Project Update Report: Extending PRRP for Spatial and Graph Partitioning

Team 17: Aditya Gambhir (862466997) & Aditya Bhardwaj (862548633)

February 10, 2025

1 Project Overview

The project aims to implement and extend the PRRP (P-Regionalization through Recursive Partitioning) algorithm for both spatial regionalization and graph partitioning. The goal is to partition spatial and graph datasets while maintaining statistical significance and spatial contiguity. The implementation is structured into four phases:

1. **PRRP for Spatial Regionalization** (Completed)

2. **Extending PRRP to Graph Partitioning** (Ongoing, ∼60% complete)

3. **Experimental Evaluation and Comparison** (Upcoming)

4. **Project Report and Documentation** (Final Stage)

The codebase consists of several modules (e.g., `src/utils.py`, `src/spatial_prrp.py`, `src/metis_parser.py`) that together implement seed selection, region growing, merging, and splitting techniques for achieving contiguous regions under cardinality constraints.

2 Progress Since Project Inception

- **Phase 1 (Spatial PRRP):**

  - Analyzed the *Statistical Inference for Spatial Regionalization* paper [2].
  - Implemented region growing, merging, and splitting algorithms in Python.
  - Developed parallel execution using multiprocessing and conducted unit tests on synthetic spatial datasets.

- **Phase 2 (Graph Partitioning Extension):**

  - Studied the METIS manual [5] and PyMETIS documentation [6].
  - Modified the original PRRP logic to work with graph data and adjusted seed selection to avoid isolated subgraphs.
  - Completed approximately 60% of the code and initial tests on synthetic graph datasets.

- **Phases 3 & 4:**

  - Experimental evaluation and comprehensive documentation are pending. Detailed experiments are planned using both the datasets from the paper and the dataset provided by the professor.

3 Team Contributions

Our group comprises two members, each contributing equally to all aspects of the project:

- **Code Development:** Joint work on designing and implementing the PRRP algorithm, including spatial and graph partitioning modules.

- **Testing and Debugging:** Collaborative efforts on unit testing and integration testing across different datasets.

- **Documentation and Reporting:** Equal participation in drafting this report, preparing code documentation, and organizing the GitHub repository.

4 Future Timeline and Milestones

The final deliverable is due on **March 12, 2025**. The upcoming milestones are as follows:

- **February 24, 2025:** Submit the next update report.

- **February 10 − February 24, 2025:** Complete testing on METIS-format graph datasets and fine-tune performance of the graph partitioning extension.

- **February 24 − March 5, 2025:** Conduct extensive experiments and comparative evaluation of PRRP versus METIS and PyMETIS. Generate performance metrics and visualizations.

- **March 5 − March 12, 2025:** Finalize the project report and documentation. Prepare and verify final deliverables.

5 Pending Tasks and Challenges

Despite the progress made, several challenges remain:

a. **Completion of Graph PRRP:** Finalize and rigorously test the modifications for graph partitioning.

b. **Extensive Experimental Evaluation:** Running large-scale experiments on both the synthetic and professor-provided datasets remains the biggest challenge.

c. **Performance Tuning:** Optimization for scalability and memory usage during extensive testing.

d. **Documentation:** Ensuring that code documentation and the final report are comprehensive and clear.

6 Related Work

Additional relevant works that informed our approach include:

- The scalable max-P regionalization technique proposed by Alrashid et al. [1].

- The formulation of the max-P-regions problem by Duque et al. [4].

- Approaches to balanced graph partitioning as studied by Andreev and Räcke [3].

7 Conclusion

In summary, the project is on track with the spatial PRRP phase completed and the graph extension well underway. Our next focus is on completing the graph-based PRRP tests and running extensive experiments to validate the approach. With a clear timeline and equal team contributions, we are confident of meeting the final deadline of March 12, 2025.

References

[1] ALRASHID, H., LIU, Y., AND MAGDY, A. Smp: Scalable max-p regionalization. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems* (New York, NY, USA, 2022), SIGSPATIAL '22, Association for Computing Machinery.

[2] ALRASHID, H., MAGDY, A., AND REY, S. Statistical inference for spatial regionalization. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems* (New York, NY, USA, 2023), SIGSPATIAL '23, Association for Computing Machinery.

[3] ANDREEV, K., AND RÄCKE, H. Balanced graph partitioning. In *Proceedings of the Sixteenth Annual ACM Symposium on Parallelism in Algorithms and Architectures* (New York, NY, USA, 2004), SPAA '04, Association for Computing Machinery, pp. 120–124.

[4] DUQUE, J. C., ANSELIN, L., AND REY, S. J. The max-p-regions problem. *Journal of Regional Science 52*, 3 (2012), 397–419.

[5] KARYPISLAB. Metis 5.1.x manual. `https://github.com/KarypisLab/METIS/blob/master/manual/manual.pdf`, 2023. Accessed: 2025-02-10.

[6] PYMETIS TEAM. Pymetis documentation. `https://documen.tician.de/pymetis/functionality.html`, 2023. Accessed: 2025-02-10.