

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 unsupervise. Thus การเลือกวิธีการจัดกลุ่มข้อมูลจึงเป็นเรื่องสำคัญ สำหรับเป้าหมายที่ต้องการ ในผลงานนี้เลือกใช้ **highly connected graph** และใช้ **Clique** เป็นจุดเริ่มต้นของการสร้างคริสเตอร์ และวิธีการวัดความเหมาะสมของแต่ละคริสเตอร์ที่นิยมใช้คือ การใช้ **Density** ซึ่งผลงานนี้ได้ทำการวัด **Density** ออกมาทั้ง 3 ระดับเพื่อให้เห็นความแตกต่างอย่างชัดเจน รวมถึงในขั้นตอนสุดท้ายยังมีการวัดคุณภาพของทั้งกราฟเมื่อทำการคริสเตอร์ออกมาเรียบร้อยแล้ว คือ **Coverage matrix and Conductance Matrix**.

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

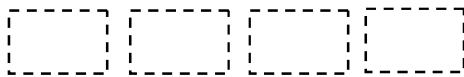
I. INTRODUCTION

ความเป็นมาของการจัดกลุ่ม เทคนิคทั่วไปที่ใช้กัน ลักษณะข้อมูลที่ต้องการ การจัดกลุ่ม bio, network การนำเสนอของการจัดกลุ่มข้อมูล ประเภทของการจัดกลุ่ม ความท้าทายในการจัดกลุ่มกับข้อมูลที่มีปริมาณมาก ๆ การใช้ตัววัดคุณภาพ

II. DATASET DESCRIPTION

A. Graph

Graph ประกอบด้วย โหนด กิ่ง และแบ่งประเภทออกเป็น กราฟมีทิศทาง ไม่มีทิศทาง มีน้ำหนัก และไม่มีน้ำหนัก

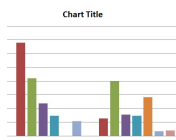


(รูปรวม ทั้ง 4 แบบ เล็ก ๆ)

ซึ่งในผลงานนี้ได้เลือกใช้ ไม่มีทิศทางและน้ำหนัก เพราะว่ามันเป็นความท้าทายมากกว่าในการที่จะจัดกลุ่มข้อมูลออกมาในรูปแบบของกราฟ

B. Graph Clustering

การแสดงการกลุ่มข้อมูลแบ่งออกได้หลายลักษณะ เช่น การใช้ตาราง การใช้ข้อมูลสถิติ การใช้กราฟแท่ง else,. ในผลงานนี้เลือกใช้การแสดงผลผ่านการใช้กราฟ เนื่องจากในปัจจุบันนี้การแสดงผลผ่านกราฟจะได้รับความนิยมมากในด้าน bio and network. ไม่ว่าจะเป็นการใช้แสดงผลของชนิด ของโรคระบาด หรือของ โฆษณาสื่อต่าง ๆ ดังนั้นผลงานนี้จึงเลือกการแสดงผลแบบกราฟ

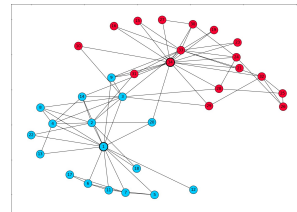


(รูปกราฟแท่งกับกราฟที่เราใช้)

การจัดกลุ่มข้อมูลในรูปแบบของกราฟโดยทั่วไปสามารถแบ่งออกได้ 3 ประเภทคือ 1. Partitioning Clustering 2. Overlapping clustering 3. Hierarchical clustering ซึ่งทั้งหมดมีการจัดรูปแบบของข้อมูลที่แตกต่างกันตามลักษณะของชื่อเรียก **Overlap** คือการจัดกลุ่มข้อมูลที่มีการ “ทับซ้อน” กันของข้อมูลของคริสเตอร์ตั้งแต่ 2 คริสเตอร์ขึ้นไป **Hierarchical** คือ การจัดกลุ่มข้อมูลที่มี “ลำดับชั้น” โดยใช้ลำดับชั้นนั้นเป็นคริสเตอร์ของข้อมูล และ **Partitioning** เป็นการจัดกลุ่มข้อมูลทีละคริสเตอร์มีการแบ่งแยกออกจากกัน และมีความแตกต่างกันอย่างชัดเจนทั้งภายนอกและภายในของคริสเตอร์ ซึ่งในผลงานชิ้นนี้ได้เลือกใช้ในรูปแบบ **Partitioning**

C. Partitioning Clustering

การจัดกลุ่มข้อมูลในประเภท **Partitioning** คือ การจัดกลุ่มข้อมูลทีละคริสเตอร์มีการแบ่งแยกออกจากกันชัดเจน ไม่มีการ “ทับซ้อน” หรือ “ลำดับชั้น” ใน 1 ข้อมูลจะอยู่กลุ่มใดกลุ่มหนึ่ง ข้อมูลภายในคริสเตอร์และภายนอกคริสเตอร์จะมีความแตกต่างกันอย่างชัดเจน ยกตัวอย่าง บลา ๆ ๆ



(รูปกราฟ Partition)

D. Spanning Graph

อธิบายคร่าวๆ เป็น 1 ในเทคนิคการค้นหาไหลในกราฟ

E. Density of Graph

ความหนาแน่นคือ ความหนาแน่นของกราฟคือ..... สูตรความหนาแน่น

III. LITERATURE REVIEW

การจัดกลุ่มที่เกิดขึ้นมาหลายยุคหลายสมัย และปริมาณของข้อมูลที่เพิ่มขึ้นในระดับเพตะไบต์ (Petabyte) การอัปเดตการจัดกลุ่มจึงมีความสำคัญมากเพื่อรองรับปริมาณข้อมูลที่มีการเปลี่ยนแปลงในระดับมหาศาล []. อ้างผลงานสัก 4 ฉบับ 1. การใช้กราฟจัดกลุ่มข้อมูล 2. การใช้กราฟแบบไม่มีทิศทาง ไม่มีน้ำหนัก มาจัดกลุ่ม 3. การใช้ความหนาแน่นมาจัดกลุ่ม 4. การใช้ **matric** มาตรวจสอบคุณภาพของทั้งกราฟ + Spanning graph

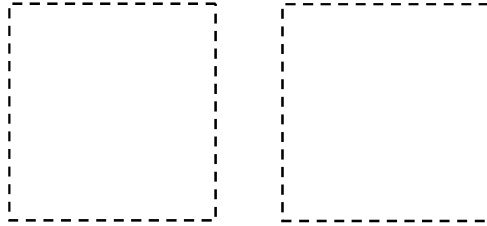
IV. METHODOLOGY

หลักการดำเนินงานวิจัย ดังนี้

A. Flow simulation method
Pseudo code

B. Clustering Graph

วิธีการครัสเตอร์มีหลายรูปแบบ



(กราฟอันนี้ที่สามารถจัดกลุ่มได้หลายวิธี)

อธิบายว่าการจัดกลุ่มกราฟสามารถทำได้หลายแบบ แบบนี้ก็ได้ แบบนั้นก็ได้

V. EXPERIMENTAL RESULTS

A. เตรียมข้อมูล

ใช้ highly Connected graph

สูตร

B. Spanning graph ยกตัวอย่าง 1 กราฟ

1) Spanning tree (รูป)

(รูปขั้นตอน)

2) Fundamental cycles

(รูปขั้นตอน)

3) Remaining cycles

(รูปขั้นตอน)

C. Clique

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

D. Clustering

ขั้นตอนการจัดกลุ่ม ใช้ Greedy Strategy

1) Find All Clique has same 2 node

2) Find inside edges

E. Matric of Cluster

ใช้ Difference Density อธิบายวิธีการหา...

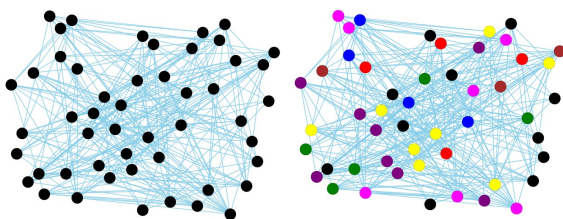
สูตร

อธิบายตัวแปร...

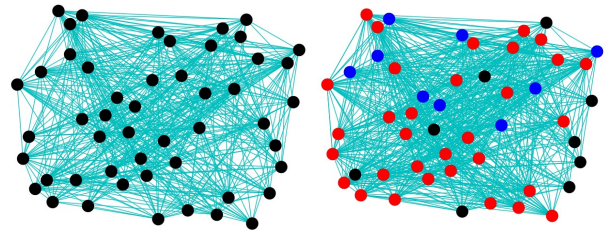
a) If a value has more than set value

b) If not

ยกตัวอย่างกราฟ



รูปกราฟ Original and before Clustering 50_300



รูปกราฟ Original and before Clustering 50_500

F. Matric of Graph

ใช้ 2 matric ดังนี้

a) Coverage

สูตร

b) Conductance

สูตร

VI. RESULT

A. 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

B. 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

C. 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

VII. CONCLUSION

..... สรุปผลจากการทดลอง ความแตกต่างของแต่ละค่า DD ผลที่ได้ อะไรให้ผลเป็นอย่างไร

VIII. FUTURE WORKS

มีแผนจะพัฒนาด้านนี้.....

1)

2)

REFERENCE