Determination of Different Sizes of Partitioning Clusters in a Highly Connected Graph Using Minimum Sub-cycles

No Author Given

No Affiliation Given

No Author Given

No Affiliation Given

Abstract— According to Content of Abstract

Keywords-component; Partiotioning Clustering; Clipue; Difference Density; Greedy Strategy;

# Introduction

ความเป็นมาของการจัดกลุ่ม เทคนิคทั่วไปที่ใช้กัน ลักษณะข้อมูลที่ต้องการการจัดกลุ่ม bio, network การนำเสนอของการจัดกลุ่มข้อมูล ประเภทของการจัดกลุ่ม ความท้าทายในการจัดกลุ่มกับข้อมูลที่มีปริมาณมาก ๆ การใช้ตัววัดคุณภาพ

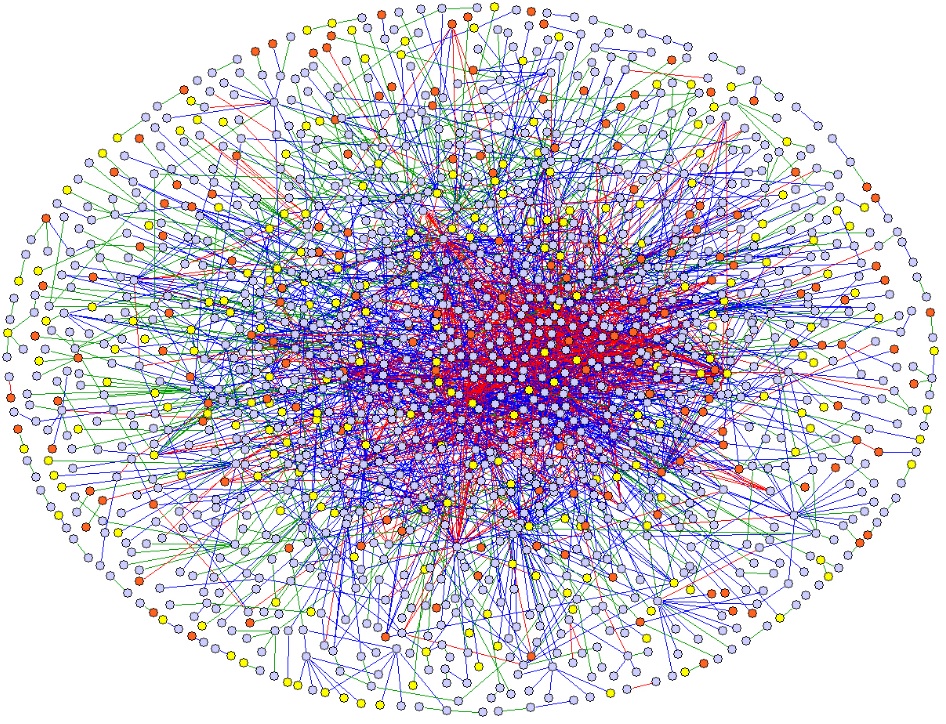
This research is concerned with partitioning clustering of a highly connected graph using minimum sub-cycles.

Organization of the paper

# Preliminary Discussion

## Graph Clustering

การจัดกลุ่มข้อมูลในรูปแบบของกราฟโดยทั่วไปสามารถแบ่งออกได้ 3 ประเภทคือ 1. Partitioning Clustering 2. Overlapping clustering 3. Hierarchical clustering ซึ่งทั้งหมดมีการจัดรูปแบบของข้อมที่แตกต่างกันตามลัดษณะของชื่อเรียก Overlap คือการจัดกลุ่มข้อมูลที่มีการ “ทับซ้อน” กันของข้อมูลของครัสเตอร์ตั้งแต่ 2 ครัสเตอร์ขึ้นไป Hierarchical คือ การจัดกลุ่มข้อมูลที่มี “ลำดับขั้น” โดยใช้ลำดับขั้นนั้นเป็นครัสเตอร์ของข้อมูล และ Partitioning เป็นการจัดกลุ่มข้อมูลที่แต่ละครัสเตอร์มีการแบ่งแยกออกจากกัน และมีความแตกต่างกันอย่างชัดเจนทั้งภายนอกและภายในของครัสเตอร์ ซึ่งในผลงานชิ้นนี้ได้เลือกใช้ในแบบ Partitioning



(รูปกราฟ)

## Partitioning Clustering

การจัดกลุ่มข้อมูลในประเภท Partitioning คือ การจัดกลุ่มข้อมูลที่ละครัสเตอร์มีการแบ่งแยกออกจากกันชัดเจน ไม่มีการ ”ทับซ้อน” หรือ “ลำดับขั้น” ใน 1 ข้อมูลจะอยู่กลุ่มใดกลุ่มหนึ่ง ข้อมูลภายในครัสเตอร์และภายนอกครัสเตอร์จะมีความแตกต่างกันอย่างชัดเจน

In general, partitioning clustering of a graph can have many forms (show different forms possible partitioning clustering of a graph – say 3 clusters)

Therefore, the task is an NP class problem.

# Density of Graph

เราจะใช้อันนี้ในการวัด 0-1 ค่าระหว่างนี้ต้องมีการกำหนด การครัสเตอร์จะทำแทบไม่ได้เลย จึงต้องมีความจำเป็นในการกำหนด ในทางปฏิบติแล้วเราอาจจะสนใจที่ต่างค่ากัน สนใจ highly con. ไปด้วย

# Literature Review

การจัดกลุ่มที่เกิดการพัฒนาขึ้นมาหลายยุคหลายสมัย และปริมาณของข้อมูลที่เพิ่มขึ้นในระดับเพตะไบต์ (Petabyte) การอัพเดทการจัดกลุ่มจึงมีความสำคัญมากเพื่อรองรับปริมาณข้อมูลที่มีการเปลี่ยนแปลงในระดับมหาศาล []. อ้างผลงงานสัก 4 ฉบับ 1. การใช้กราฟจัดกลุ่มข้อมูล 2. การใช้กราฟแบบไม่มีทิศทาง ไม่มีน้ำหนัก มาจัดกลุ่ม 3. การใช้ความหนาแน่นมาจัดกลุ่ม 4. การใช้ metric มาตรวจสอบคุณภาพของทั้งกราฟ + Spanning graph

# Methodology

หลักการดำเนินงานวิจัย ดังนี้

## Partitioning Clustering Algorithm

### Minimum Sub

### Partitioning Algorithm

# Metric Of Graph

ทดลองไปกี่แบบ.....

ใช้ Difference Density อธิบายวิธีการหา....

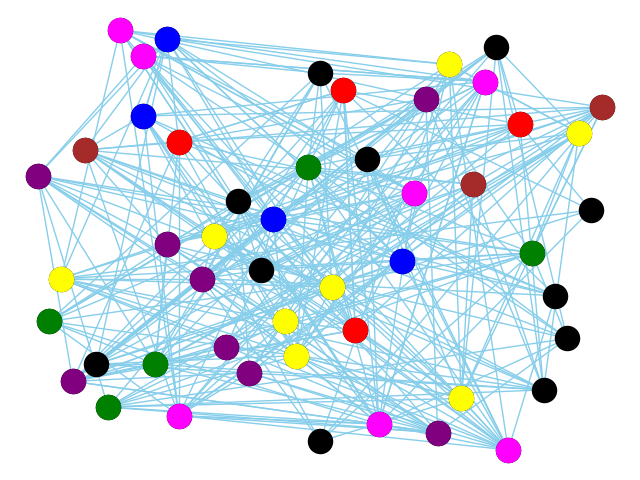
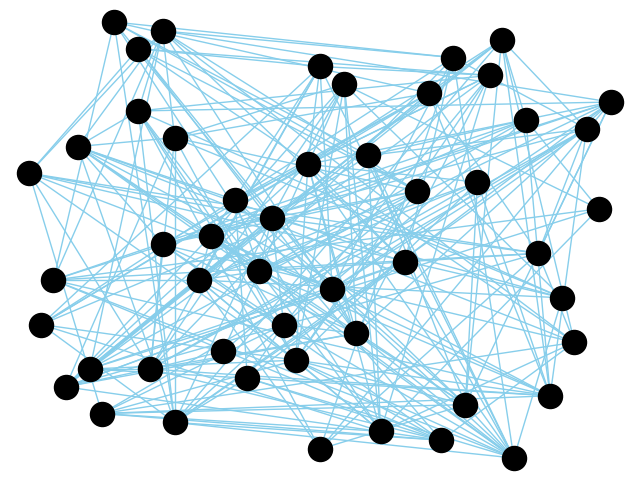
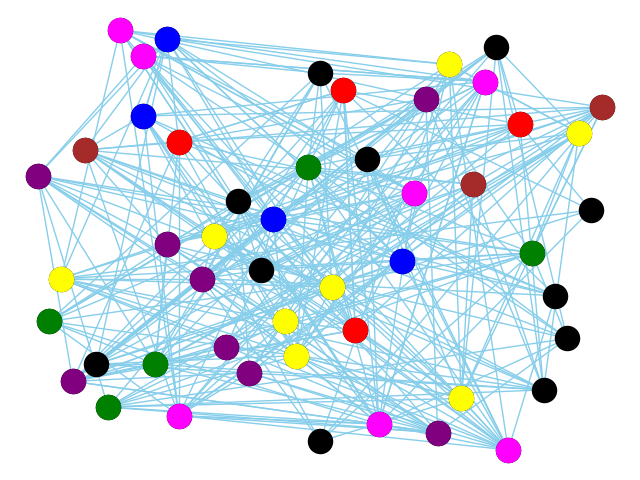
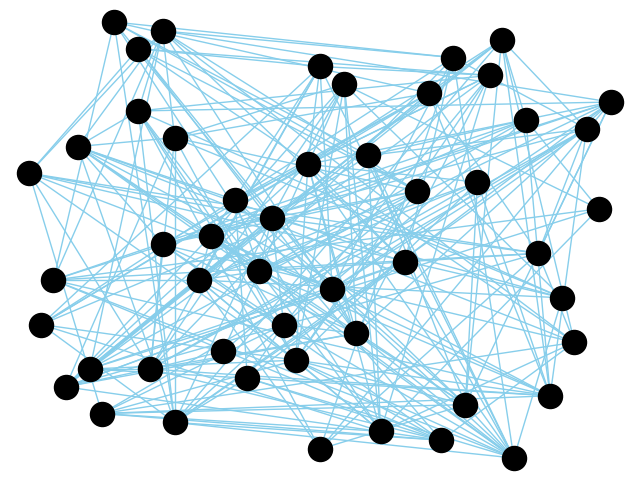
สูตร ของ Coverage, Conductance

# Experiment and Result

จากการทดลองที่ผ่านมากับกราฟที่มีโหนดเป็นจำนวนมาก แต่เพื่อความเข้าใจที่มองเห็นได้ชัดเจน จึงขอแสดงผลกับกราฟจำนวนโหนด 50 โหนด ตามตัวอย่างด้านล่างนี้จำนวน 2 ตัวอย่างคือ

## Example 1 50 Nodes 300 Edges

### Difference Density = 0.5



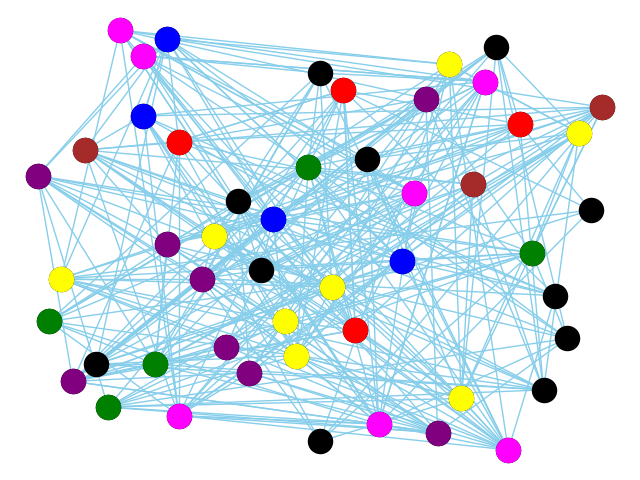
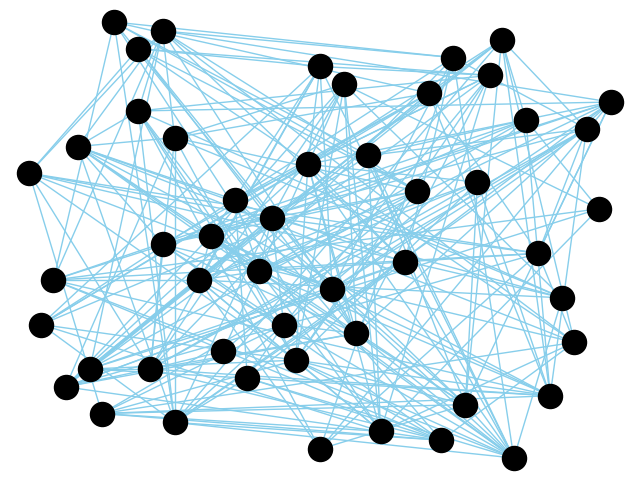
ใส่รูป

Result 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Graph | Node | No. of edges within the clusters | Total degree of all nodes in the cluster | ค่า Difference Density | Un - Cluster Node |
|  |  |  |  |  |  |
|  |  | ใช้กราฟ Full Detail | |  |  |
|  |  |  |  |  |  |

Table 1

### Difference Density = 0.6

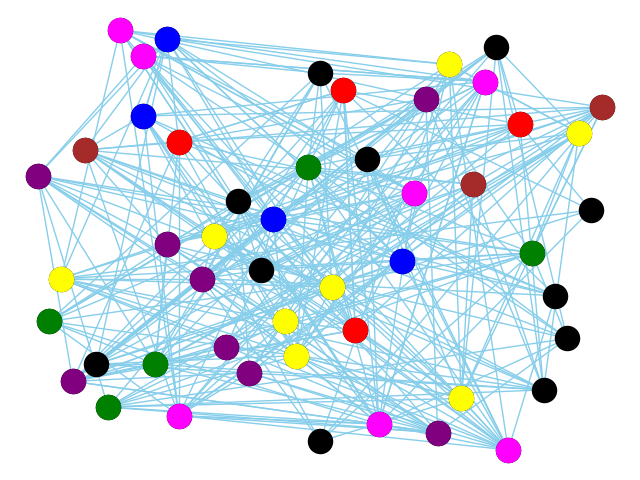
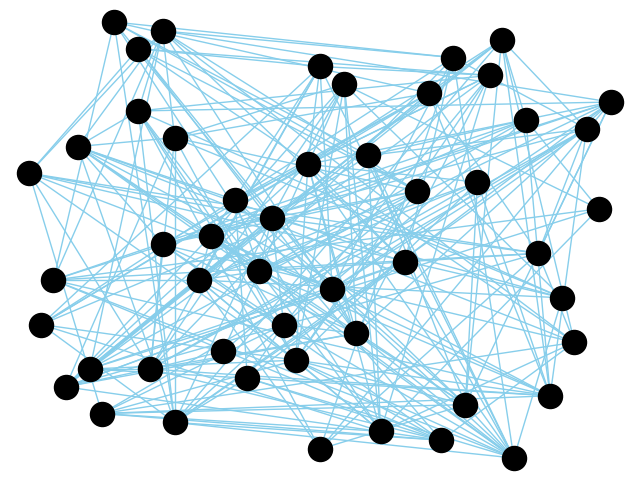


ใส่รูป

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Graph | Node | No. of edges within the clusters | Total degree of all nodes in the cluster | ค่า Difference Density | Un - Cluster Node |
|  |  |  |  |  |  |
|  |  | ใช้กราฟ Full Detail | |  |  |
|  |  |  |  |  |  |

Table 2

### Difference Density = 0.7



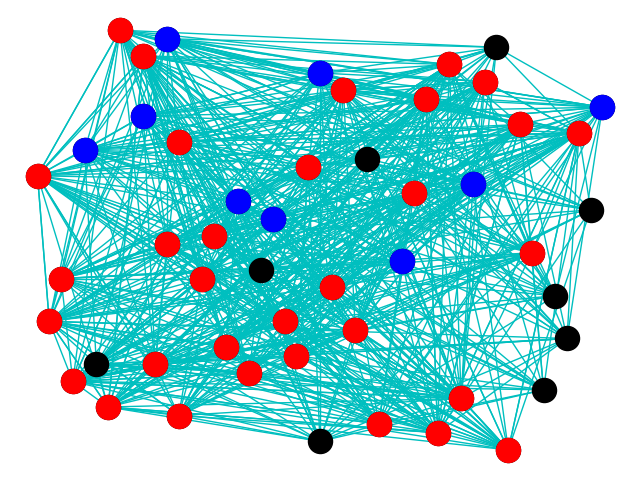
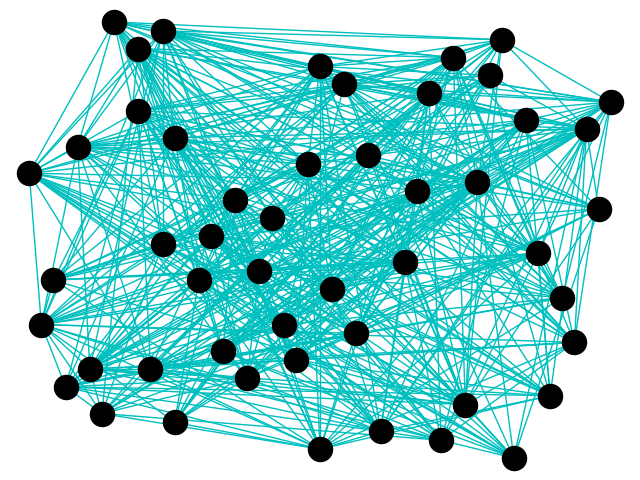
ใส่รูป

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Graph | Node | No. of edges within the clusters | Total degree of all nodes in the cluster | ค่า Difference Density | Un - Cluster Node |
|  |  |  |  |  |  |
|  |  | ใช้กราฟ Full Detail | |  |  |
|  |  |  |  |  |  |

Table 3

## Example 2 50 Nodes 500 Edges

### Difference Density = 0.5



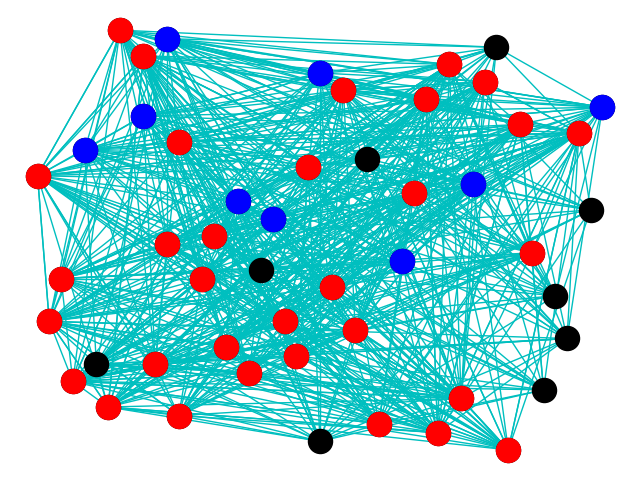
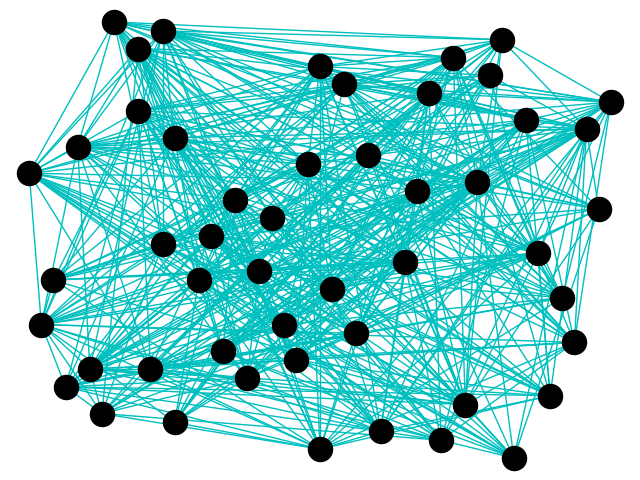
ใส่รูป

Result 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Graph | Node | No. of edges within the clusters | Total degree of all nodes in the cluster | ค่า Difference Density | Un - Cluster Node |
|  |  |  |  |  |  |
|  |  | ใช้กราฟ Full Detail | |  |  |
|  |  |  |  |  |  |

Table 4

### Difference Density = 0.6

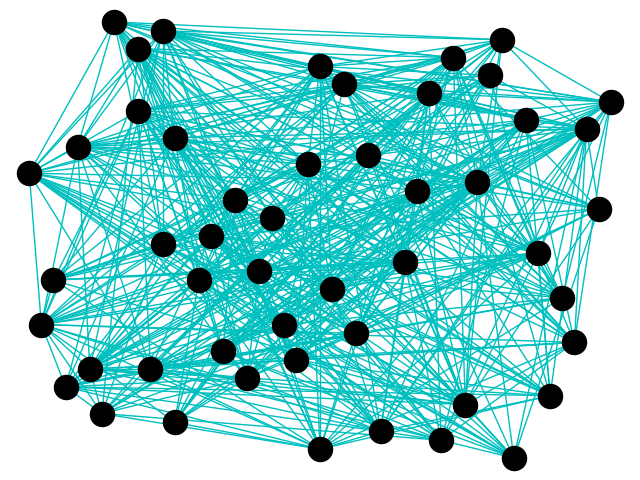


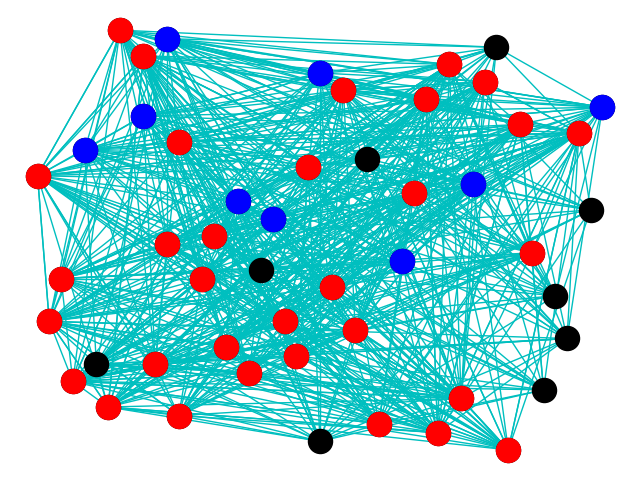
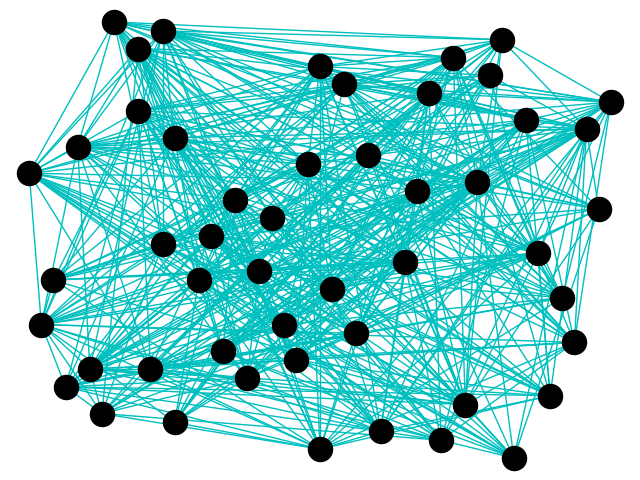
ใส่รูป

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Graph | Node | No. of edges within the clusters | Total degree of all nodes in the cluster | ค่า Difference Density | Un - Cluster Node |
|  |  |  |  |  |  |
|  |  | ใช้กราฟ Full Detail | |  |  |
|  |  |  |  |  |  |

Table 5

### Difference Density = 0.7





ใส่รูป

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Graph | Node | No. of edges within the clusters | Total degree of all nodes in the cluster | ค่า Difference Density | Un - Cluster Node |
|  |  |  |  |  |  |
|  |  | ใช้กราฟ Full Detail | |  |  |
|  |  |  |  |  |  |

Table 6

จากผลลัพท์แสดงให้เห็นว่าจะเจอขนาดใหญ่ไล่ลงมาก่อน เนื่องจากเราใช้ Greedy Strategy เพื่อเลือกสิ่งดีที่สุดก่อนเสมอ ดังนั้นครัสเตอร์ก้อนแรกที่ได้จึงเป็นก้อนที่ใหญ่ที่สุด แม้ว่าจะมีก้อนอื่นที่มีขนาดเท่ากันบ้าง แต่ส่วนใหญ่ผลลัพท์ที่ได้จะเริ่มจากขนาดใหญ่เสมอ

# Discussion and Conclusion

……… สรุปผลจากการทดลอง ความแตกต่างของแต่ละค่า DD ผลที่ได้อะไรให้ผลเป็นอย่างไร

# Future Works

มีแพลนจะพัฒนาตามนี้.....

### ลองกับ Randoms แทนที่ Greedy

### ใช้แนวทางเดิมแต่เปลี่ยนเป็น Overlapping , Hirarchical

##### Reference

