

บทที่ 3 ปัญหาและวิธีการแก้ปัญหา

แม้ว่าขั้นตอนวิธี JXP จะทำงานได้อย่างมีประสิทธิภาพ แต่การเปลี่ยนแปลงโครงสร้างของเว็บลิงค์กราฟ ทำให้ขั้นตอนวิธี JXP ต้องเริ่มต้นคำนวณค่าเพจแรงค์ของทุกๆเว็บเพจในเว็บลิงค์กราฟใหม่ทุกครั้ง ซึ่งในบางกรณีไม่จำเป็นต้องคำนวณค่าเพจแรงค์ของเว็บเพจใหม่ทั้งหมด ดังแสดงในตารางที่ 3.1

L-WLG	P1	P2	P3	P4	P5	P6	P7	P8	P9
เริ่มต้น	0.037	0.037	0.399	0.037	0.399	0.037	0.053	-	-
เวลา T1	0.033	0.033	0.308	0.033	0.238	0.033	0.174	0.150	-
เวลา T2	0.029	0.029	0.246	0.029	0.189	0.029	0.131	0.211	0.106
เวลา T3	0.029	0.029	0.200	0.029	0.154	0.029	0.141	0.234	0.156

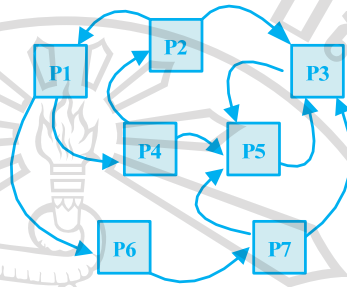
ตารางที่ 3.1 ค่าเพจแรงค์ของเว็บลิงค์กราฟเมื่อมีการเปลี่ยนแปลงของโครงสร้างของเว็บลิงค์กราฟ

จากตารางที่ 3.1 เป็นค่าเพจแรงค์ที่ได้จากเว็บลิงค์กราฟในรูปที่ 3.1 ซึ่งเริ่มต้นด้วยเว็บลิงค์กราฟที่มีเว็บเพจเพียง 7 โหนด ดังแสดงในรูปที่ 3.1a และเมื่อเวลาผ่านไปเป็น T1 มีเว็บเพจ P8 เพิ่มขึ้นในเว็บลิงค์กราฟ ดังแสดงในรูปที่ 3.1b จากนั้นเมื่อเวลาเปลี่ยนเป็น T2 มีเว็บเพจ P9 เพิ่มขึ้นในเว็บลิงค์กราฟ ดังแสดงในรูปที่ 3.1c ส่วนเมื่อเวลาเปลี่ยนเป็น T3 จำนวนเว็บเพจในเว็บลิงค์กราฟไม่เปลี่ยนแปลง เพียงแต่มีการเพิ่มขึ้นของเว็บลิงค์ ดังแสดงในรูป 3.1d

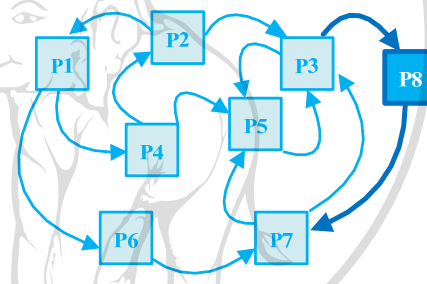
จากค่าของเพจแรงค์ของแต่ละเว็บเพจในตารางที่ 3.1 จะเห็นว่าแม้โครงสร้างของเว็บลิงค์กราฟจะเปลี่ยนแปลงไป แต่ค่าเพจแรงค์ของบางเว็บเพจในเว็บลิงค์กราฟเปลี่ยนแปลงไปเพียงเล็กน้อยเท่านั้น ดังตัวอย่างค่าเพจแรงค์จากเว็บเพจ P1, P2, P4 และ P6 ในตารางที่ 3.1 เมื่อเปรียบเทียบการเปลี่ยนแปลงของค่าเพจแรงค์ของเว็บเพจดังกล่าว กับการเปลี่ยนแปลงของค่าเพจแรงค์ของเว็บเพจอื่นๆ ที่เหลือในเว็บลิงค์กราฟแล้ว ถือว่าค่าเพจแรงค์เปลี่ยนแปลงไปเพียงเล็กน้อย ซึ่งไม่มีผลกระทบต่อการจัดลำดับของเว็บเพจมากนัก

ดังนั้นบางโครงสร้างของเว็บลิงค์กราฟ เมื่อมีการเปลี่ยนแปลงโครงสร้างใหม่ ไม่ว่าจะเป็นการเพิ่มขึ้นของเว็บเพจหรือการเพิ่มขึ้นของเว็บลิงค์ ไม่จำเป็นต้องเริ่มต้นคำนวณค่าเพจแรงค์ของทุกเว็บเพจใหม่คำนวณแค่บางส่วนก็ได้ ซึ่งการคำนวณเฉพาะบางส่วนนี้จะทำให้โครงสร้างเว็บลิงค์กราฟที่นำมาคำนวณค่าเพจแรงค์มีขนาดเล็กลงจึงใช้เวลาในการปรับค่าเพจแรงค์น้อยลงไปด้วย

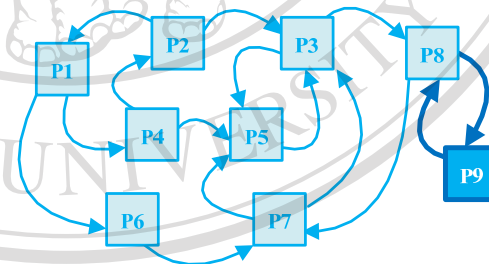
a. โดคอลเว็บลิงค์กราฟเริ่มต้น



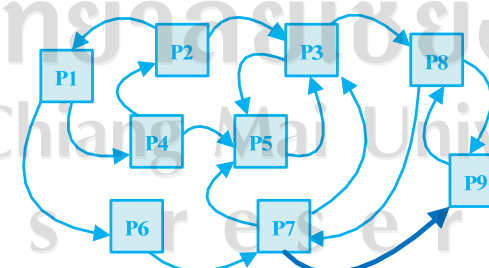
b. โดคอลเว็บลิงค์กราฟเมื่อเวลา T1



c. โดคอลเว็บลิงค์กราฟเมื่อเวลา T2



d. โดคอลเว็บลิงค์กราฟเมื่อเวลา T3



รูปที่ 3.1 แสดงการเปลี่ยนแปลงของโครงสร้างโดคอลเว็บลิงค์กราฟ

สำหรับในงานวิจัยฉบับนี้ ผู้วิจัยได้ศึกษาเพื่อแก้ปัญหาของการเริ่มต้นคำนวณค่าเพจแรงค์ใหม่หมดในทุกเว็บเพจของเว็บลิงค์กราฟเมื่อโครงสร้างของเว็บลิงค์กราฟเปลี่ยนแปลง เพื่อปรับปรุงประสิทธิภาพในเรื่องเวลาการคำนวณค่าเพจแรงค์ของขั้นตอนวิธี JXP โดยใช้หลักการของการจัดลำดับเว็บเพจแบบเพิ่มขึ้น (Incremental Web Ranking) เพื่อพิจารณาค่าเพจแรงค์เฉพาะบางส่วนของเว็บเพจในเว็บลิงค์กราฟ แทนการจัดลำดับเว็บเพจแบบธรรมดาในขั้นตอนวิธี JXP ดังแสดงส่วนของขั้นตอนวิธี JXP ที่ถูกปรับปรุงในรูปแบบที่ 3.2

Algorithm1 Initial PageWeight Computation

```

1: input: local graph G and est. size of global graph N
2: n <-- size(G)
3: Create world node W
4: PageWeight(plp  $\in$  G) <--  $\frac{1}{N}$ 
5: PageWeight(W) <--  $\frac{N-n}{N}$ 
6: add W to G
7: run PageRank power iteration on G
8: Update(PageWeights)

```

Algorithm2 the JXP Algorithm

```

1: input: local graph  $G_A$ , world node  $W_A$ , score list  $L_A$ 
2: repeat
3: Contact a random peer B in the network and exchange information
4:  $G_M$  <-- mergeGraphs( $G_A, G_B$ )
5:  $W_M$  <-- mergeWorldNodes( $W_A, W_B$ )
6:  $G'_M$  <-- ( $G_M + W_M$ )
7:  $L_M$  <-- combineLists( $L_A, L_B$ )
8:  $PR$  <-- PageRank( $G'_M$ )
9:  $L'_M$  <-- updateScoresList( $L_M, PR$ )
10: update( $L_A$ )
11: update( $W_A$ )
12: Discard( $G_M, W_M, L_M, G_B, W_B, L_B$ )

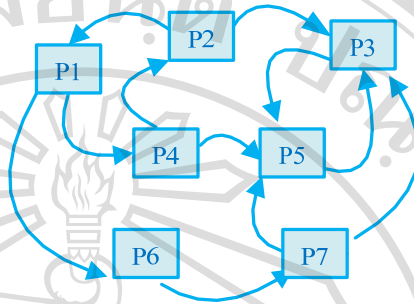
```

รูปที่ 3.2 ส่วนของขั้นตอนวิธี JXP ที่ถูกปรับปรุง

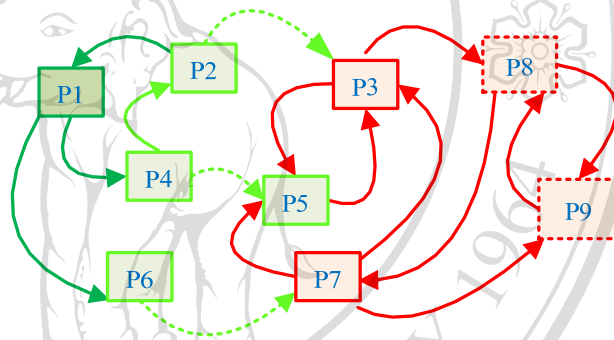
โดยผู้วิจัยจึงได้แบ่งส่วน (Partition) ของกราฟออกเป็น 3 ส่วนคือ ส่วนแรกเป็นส่วนของเว็บเพจที่ถูกผลกระทบหลังจากมีเว็บเพจเพิ่มเข้ามาในเว็บลิงค์กราฟ ส่วนที่สองคือส่วนของเว็บเพจ

ที่ไม่ถูกผลกระทบหลังจากมีเว็บเพจเพิ่มเข้ามาในเว็บลิงค์กราฟ และส่วนสุดท้าย คือส่วนของเว็บเพจในส่วนแรกที่มีเพียงลิงค์ออกไปยังเว็บเพจในส่วนที่สอง ดังตัวอย่างในรูปที่ 3.3 แสดงเว็บลิงค์กราฟที่ถูกแบ่งส่วน และจำแนกจำนวนค่าเพจแรงค์เฉพาะส่วนแรกเท่านั้น

a. โดคอลเว็บลิงค์กราฟเริ่มต้น



b. โดคอลเว็บลิงค์กราฟเมื่อเวลา T1



รูปที่ 3.3 เว็บลิงค์กราฟที่ถูกแบ่งส่วน

ส่วนแรกของเว็บลิงค์กราฟ เป็นส่วนของกลุ่มของเว็บเพจที่มีการเปลี่ยนแปลงค่าเพจแรงค์เป็นอย่างมาก เมื่อมีเว็บเพจใหม่เพิ่มเข้ามาในเว็บลิงค์กราฟ กล่าวคือกลุ่มของเว็บเพจทั้งหมดที่เป็นเดสเซนเด้นต์ (Descendent) ของเว็บเพจที่เพิ่มเข้ามาใหม่ ดังแสดงตัวอย่างใน รูปที่ 3.3b จะเห็นได้ว่าจากเว็บลิงค์กราฟเริ่มต้นในรูปที่ 3.3a เมื่อมีเว็บเพจ P8 และ P9 เพิ่มเข้ามาใหม่จะมีโครงสร้างของเว็บลิงค์กราฟเปลี่ยนแปลงไปเป็นเว็บลิงค์กราฟในรูป 3.3b ซึ่งแสดงให้เห็นว่าทั้งเว็บเพจ P8 และ P9 ส่งผลกระทบบต่อค่าเพจแรงค์ของเว็บเพจ P3, P5 และ P7

ส่วนที่สอง เป็นส่วนของกลุ่มของเว็บเพจที่มีการเปลี่ยนแปลงของค่าเพจแรงค์เพียงเล็กน้อย เมื่อมีเว็บเพจใหม่เพิ่มเข้ามาในเว็บลิงค์กราฟ ซึ่งส่วนดังกล่าวเป็นกลุ่มของเว็บเพจที่ไม่ใช่กลุ่มของเว็บเพจในส่วนแรกทั้งหมด ดังรูปที่ 3.3b เว็บเพจ P1, P2, P4 และ P6 เป็นกลุ่มของเว็บเพจในส่วนที่สอง

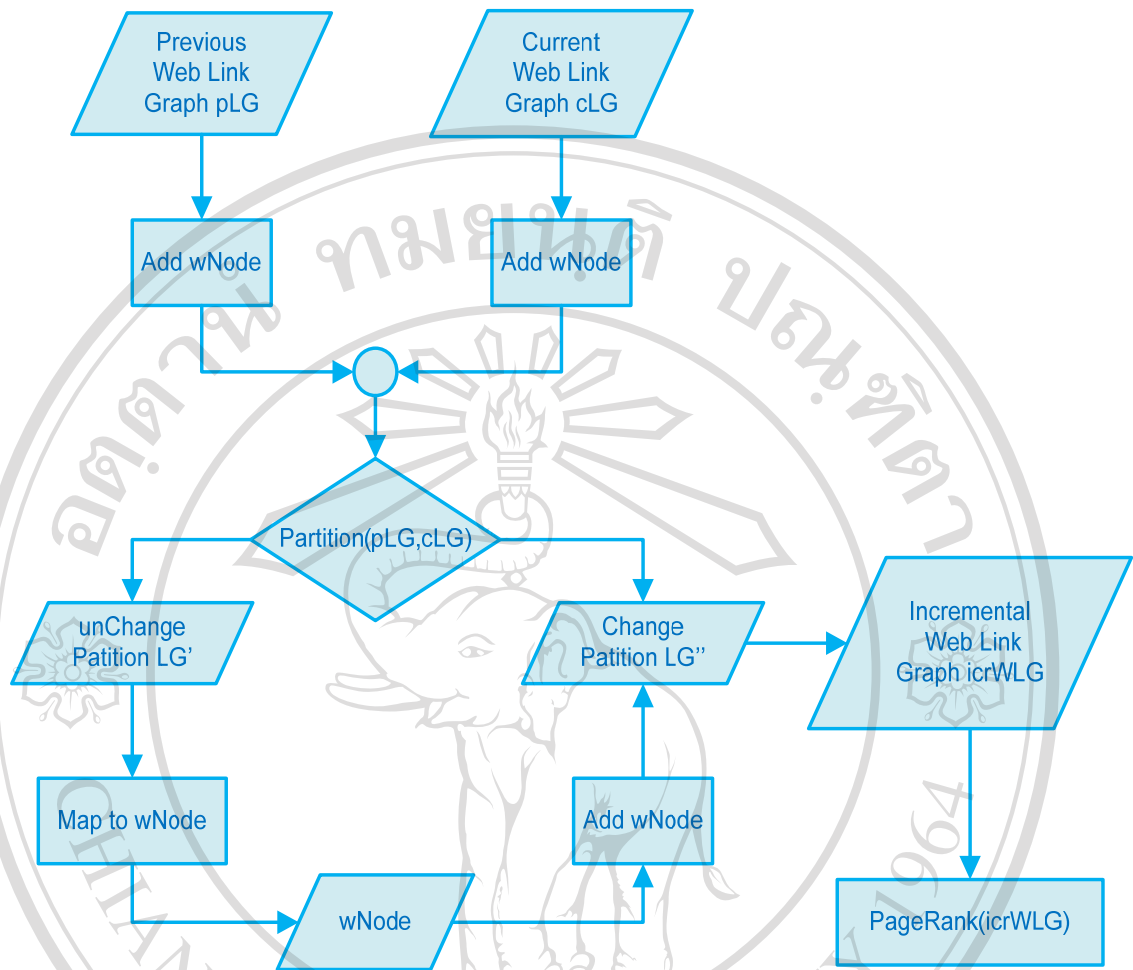
ส่วนที่สาม เป็นกลุ่มของเว็บเพจในส่วนที่สองที่มีลิงค์ออกไปยังกลุ่มของเว็บเพจในส่วนแรก ดังรูปที่ 3.3b เว็บเพจ P2, P4 และ P6 เป็นเว็บเพจในส่วนที่สาม เรียกกลุ่มของเว็บเพจในส่วนดังกล่าวว่าโหนดขอบ (Border Nodes) แม้ว่าโหนดขอบจะมีการเปลี่ยนแปลงของค่าเพจแรงค์เพียงเล็กน้อยเมื่อมีเว็บเพจใหม่เพิ่มเข้ามาในเว็บลิงค์กราฟ แต่ถือว่ามีความสำคัญต่อการจัดลำดับเว็บเพจแบบเพิ่มขึ้นเป็นอย่างมาก เนื่องจากโหนดขอบจะให้ค่าเพจแรงค์จากกลุ่มของเว็บเพจในส่วนที่สองไปยังกลุ่มของเว็บเพจในส่วนแรก

ซึ่งหลักการของการจัดลำดับแบบเพิ่มขึ้นจะคำนวณค่าเพจแรงค์เฉพาะในกลุ่มของเว็บเพจในส่วนแรกของเว็บลิงค์กราฟเท่านั้น เนื่องจากเป็นกลุ่มของเว็บเพจที่ได้รับผลกระทบจากการเพิ่มเข้ามาของเว็บเพจใหม่ ในเว็บลิงค์กราฟ ส่งผลให้ขนาดของเว็บลิงค์กราฟที่นำมาคำนวณค่าเพจแรงค์มีขนาดเล็กลง

อย่างไรก็ตามด้วยหลักการของขั้นตอนวิธี IPR สามารถคำนวณค่าเพจแรงค์ในส่วนแรกได้ทันที เนื่องจากเว็บเพจทั้งหมดถูกเก็บรวมกันไว้ในที่เก็บเดียวกัน

แต่สำหรับขั้นตอนวิธี JXP เว็บลิงค์กราฟถูกแยกเก็บไว้ในแต่ละเพียร์ และมีการนำเวกเตอร์โหนดเข้ามาพิจารณาด้วย ผู้วิจัยจึงได้นำแนวคิดของขั้นตอนวิธี IPR มาประยุกต์ใช้ โดยจะนำเว็บเพจในส่วนที่สองของแต่ละโดคอลเว็บลิงค์กราฟไปรวมกับเวกเตอร์โหนด ดังแสดงขั้นตอนวิธีอย่างง่ายในการคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ โดยใช้เป็นแผนภาพประกอบการบรรยายในรูปที่ 3.4

จากรูปที่ 3.4 เมื่อมีการแบ่งส่วนของเว็บลิงค์กราฟ cLG เป็น 2 ส่วนคือ LG' ซึ่งเป็น ส่วนของเว็บเพจที่ไม่ถูกผลกระทบหลังจากมีเว็บเพจเพิ่มเข้ามาใหม่ในเว็บลิงค์กราฟ และ LG'' ซึ่งเป็น ส่วนของเว็บเพจที่ถูกผลกระทบหลังจากมีเว็บเพจเพิ่มเข้ามาในเว็บลิงค์กราฟ จากนั้นจะรวมส่วน LG' เข้ากับเวกเตอร์โหนด และเชื่อมต่อเวกเตอร์โหนดเข้ากับส่วน LG'' ได้เป็นเว็บลิงค์กราฟใหม่และคำนวณค่าเพจแรงค์ในเว็บลิงค์กราฟ icrWLG



รูปที่ 3.4 แผนภาพการทำงานของขั้นตอนการคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่

สำหรับการคำนวณค่าเพจแรงค์ในขั้นตอนวิธี JXP จะประกอบไปด้วยการคำนวณ 2 ส่วนด้วยกันคือ การคำนวณค่าเพจแรงค์ในโกลบอลเว็บลิงค์กราฟที่ถูกเก็บไว้ในแต่ละเพียร์ และการคำนวณค่าเพจแรงค์ในขณะรวมสองโกลบอลเว็บลิงค์กราฟ ให้เป็นกราฟใหม่ในกระบวนการจับคู่ระหว่างเพียร์ ดังนั้นจึงมีการใช้ขั้นตอนวิธีการคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ 2 ครั้งด้วยกัน ในการทำงานของขั้นตอนวิธี JXP คือการคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ในโกลบอลเว็บลิงค์กราฟ และการคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ในกระบวนการจับคู่ระหว่างเพียร์

การคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ในแต่ละโกลบอลเว็บลิงค์กราฟ ทำงานในกระบวนการแรกของขั้นตอนวิธี JXP ซึ่งเป็นการคำนวณค่าเพจแรงค์เฉพาะในเพียร์ ดังรายละเอียดการทำงานในขั้นตอนวิธี ในรูปที่ 3.5

Algorithm ICR_LocalPR

```

1  Let  $p$  be a peer in the P2P network.
2  Let  $l\_wlg$  be a local web link graph of size  $n$  nodes.
3  Let  $g\_wlg$  be the global web link graph of size  $N$  nodes.
4  Let  $w$  be the world node in the peer.
5  If  $\text{Pagerank}(w) = \text{Null}$  then
6    Set  $\text{Pagerank}(w) = \frac{N-n}{N}$ 
7  End If
8  For  $i = 1$  to  $n$ 
9    If  $\text{Pagerank}(n_i) = \text{Null}$  then
10     Set  $\text{Pagerank}(n_i) = \frac{1}{N}$ ,  $n_i$  is the  $i$ -node in  $l\_wlg$ .
11   End If
12 End For
13 Find the changed partition of  $l\_wlg$ , denoted as  $cl\_wlg$ .
14 Find the unchanged partition of  $l\_wlg$ , denoted as  $ul\_wlg$ .
15 Find the set of boundary nodes between  $cl\_wlg$  and  $ul\_wlg$ , denoted as  $adjl\_wlg$ .
16 Add  $w$  into  $adjl\_wlg$ .
17 Determine the Pagerank in  $adjl\_wlg$ .

```

รูปที่ 3.5 แสดงการคำนวณค่าเพจเรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ในโลคอลเว็บลิงก์กราฟ

บรรทัดที่ 1 กำหนดให้ p แทนเพียร์ในเพียร์ทูเพียร์เน็ตเวิร์ก

บรรทัดที่ 2 กำหนดให้ l_wlg เป็นโลคอลเว็บลิงก์กราฟขนาด n โหนด

บรรทัดที่ 3 กำหนดให้ g_wlg เป็นโลคอลเว็บลิงก์กราฟขนาด N โหนด

บรรทัดที่ 4 กำหนดให้ w เป็นเวปด์โหนดในเพียร์

บรรทัดที่ 5, 6, 7 เป็นการกำหนดค่าเพจเรงค์เริ่มต้นให้กับเวปด์โหนดในแต่ละเพียร์ โดยกำหนดให้ ค่าเพจเรงค์ของเวปด์โหนดเป็น $\frac{N-n}{N}$

บรรทัดที่ 8, 9, 10, 11, 12 เป็นการกำหนดค่าเพจเรงค์เริ่มต้นให้กับเว็บเพจในโลคอลเว็บลิงก์กราฟของแต่ละเพียร์ โดยกำหนดให้ ค่าเพจเรงค์ของเว็บเพจในโลคอลเว็บลิงก์กราฟเป็น $\frac{1}{N}$

บรรทัดที่ 13 หาส่วนของเว็บเพจที่ได้รับผลกระทบจากเว็บเพจที่เพิ่มเข้ามาใหม่ในเว็บลิงค์กราฟ l_wlg แทนด้วย cl_wlg

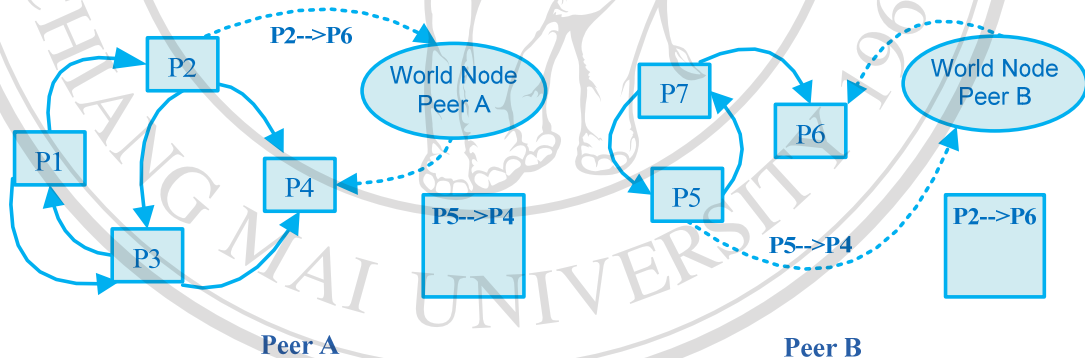
บรรทัดที่ 14 หาส่วนของเว็บเพจที่ไม่ได้รับผลกระทบจากเว็บเพจที่เพิ่มเข้ามาใหม่ในเว็บลิงค์กราฟ l_wlg แทนด้วย ul_wlg

บรรทัดที่ 15 หาเซตของเว็บเพจใน ul_wlg ที่มีเฉพาะลิงค์ออกไปยังเว็บเพจใน cl_wlg เพื่อหาเซตของโหนดขอบ เพื่อนำไปรวมกับเว็บเพจในส่วน cl_wlg แทนด้วย $adjl_wlg$

บรรทัดที่ 16 เพิ่มเวกซ์โหนด w ให้กับ $adjl_wlg$ ซึ่งจะรวมเว็บเพจในส่วน ul_wlg ที่เหลือจากการตัดโหนดขอบทิ้งเข้าไปกับเวกซ์โหนด w ด้วย

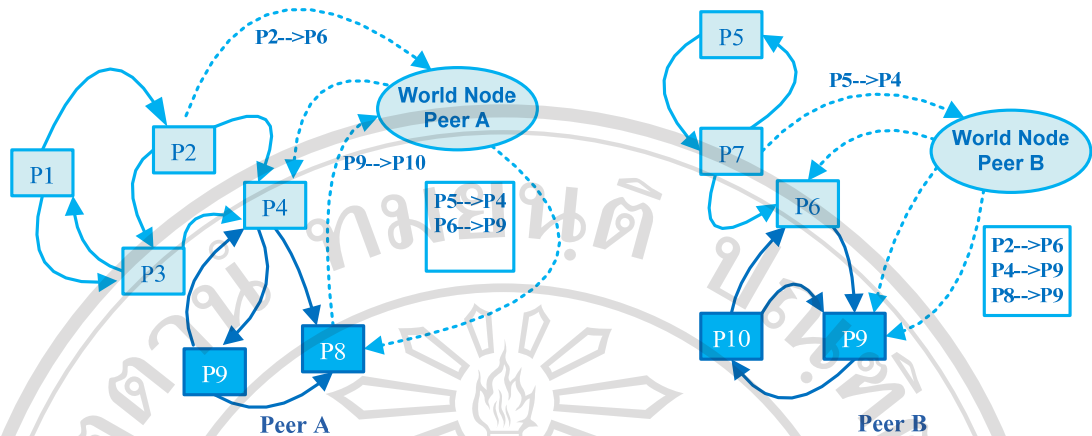
บรรทัดที่ 17 คำนวณค่าเพจแรงค์ใน $adjl_wlg$ ที่ถูกเก็บไว้ในแต่ละเพียร์

ตัวอย่างการคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ในแต่ละ โคลดเว็บลิงค์กราฟ กำหนดให้เครือข่ายเพียร์ทูเพียร์มีเพียร์ทั้งหมด 2 เพียร์ ซึ่งเก็บโคลดเว็บลิงค์กราฟได้ดังรูปที่ 3.6



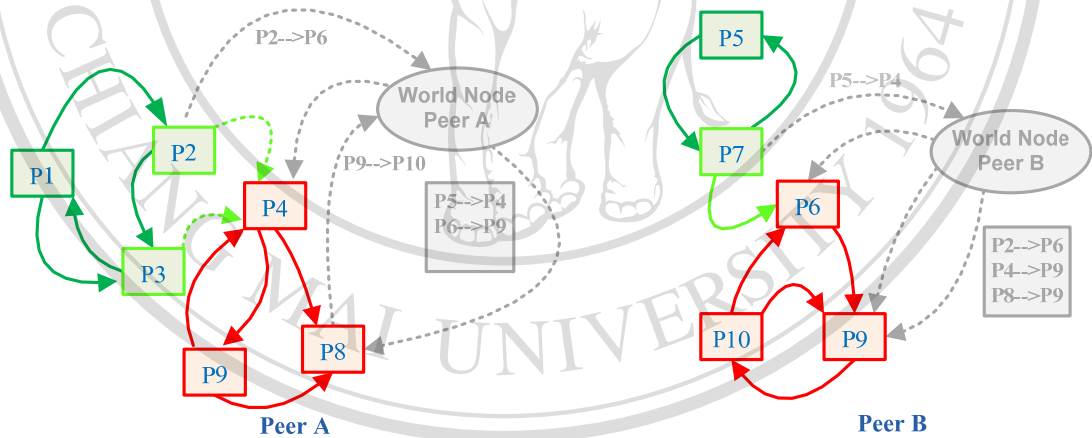
รูปที่ 3.6 โคลดเว็บลิงค์กราฟที่ถูกเก็บไว้ในเพียร์ A และเพียร์ B

จากเว็บลิงค์กราฟในรูปที่ 3.6 เมื่อมีเว็บเพจ P8, P9 และ P10 เพิ่มเข้ามาใหม่ ทำให้โครงสร้างของเว็บลิงค์กราฟในแต่ละเพียร์เปลี่ยนแปลงไป ซึ่งแสดงโครงสร้างของเว็บลิงค์กราฟที่เปลี่ยนแปลงไปของแต่ละเพียร์ในรูปที่ 3.7



รูปที่ 3.7 โครงสร้างของโลคอลเว็บลิงค์กราฟที่เปลี่ยนแปลงในเพียร์ A และเพียร์ B

จากรูปที่ 3.7 จะเห็นได้ว่าที่โลคอลเว็บลิงค์กราฟในเพียร์ A จะมีเว็บเพจ P8 และ P9 เพิ่มเข้ามาใหม่ และที่โลคอลเว็บลิงค์กราฟในเพียร์ B จะมีเว็บเพจ P9 และ P10 เพิ่มเข้ามาใหม่ ซึ่งการเพิ่มขึ้นมาของโหนดดังกล่าว สามารถแบ่งโลคอลเว็บลิงค์กราฟในแต่ละเพียร์ได้ดังรูปที่ 3.8

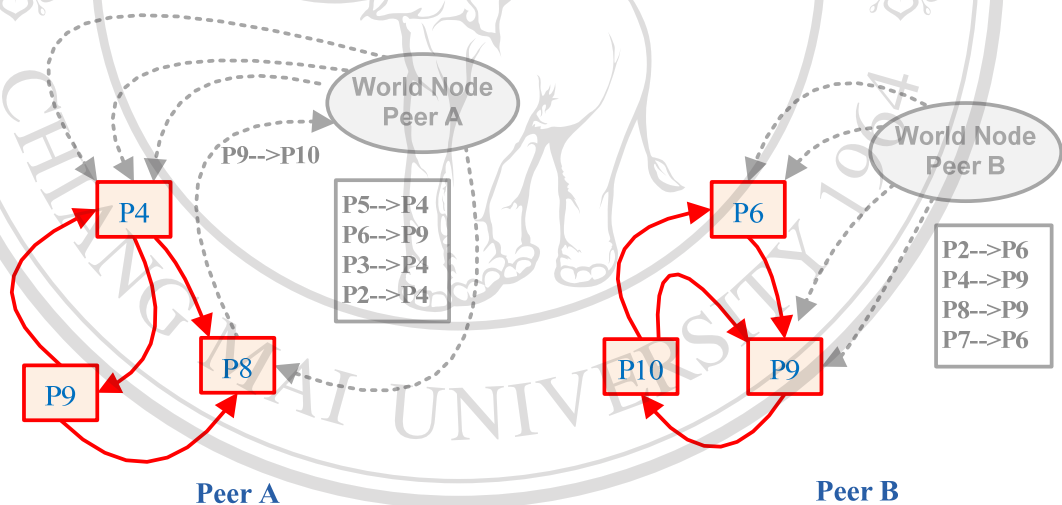


รูปที่ 3.8 การแบ่งส่วนของโลคอลเว็บลิงค์กราฟในเพียร์ A และเพียร์ B

การแบ่งส่วนที่เพียร์ A จากรูปที่ 3.8 เมื่อมีเว็บเพจ P8 และ P9 เพิ่มเข้ามาใหม่จะส่งผลกระทบต่อค่าเพจแรงค์ของเว็บเพจที่เป็นเคสเซ็นเด็นทั้งหมด กล่าวคือเว็บเพจ P4, P8 และ P9 ลักษณะดังกล่าวจะทำให้เว็บลิงค์กราฟในเพียร์ A ถูกแบ่งเป็น 2 ส่วนอย่างชัดเจน ดังรูปที่ 3.8 คือกลุ่มของเว็บเพจที่เป็นเคสเซ็นเด็นของเว็บเพจที่เพิ่มเข้ามาใหม่ และกลุ่มของเว็บเพจที่มีผลกระทบต่อค่าเพจแรงค์เพียงเล็กน้อยเมื่อมีเว็บเพจใหม่เพิ่มขึ้นมา ซึ่งได้แก่เว็บเพจ P1, P2 และ P3 ซึ่งเว็บเพจทั้งสองกลุ่มจะเชื่อมต่อกันด้วยโหนดขอบ P2 และ P3

ในทำนองเดียวกันการแบ่งส่วนที่เพียร์ B จากรูปที่ 3.8 เมื่อมีเว็บเพจ P9 และ P10 เพิ่มเข้ามาใหม่ จะส่งผลกระทบต่อค่าเพจแรงค์ของโหนดเดสเซ็นด์นั้นกล่าวคือเว็บเพจ P6, P9 และ P10 ซึ่งเว็บลิงก์กราฟจะถูกแบ่งเป็น 2 ส่วนเช่นเดียวกับเพียร์ A คือ กลุ่มของเว็บเพจที่เป็นเดสเซ็นด์ของเว็บเพจที่เพิ่มเข้ามาใหม่ และกลุ่มของเว็บเพจที่มีผลกระทบต่อค่าเพจแรงค์เพียงเล็กน้อย เมื่อมีเว็บเพจใหม่เพิ่มขึ้นมา ซึ่งได้แก่เว็บเพจ P5 และ P7 โดยมีการเชื่อมต่อกันที่โหนดขอบ P7

อย่างไรก็ตาม จะเห็นได้ว่าแต่ละเพียร์การแบ่งส่วนของโลกเว็บลิงก์กราฟจะไม่พิจารณาเว็บลิงก์ที่เชื่อมต่อไปยังเว็ลด์โหนด และส่วนของกลุ่มของเว็บเพจที่มีผลกระทบต่อค่าเพจแรงค์เมื่อมีเว็บเพจใหม่เพิ่มเข้ามา จะถูกนำไปรวมกับเว็ลด์โหนด ดังแสดงตัวอย่างในรูปที่ 3.9 ซึ่งเป็นเว็บลิงก์กราฟที่เกิดจากการพิจารณาเฉพาะเดสเซ็นด์ของเว็บเพจที่เพิ่มเข้ามาใหม่ ซึ่งจะนำโลกเว็บลิงก์กราฟที่ได้นี้ไปคำนวณค่าเพจแรงค์ตามวิธีการของขั้นตอนวิธี JXP ที่ได้กล่าวถึงไปแล้วในบทที่ 2



รูปที่ 3.9 โลกเว็บลิงก์กราฟในเพียร์ A และเพียร์ B ที่นำมาคำนวณค่าเพจแรงค์

หลังจากคำนวณค่าเพจแรงค์แบบเฉพาะส่วนที่เพิ่มขึ้นมาใหม่ จนครบในทุกๆ เพียร์ในเครือข่ายเพียร์ทูเพียร์แล้ว จากนั้นจะเป็นการคำนวณค่าเพจแรงค์ในขณะรวมสองโลกเว็บลิงก์กราฟ ให้เป็นกราฟใหม่ในกระบวนการจับคู่ระหว่างเพียร์ ซึ่งมีขั้นตอนวิธีดังรูปที่ 3.10

Algorithm ICR_MergePR

- 1 Let *largest_wlg* be the largest local web link graph.
- 2 Let *g_weight* be the collection of the weight of all webs.
- 3 Let *m* be the number of peers in the P2P network.
- 4 For *i* = 1 to *m*
- 5 Select a local web link graph *b_wlg* from a peer randomly.
- 6 Let *m_wlg* = Merge(*largest_wlg*, *b_wlg*).
- 7 Find the changed partition of *m_wlg*, denoted as *cm_wlg*.
- 8 Find the unchanged partition of *m_wlg*, denoted as *um_wlg*.
- 9 Find the set of boundary nodes between *cm_wlg* and *um_wlg*, denoted as *adjm_wlg*.
- 10 Add *w* into *adjm_wlg*.
- 11 Determine the Pagerank in *adjm_wlg*.
- 12 Update the processed Pagerank in *g_weight*.
- 13 Disconnect *b_wlg* from *m_wlg*.
- 14 End For

รูปที่ 3.10 การคำนวณค่าเพจแรงค์เฉพาะส่วนที่เพิ่มเข้ามาใหม่ในกระบวนการจับคู่ระหว่างเพียร์

บรรทัดที่ 1 กำหนดให้ *largest_wlg* แทนเว็บลิงค์กราฟที่มีขนาดใหญ่ที่สุด

บรรทัดที่ 2 กำหนดให้ *g_weight* แทนเซตของเว็บเพจทั้งหมดในเครือข่าย

บรรทัดที่ 3 กำหนดให้ *m* เป็นขนาดของเครือข่ายแบบเพียร์ทูเพียร์

บรรทัดที่ 4 เริ่มขั้นตอนการจับคู่ระหว่างเพียร์

บรรทัดที่ 5 สุ่มเพียร์ที่เหลืออยู่ในเครือข่ายขึ้นมาทีละเพียร์ให้ชื่อว่า *b_wlg*

บรรทัดที่ 6 รวมเว็บลิงค์กราฟ *largest_wlg* และ เว็บลิงค์กราฟ *b_wlg* เข้าด้วยกันเป็นเว็บลิงค์กราฟใหม่ที่ชื่อว่าเว็บลิงค์กราฟ *m_wlg*

บรรทัดที่ 7 หาส่วนของเว็บเพจที่ได้รับผลกระทบจากเว็บเพจที่เพิ่มเข้ามาใหม่ในเว็บลิงค์กราฟ *m_wlg* แทนด้วย *cm_wlg*

บรรทัดที่ 8 หาส่วนของเว็บเพจที่ไม่ได้รับผลกระทบจากเว็บเพจที่เพิ่มเข้ามาใหม่ในเว็บลิงค์กราฟ *m_wlg* แทนด้วย *um_wlg*

บรรทัดที่ 9 หาเซตของเว็บเพจใน um_wlg ที่มีเฉพาะลิงค์ออกไปยังเว็บเพจใน cm_wlg เพื่อหาเซตของโหนดขอบ เพื่อนำไปรวมกับเว็บเพจในส่วน cm_wlg แทนด้วย $adjm_wlg$

บรรทัดที่ 10 เพิ่มเวกเตอร์โหนด w ให้กับ $adjm_wlg$

บรรทัดที่ 11 คำนวณค่าเพจแรงค์ที่เว็บลิงค์กราฟ $adjm_wlg$

บรรทัดที่ 12 ปรับปรุงค่าเพจแรงค์ใน g_weight

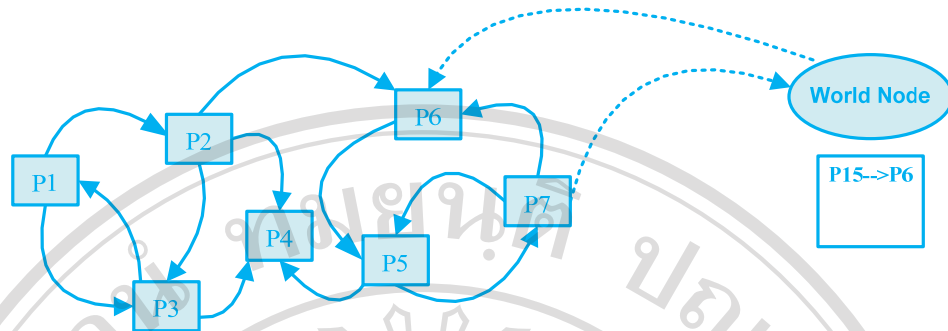
บรรทัดที่ 13 แยกเว็บลิงค์กราฟ b_wlg ออกจากเว็บลิงค์กราฟ m_wlg

บรรทัดที่ 14 ทำซ้ำในบรรทัดที่ 4-13 จนครบหมดทุกเพียร์ในเครือข่าย

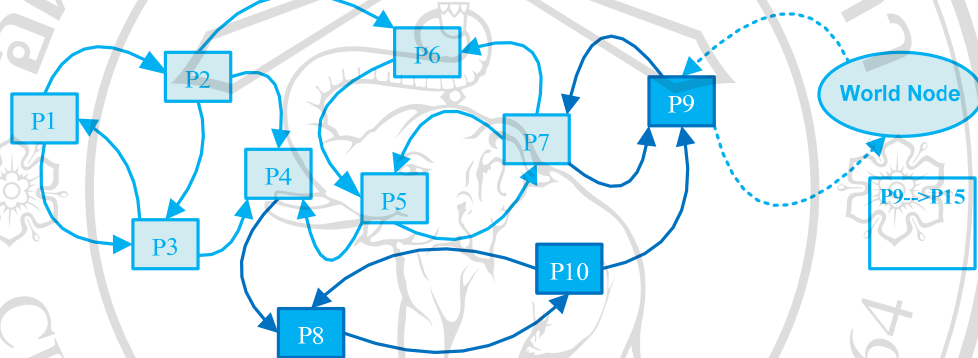
จากขั้นตอนวิธีในรูป 3.10 ที่กล่าวมาจะเป็นการสุ่มเลือกเพียร์ขึ้นมาทีละหนึ่งเพียร์ เพื่อนำโลคอลเว็บลิงค์กราฟมารวมกับโลคอลเว็บลิงค์กราฟในเพียร์หลัก กลายเป็นเว็บลิงค์กราฟใหม่ ขึ้นมาดังแสดงในรูปที่ 3.11 ซึ่งเป็นตัวอย่างของกราฟที่เกิดจากการรวมกันของโลคอลเว็บลิงค์กราฟใน 2 เพียร์

ก่อนการเปลี่ยนแปลงโครงสร้างของโลคอลเว็บลิงค์กราฟในแต่ละเพียร์ เมื่อนำโลคอลเว็บลิงค์กราฟที่ถูกเก็บไว้ในเพียร์มารวมกันจะได้เว็บลิงค์กราฟใหม่ ดังตัวอย่างที่แสดงในรูป 3.11a เมื่อมีเว็บเพจเพิ่มขึ้นมาใหม่ในแต่ละเพียร์ จะทำให้โครงสร้างเว็บลิงค์กราฟในแต่ละเพียร์เปลี่ยนไป เมื่อเกิดการรวมกราฟทำให้ได้เว็บลิงค์กราฟใหม่ดังแสดงตัวอย่างในรูป 3.11b

จากรูป 3.11b จะเห็นว่าเว็บเพจ P8, P9 และ P10 เพิ่มขึ้นมาใหม่ ซึ่งเว็บเพจดังกล่าว เป็นเว็บเพจที่เพิ่มขึ้นมาใหม่ในแต่ละโลคอลเว็บลิงค์กราฟ ซึ่งเว็บเพจที่เพิ่มขึ้นมาใหม่เหล่านี้ จะเป็นส่วนที่เพิ่มขึ้นมาใหม่ของเว็บลิงค์กราฟที่เกิดจากการรวมสองโลคอลเว็บลิงค์กราฟเข้าด้วยกันด้วย



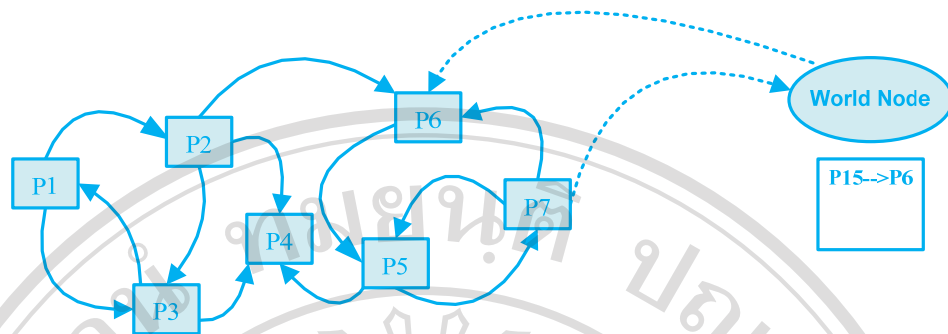
a. Merge Graph เริ่มต้น



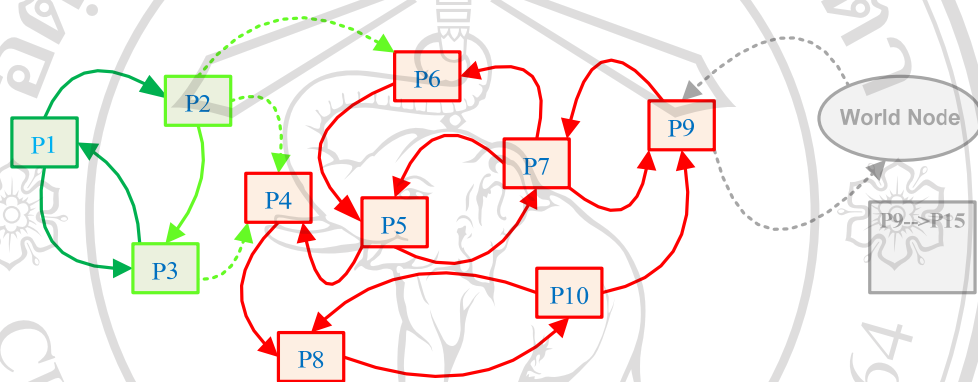
b. Merge Graph ที่มีโครงสร้างเปลี่ยนแปลงไป

รูปที่ 3.11 แสดงตัวอย่าง Merge Graph

สำหรับการคำนวณเฉพาะส่วนที่เพิ่มเข้ามาใหม่ในขั้นตอนของการรวมกราฟ มีหลักการเช่นเดียวกับการคำนวณเฉพาะส่วนที่เพิ่มเข้ามาใหม่ของแต่ละ โลกอลเว็บลิงค์กราฟ ซึ่งทำได้โดยการหาเว็บเพจเดสเซนเดนซ์ของเว็บเพจที่เพิ่มเข้ามาใหม่ เพื่อทำการแบ่งกราฟออกเป็น 2 ส่วน คือ ส่วนที่ได้รับผลกระทบต่อค่าเพจแรงค์เมื่อมีเว็บเพจใหม่เพิ่มเข้ามา และส่วนที่ไม่ได้รับผลกระทบของค่าเพจแรงค์เมื่อมีเว็บเพจใหม่เพิ่มเข้ามา ดังแสดงตัวอย่างของการแบ่งเว็บลิงค์กราฟที่เกิดจากการรวมกันของโลกอลเว็บลิงค์กราฟใน 2 เพียร์ดังรูปที่ 3.12



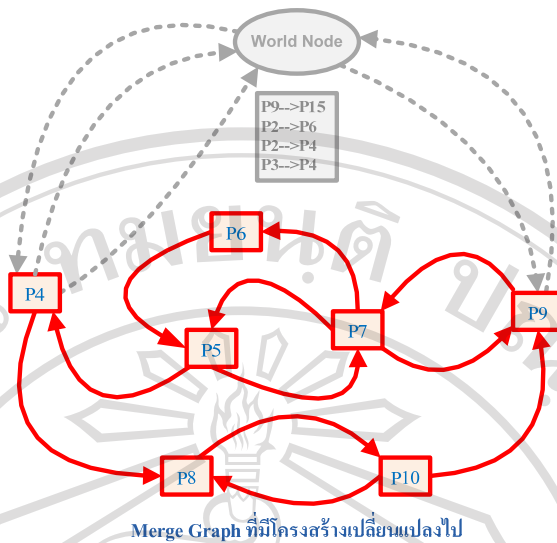
a. Merge Graph เริ่มต้น



b. Merge Graph ที่มีโครงสร้างเปลี่ยนแปลงไป

รูปที่ 3.12 แสดงตัวอย่าง Merge Graph ที่ถูกแบ่งส่วน

จากรูป 3.12 เมื่อมีเว็บเพจ P8, P9 และ P10 เพิ่มขึ้นมาใหม่จะทำให้ค่าเพจแรงค์ของโหนดเดสเซ็นเด็น ซึ่งได้แก่ เว็บเพจ P4, P5, P6, P7, P8, P9 และ P10 ซึ่งจะคำนวณค่าเพจแรงค์ใหม่เฉพาะเว็บเพจในส่วนดังกล่าว และเว็บเพจในส่วนที่เหลือซึ่งได้แก่ เว็บเพจ P1, P2 และ P3 จะถูกนำไปรวมกับโหนด โหนด จะได้เว็บลิงค์กราฟใหม่ดังแสดงในรูปที่ 3.13 ซึ่งเป็นเว็บลิงค์กราฟที่จะพิจารณาเฉพาะส่วนที่ได้รับผลกระทบต่อค่าเพจแรงค์เมื่อมีเว็บเพจใหม่เพิ่มเข้ามา สำหรับการคำนวณค่าเพจแรงค์นั้นทำได้โดยวิธีการที่กล่าวไปแล้วในบทที่ 2



รูปที่ 3.13 แสดง Merge Graph ที่นำมาคำนวณค่าเพจแรงค์

ดังที่กล่าวมาจะพบว่า ทั้งการพิจารณาค่าเพจแรงค์เฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่ในโกลบอลเว็บลิงค์กราฟของแต่ละเพียร์ และการพิจารณาค่าเพจแรงค์เฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่ในเว็บลิงค์กราฟที่เกิดจากการรวมโกลบอลเว็บลิงค์กราฟในเพียร์ เว็บลิงค์กราฟที่ถูกพิจารณา จะมีโครงสร้างที่เล็กลงเมื่อเทียบกับเว็บลิงค์กราฟที่ไม่มีการพิจารณาเฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่ ดังตารางที่ 3.2 แสดงตัวอย่างการเปรียบเทียบขนาดและเวลาการคำนวณค่าเพจแรงค์ ระหว่างโกลบอลเว็บลิงค์กราฟที่มีการพิจารณาเฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่ และไม่มีการพิจารณาเฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่

เพียร์	เว็บลิงค์กราฟที่ไม่มีการพิจารณาเฉพาะส่วนที่เพิ่มเข้ามาใหม่		เว็บลิงค์กราฟที่มีการพิจารณาเฉพาะส่วนที่เพิ่มเข้ามาใหม่	
	ขนาด (โหนด)	เวลา (รอบ)	ขนาด (โหนด)	เวลา (รอบ)
Peer 1	1960	17	1897	17
Peer 2	1957	5	1450	5
Peer 3	1957	5	875	4
Peer 4	1957	4	965	3
Peer 5	1947	7	42	1

ตารางที่ 3.2 การเปรียบเทียบขนาดและเวลาในการคำนวณค่าเพจแรงค์ของโกลบอลเว็บลิงค์กราฟที่มีการพิจารณาเฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่ และไม่มีการพิจารณาเฉพาะเว็บเพจที่เพิ่มเข้ามาใหม่

จากตารางที่ 3.2 จะเห็นว่าเว็บลิงค์กราฟที่พิจารณาเฉพาะส่วนที่เพิ่มเข้ามาใหม่ จะมีจำนวนเว็บเพจที่นำมาคำนวณค่าเพจแรงค่าน้อยกว่าเว็บลิงค์กราฟที่ไม่มีการพิจารณาเฉพาะส่วนที่เพิ่มเข้ามาใหม่ แต่จะสังเกตเห็นว่า โคลอเว็บลิงค์กราฟที่เพียร์ 3, เพียร์ 4 และเพียร์ 5 มีจำนวนของเว็บเพจที่จะนำมาพิจารณาค่าเพจแรงค่าน้อยกว่าครึ่งของขนาดโคลอเว็บลิงค์กราฟ ส่งผลให้เวลาที่ใช้คำนวณค่าเพจแรงค่าน้อยกว่าแบบเห็นได้ชัดเจน

ดังนั้นขั้นตอนวิธีการพิจารณาเฉพาะเว็บเพจที่เพิ่มขึ้นมาใหม่ในโคลอเว็บลิงค์กราฟ ดังที่ได้เสนอไปในงานวิจัยนี้ จะมีประสิทธิภาพสูง ก็ต่อเมื่อประมวลผลกับเว็บลิงค์กราฟสามารถแบ่งส่วนให้เป็น 2 ส่วนได้อย่างเหมาะสม ซึ่งประสิทธิภาพของขั้นตอนวิธีของการพิจารณาเพจแรงค่านเฉพาะส่วนที่เพิ่มเข้ามาใหม่นี้จะแสดงรายละเอียดในบทต่อไป