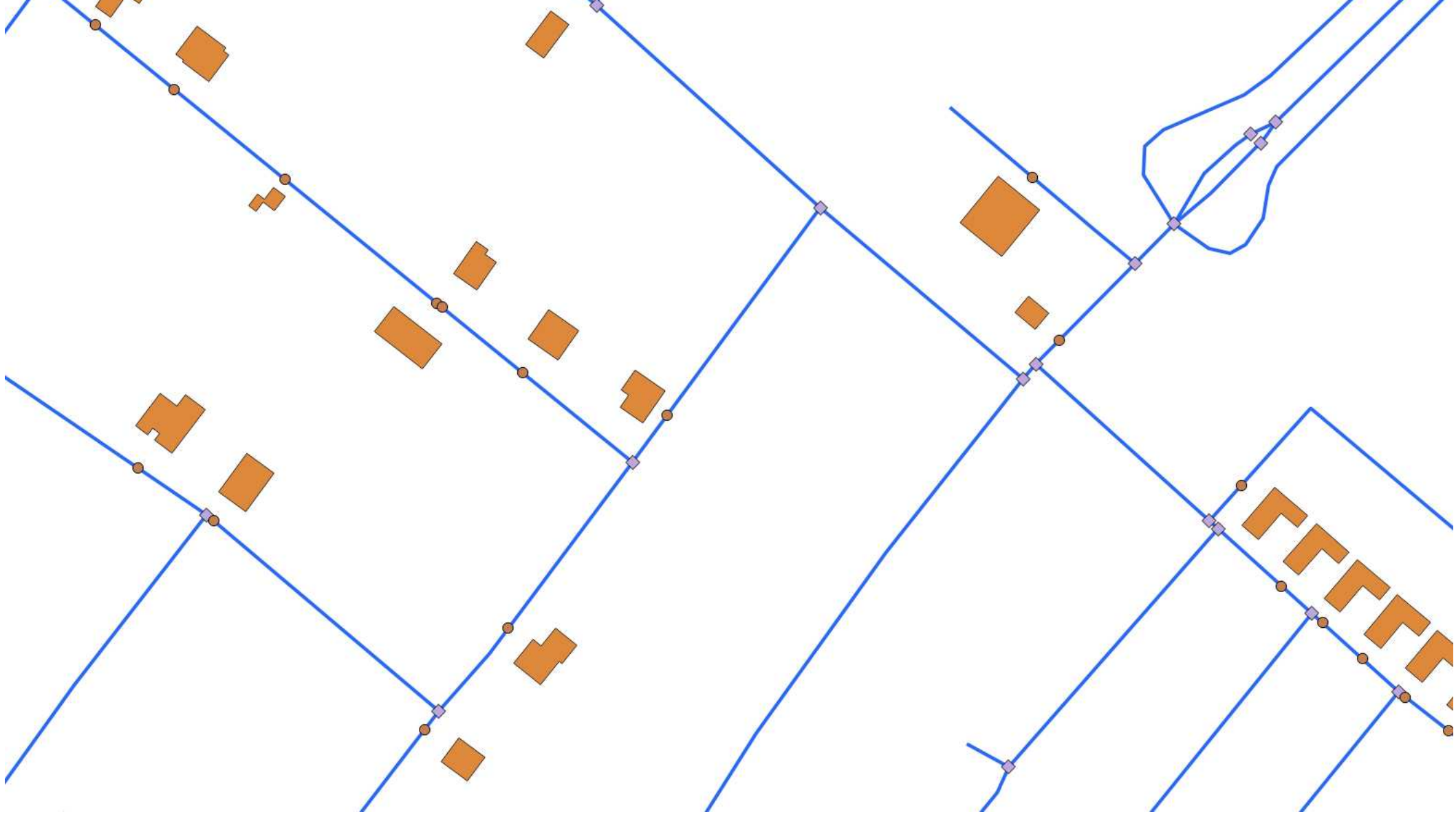


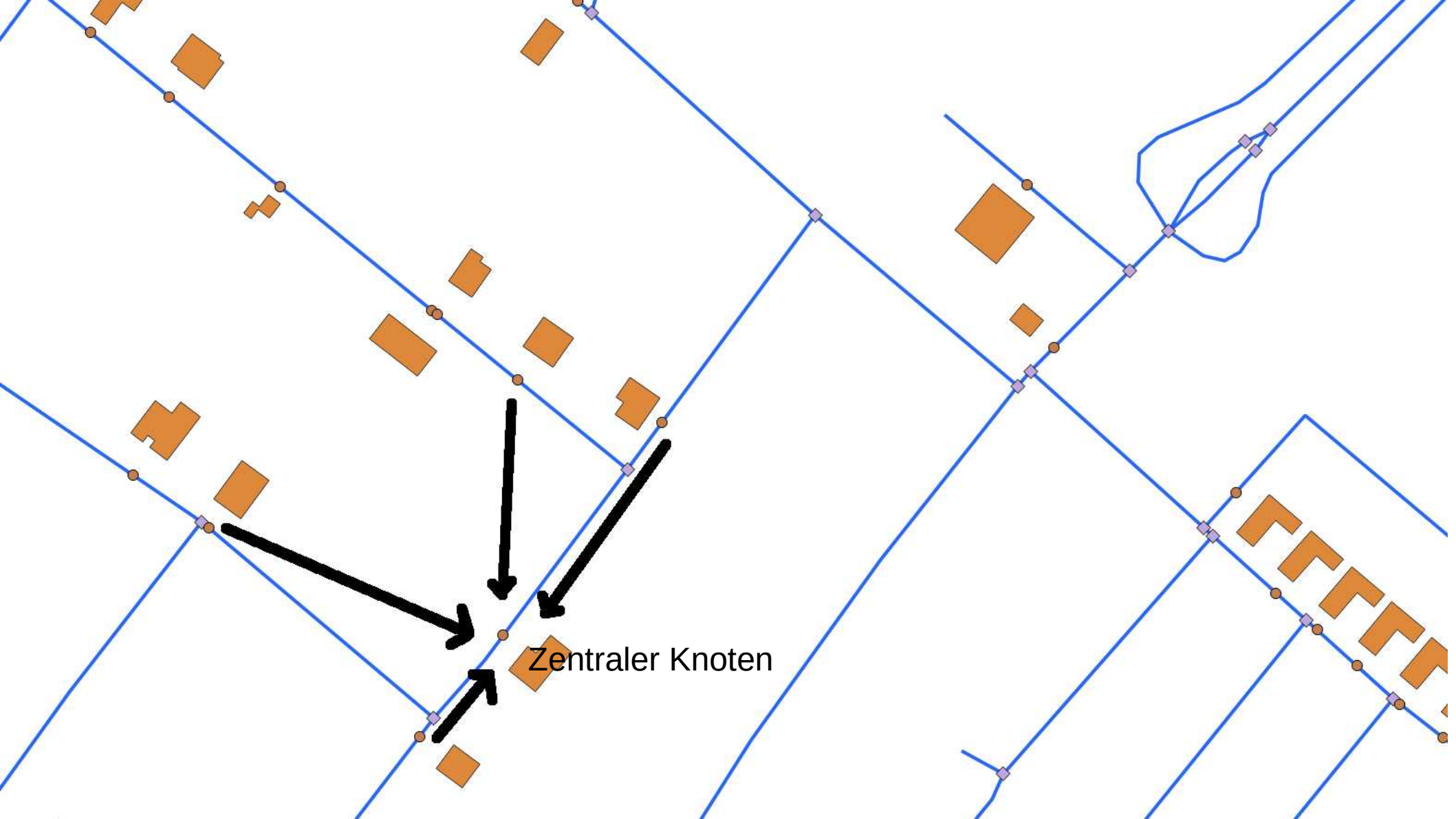




Ideen / Vorüberlegung

- Straßennetz einbinden
- Auswahl optimieren für bessere Lösungen
- Einbinden von Residuen





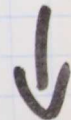
\mathbb{R}^2

G

Geoobjekte
+ Straßennetz



GSD-Graph



GOA



Territorien

Datenquellen / Tools



Testdatensätze

| | Neukirchen | Purbach | Wien 15ter |
|--|------------|---------|------------|
| Anz. Knoten | 1956 | 1280 | 3880 |
| Anz. Kanten | 2151 | 1473 | 4209 |
| Graph Dichte [ppm] | 1125 | 1799 | 559 |
| Anz. Geoobjekte | 792 | 622 | 3158 |
| Anz. Kreuzungen ($deg > 2$) | 746 | 406 | 528 |
| Anz. unechter Kreuzungen ($deg = 2$) | 418 | 252 | 194 |
| durchschnittlicher Grad von Kreuzungen | 3.04 | 3.11 | 3.32 |
| Straßennetz [km] | 134.13 | 161.89 | 77.42 |

Tab. 2: Eckdaten der Testdatensätze

GOA Algortihmus (vereinfacht)

1. Teilgraphen finden
2. Eindeutig bestimmte TG auswählen
3. **GA zur optimierten Auswahl von TG**
4. Residuen zuordnen
5. Zentren neu berechnen
6. **Territorien in der Ebene erstellen.**

KNN im Graph

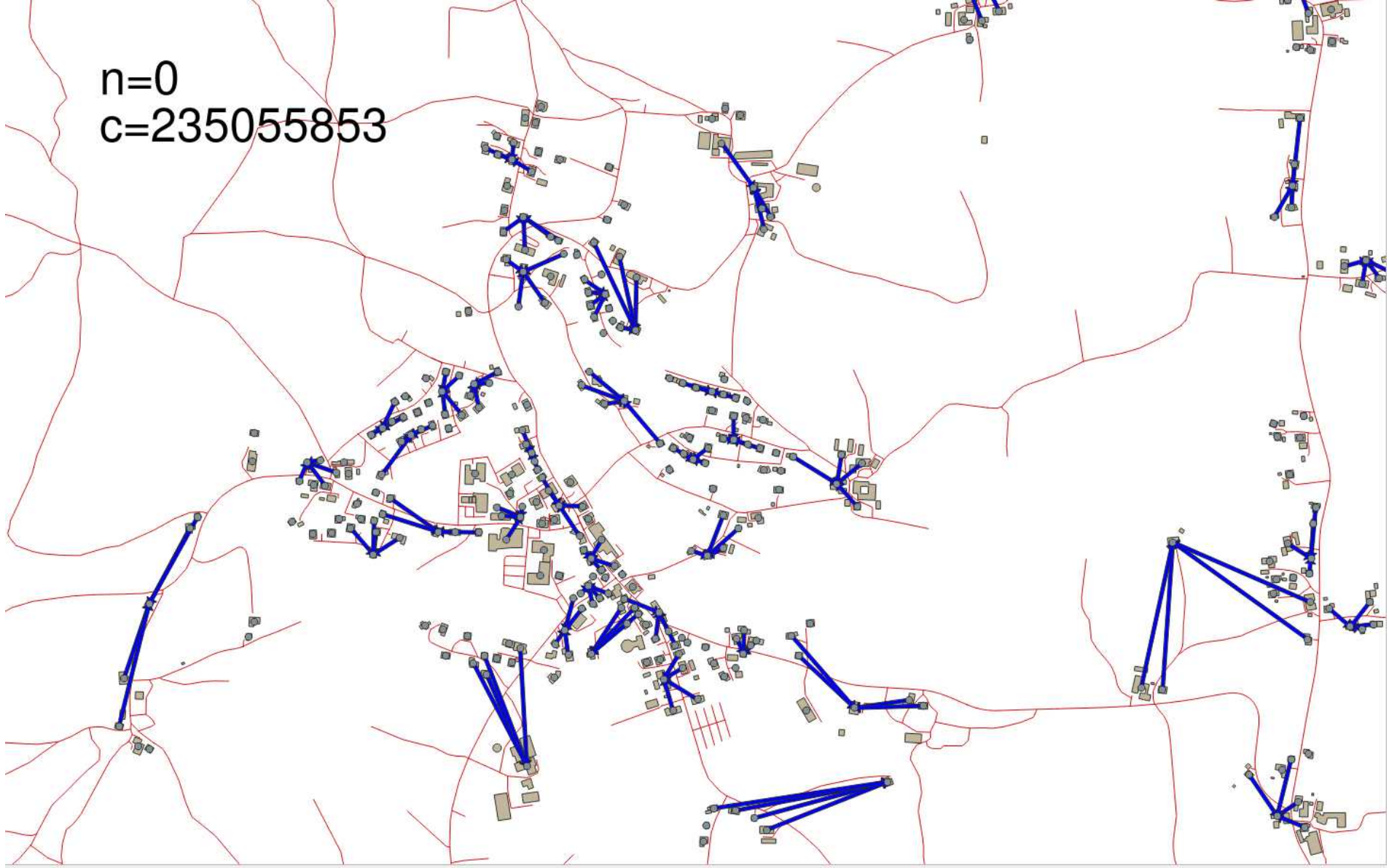


- Geodesic Knn
- Eingangsdaten: kantengewichteter Graph mit gelabelte und ungelabelte Knoten
- Ausgangsdaten: Listen der kNN aus der Menge der gelabelten Daten für alle Knoten (+Distanzen)
- Algorithmus: Dijkstra-artig mit gleichzeitigem Start an allen gelabelten Knoten.

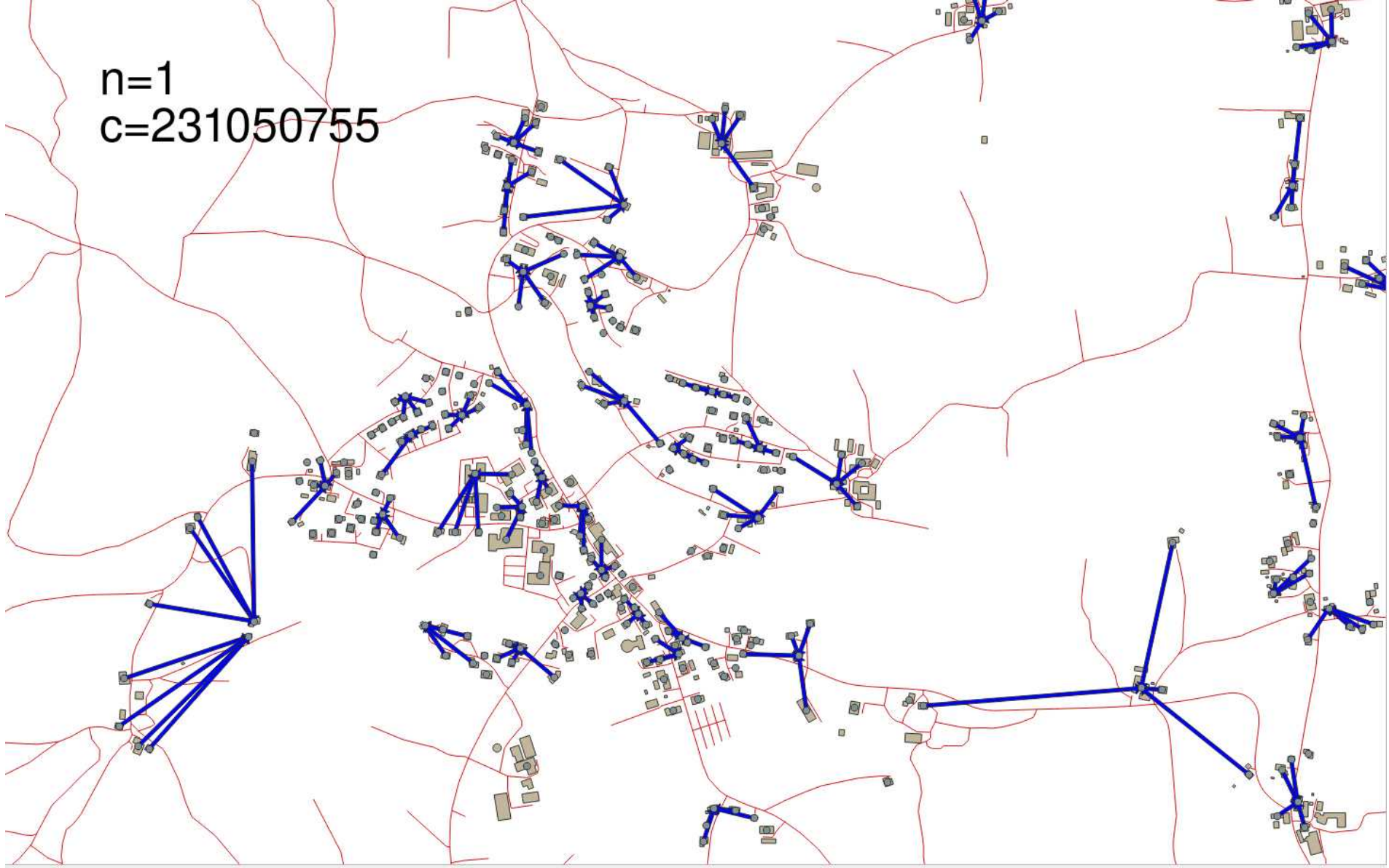
GA zur Auswahl der Teilgraphen

- Lösungsvektor $s = [0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ \dots\ 0]$
- Population $p = 100$; Runden $n = 100$
- Tournament Selection
- 50:50 Crossover
- Bereinigungsschritt
- Elitism

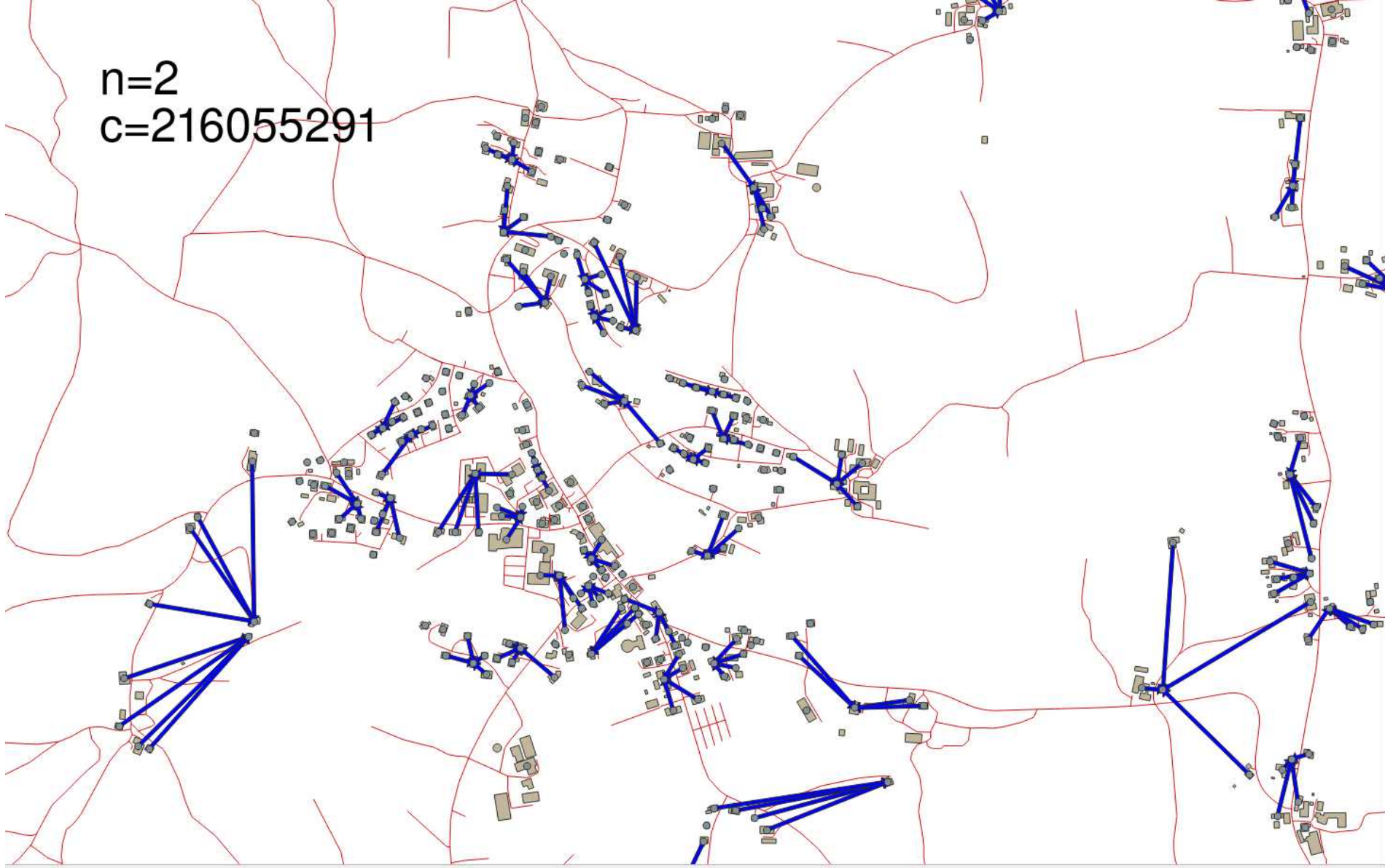
$n=0$
 $c=235055853$



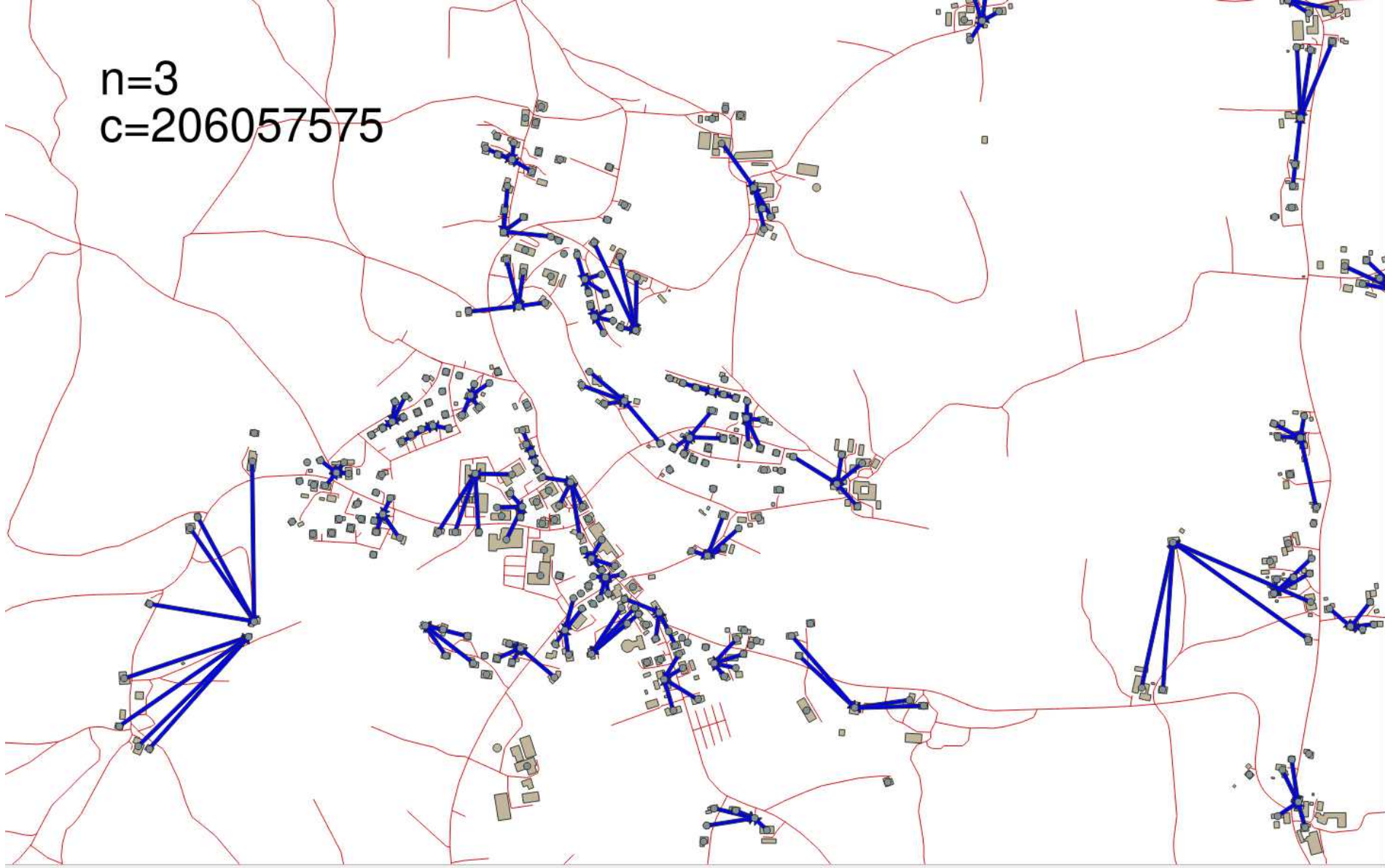
$n=1$
 $c=231050755$



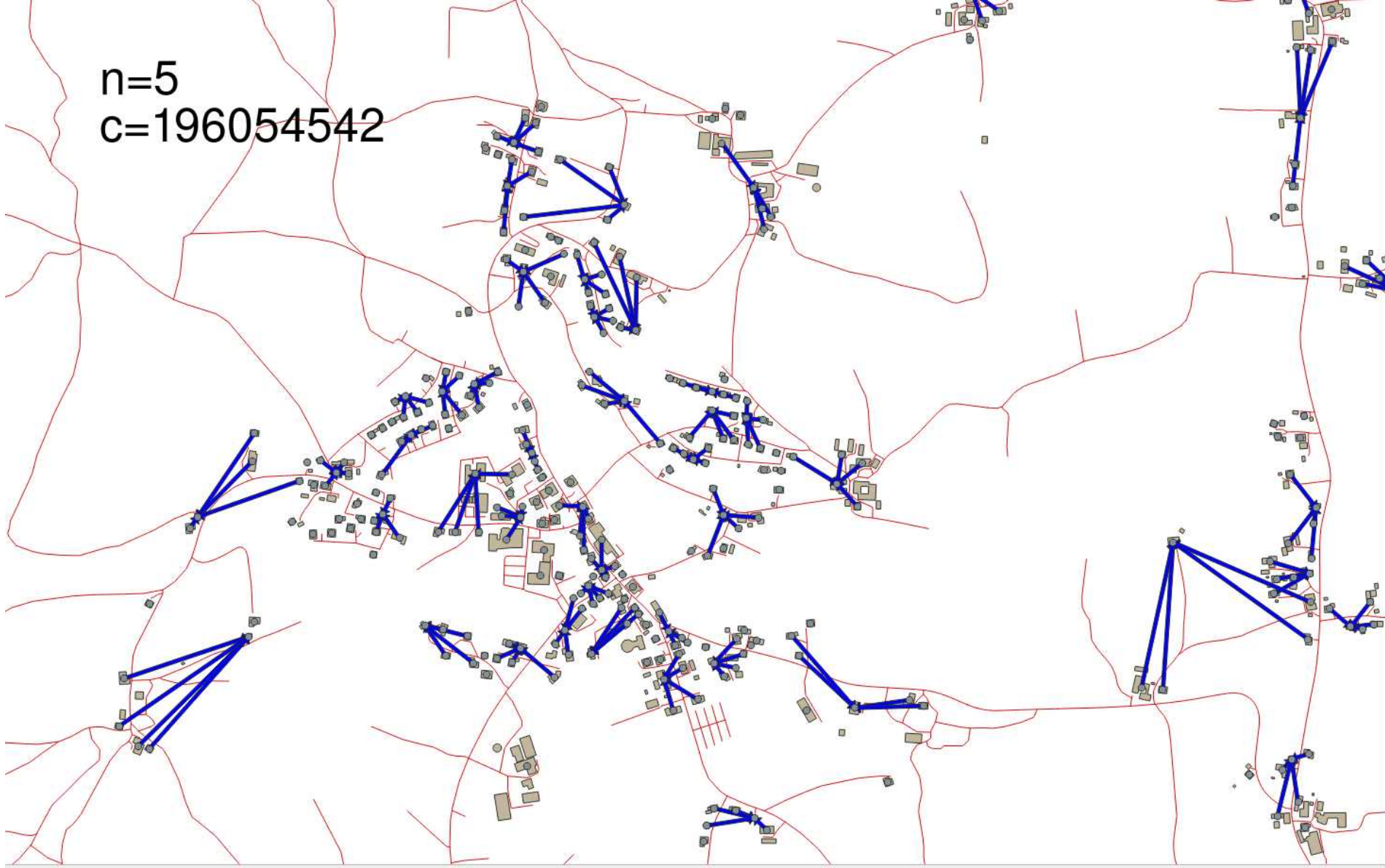
$n=2$
 $c=216055291$



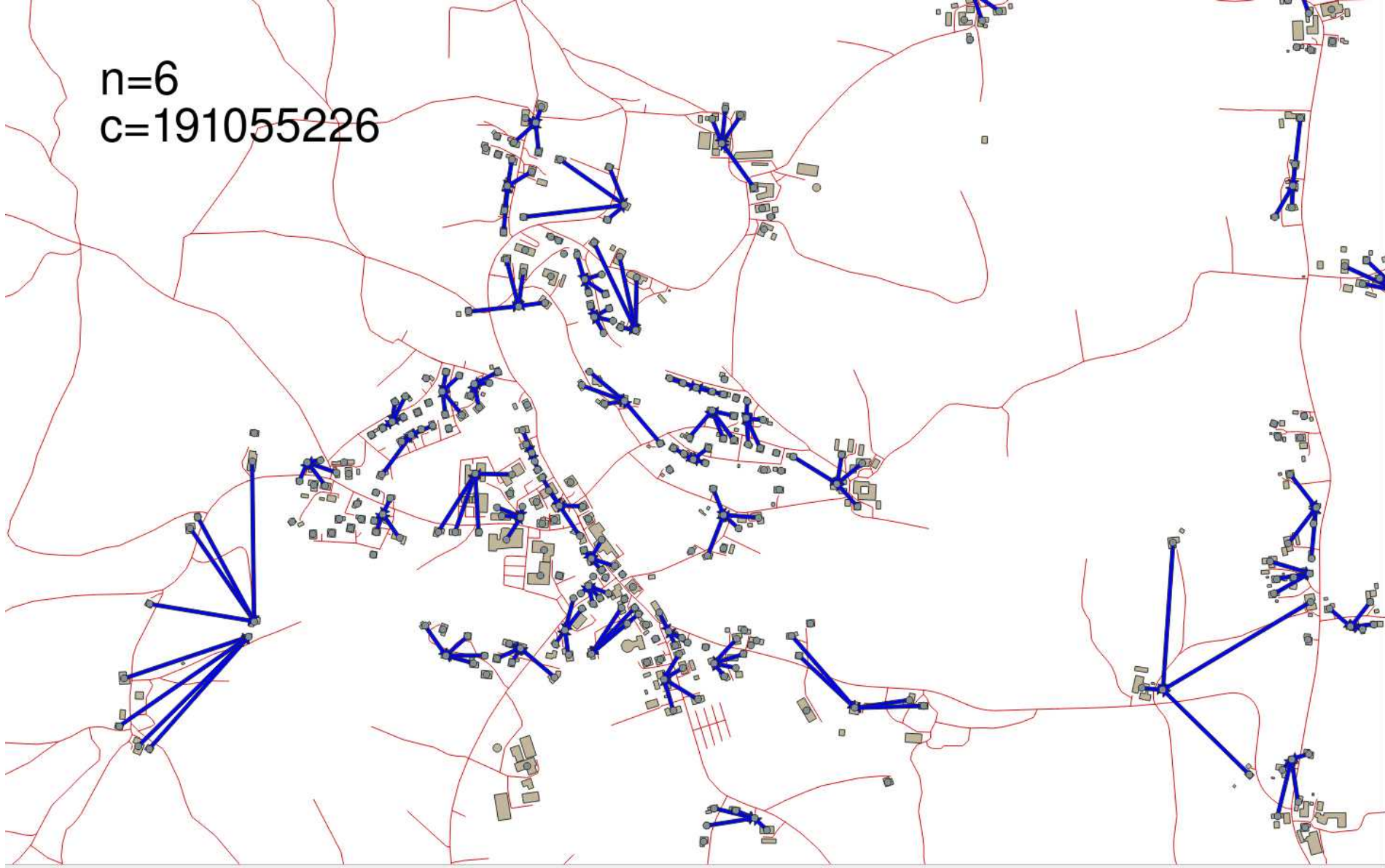
$n=3$
 $c=206057575$



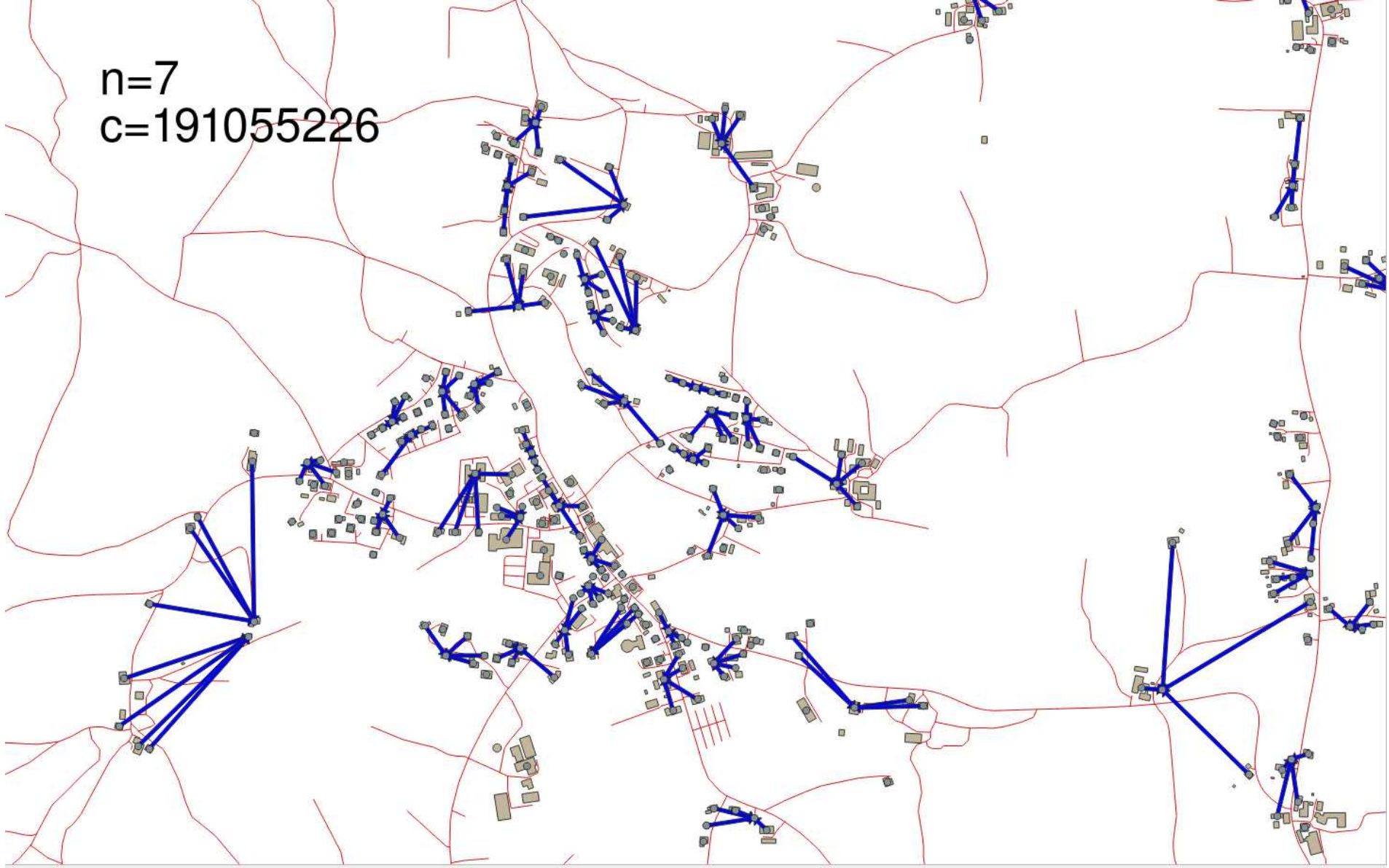
$n=5$
 $c=196054542$



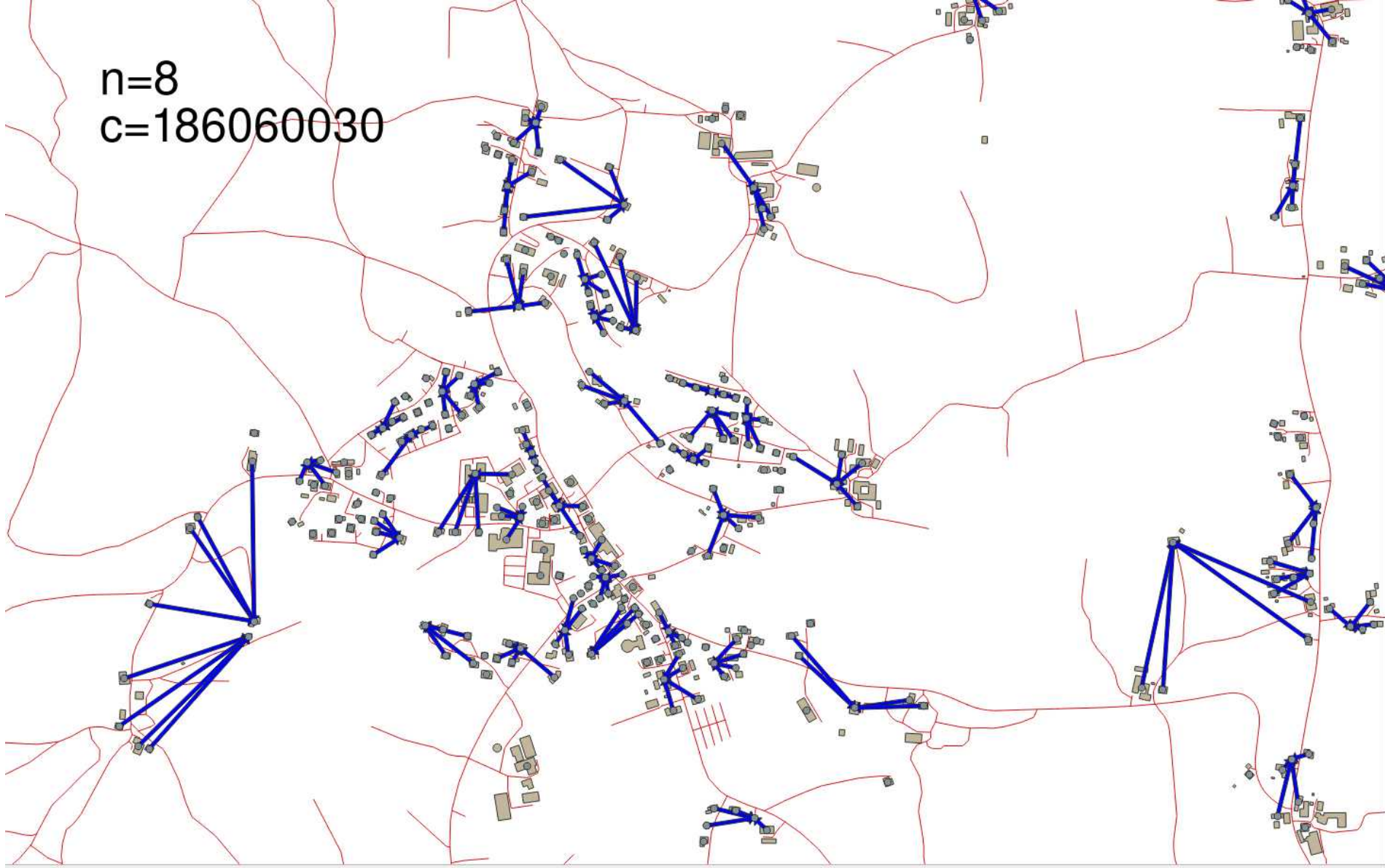
$n=6$
 $c=191055226$



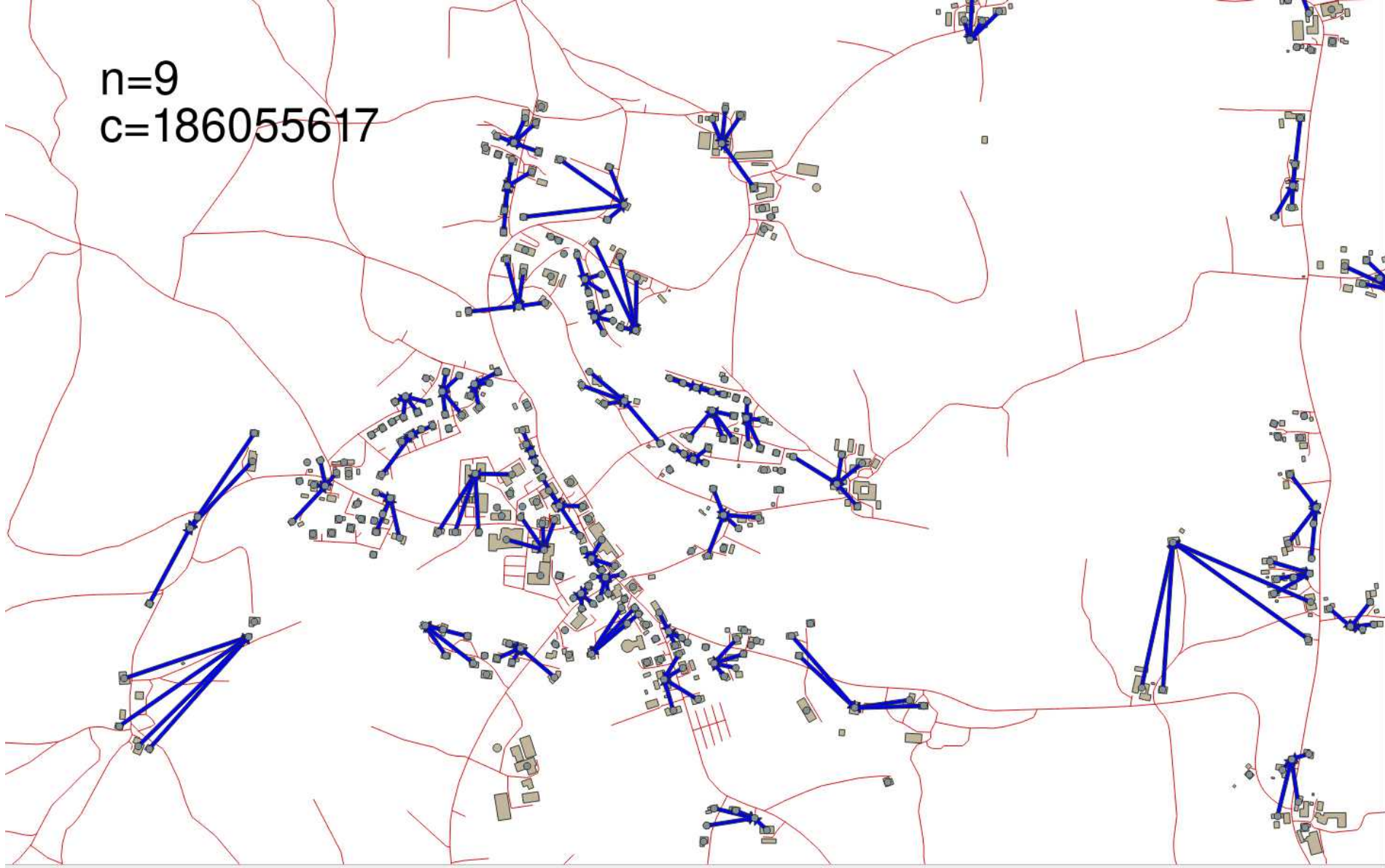
$n=7$
 $c=191055226$



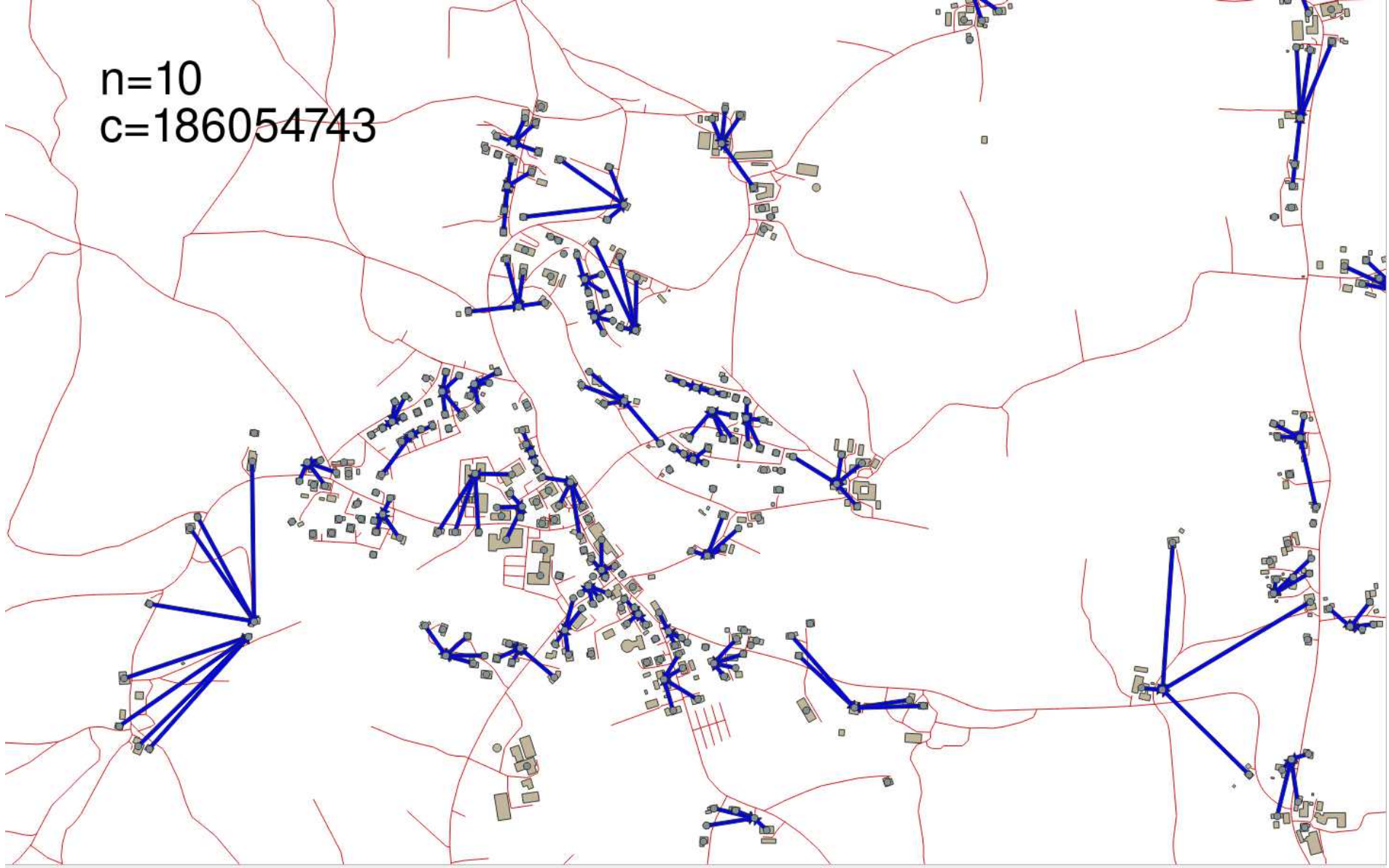
$n=8$
 $c=186060030$



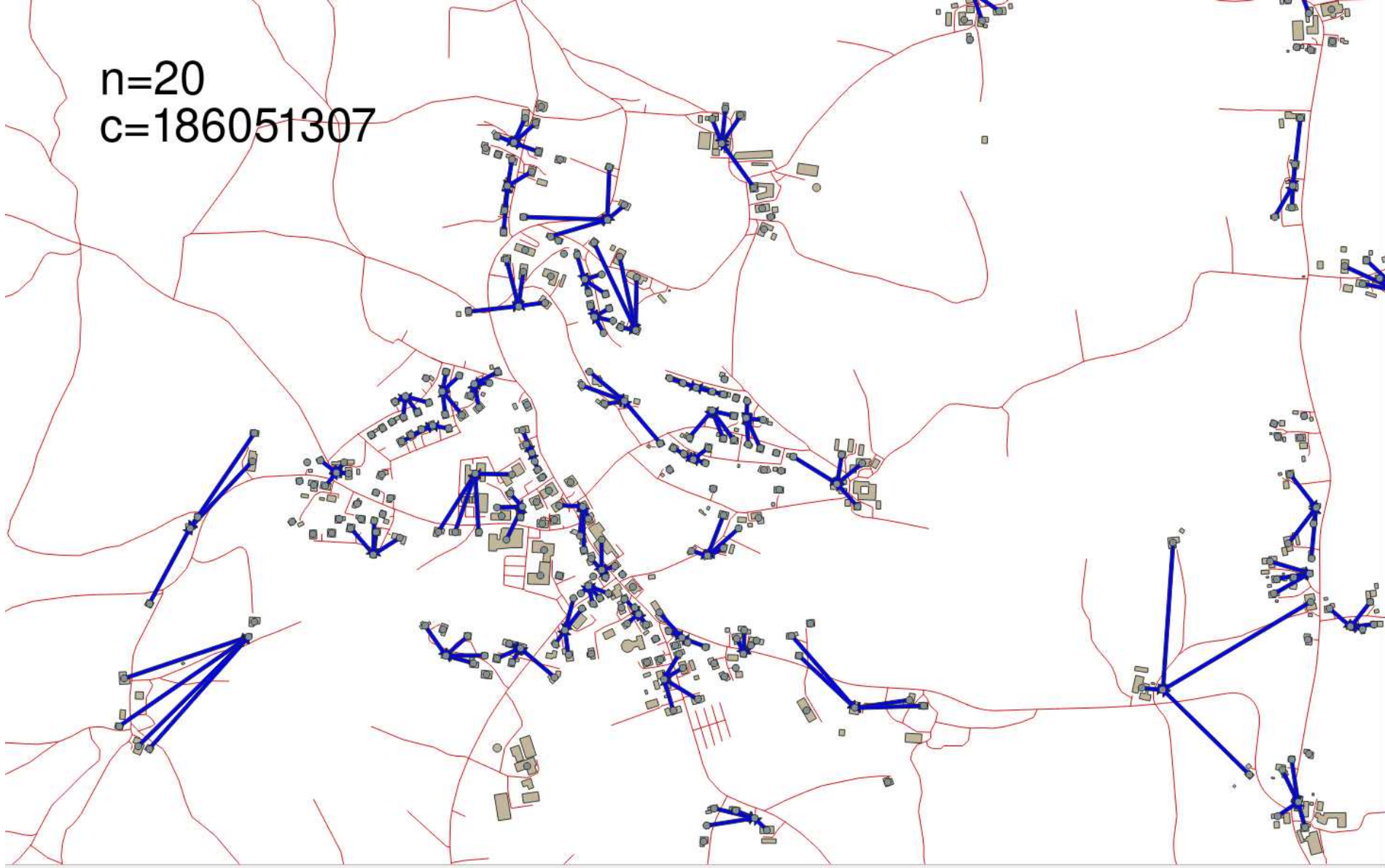
$n=9$
 $c=186055617$



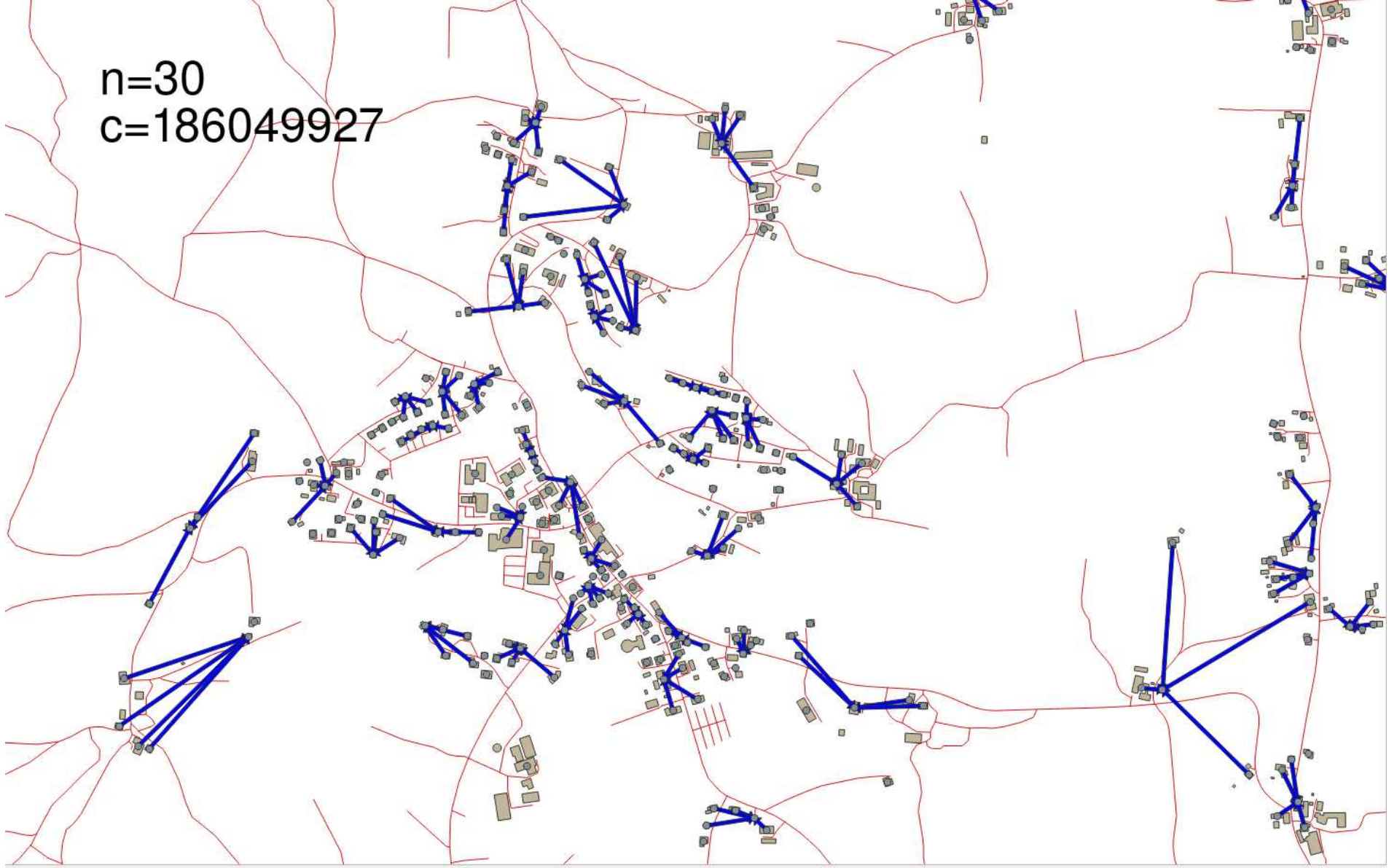
$n=10$
 $c=186054743$



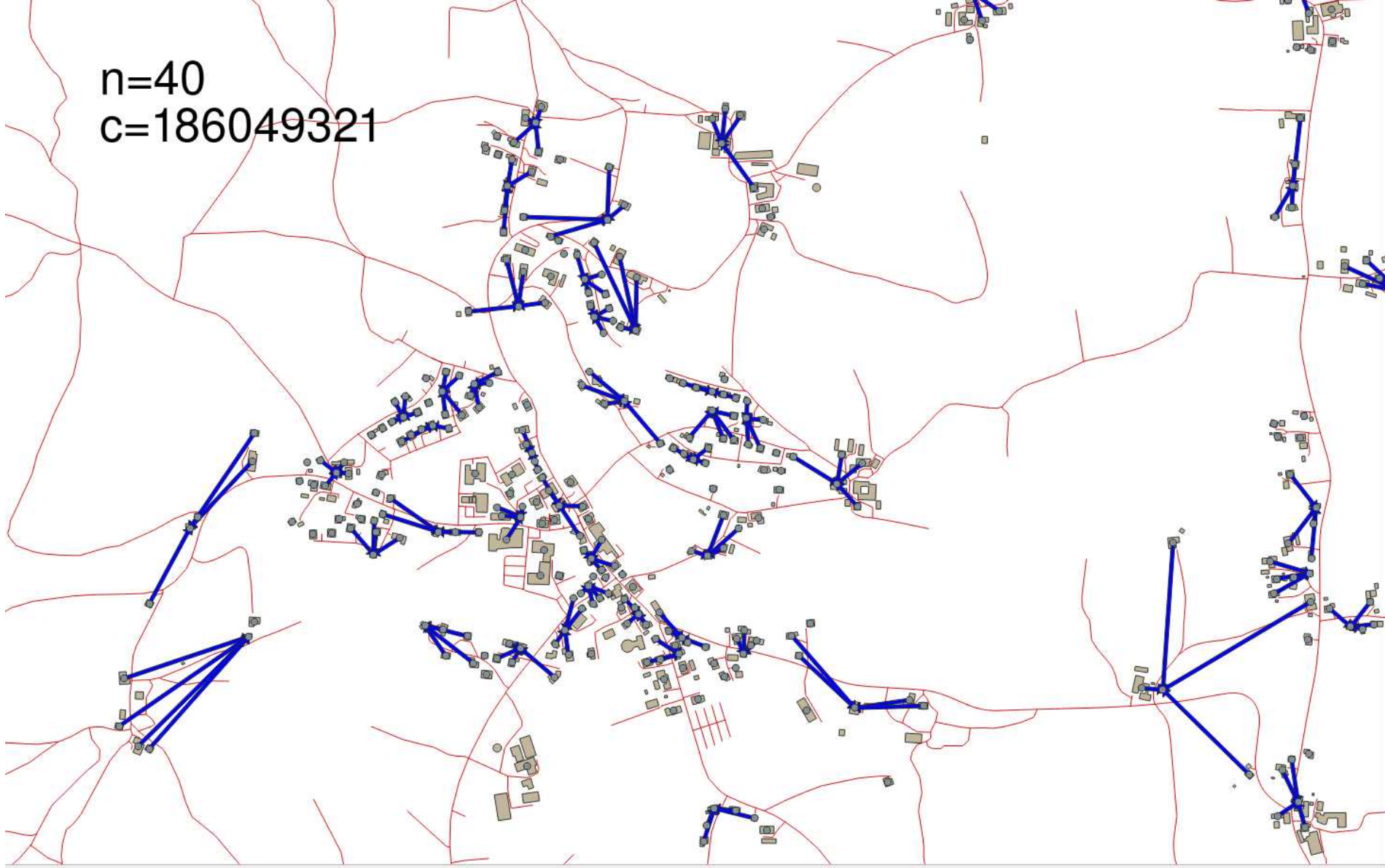
$n=20$
 $c=186051307$



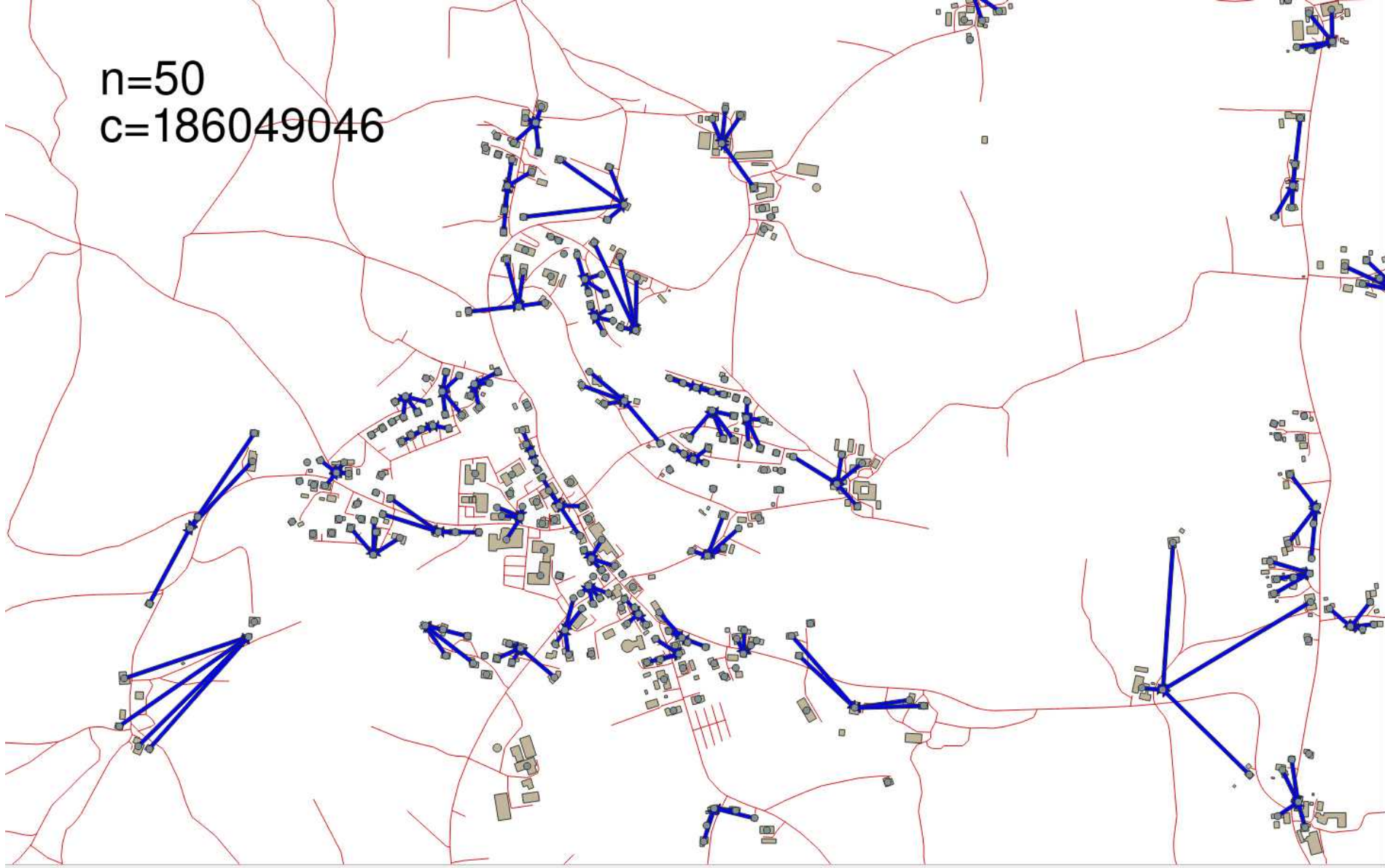
$n=30$
 $c=186049927$



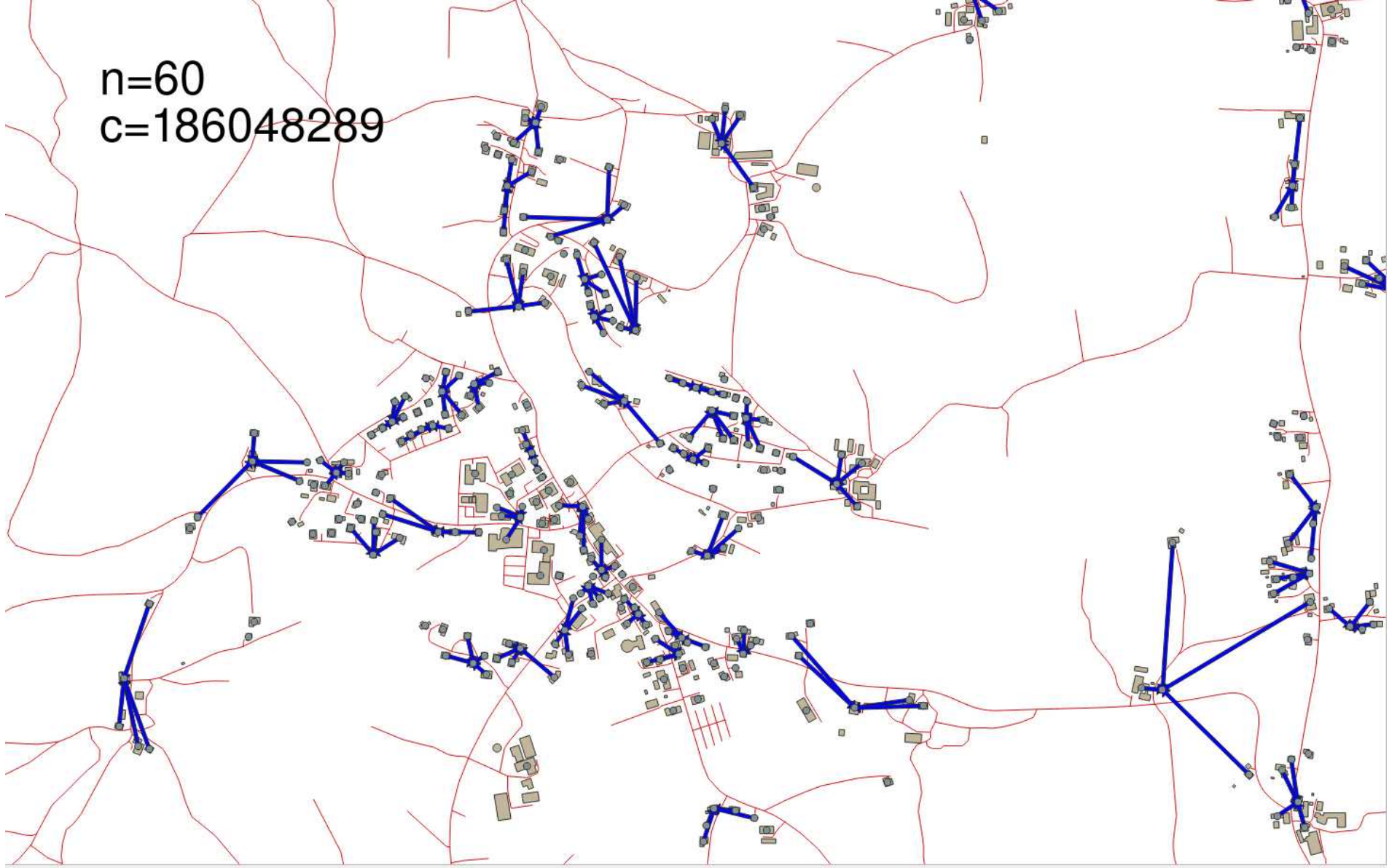
$n=40$
 $c=186049321$



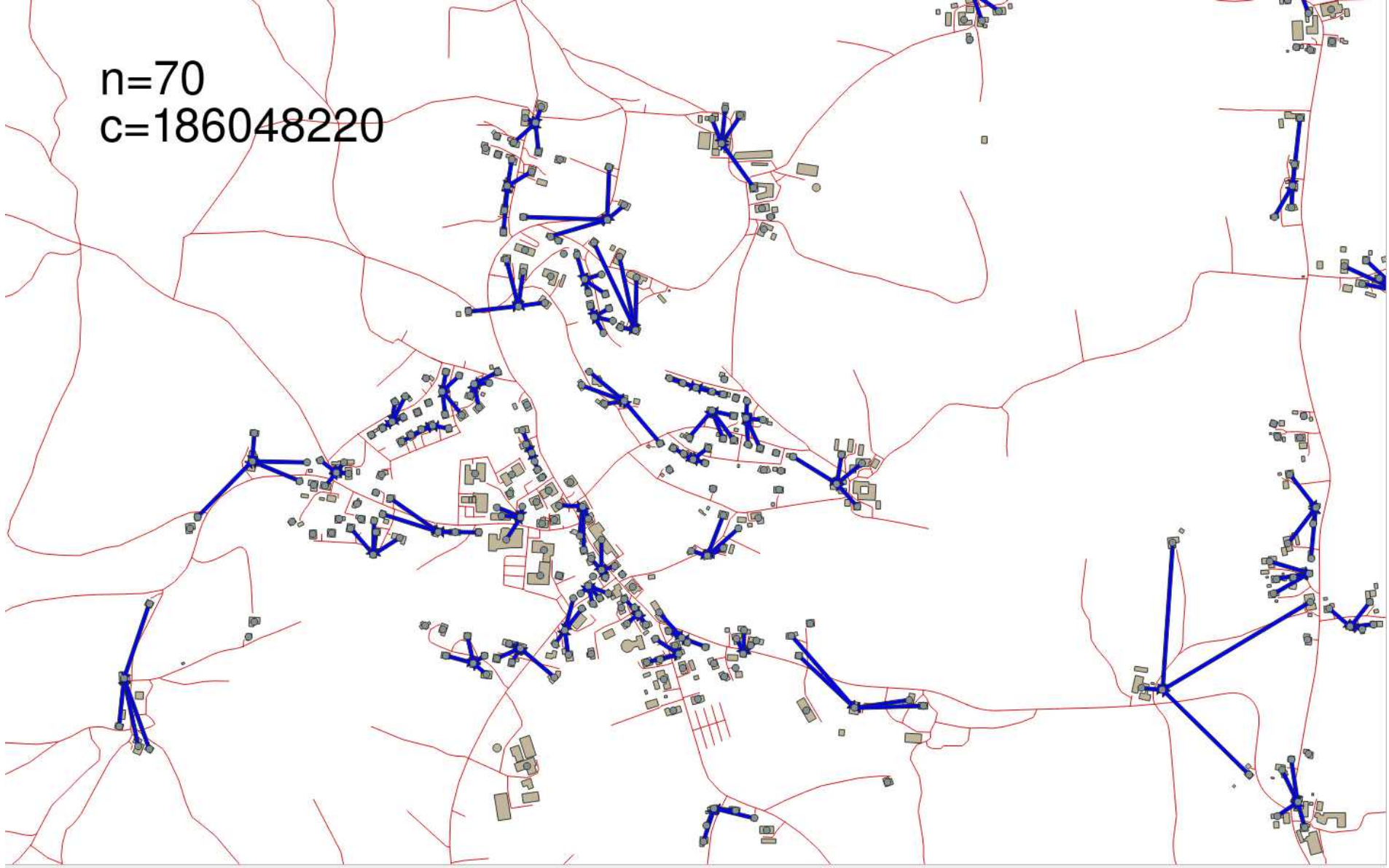
$n=50$
 $c=186049046$



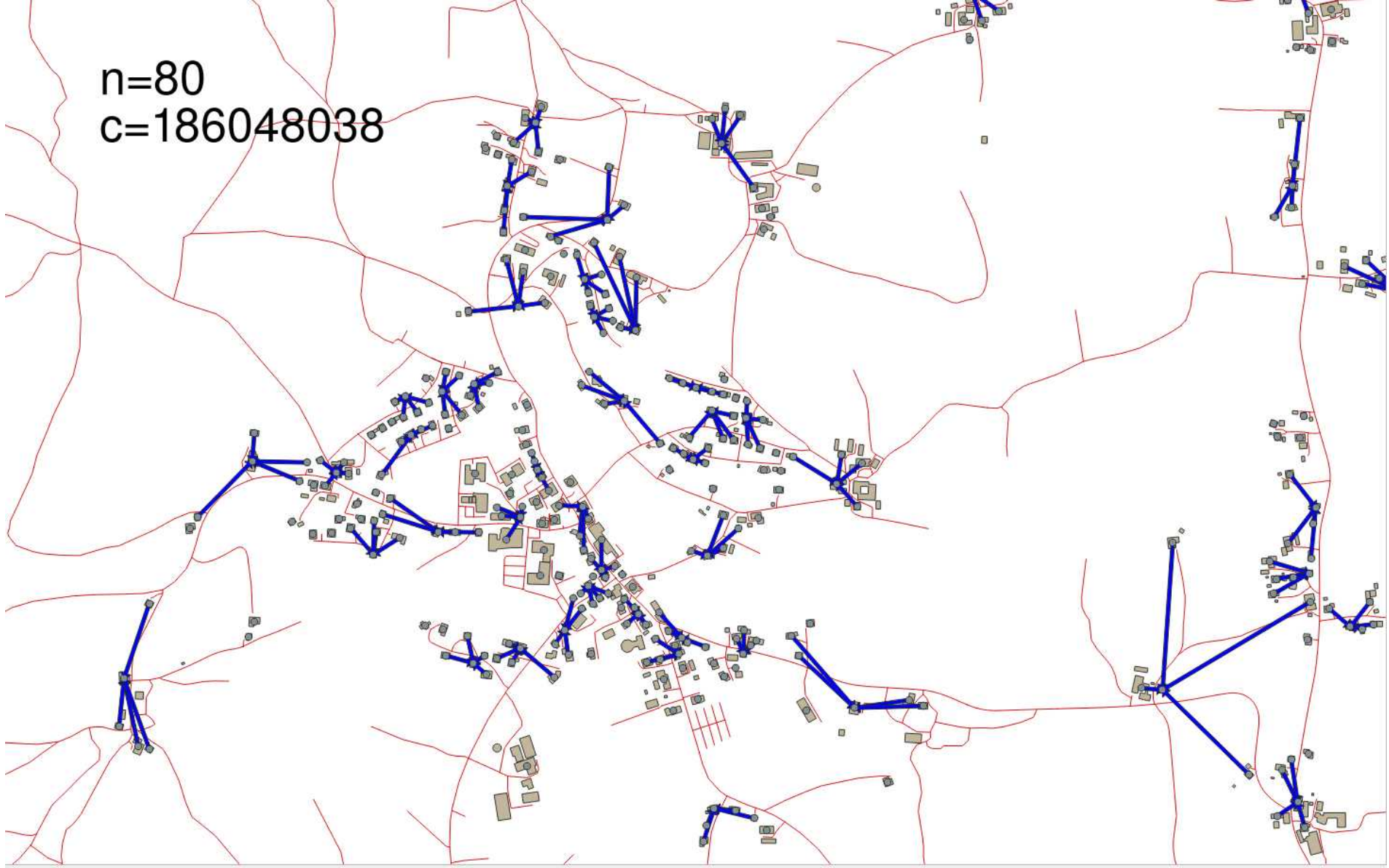
$n=60$
 $c=186048289$



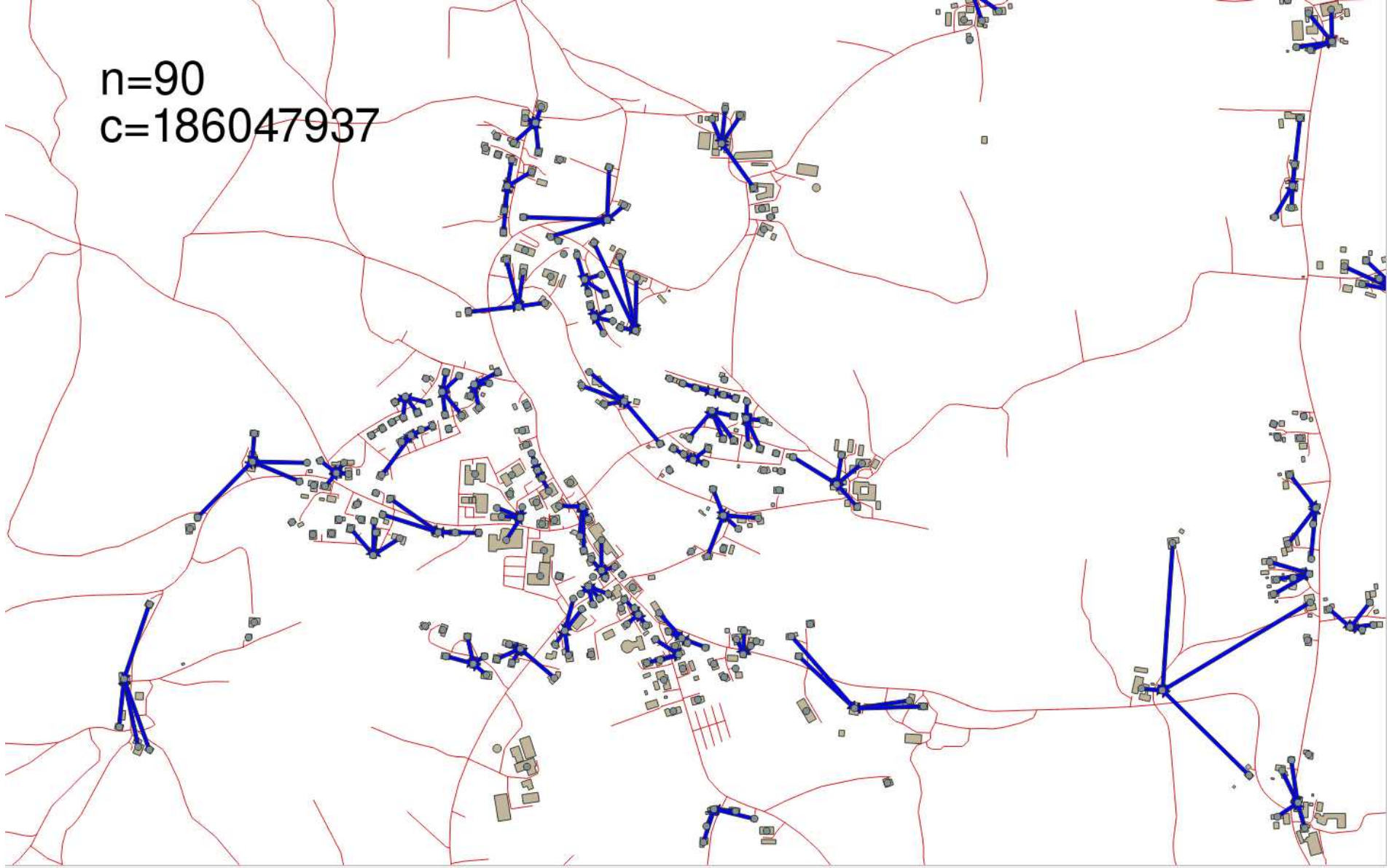
$n=70$
 $c=186048220$



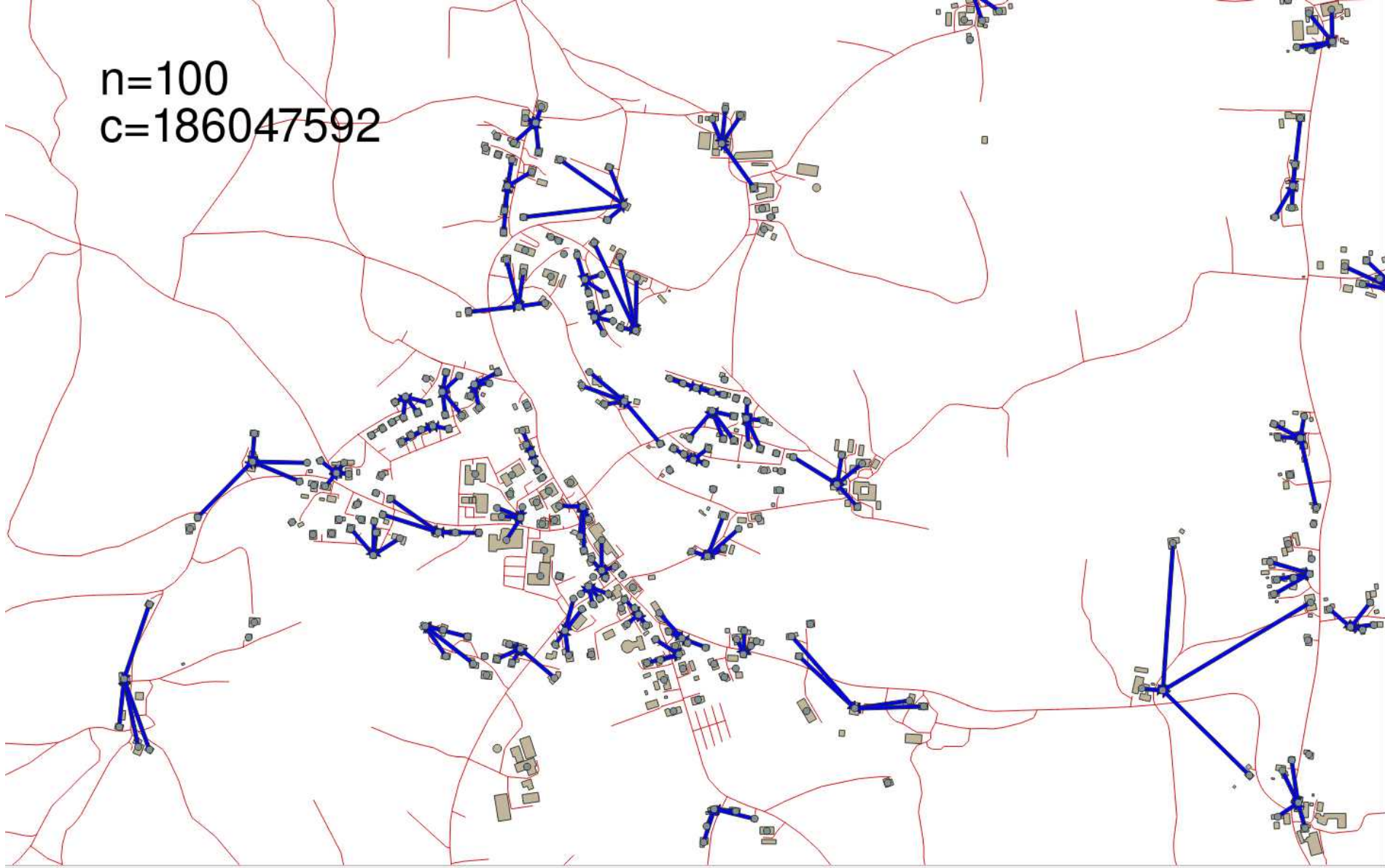
$n=80$
 $c=186048038$

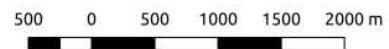
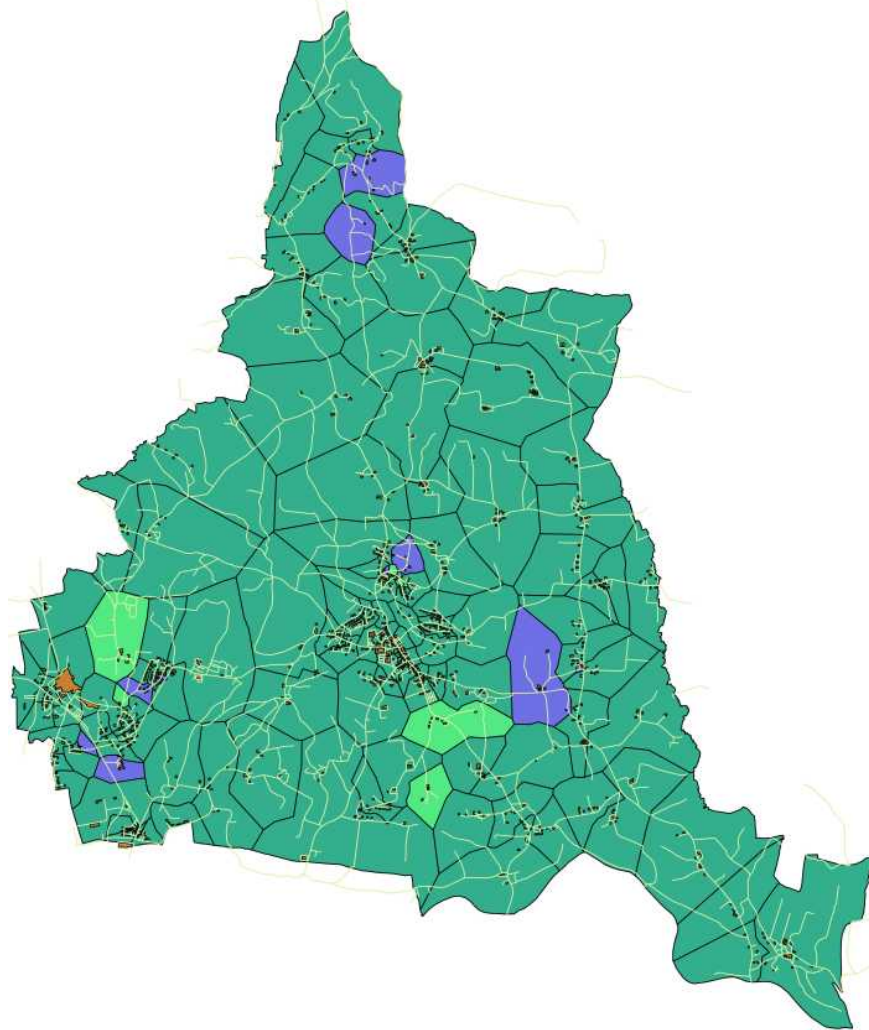


$n=90$
 $c=186047937$



$n=100$
 $c=186047592$





| | Neukirchen | Purbach | Wien 15ter |
|-------------------------------------|------------|---------|------------|
| geodesic kNN [s] | 1.287 | 0.706 | 3.329 |
| Select Sets [s] | 0.004 | 0.003 | 0.017 |
| GA Select [s] | 17.042 | 13.809 | 195.134 |
| Residuals [s] | 0.001 | 0.001 | 0.004 |
| Recals Centers (central vertex) [s] | 0.425 | 0.238 | 1.247 |
| Steal Nodes [s] | 0.003 | 0.002 | 0.018 |
| Recalc Centers (minsum) [s] | 0.402 | 0.201 | 1.246 |
| GOA Total Time [s] | 19.180 | 14.970 | 201.039 |

Tab. 6: Laufzeit von runGOA.py