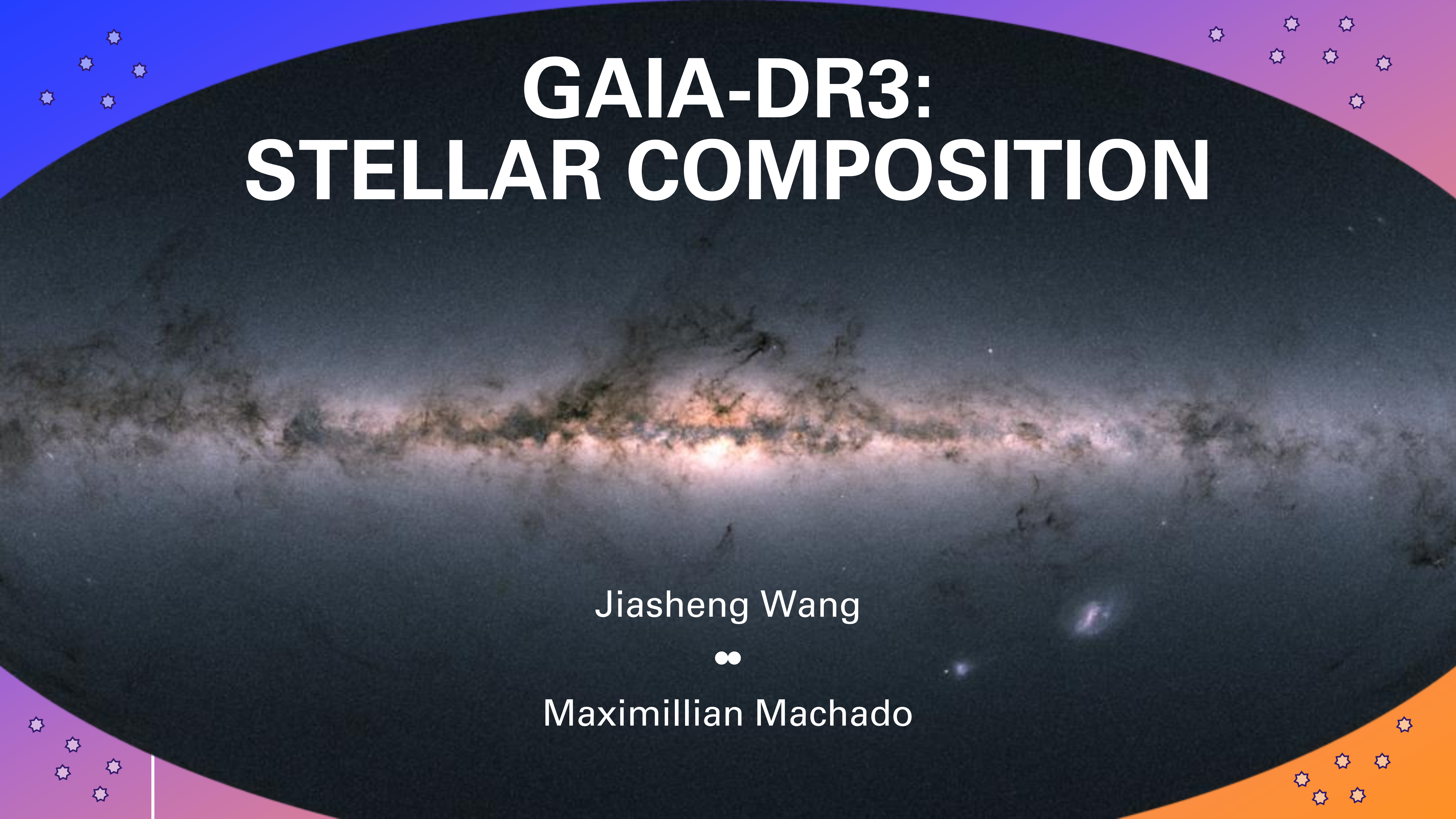
GAIA-DR3: STELLAR COMPOSITION

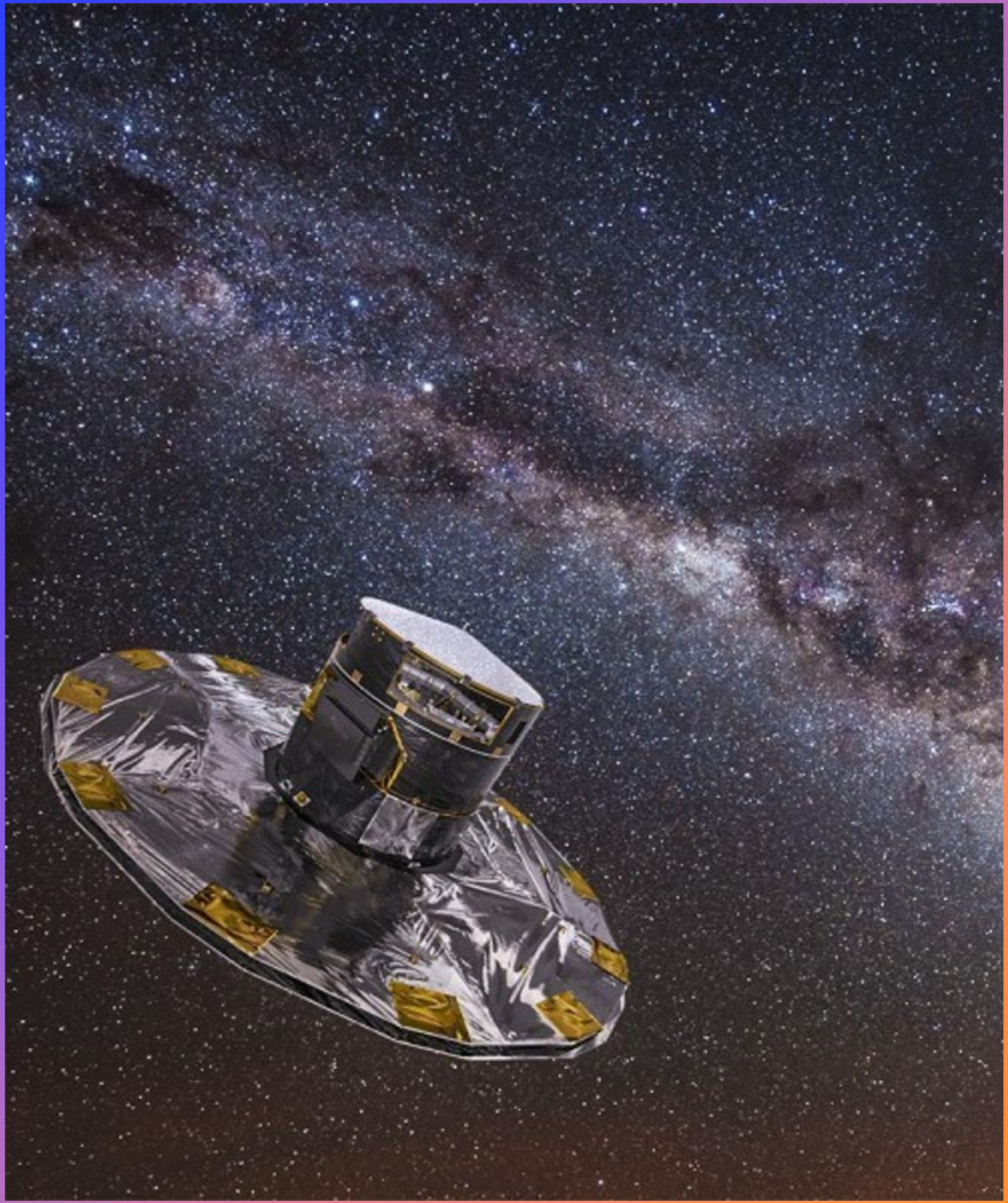


Jiasheng Wang

Maximillian Machado

GLOBAL ASTROMETRIC INTERFEROMETER (FOR) ASTROPHYSICS

- Gaia is a space observatory
  - Launched by the **E**uropean **S**pace **A**gency
- Data Release (DR) 3
  - Released on **June 13, 2022**
  - Collected over a **34-month** period
  - Contains around **1.81 billion sources**
  - Position expressed by **right ascension**, **declination**, **parallax**, and **proper motion**



AGENDA

- Research Question
- Methodology
- K-Nearest Neighbors
- R-Nearest Neighbors
- Graph Analysis
- Key Takeaways

RESEARCH MOTIVATION

Research Question

1. How can we uncover the **underlying structure of stellar** distributions in stars around us?

2. Can graph-based approaches **highlight different scales** of structure in the stellar population?

3. Are there **distinct communities** or "clusters" of stars that emerge naturally from these graph models?

💫AGENDA💫

Research Question ✅

Methodology ☑️

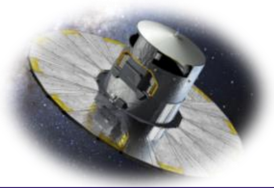
K-Nearest Neighbors ☑️

R-Nearest Neighbors ☑️

Graph Analysis ☑️

Key Takeaways ☑️

METHODOLOGY



Data Query Engine

Description:

Collection of Gaia observational data

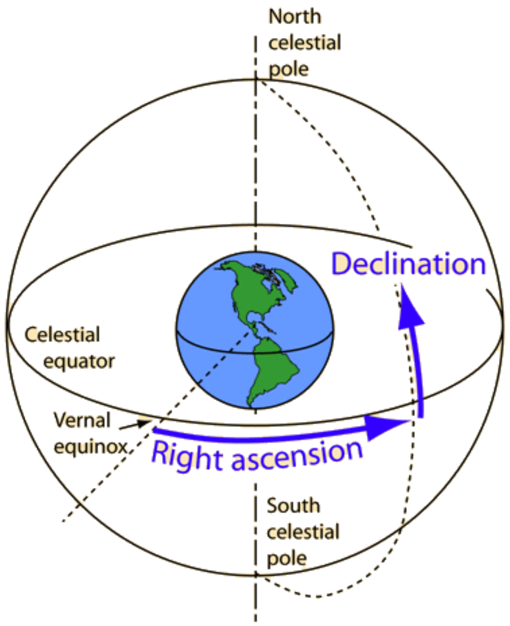
Input:

SQL statement (parallax > 10)

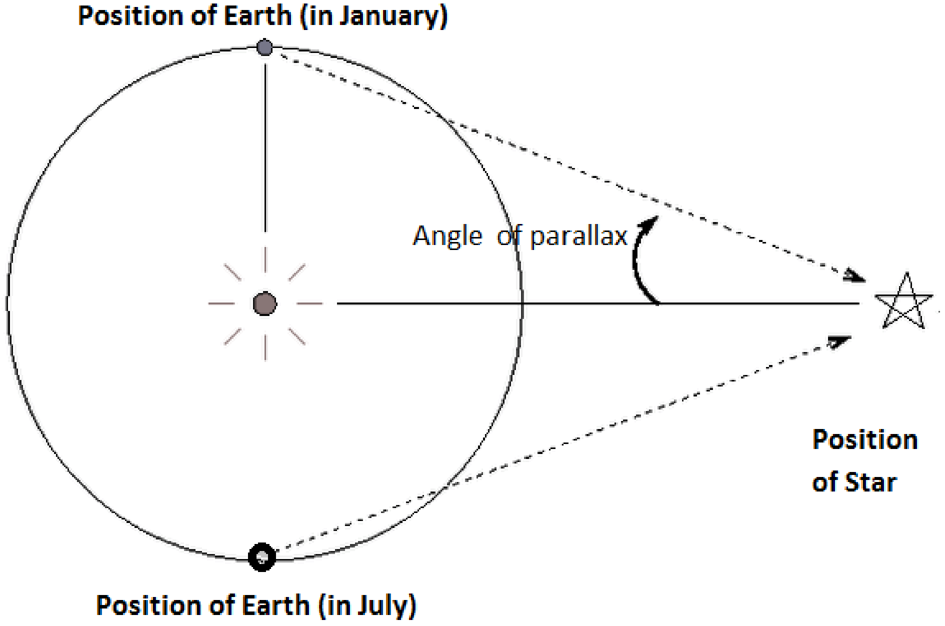
Output:

- Right ascension
- Declination
- Parallax
- Proper motion

Ascension vs Declination

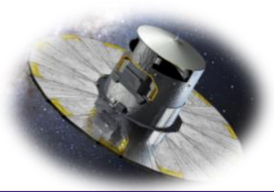


Parallax Method



METHODOLOGY

Methodology



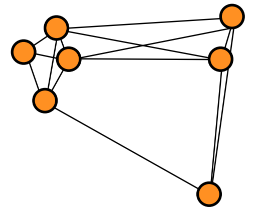
Data Query Engine	Graph Generation
<u>Description:</u> Collection of Gaia observational data	<u>Description:</u> Construction of graphs (KNN & RNN)
<u>Input:</u> SQL statement (parallax > 10)	<u>Input:</u> Spatial Coordinates (X, Y, Z)
<u>Output:</u> • Right ascension • Declination • Parallax • Proper motion	<u>Output:</u> • KNN Graph • RNN Graph

Input Coord:
$(\alpha, \delta) \rightarrow (x, y, z)$

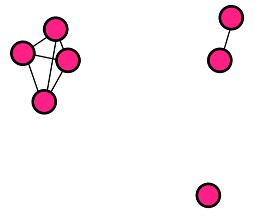
Pairwise distances:
$d(star_i, star_j)$

Node Definition: $\mathcal{N} := \{star_i\}_{i=1}^{n}$

KNN Graph:



RNN Graph:

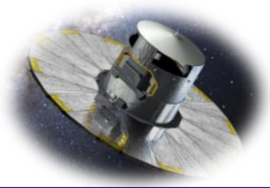


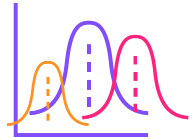
Edge Definition:
$E := (star_i, star_j)$,
top K from
$\min_{star_j} \; d(star_i, starj)$

Edge Definition:
$E := \left(star_i, star_j\right)$,
where
$d(star_i, starj) \leq R$

METHODOLOGY

Methodology



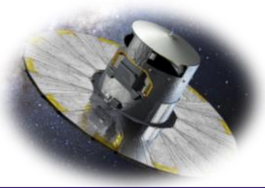


Data Query Engine	Graph Generation	Summary Statistics
<u>Description:</u> Collection of Gaia observational data	<u>Description:</u> Construction of graphs (KNN & RNN)	<u>Description:</u> Exploratory statistics over underlying graphs
<u>Input:</u> SQL statement (parallax > 10)	<u>Input:</u> Spatial Coordinates (X, Y, Z)	<u>Input:</u> KNN and RNN network graphs
<u>Output:</u> • Right ascension • Declination • Parallax • Proper motion	<u>Output:</u> • KNN Graph • RNN Graph	<u>Output:</u> • Degree distributions • Density measures • Volume

...

METHODOLOGY

Methodology



Data Query Engine

Description:

Collection of Gaia observational data

Input:

SQL statement (parallax > 10)

Output:

- Right ascension
- Declination
- Parallax
- Proper motion

Graph Generation

Description:

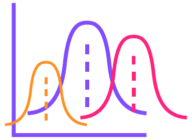
Construction of graphs (KNN & RNN)

Input:

Spatial Coordinates (X, Y, Z)

Output:

- KNN Graph
- RNN Graph



Summary Statistics

Description:

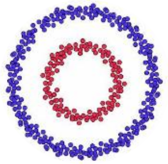
Exploratory statistics over underlying graphs

Input:

KNN and RNN network graphs

Output:

- Degree distributions
- Density measures
- Volume



Graph Analysis

Description:

Spectral embedding & community detection

Input:

KNN and RNN network graphs

Output:

- Spectral embedding
- Community clusters
- Partitioned Adjacency Matrix

💫 AGENDA 💫

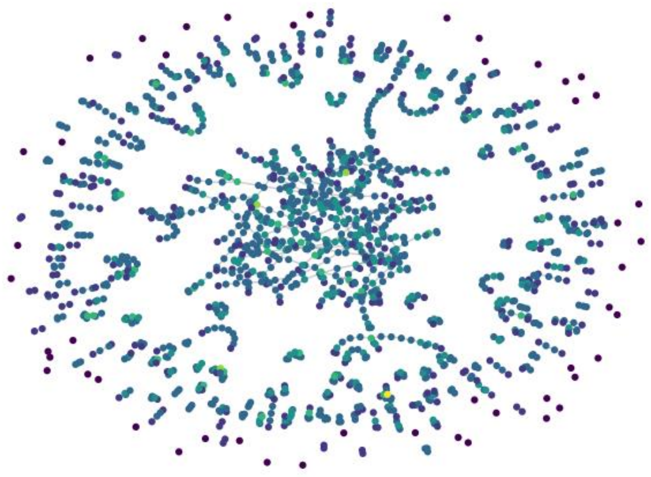
- Research Question ✅
- Methodology ✅
- K-Nearest Neighbors ☑️
- R-Nearest Neighbors ☑️
- Graph Analysis ☑️
- Key Takeaways ☑️

VISUALIZATION

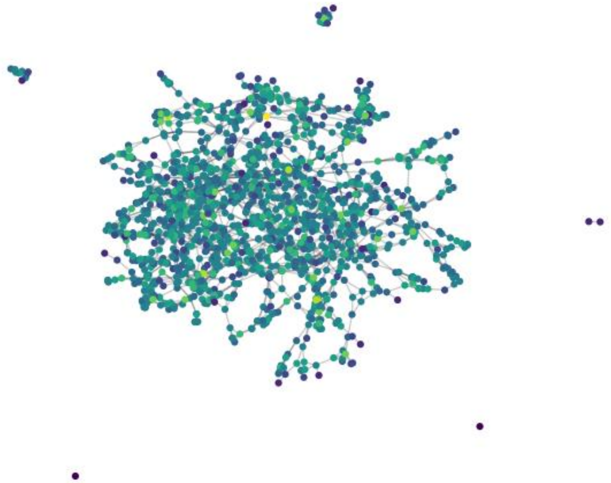
K-Nearest Neighbors

- Network plots for k = 2, 4, and 5.
- Graph connectivity rapidly increasing until k = 5.
- The nodes are colored by degree, the lighter the color, the higher the degree.

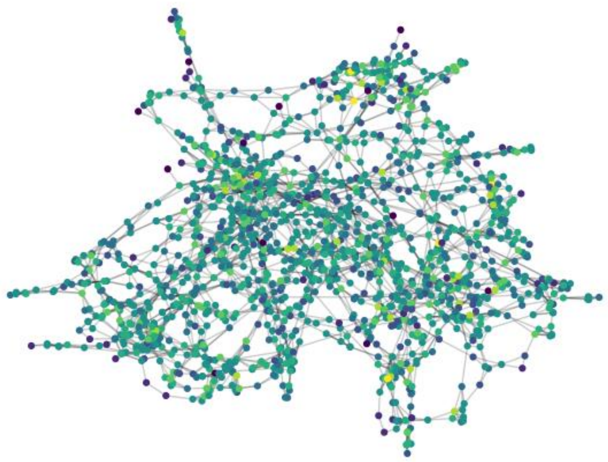
KNN (k=2) - Graph Visualization (Colored by Degree)



KNN (k=4) - Graph Visualization (Colored by Degree)



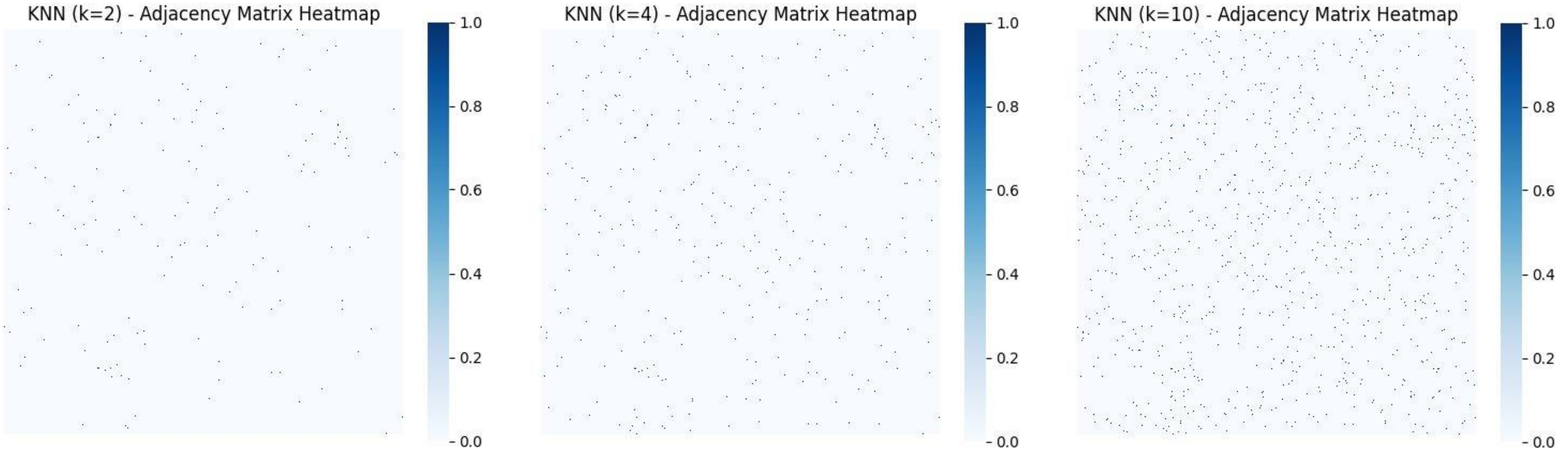
KNN (k=5) - Graph Visualization (Colored by Degree)



ADJACENCY MATRIX HEATMAPS

K-Nearest Neighbors

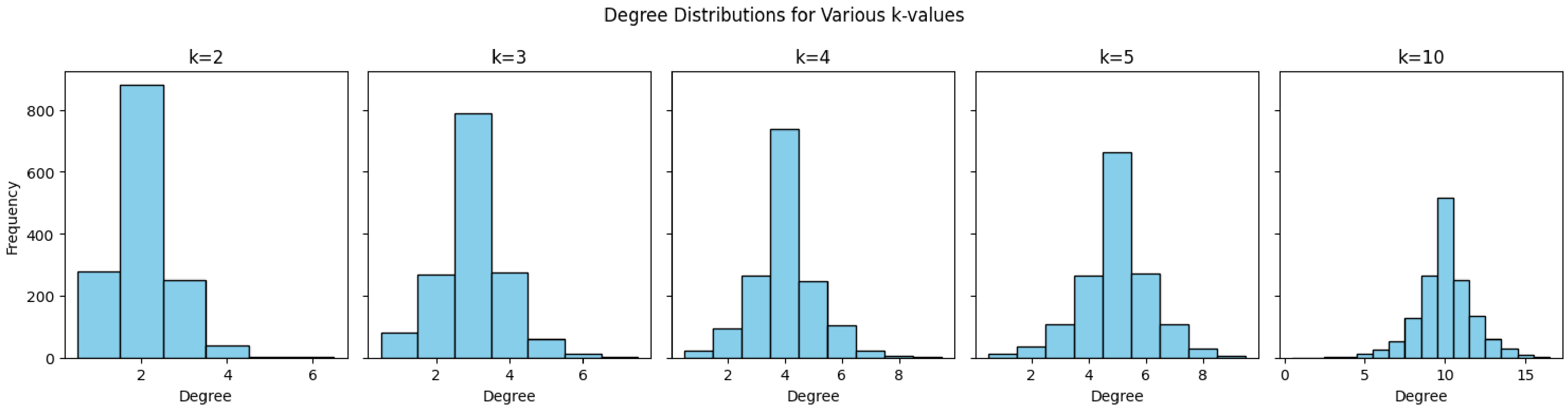
- Heatmaps of the adjacency matrix for k = 2, 4, and 10.
- For lower k, the adjacency matrix is sparser.
- For higher k it's more filled in, demonstrating higher global connectivity.



DEGREE DISTRIBUTION

K-Nearest Neighbors

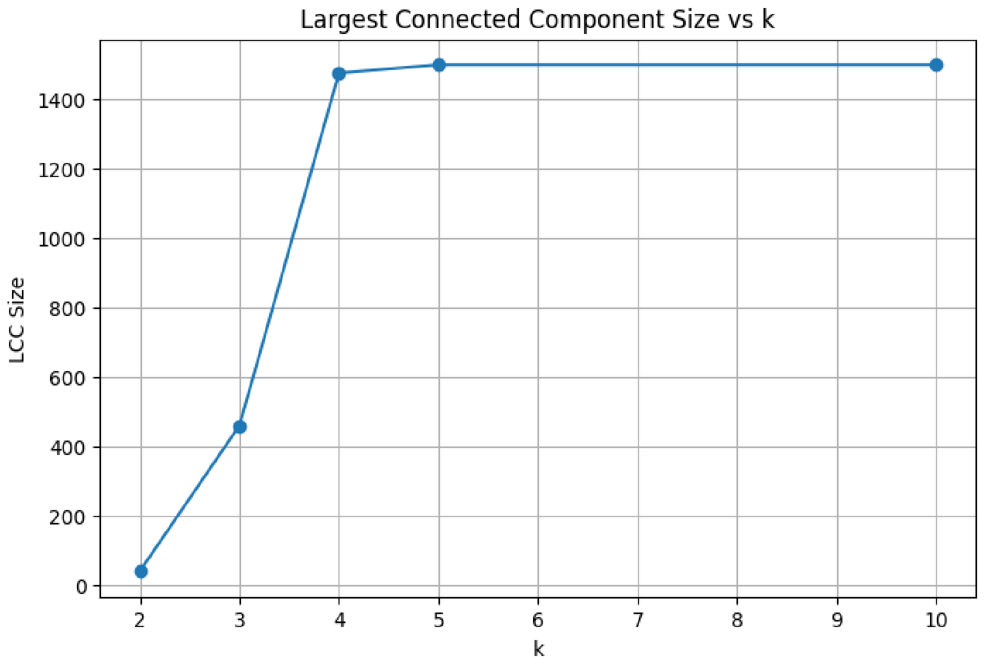
- Below are multiple histograms of KNN degree distributions for k = 2,3,4,5,10.
- low k gives a more skewed distribution, high k yields a more normal distribution (around k).
- As k increases, the graph becomes denser and more uniformly connected.

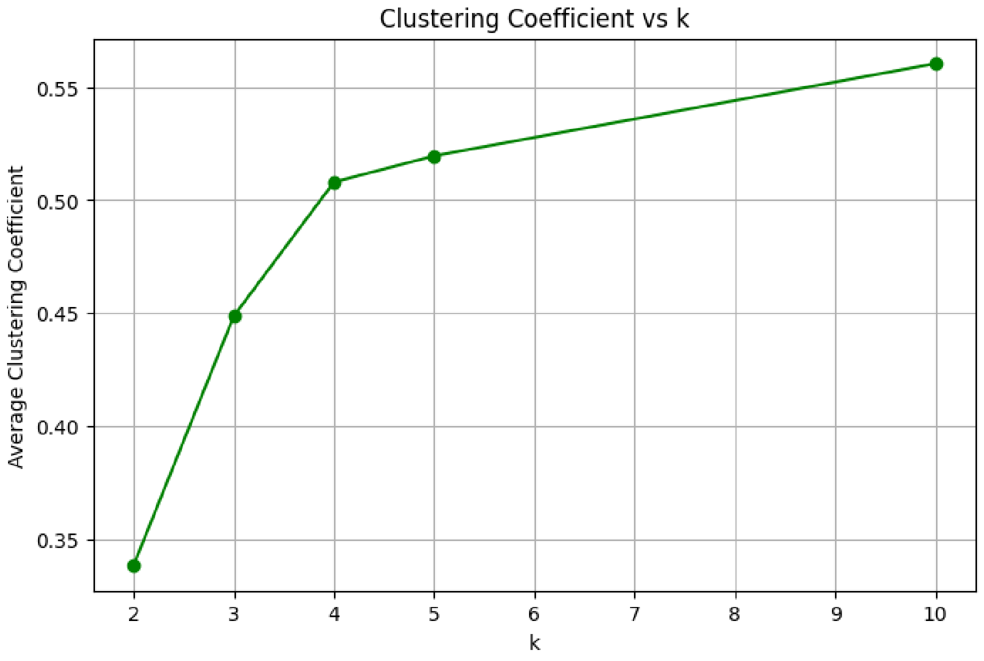


CONNECTIVITY AND CLUSTERING

K-Nearest Neighbors

- The whole graph becomes connected after k ≥ 5
- As k grows, the largest component often includes more nodes, indicating better global connectivity.
- More edges can either homogenize neighborhoods or highlight certain cliques





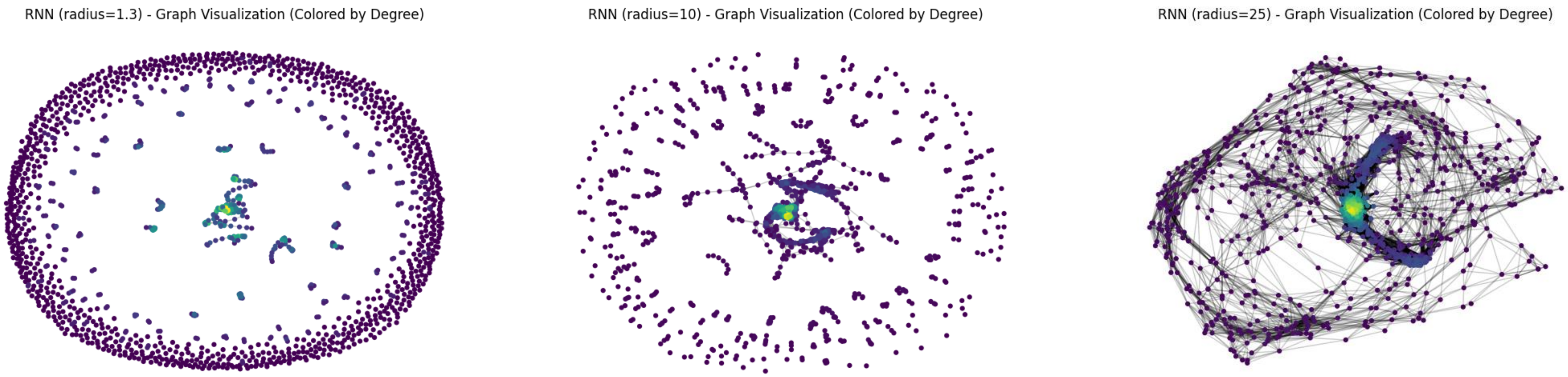
AGENDA

- Research Question ✅
- Methodology ✅
- K-Nearest Neighbors ✅
- R-Nearest Neighbors ☑️
- Graph Analysis ☑️
- Key Takeaways ☑️

VISUALIZATION

R-Nearest Neighbors

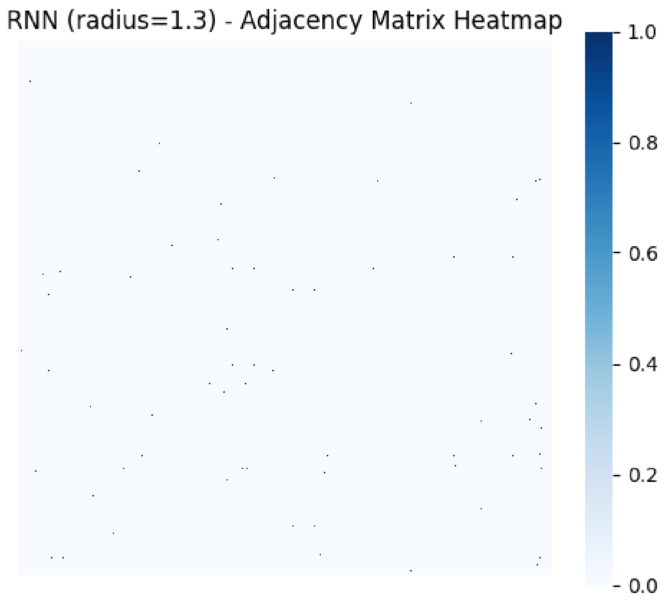
- Network plots for r = 1.3, 10, and 25 parsec.
- Degree-based coloring reveals isolated nodes at smaller radii and the emergence of high-degree hubs (yellow) in denser graphs.
- Unlike KNN, graph density quickly concentrates as radius increases.

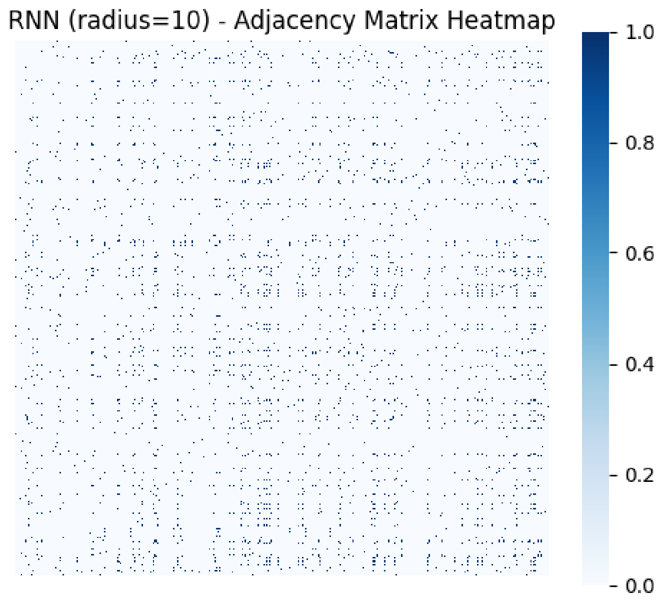


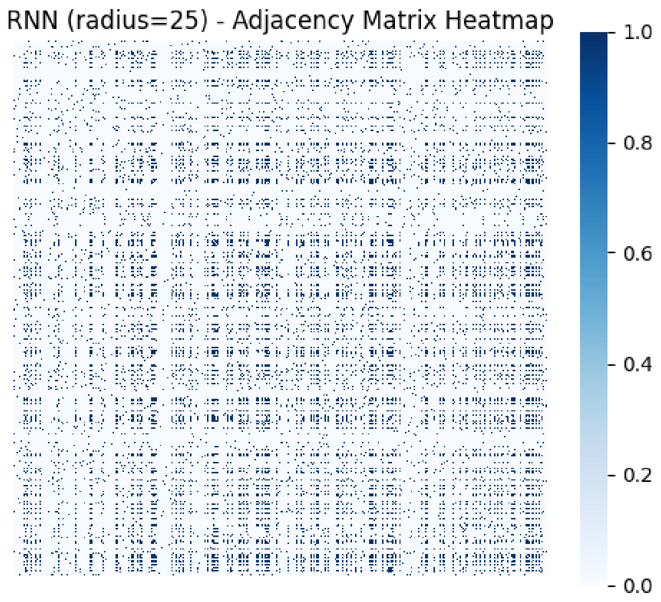
ADJACENCY MATRIX HEATMAPS

R-Nearest Neighbors

- Heatmaps of the adjacency matrix for r = 1.3, 10, and 25.
- Similar to KNN, the adjacency is mostly sparse at lower radii.
- The concentration of degree is not readily apparent.



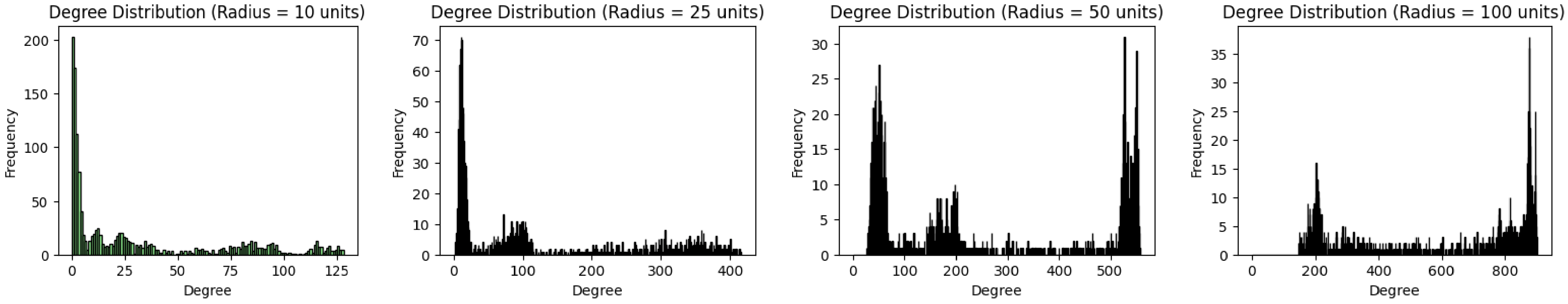




DEGREE DISTRIBUTION

R-Nearest Neighbors

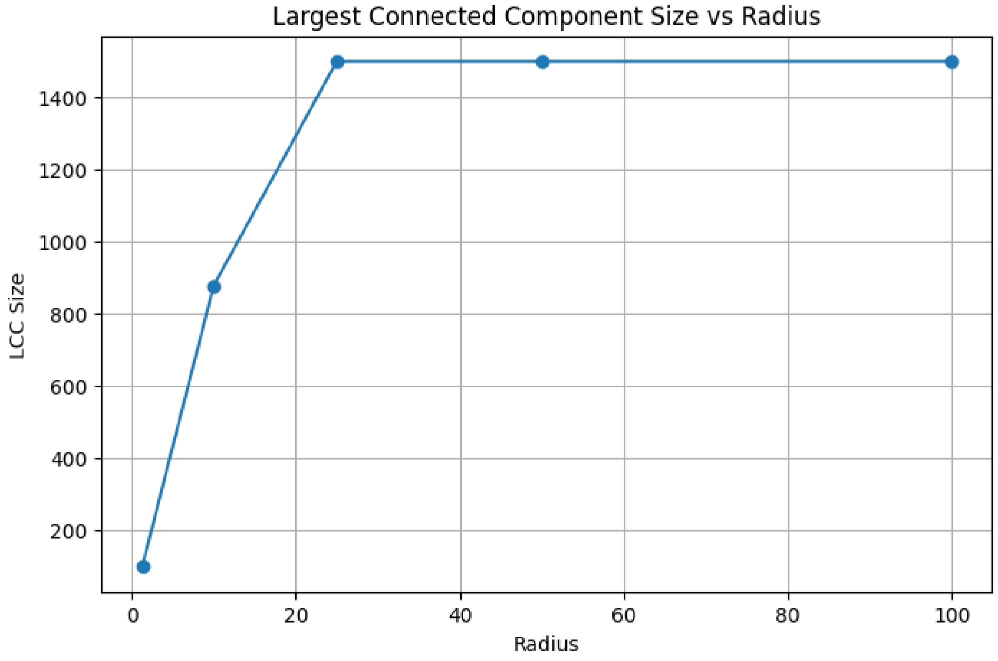
- RNN degree distributions for r = 10, 25, 50, and 100.
- The range is larger than the KNN counter part, especially at larger radii.
- At larger radii (e.g., radius = 50 or 100), the distributions become multi-modal, indicating groups of nodes with distinct connectivity characteristics, potentially tied to spatial clustering.
- The low degree node concentration rapidly shifts to high degree as radius increases.

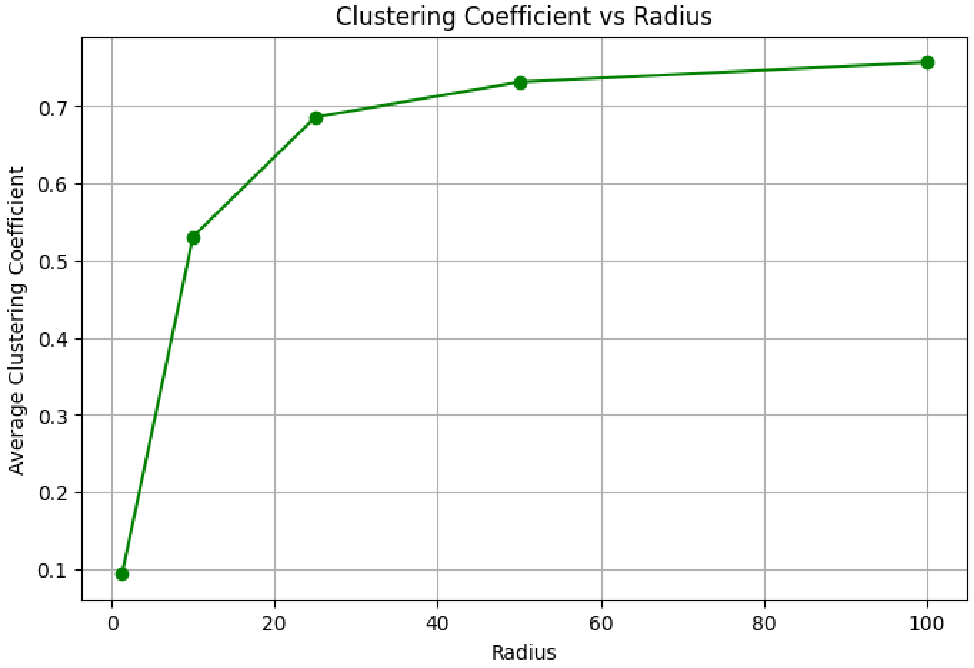


CONNECTIVITY AND CLUSTERING

R-Nearest Neighbors

- The LCC size increases sharply with the radius reaching near-maximum size by radius = 20, indicating a quick transition to a globally connected graph.
- The rate of increase slows beyond radius = 40, indicating diminishing returns in forming tightly-knit local groups as the radius grows.
- The average clustering coefficient rises steadily with the radius, reflecting stronger local connectivity as the neighborhood radius expands.



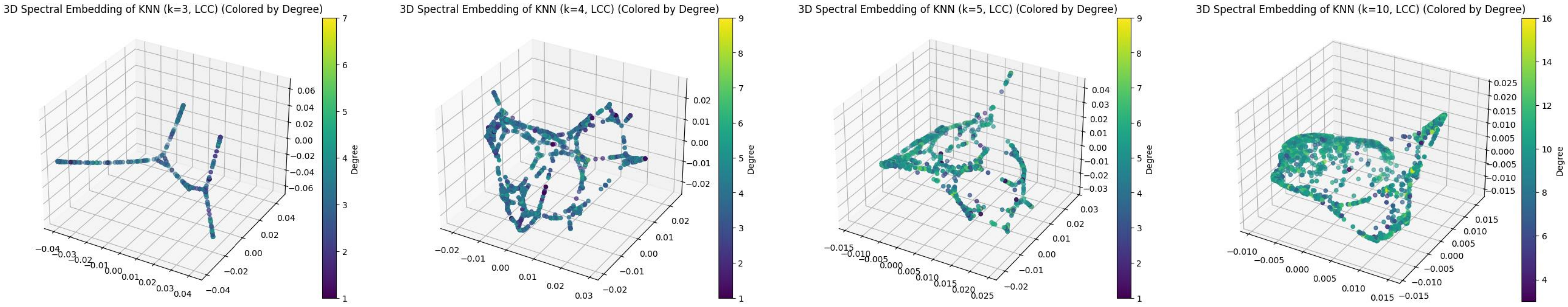


AGENDA

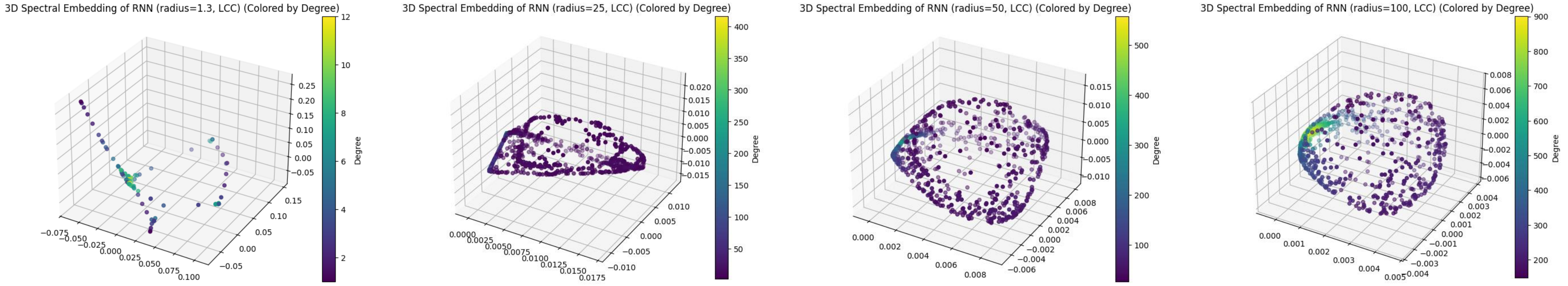
- Research Question ✅
- Methodology ✅
- K-Nearest Neighbors ✅
- R-Nearest Neighbors ✅
- Graph Analysis ☑️
- Key Takeaways ☑️

SPECTRAL ANALYSIS

K-Nearest Neighbors

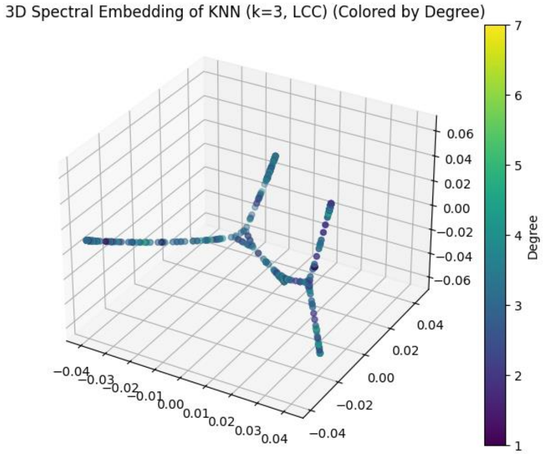


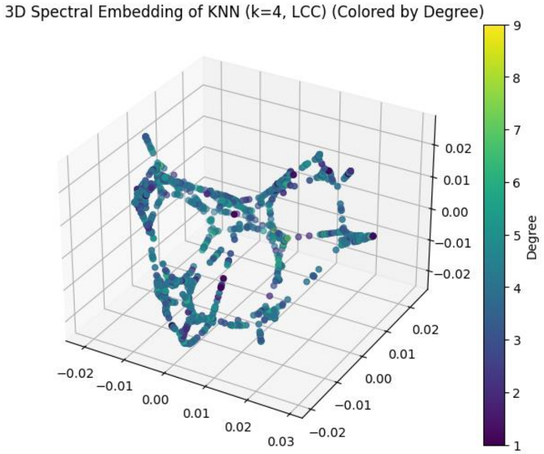
Radius-Nearest Neighbors

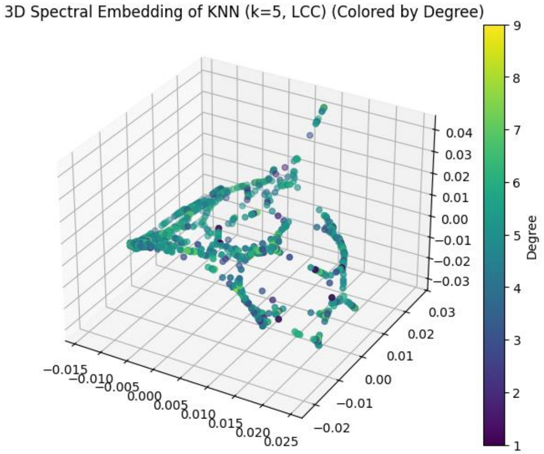


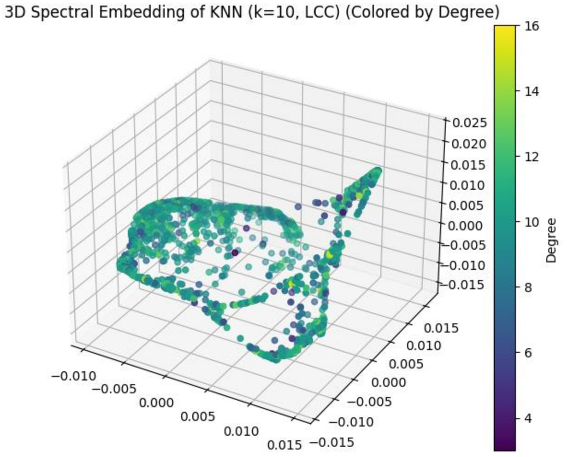
SPECTRAL ANALYSIS

K-Nearest Neighbors

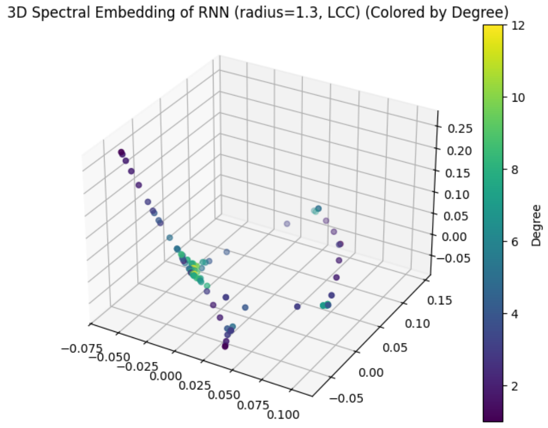


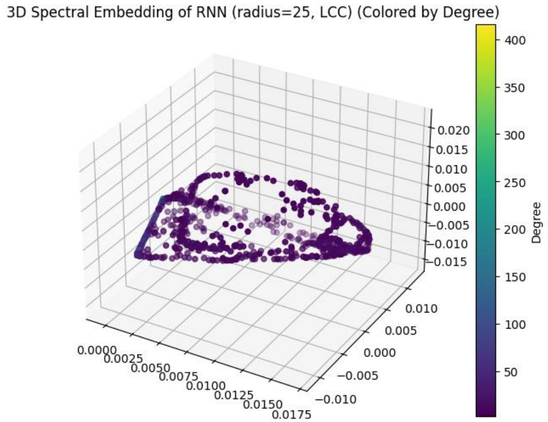


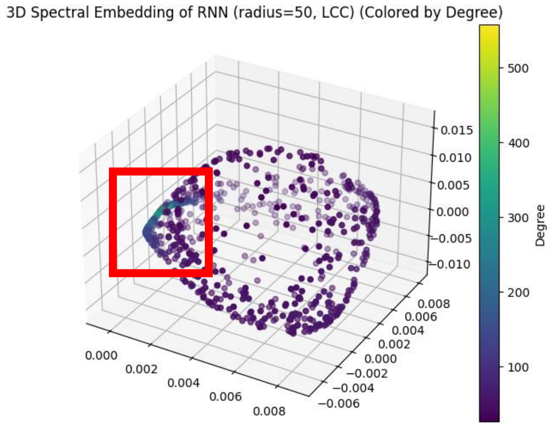


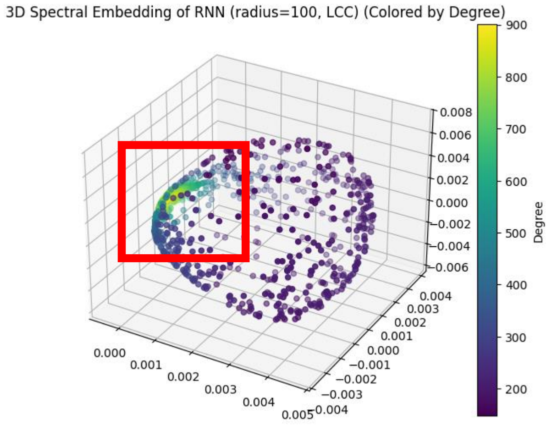


Radius-Nearest Neighbors







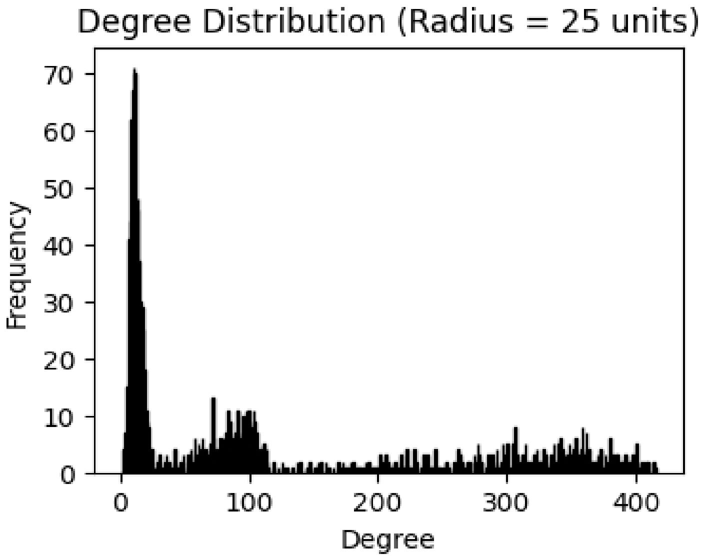


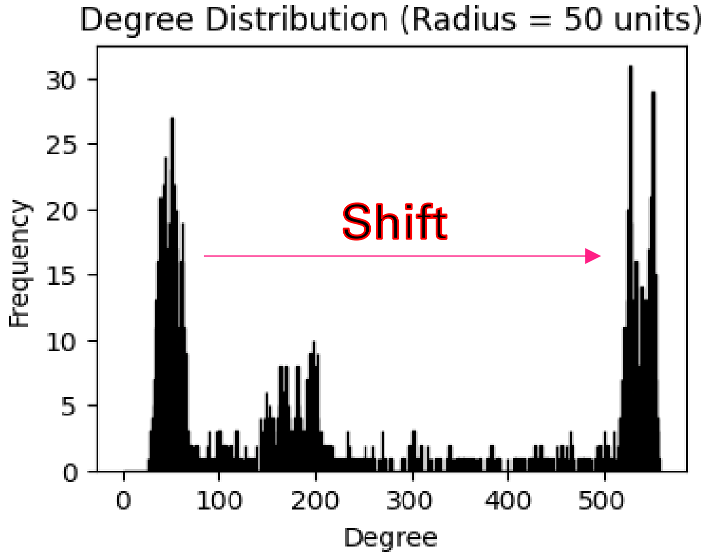
SPECTRAL ANALYSIS

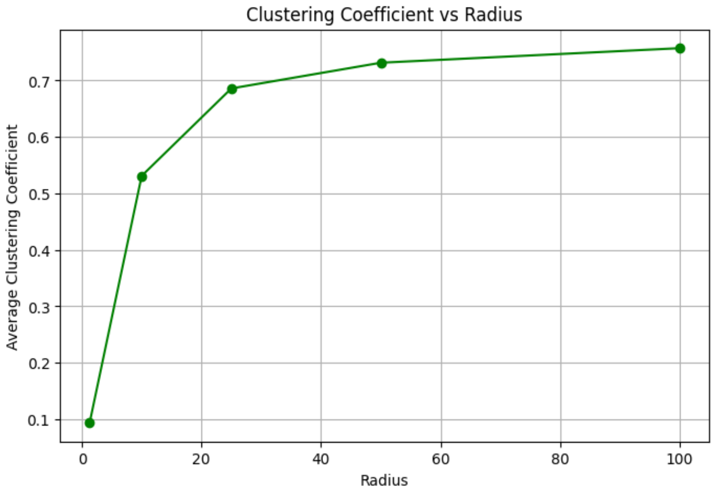
Graph Analysis

K-Nearest Neighbors

Previous Evidence



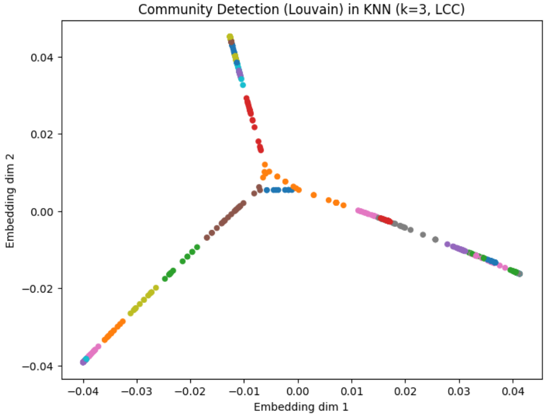


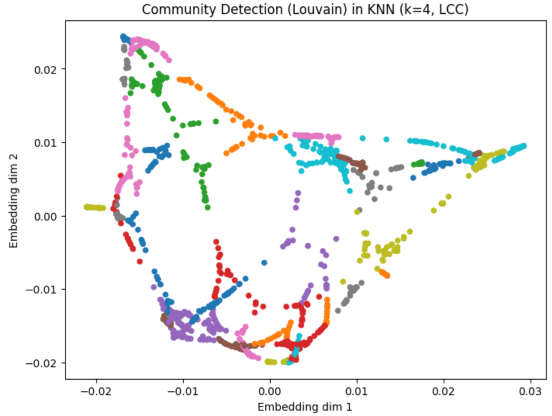


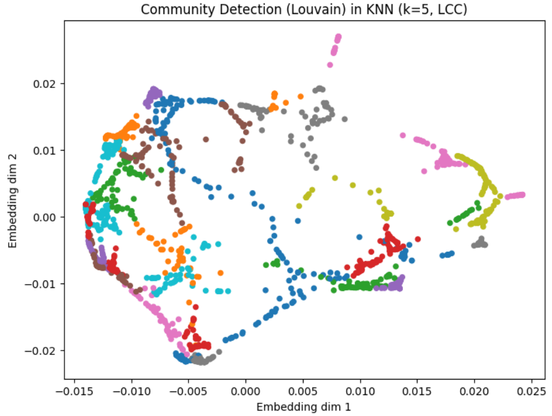
Potential Cause:

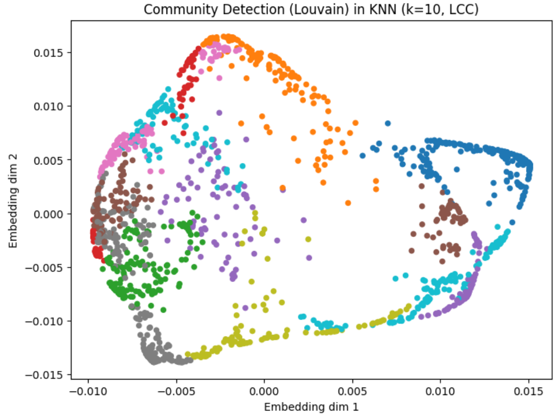
COMMUNITY ANALYSIS

K-Nearest Neighbors

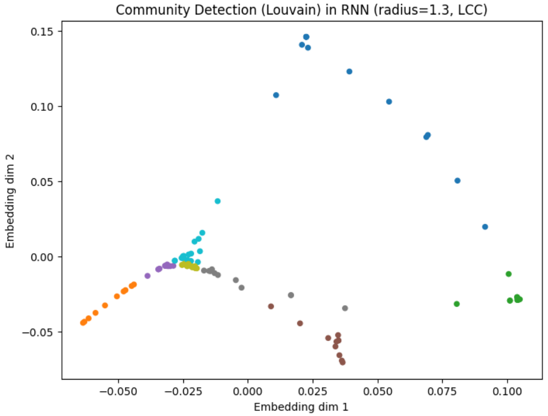


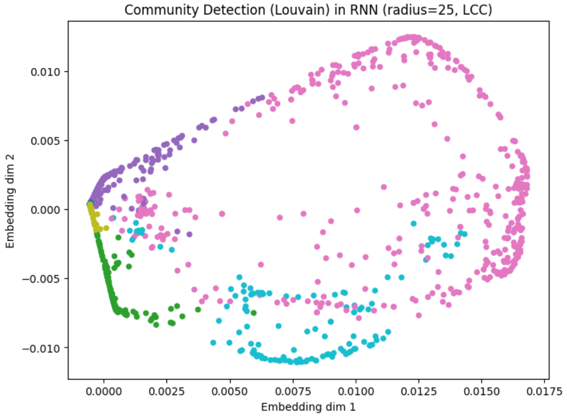


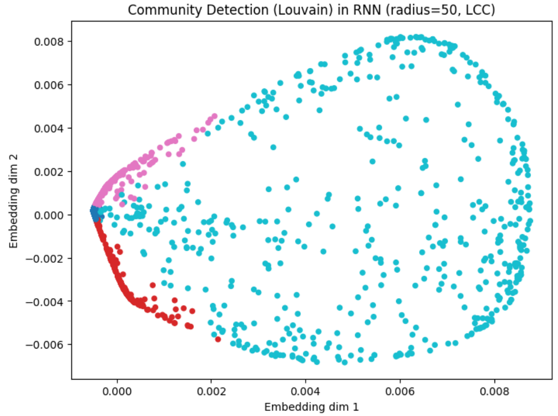


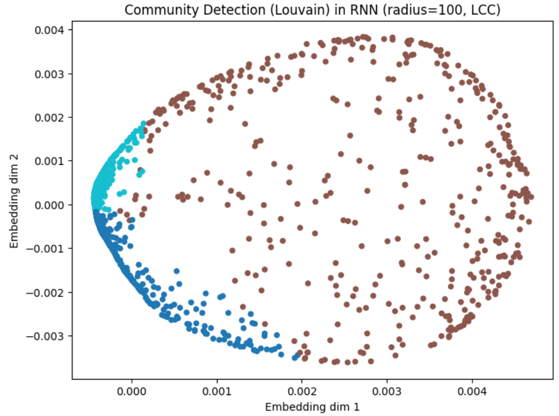


Radius-Nearest Neighbors



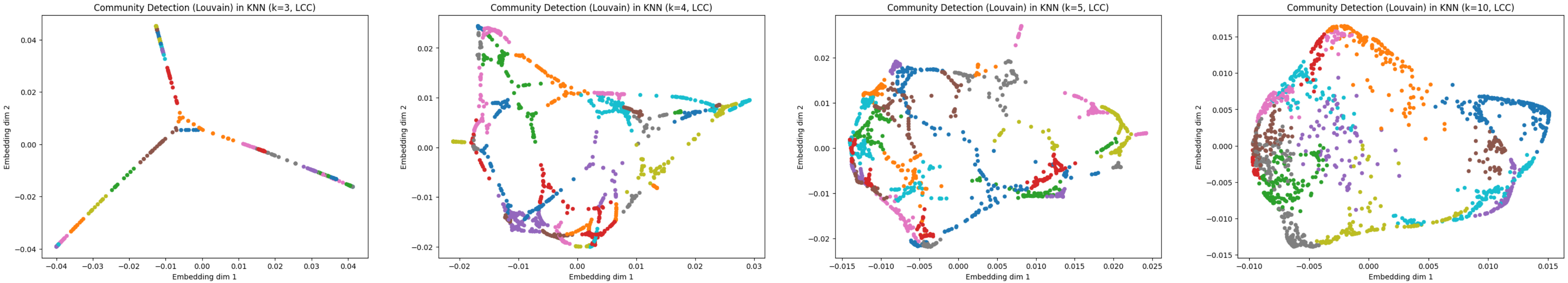




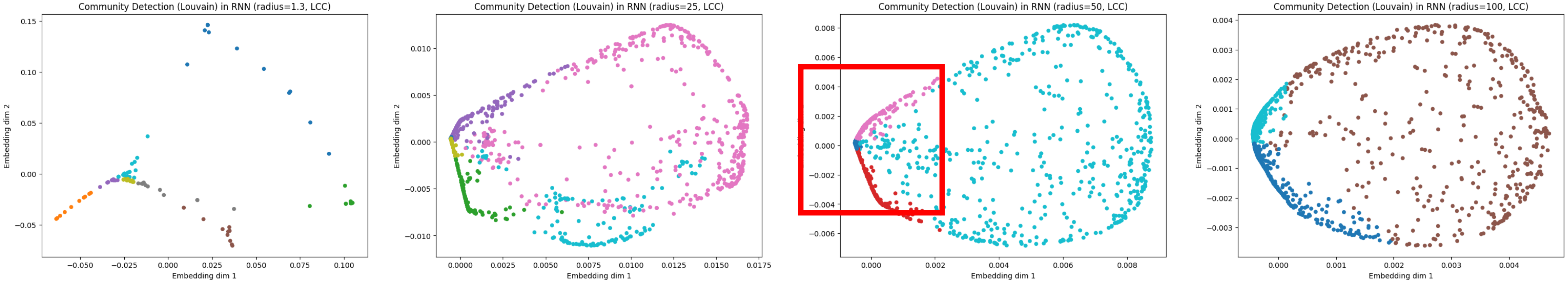


COMMUNITY ANALYSIS

K-Nearest Neighbors



Radius-Nearest Neighbors



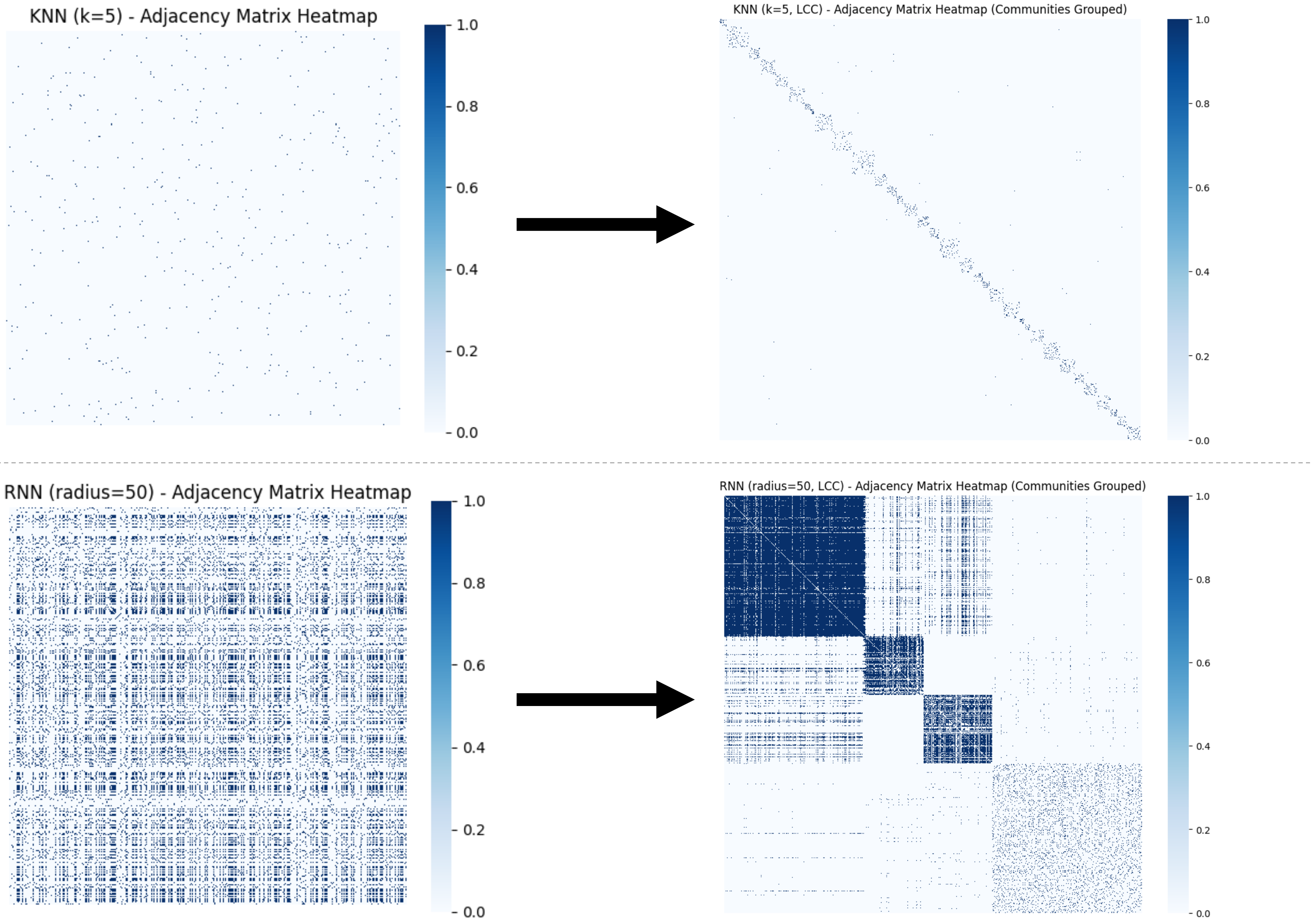
Graph Analysis

PARTITIONED ADJACENCY MATRIX

Original Adjacency Matrix | **After Community Detection**

KNN Graph (k=5)

RNN Graph (r=50)



AGENDA

- Research Question ✅
- Methodology ✅
- K-Nearest Neighbors ✅
- R-Nearest Neighbors ✅
- Graph Analysis ✅
- Key Takeaways ☑️

ANALYSIS SUMMARY

Characteristics	K-NN	R-NN
Graph Characteristics	• **Stays sparse** during connectivity growth • Degree distribution stays balanced as **k** increases	• **Density concentrates** during connectivity growth • Degree distribution inverts from low to high as **radius** increases
Spectral & Community Analysis	• **Small communities** but higher quantity • Embedding takes on a **web shape** – potentially hinting at the efficiently in connecting the stars	• **Communities are large**, fewer in numbers and vary in density • Embedding appears to be **concentrated**

💫 AGENDA 💫

- Research Question ✅
- Methodology ✅
- K-Nearest Neighbors ✅
- R-Nearest Neighbors ✅
- Graph Analysis ✅
- Key Takeaways ✅

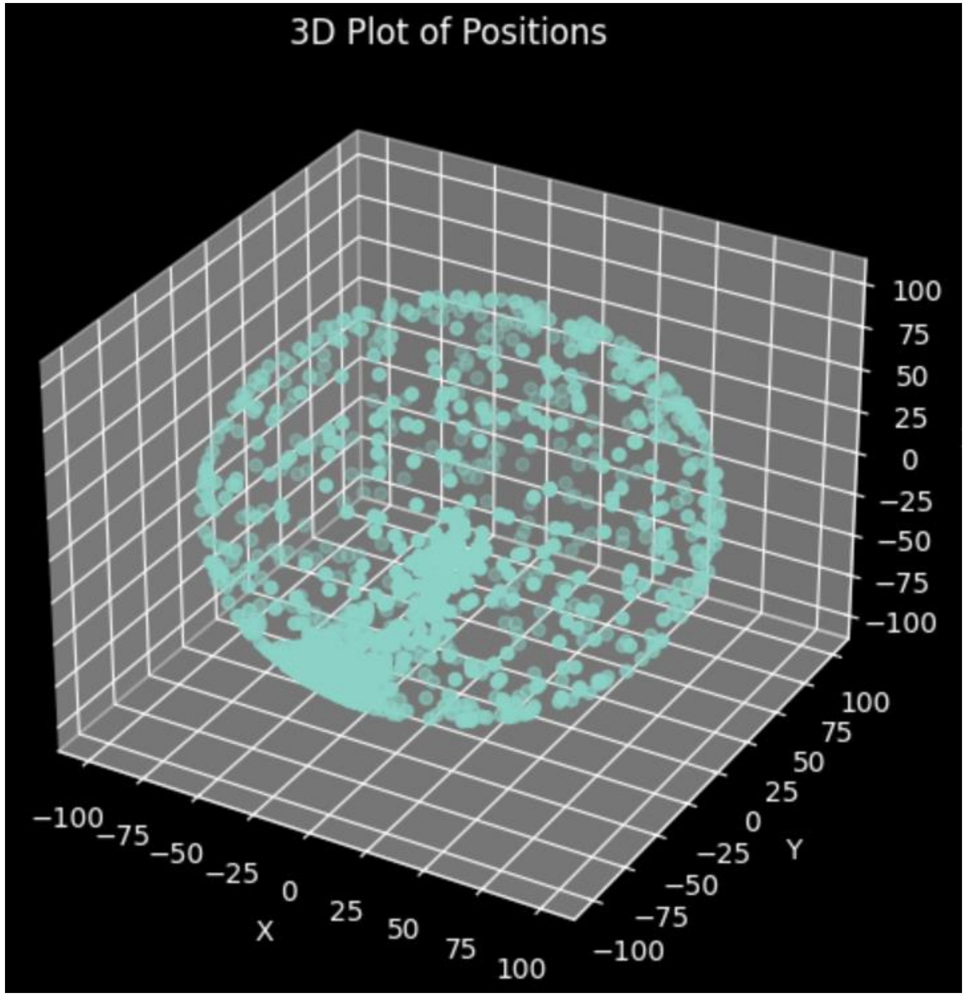
FINAL TAKEAWAY…

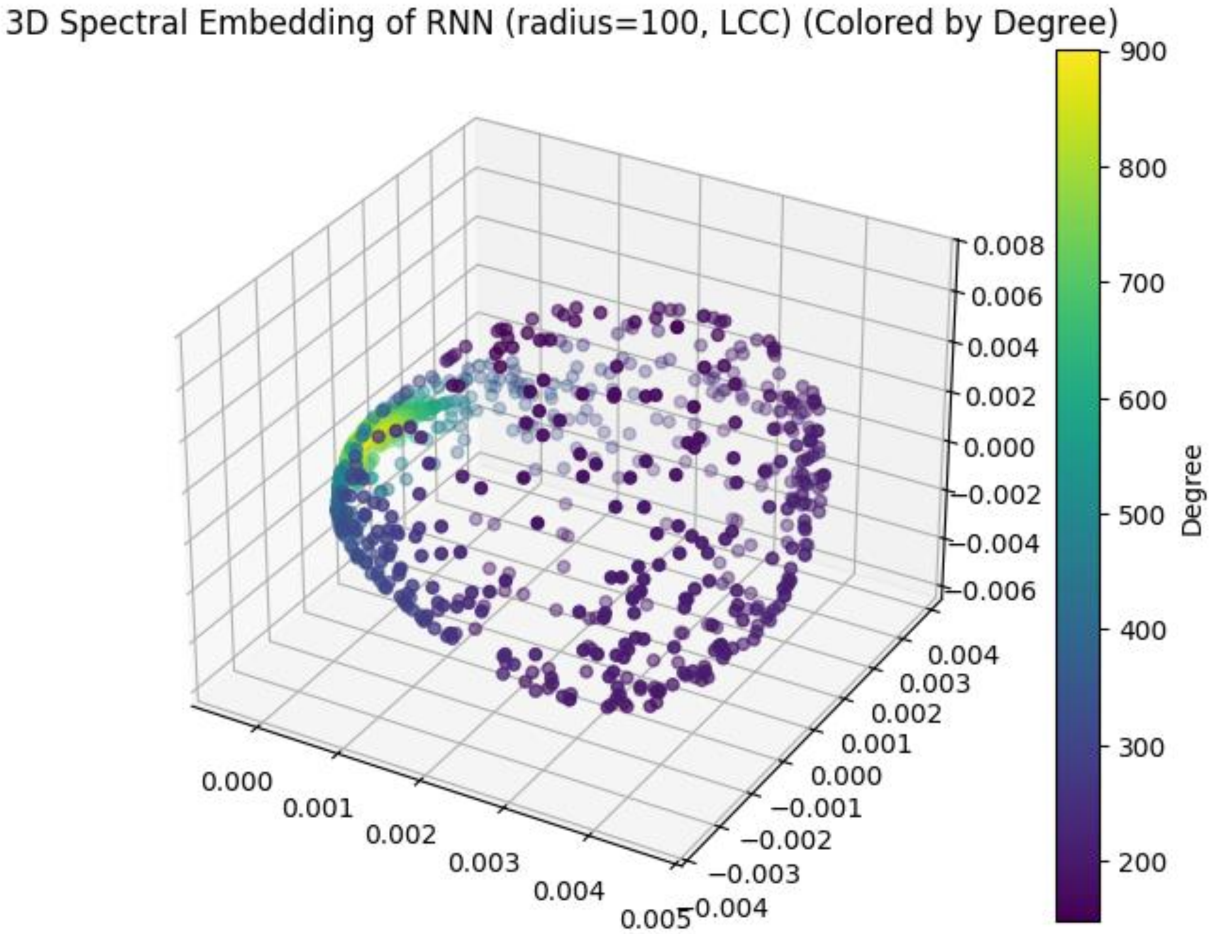
Key Takeaways

FINAL TAKEAWAY…

While you’re looking at the New Years fireworks, remember that somewhere far away is a star making our lightshow look like an ant pile.

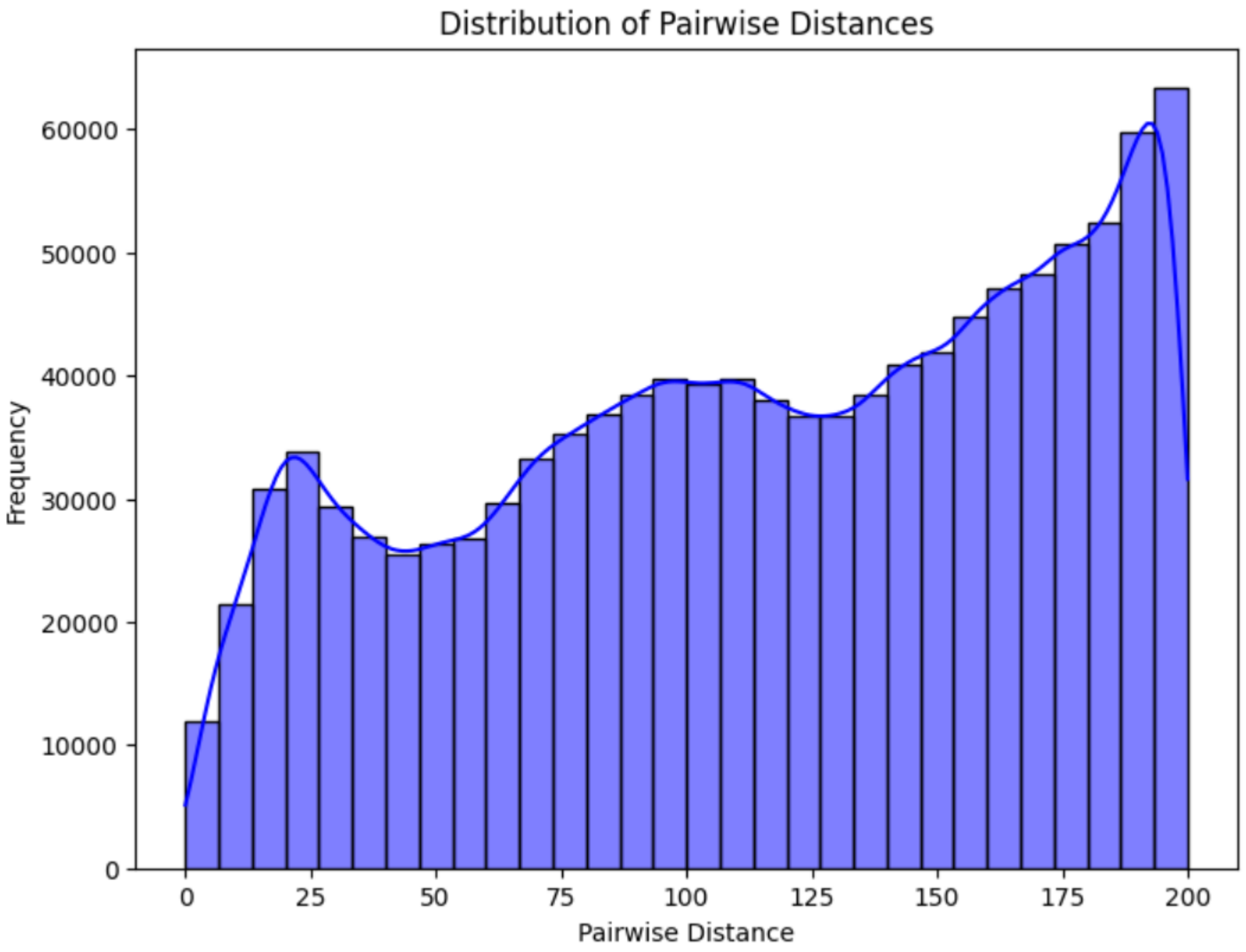
3D POSITION OF STARS





Appendix

DISTRIBUTION OF PAIRWISE DISTANCES



Appendix

LOUVAIN COMMUNITIES

Appendix

Louvain Community Detection Algorithm is a simple method to extract the community structure of a network. This is a heuristic method based on modularity optimization. [1]

The algorithm works in 2 steps. On the first step it assigns every node to be in its own community and then for each node it tries to find the maximum positive modularity gain by moving each node to all of its neighbor communities. If no positive gain is achieved the node remains in its original community.

The modularity gain obtained by moving an isolated node i into a community C can easily be calculated by the following formula (combining [1] [2] and some algebra):

$$\Delta Q = \frac{k_{i,in}}{2m} - \gamma \frac{\Sigma_{tot} \cdot k_i}{2m^2}$$

where m is the size of the graph, $k_{i,in}$ is the sum of the weights of the links from i to nodes in C, k_i is the sum of the weights of the links incident to node i, Σ_{tot} is the sum of the weights of the links incident to nodes in C and γ is the resolution parameter.

WORKS CITED

https://www.cosmos.esa.int/web/gaia/dr3 (GAIA-D3 Documentation)

https://www.aanda.org/articles/aa/full_html/2023/06/aa43940-22/aa43940-22.html (White Paper)

https://en.wikipedia.org/wiki/Gaia_(spacecraft) (Gaia background and image)

http://hyperphysics.phy-astr.gsu.edu/hbase/eclip.html (Astrophysics Methods)

https://www.vedantu.com/question-answer/explain-parallax-method-for-measuring-distance-class-11-physics-cbse-5f847566766fc5381b1d7c0b (Parallax Method)

https://en.wikipedia.org/wiki/Louvain_method (Louvain Community Detection)

Appendix