# Abstract

**According to …..**

# Introduction

(ข้อมูลมีจำนวนมากขึ้นเรื่อย ๆ และเป็น highly connected graph […], การจัดการข้อมูลทำได้หลายวิธี […], ส่วนใหญ่เป็น NP hard problem […], NP-HP is… […], วิธีการจัดการ NP-HP วิธีทั่วไปคือ, Datamining […], Clustering, Graph…. […])

# Preliminary Discussion

## Graph Clustering (Graph is…[…], GC is…[…], Kind of GC.[…], Partitiong, Overlapping, Hirarchicale)

## Partitioning clustering (Graph Partitioning is…., Many method of Clustering ex.1,2,3…)

# Density of Graph

(Measure of Cluster, Density of Graph, Inter edges, Intra edges, Inter Cluster Density (formulas), Intra Cluster Density (formulas), Different Density)

# Determinationtion of Different Sizes of Clusters.

## Determination of Minimum Sib-cycles (Start of …, Highly connected, Minimum Sub-cycles, Sub-graph is spanning tree., Greedy Strategy, Clique, Max degree, Sum of degree of clique )

ในปัจจุบันกราฟที่มีข้อมูลจำนวนมากจนเรียกได้ว่าเป็น Highly Connected Graph […] ในผลงานวิจัยนี้ได้มีการให้คำนิยามของ highly connected Graph is …… […] , เพื่อนำมาใช้เป็นข้อมูลในการจัดกลุ่ม จริง ๆ แล้ว มีคำนิยามมากมายที่กำหนดการเป็น HCG. แต่ในผลงานนี้ทางผู้วิจัยได้ใช้เพียงนิยามเดียว เพื่อไม่ให้จำกัดการใช้ข้อมูลที่หลากหลายมากเกินไป เช่น คำนิยาม [Finding Highly Connected Sub-Graph] ที่กำหนดให้ HCG. Are very similar to 0.5-quasi-complete graphs. [17] เป็นต้น ซึ่งจะทำให้กราฟข้อมูลที่เข้านิยามมีจำนวนน้อยลงมากกว่าการที่เราจะเลือกใช้นิยามที่เหมาะสมที่สุดเพียงไม่กี่นิยาม

การเริ่มต้นการจัดกลุ่มกราฟคือ การเริ่มต้นการสร้างครัสเตอร์ในแต่ละครัสเตอร์ ในผลงานวิจัยนี้ทางผู้วิจัยได้พบว่า ไม่จำเป็นต้องเริ่มต้นจากเพียงโหนด 1 โหนด แต่สามารถเริ่มต้นจาก Minimum Sub-cycles คือ การค้นหา Sub-graph หรือ Clique (จากนนี้จะใช้ Clique ในการเรียก Minimum Sub-cycles) ทั้งหมดภายใน Original graph. ซึ่งใช้วิธีการดังต่อไปนี้

### Spanning Tree.

เริ่มต้นการค้นหา Clique ทั้งหมดในกราฟ เริ่มจากวิธี Spanning tree. คือ .......

### Fundamantal Cycles.

### Remaining Cycles.

ในผลงานวิจัยนี้ ทางผู้วิจัยได้ค้นพบว่า การเริ่มต้นจุดที่แตกต่างกันนั้น ส่งผลต่อการสร้างครัสเตอร์ในแต่ละก้อน เช่น เทคนิค K-means ที่ใช้การตัดสินใจของผู้ใช้ในการกำหนดจุด centroid ในแต่ละครั้ง ดังนั้นทางผู้วิจัยจึงได้เลือกวิธีการ Greedy Strategy เพื่อค้นหาสิ่งที่ดีที่สุดในกราฟมาใช้ในการเริ่มต้น ซึ่งในที่นี้ทางผู้วิจัยได้ให้ความหมายของคำว่า ‘’The best way’ คือ Sum of Maximum degree of Clique. คือ การหาผลรวมดีกรีของแต่ละโหนดใน Clique ทุก ๆ Clique ในกราฟ มาเปรียบเทียบหาค่าผลรวมสูงสุดแล้วเริ่มต้นการสร้างครัสเตอร์

## Determination of Cluster

เมื่อได้จุดเริ่มต้นของการสร้างครัสเตอร์เรียบร้อยแล้ว การเริ่มต้นการสร้างครัสเตอร์ ในผลงานนี้จะเริ่มต้นจาก นำ Clique ที่มีผลรวมดีกรีสูงสุดเป็นจุดเริ่มต้นของครัสเตอร์ก้อนแรก และทำการค้นหา Clique ที่เหลือ ที่มีโหนดเหมือนกับ Clique เริ่มต้นจำนวน 2 โหนด และมีโหนดที่เหลือ 1 โหนด มีผลรวมดีกรีสูงที่สุด (รูปอธิบาย) นำมาต่อเพิ่มกับ Clique เริ่มต้น กลายเป็น Sub-Cluster ก้อนแรก จากนั้นนำ Sub-Cluster มาหากิ่งภายใน (รูปอธิบาย) เพื่อไม่ให้พลาดกิ่งที่มีอยู่จริงใน Original Graph. แล้วนำโหนดและกิ่งทั้งหมดมาคำนวนหาค่า Inter Cluster Density , Intra Cluster Density , และ Different Density (User -determined)

## The Algorithm for Determining Different Sizes of Partitioning Clusters (User-determined, Bigger down to smaller)

อัลกอริทึมการจัดกลุ่มข้อมูลในผลงานวิจัยนี้ ยังจำเป็นต้องให้ User เป็นคนกำหนดค่า DD ในการจัดกลุ่มครัสเตอร์ เนื่องจาก...

# The Graph Metrics

การวัดคุณภาพของการจัดกลุ่มข้อมูล นอกจากจะมีการวัดในระดับครัสเตอร์ที่ละครัสเตอร์แล้ว ยังมีการวัดคุณภาพของทั้งกราฟหลังจากการจัดกลุ่มของข้อมูล สามารถบอกได้ว่าหลังจากที่ใช้เทคนิคต่าง ๆ ในการจัดกลุ่มแล้ว ประสิทธิภาพทั้งหมดนั้นออกมาเหมาะสมมากเพียงใด ซึ่งตัววัดที่จะกล่าวถึงนี้คือ Coverage และ Conductance Metrics ซึ่งอ้างอิงมาจากเปเปอร์ [..] ที่มีการใช้ตัววัดทั้ง 3 ตัว ที่เพิ่มมาคือตัววัด Modularity ทางผู้วิจัยได้ดำเนินการทดลองตัววัดนี้ในผลงานวิจัยแล้ว แต่พบว่าไม่ได้สามารถจำกัดพิกัดได้ จึงนำมาใช้แค่ 2 ตัววัด ดังนี้

## Coverage Metric (formulas, Detail of formules)

## Conductance Metric (formulas, Detail of formules)

การจัดกลุ่ม

# Discussion and Conclusion

### หากจุดเริ่มต้นแต่ละ clique มีผลรวมเท่ากัน จะเลือกอันไหนเป็นจุดเริ่มต้น