# Fast Deterministic Single-Linkage 2D-Spatial Cluster Analysis

Daniel Goldbach
University of New South Wales

## ABSTRACT

Cluster analysis is a common task in data mining, machine learning and related fields. There exist a plethora of clustering algorithms designed for this purpose, but many are prohibitively inefficient (e.g. quality-threshold clustering), non-deterministic (k-means) or utilise inherently lossy partitioning models (k-d tree clustering). Single-linkage hierarchical clustering is a form of cluster analysis which unites clusters based on the minimum distance between them, using a given distance metric. Though more complex clustering methods exist, the intuitive nature and ease of implementation of single-linkage hierarchical clustering makes it a reasonably common choice for cluster analysis. However, the general case of single-linkage clustering is $O(n^3)$ (though the SLINK algorithm runs in $O(n^2)$ time for some special cases [1]).

A specific case – and likely the most intuitive case – of cluster analysis is that which is performed on a two-dimensional Euclidean plane. This has many real-world applications, including image analysis/segmentation and medical imaging. This paper presents a quasi-linear time algorithm for single-linkage hierarchical clustering of points in two-dimensional Euclidean space. This concept is not itself novel (see [2, 3]); however, the use of an agglomerative approach as opposed to a fixed-threshold edge filtering provides a concise and effective way to extract a specific number of clusters. The algorithm also guarantees that the maximum distance between any pair of points in a cluster is minimised.

## BODY

*Find the Delaunay triangulation of the points with Sweephull, then Kruskal's MST until the required cluster count is reached.* $O(n \log n)$.

## REFERENCES

[1] R. Sibson. "SLINK: an optimally efficient algorithm for the single-link cluster method." In *The Computer Journal (British Computer Society) 16 (1)*, 1973.

[2] X. Yang and W. Cui. "A Novel Spatial Clustering Algorithm Based on Delaunay Triangulation." In *Journal of Software Engineering and Applications, Vol. 3 No. 2*, 2010.

[3] In-Soo Kang, Tae-wan Kim and Ki-Joune Li. "A Spatial Data Mining Method by Delaunay Triangulation." In *Proceedings of the 5th ACM International Workshop on Advances in Geographic Information Systems*, 1997.

*Volume 1 of Tiny Transactions on Computer Science*