

Final Exam (2021/2) Part3: Data Science ส่วน Visualization (คะแนนเต็ม 50 คะแนน คิดเป็น 10%)

All files must be renamed as “{student_id}_{firstname}_Part3”, e.g.,

- 6030133421_Chaiyatad_Part3.docx (สำหรับใส่รูป และคำตอบ)
 - o สำหรับรูปภาพที่ Capture มาได้ หากเห็นตำแหน่งเป้าหมายที่โจทย์ระบุไว้ไม่ชัด ก็จะทำให้ไม่สามารถให้คะแนนได้ ส่งให้ผลคะแนนในข้อนั้นเป็น 0
- 6030133421_Chaiyatad_Part3.gephi (for Module1)
- 6030133421_Chaiyatad_Part3.pbix (for Module2)
- สุดท้าย zip เป็นชื่อไฟล์ 6030133421_Chaiyatad_Part3.zip

Module 1. Network Analysis and Visualization with Gephi

1. ให้นักศึกษาใช้โปรแกรม Gephi เปิดไฟล์ hero-social-network.gephi ซึ่งเป็นข้อมูลเครือข่ายความสัมพันธ์ของตัวละครในนิยายภาพแนวซูเปอร์ฮีโร่ โดยแต่ละ node เป็นตัวละครแต่ละตัว และ edge แสดงความสัมพันธ์ระหว่างตัวละคร ให้นักศึกษาใช้เครื่องมือวิเคราะห์ในโปรแกรม Gephi และสร้าง network visualization เพื่อตอบคำถามข้างล่างนี้ โดยให้ capture หน้าจอที่แสดงผลลัพธ์ที่เกี่ยวข้องมาแปะที่ภาพก็ได้ พร้อมทั้งเขียนคำอธิบายประกอบภาพ

- 1.1. (10 คะแนน) เน็ตเวิร์คนี้มีคุณลักษณะใกล้เคียงกับเน็ตเวิร์คโมเดลประเภทใด (random, small-world, scale-free) พิจารณาจากอะไร

- 1.2. (10 คะแนน) ถ้าต้องการหาว่าใครคือตัวละครที่มีความสำคัญที่สุด 3 อันดับแรก โดยพิจารณาจาก centrality measure ที่ให้ความสำคัญกับโหนดที่เป็นตัวเชื่อมโยงระหว่างกลุ่มต่างๆ ในเครือข่าย ควรใช้ค่า centrality ตัวไหน และตัวละคร 3 อันดับแรกคือใครบ้าง

ให้แสดงภาพ 1) กราฟ distribution ของ centrality ที่เลือกใช้

2) ตารางในแท็บ Data Laboratory ที่เรียงลำดับตามค่า centrality ตัวที่เลือกใช้จากมากไปน้อย

1.3. (10 คะแนน) ให้รัน Modularity detection แล้วให้สร้าง network visualization ที่

- 1) ปรับสีของ node ตาม modularity class
- 2) แสดง label โดยให้มีขนาดตัวอักษรขึ้นอยู่กับค่า centrality ที่เลือกใช้ในข้อที่ 1.2 เพื่อให้เห็นชื่อตัวละครสำคัญได้เด่นชัดกว่าตัวละครอื่น
- 3) เลือกใช้ layout ที่แสดงให้เห็นการจับกลุ่มของโหนดที่อยู่ใน community เดียวกันได้ชัดเจน โดยอาจปรับค่าพารามิเตอร์ของ layout และตามด้วยการใช้ layout แบบ Noverlap หรือ Label Adjust, Expansion, Contraction ได้ตามความเหมาะสม

Module 2. Spatial Analysis and Visualization with Power BI

2. ให้นักศึกษาใช้โปรแกรม PowerBI เปิดไฟล์ taxi_od.pbix (สำหรับ Power BI Online ให้ Get Data → Files, Get → Local File → select file → Open → click on file “taxi_od” with type “Report” → Edit)
ไฟล์นี้ได้มีการนำข้อมูลไว้แล้วในตารางชื่อ taxi_od เป็นข้อมูลการเดินทางด้วยรถแท็กซี่จำนวนหนึ่งภายในหนึ่งวัน ข้อมูลแต่ละแถวคือบันทึกเหตุการณ์การรับหรือส่งผู้โดยสาร ประกอบด้วยคอลัมน์ดังต่อไปนี้

ชื่อคอลัมน์	คำอธิบาย
date	วันที่ (ค่าคงที่ เป็นข้อมูลวันเดียว)
distance	ระยะทางของทริป หน่วยเป็นเมตร มีค่าเฉพาะเมื่อ status เป็น D
duration	เวลาที่ใช้ในทริป หน่วยเป็นวินาที มีค่าเฉพาะเมื่อ status เป็น D
hour	timestamp เฉพาะส่วนชั่วโมง
latitude	พิกัดตำแหน่งละติจูดแบบทศนิยม
longitude	พิกัดตำแหน่งลองจิจูดแบบทศนิยม
rounded_latitude	พิกัดตำแหน่งละติจูดแบบทศนิยม ที่ลดความละเอียดเหลือเพียงทศนิยมสองตำแหน่ง เพื่อให้จุดที่อยู่ในพื้นที่เดียวกันมีพิกัดเท่ากัน เหมือนแบ่งพื้นที่เป็น grid
rounded_longitude	พิกัดตำแหน่งลองจิจูดแบบทศนิยม ที่ลดความละเอียดเหลือเพียงทศนิยมสองตำแหน่ง เพื่อให้จุดที่อยู่ในพื้นที่เดียวกันมีพิกัดเท่ากัน เหมือนแบ่งพื้นที่เป็น grid
status	O คือ origin หรือ pick-up แสดงว่าเป็นการรับผู้โดยสาร D คือ destination หรือ drop-off แสดงเป็นการส่งผู้โดยสาร
taxi_ID	หมายเลขของแท็กซี่
time	timestamp อยู่ในรูปแบบ yyyy-mm-dd hh:mm:ss เช่น 2016-09-23 06:03:46
trip_ID	หมายเลข trip แท็กซี่หนึ่งคันมีได้หลาย trip

taxi_ID	trip_ID	date	time	hour	status	latitude	longitude	rounded_latitude	rounded_longitude	distance	duration
AK3535	AK3535_T_7474	Friday, September 23, 2016	6:29:00 AM	6	O	13.6879034	100.5571594	13.69		100.56	
AK3535	AK3535_T_7474	Friday, September 23, 2016	6:42:00 AM	6	D	13.68603134	100.5688477	13.69		100.57	1337
AK3535	AK3535_T_7476	Friday, September 23, 2016	7:03:00 AM	7	O	13.67806625	100.5021973	13.68		100.5	
AK3535	AK3535_T_7476	Friday, September 23, 2016	7:06:00 AM	7	D	13.674613	100.5041962	13.67		100.5	535
AK3535	AK3535_T_7480	Friday, September 23, 2016	7:54:00 AM	7	O	13.73393345	100.5027695	13.73		100.5	
AK3535	AK3535_T_7480	Friday, September 23, 2016	7:56:00 AM	7	D	13.73385334	100.5028534	13.73		100.5	402

ตัวอย่างข้อมูล

นาย ก ต้องการหารายได้ด้วยการขับรถแท็กซี่ โดยใช้เวลารวันละ 12 ชั่วโมงติดต่อกัน จึงต้องการใช้ข้อมูลช่วยในการตัดสินใจ โดยจะเลือกช่วงเวลาที่มีคนเดินทางด้วยรถแท็กซี่มากที่สุด และจะไปหาผู้โดยสารในบริเวณที่จะมีโอกาสได้ผู้โดยสารมากที่สุด

ให้นิสิตสร้าง **visualization** ใน **Power BI** จากข้อมูล เพื่อตอบคำถามต่อไปนี้ นิสิตสามารถเลือกใช้ **visualization** ที่คิดว่าช่วยตอบคำถามและสื่อสารได้อย่างมีประสิทธิภาพที่สุด ที่ **visualization** ก็ได้ โดยให้ **capture** หน้าจอมาแปะ โดยให้มีทั้งภาพ **visualization** พร้อมทั้งส่วน **Filters, Visualizations, Fields** ที่ใช้ แล้วเขียนคำอธิบายแนวคิดและผลลัพธ์ประกอบภาพ

2.1 (5 คะแนน) นาย ก ควรจะเลือกขับรถในช่วงเวลาใด

2.2 (15 คะแนน) ในช่วงเวลาตามข้อที่แล้ว นาย ก ควรไปหาผู้โดยสารในพื้นที่บริเวณใดบ้างที่มีความต้องการหนาแน่น ให้แสดงเป็นแผนที่ โดยให้ **capture** หน้าจอมาแปะ โดยให้มีทั้งภาพ **visualization** พร้อมทั้งส่วน **Filters, Visualizations, Fields** ที่ใช้ ควรเลือกข้อมูลที่อยู่ในช่วงเวลาดังกล่าว และแสดงหรือเน้นให้เห็นพื้นที่ที่แนะนำอย่างชัดเจน