An Approach to Partitioning Graph Clustering – An Investigation

No Author Given

No Affiliation Given

No Author Given

No Affiliation Given

Abstract— the problem of clustering, Undirected and Unweight, is not allowed to effortlessly clustering because of the clustering are un-structure and un-supervise. Thus ……. การเลือกวิธีการจัดกลุ่มของข้อมูลจึงเป็นเรื่องสำคัญสำหรับเป้าหมายที่ต้องการ ในผลงานนี้เลือกใช้ highly connected graph และใช้ Clique เป็นจุดเริ่มต้นของการสร้างครัสเตอร์ และวิธีการวัดความเหมาะสมของแต่ละครัสเตอร์ที่นิยมใช้คือ การใช้ Density ซึ่งผลงานนี้ได้ทำการวัด Density ออกมาทั้ง 3 ระดับเพื่อให้เห็นความแตกต่างอย่างชัดเจน รวมถึงในขั้นตอนสุดท้ายยังมีการวัดคุณภาพของทั้งกราฟเมื่อทำการครัสเตอร์ออกมาเรียบร้อยแล้ว คือ Coverage matric and Conductance Matric.

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

# Introduction

ความเป็นมาของการจัดกลุ่ม เทคนิคทั่วไปที่ใช้กัน ลักษณะข้อมูลที่ต้องการการจัดกลุ่ม bio, network การนำเสนอของการจัดกลุ่มข้อมูล ประเภทของการจัดกลุ่ม ความท้าทายในการจัดกลุ่มกับข้อมูลที่มีปริมาณมาก ๆ การใช้ตัววัดคุณภาพ

# Dataset Description

## Graph

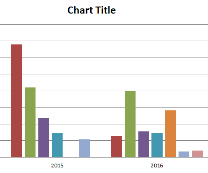
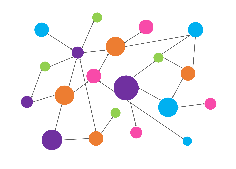
Graph ประกอบด้วย โหนด กิ่ง และแบ่งประเภทออกเป็น กราฟมีทิศทาง ไม่มีทิศทาง มีน้ำหนัก และไม่มีน้ำหนัก

(รูปรวม ทั้ง 4 แบบ เล็ก ๆ)

ซึ่งในผลงานนี้ได้เลือกใช้ ไม่มีทิศทางและน้ำหนัก เพราะว่ามันเป็นความท้าทายมากกว่าในการที่จะจัดกลุ่มข้อมูลออกมาในรูปแบบของกราฟ

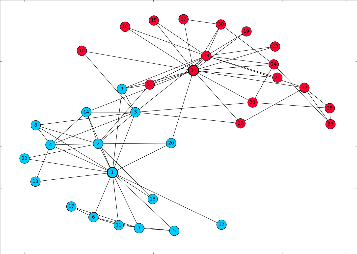
## Graph Clustering

การแสดงการกลุ่มข้อมูลแบ่งออกได้หลายลักษณะ เช่น การใช้ตาราง การใช้ข้อมูลสถิติ การใช้กราฟแท่ง else,. ในผลงานนี้เลือกใช้การแสดงผ่านการใช้กราฟ เนื่องจากในปัจจุบันนี้การแสดงข้อมูลผ่านกราฟจะได้รับความนิยมมากในด้าน bio and network. ไม่ว่าจะเป็นการใช้แสดงผลของยีนต์ ของโรคระบาด หรือของโซเชียลมีเดียต่าง ๆ ดังนั้นผลงานนี้จึงเลือกการแสดงผลแบบกราฟ



(รูปกราฟแท่งกับกราฟที่เราใช้)  
 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

## Partitioning Clustering

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

(รูปกราฟ Partition)

## Spanning Graph

อธิบายคร่าวๆ เป็น 1 ในเทคนิคการค้นหาไซเคิลภายในกราฟ

## Density of Graph

ความหนาแน่นคือ ..... ความหนาแน่นของกราฟคือ..... สูตรความหนาแน่น

# Literature Review

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

# Methodology

หลักการดำเนินงานวิจัย ดังนี้

## Flow simulation method

*Pseudo code*

## Clustering Graph

*วิธีการครัสเตอร์มีหลายรูปแบบ*

*(กราฟอันนึงที่สามารถจัดกลุ่มได้หลายวิธี)*

อธิบายว่าการจัดกลุ่มกราฟสามารถทำได้หลายแบบ แบบนี้ก็ได้ แบบนั้นก็ได้

# Experimental Results

## เตรียมข้อมูล

ใช้ highly Connected graph

สูตร ........................

## Spanning graph ยกตัวอย่าง 1 กราฟ

### Spanning tree (รูป)

### (รูปขั้นตอน)

### Fundamental cycles

### (รูปขั้นตอน)

### Remaining cycles

### (รูปขั้นตอน)

## Clique

ใช้ในรูปแบบไซเคิล 3 node และใช้ maximum degree of clique. เนื่องจากได้เลือกใช้ Greedy Strategy

## Clustering

ขั้นตอนการจัดกลุ่ม ใช้ Greedy Strategy

### Find All Clique has same 2 node

### Find inside edges

## Matric of Cluster

ใช้ Difference Density อธิบายวิธีการหา....

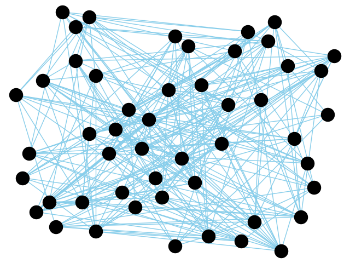
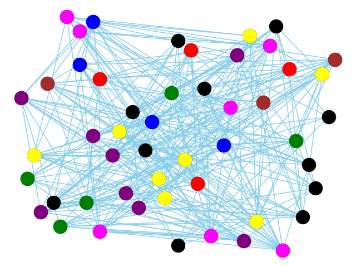
สูตร ......................................

อธิบายตัวแปร...

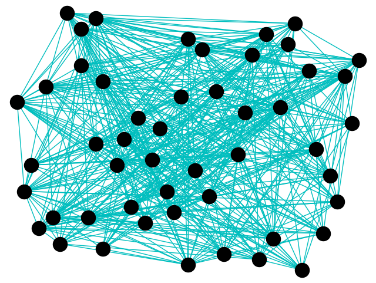
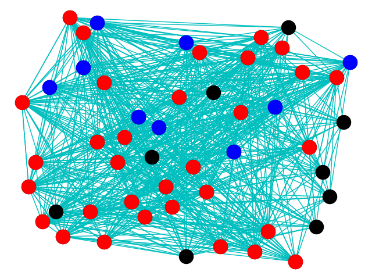
#### If a value has more than set value

#### If not

ยกตัวอย่างกราฟ



รูปกราฟ Original and before Clustering 50\_300



รูปกราฟ Original and before Clustering 50\_500

## Matric of Graph

ใช้ 2 matric ดังนี้

#### Coverage

สูตร........................

#### Conductance

สูตร........................

# Result

## Difference Density = 0.5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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

# Conclusion

……… สรุปผลจากการทดลอง ความแตกต่างของแต่ละค่า DD ผลที่ได้อะไรให้ผลเป็นอย่างไร

# Future Works

มีแพลนจะพัฒนาตามนี้.....

### ......

### .......

##### Reference