Discrete Structures Project Report

**Clustering using Spanning tree**

horizontal line

# 



**DELHI TECHNOLOGICAL UNIVERSITY**

**PROJECT REPORT**

**Submitted By: Submitted To:**

Abhishek Kumar Singh (2K19/CO/021)Ms. Tanya Malhotra

Abhishek Singh (2K19/CO/024)

**DELHI TECHNOLOGICAL UNIVERSITY**

(Formerly Delhi College of Engineering)

Bawana Road, Delhi, 110042

**CANDIDATE’S DECLARATION**

We, Abhishek Kumar Singh, Abhishek Singh Roll No(s). 2K19/CO/021, 2K19/CO/024, students of B.Tech. in Computer Science & Engineering, hereby declare that the project Dissertation titled **Clustering using Spanning tree** which is submitted by us to the Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the mid-semester component evaluation, semester-3 of Bachelor of Technology is original and not copied from any source without proper citation. This work has not previously formed a basis for the award of any degree, Diploma Associateship, Fellowship, or any similar title or recognition.

Place: Delhi Abhishek Kumar Singh

Date: 23 November 2020 Abhishek Singh

**Department of Computer Science & Engineering**

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi, 110042

**CERTIFICATE**

We hereby declare that the project Dissertation titled “**Clustering using spanning tree**” which is submitted by Abhishek Kumar Singh, Abhishek Singh Roll No(s). 2K19/CO/021, 2K19/CO/024 Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the mid-semester component evaluation, semester-3 of Bachelor of Technology, is the record of the project work carried out by the students under my supervision.

Place: Delhi **Ms. Tanya Malhotra**

Date: 23/11/2020 Delhi Technological University

(Formerly DCE)

Bawana Road, Delhi - 110042

**ACKNOWLEDGEMENT**

At the very outset of this report, we would like to extend our sincere and heartfelt obligation towards all the personages who have guided us with the project.

A special thanks to Ms. Tanya Malhotra for teaching us the subject “Discrete Structures”. She helped us visualize the subject and to find its applications in real life. She supervised us with the intricacies of this project. She also offered many relevant and productive recommendations for the project, for which we are very grateful. We would also like to extend sincere gratitude towards our Vice-Chancellor Mr. Yogesh Singh for allowing the students to improve their theoretical with crucial subjects like Discrete Structures.

**Aim**

To partition a given set of objects into subsets (or clusters) in such a way that any two objects from the same subset are close (or similar ) to each other, while any two objects from different subsets are far apart.

**Abstract**

Clustering is the process of putting together meaning-full or use-full similar object into one group. It is a common technique for statistical data, machine learning, and computer science analysis. Clustering is a kind of unsupervised data mining technique which describes general working behavior, pattern extraction and extracts useful information from unstructured data. Clustering algorithms are used for disease classification in medical science and also for customer classification in marketing research and for environmental health risk assessment in environmental engineering. It is used in many fields including pattern recognition, image analysis, information retrieval, bioinformatics, data compression and computer graphics. In this paper we have implemented clustering algorithm based on spanning tree to compute the largest possible value of d such that the given points can be partitioned into 𝑘 non-empty subsets in such a way that the distance between any two points from different subsets is at least d.

**Introduction**

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them. The purpose of clustering algorithms is to make sense of and extract value from large sets of structured and unstructured data.If working with huge volumes of unstructured data, it only makes sense to try to partition the data into some sort of logical groupings before attempting to analyze it.

Clustering allows you to take a sweeping glance of your data en masse, and then form some logical structures based on what you find there before going deeper into the nuts-and-bolts analysis.

In their simplest form, *clusters* are sets of data points that share similar attributes, and *clustering algorithms* are the methods that group these data points into different clusters based on their similarities.

**Preliminaries/Basic Knowledge**

1. **Graph**

A Graph is a non-linear data structure consisting of nodes and edges. The nodes are sometimes also referred to as vertices and the edges are lines or arcs that connect any two nodes in the graph. More formally a Graph can be defined as,

*A Graph consists of a finite set of vertices(or nodes) and set of Edges which connect a pair of nodes.*

**

*Graphs are used to solve many real-life problems. Graphs are used to represent networks. The networks may include paths in a city or telephone network or circuit network. Graphs are also used in social networks like linkedIn, Facebook.*

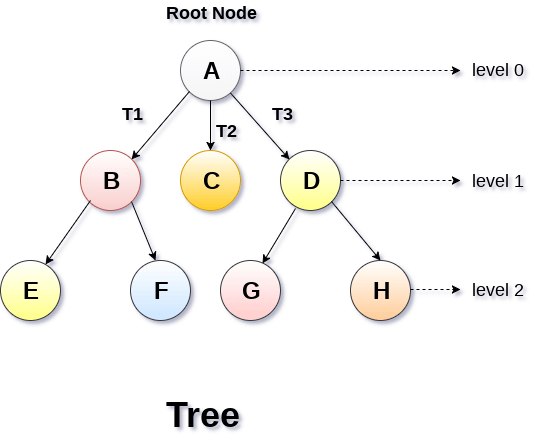
1. **Trees**

A Tree is a recursive data structure containing the set of one or more data nodes where one node is designated as the root of the tree while the remaining nodes are called as the children of the root.

The nodes other than the root node are partitioned into the non empty sets where each one of them is to be called a sub-tree.

Nodes of a tree either maintain a parent-child relationship between them or they are sister nodes.

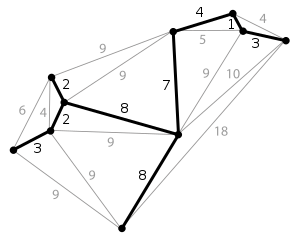
The following image shows a tree, where the node A is the root node of the tree while the other nodes can be seen as the children of A.



In a general tree, A node can have any number of children nodes but it can have only a single parent.

1. **Minimum Spanning Tree**

A minimum spanning tree (MST) or minimum weight spanning tree is a subset of the edges of a connected, edge-weighted undirected graph that connects all the vertices together, without any cycles and with the minimum possible total edge weight. That is, it is a spanning tree whose sum of edge weights is as small as possible. More generally, any edge-weighted undirected graph (not necessarily connected) has a minimum spanning forest, which is a union of the minimum spanning trees for its connected components

.

1. **C++**

C++ is a general-purpose programming language created by Bjarne Stroustrup as an extension of the C programming language, or "C with Classes. The language has expanded significantly over time, and modern C++ now has object-oriented, generic, and functional features in addition to facilities for low-level memory manipulation.

**Task**

Given 𝑛 points on a plane and an integer 𝑘, compute the largest possible value of 𝑑 such that the given points can be partitioned into 𝑘 non-empty subsets in such a way that the distance between any two points from different subsets is at least 𝑑.

**Main Work/ Implementation Detail**

**This code is used to find weight between the vertices**

double weight(int x1, int y1, int x2, int y2) {

return sqrt((x1 - x2) \* (x1 - x2) + (y1 - y2) \* (y1 - y2));

**Main code for clustering**

double clustering(vector<int> x, vector<int> y, int k) {

int n = x.size();

vector<node> nodes;

for (int i = 0; i < n; i++)

{

make\_set(i, nodes, x, y);

}

vector<edge> edges;

for (int i = 0; i < n; i++) {

for (int j = i + 1; j < n; j++)

{

edges.push\_back(edge(i, j, weight(x[i], y[i], x[j], y[j])));

}

}

**Main function**

int main() {

size\_t n;

int k;

std::cin >> n;

vector<int> x(n), y(n);

for (size\_t i = 0; i < n; i++)

{

std::cin >> x[i] >> y[i];

}

std::cin >> k;

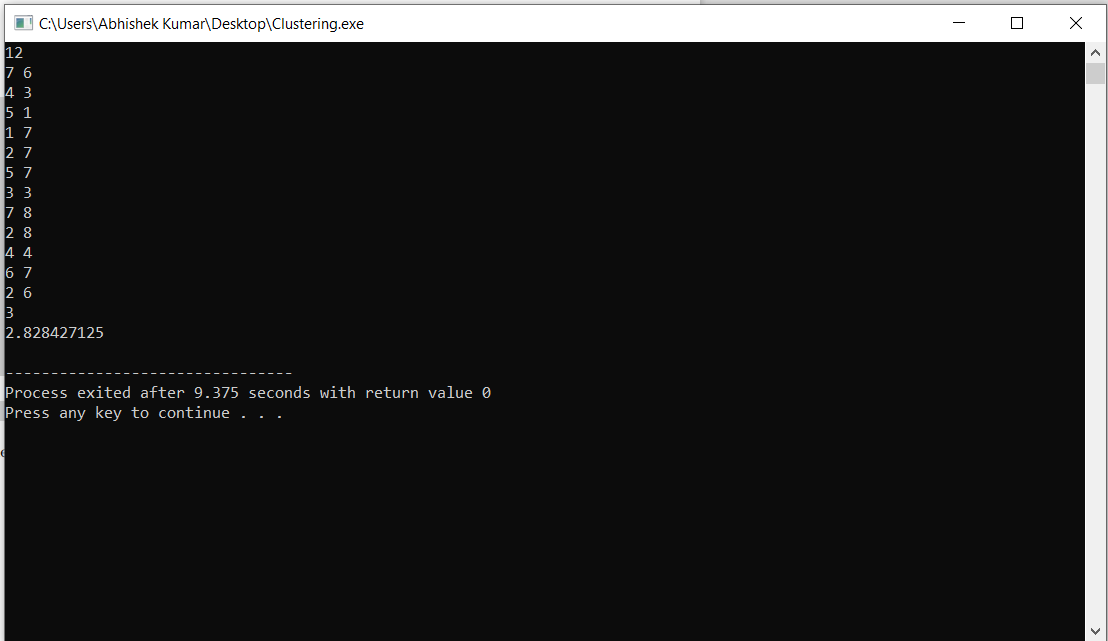
std::cout << std::setprecision(10) << clustering(x, y, k) << std::endl;

}

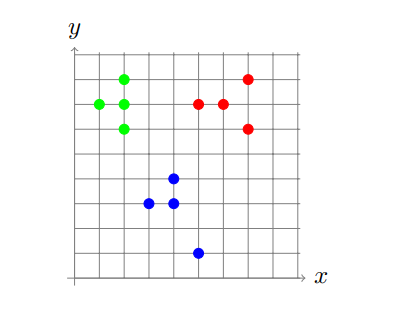
**Output**

Input : We have given the number of objects in first line i.e. 12 and coordinates of those 12 objects in the next 12 line. The last line contains the number of clusters i.e. 3.

Output : Output the largest value of d i.e. 2.828.

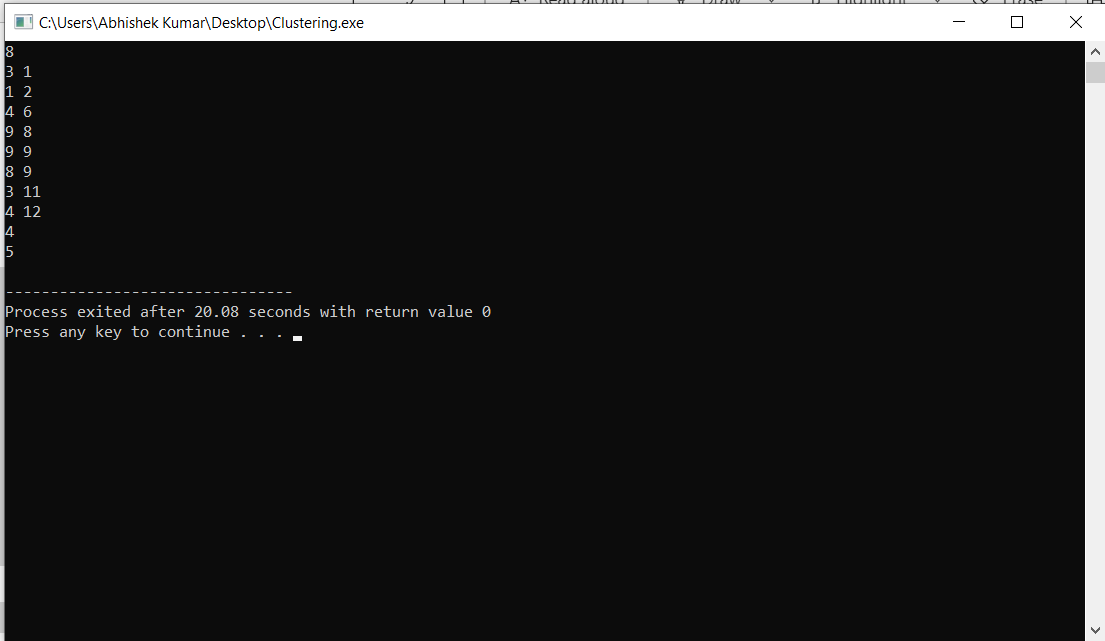


The distance between any two objects from different subsets is at least 𝑑 i.e. √ 8. The corresponding partition of the set of points into three clusters is shown below

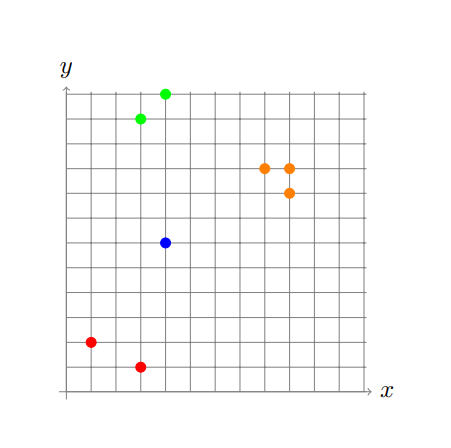
.

Input : We have given the number of objects i.e. 8 in the first line and coordinates of those 8 objects in the next 8 lines. The last line contains the number of clusters i.e. 4.

Output : Output the largest value of d i.e. 5.



The distance between any two objects from different subsets is at least 𝑑 i.e. 5. The corresponding partition of the set of points into four clusters is shown below.



**Conclusion**

We have successfully implemented clustering algorithm and partitioned a given set of objects into subsets (or clusters) in such a way that any two objects from the same subset are close (or similar ) to each other, while any two objects from different subsets are far apart using spanning tree. This can be very helpful and can be used in various real life applications like in Marketing and Sales(This is achieved by looking at specific characteristics of a person and sharing campaigns with them that have been successful with other similar people. ), Network design(*– telephone, electrical, hydraulic, TV cable, computer, road.*), etc.

**References**

1. S.R. Nanda, B. Mahanty, M.K. Tiwari “Clustering Indian stock market data for portfolio management” of Expert Systems with Applications @ 2010
2. S. Narain Sinha, Ram Lal Yadav “Analysis of Data using K-Means Clustering Algorithm with Min Max Function” @ 2018