Page Clustering

Mehrdad Mohammadian

Data

P1: t1, t4

P2: t3

P3: t2, t3, t4

P4: t3, t4

P5: t1, t5

P6:

P7: t1, t2, t3

P8:

P9:

P10: t1, t3

Calculate Distance Matrix

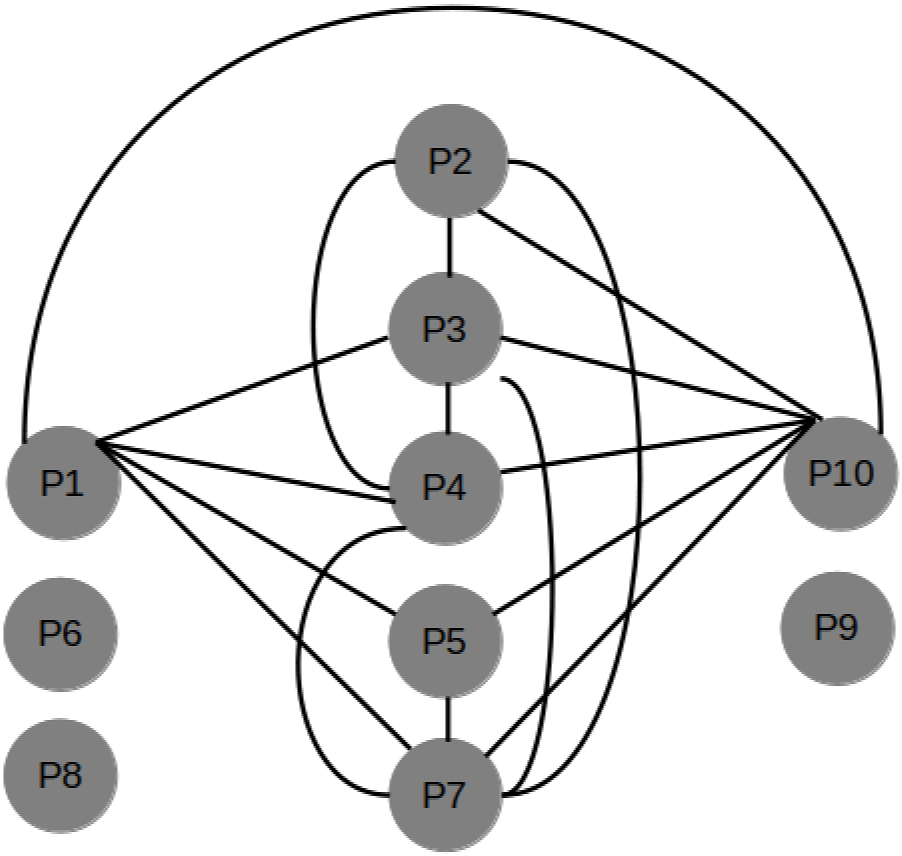
$$dis(x_i, x_j) = \frac{p - m}{p}$$

10 * 10 Matrix:

$$\begin{bmatrix}
- & & & & & & & & & \\
1 & - & & & & & & & & \\
0.8 & 0.8 & - & & & & & & & \\
0.8 & 0.8 & 0.6 & - & & & & & & \\
0.8 & 1 & 1 & 1 & - & & & & & \\
1 & 1 & 1 & 1 & 1 & - & & & & \\
0.8 & 0.8 & 0.6 & 0.8 & 0.8 & 1 & - & & & \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & - & & \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & - & \\
0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 1 & 0.6 & 1 & 1 & -
\end{bmatrix}$$

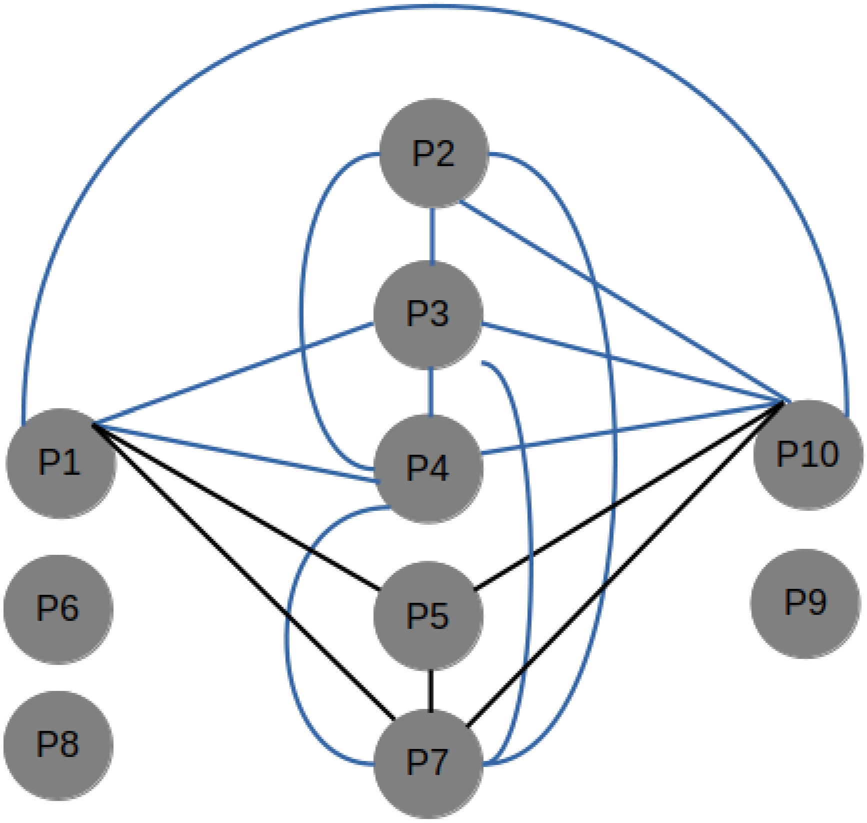
set threshold: 0.8

$$\begin{bmatrix}
- & & & & & & & & & \\
0 & - & & & & & & & & \\
1 & 1 & - & & & & & & & \\
1 & 1 & 1 & - & & & & & & \\
1 & 0 & 0 & 0 & - & & & & & \\
0 & 0 & 0 & 0 & 0 & - & & & & \\
1 & 1 & 1 & 1 & 1 & 0 & - & & & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & - & & \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & - & \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & -
\end{bmatrix}$$

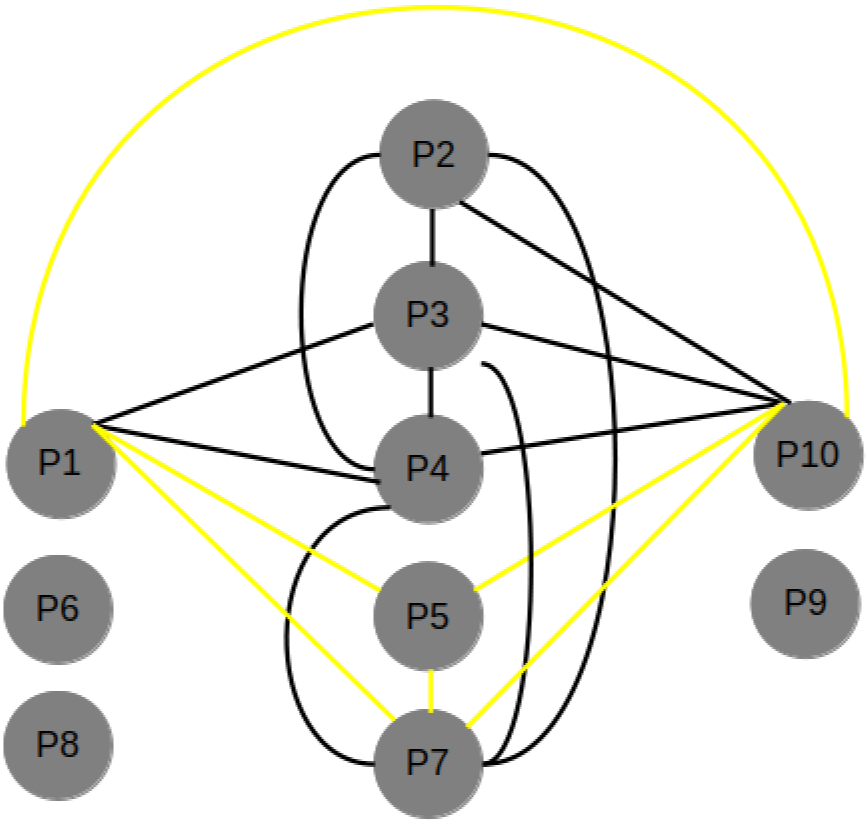


Maximal Clique Method

clique 1: {p1, p2, p3, p4, p7, p10}



clique 2: {p1, p5, p7, p10}



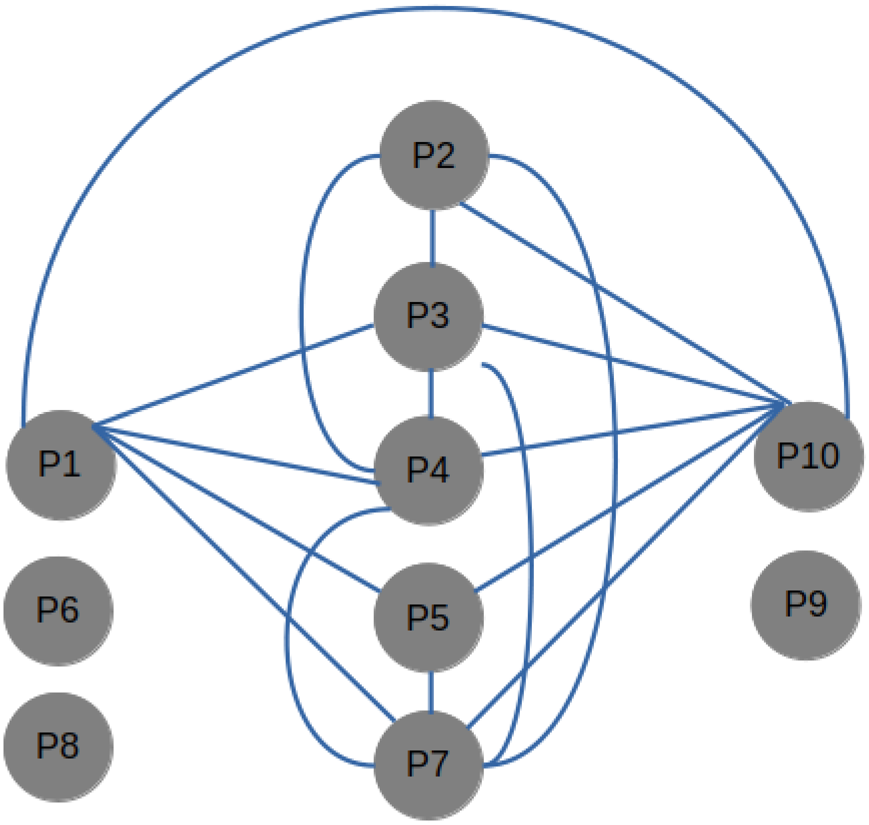
clique 3: {p6}

clique 4: {p8}

clique 5: {p9}

Single Link Method

Cluster 1: {p1, p2, p3, p4, p5, p7, p10}



Cluster 2: {p6}

Cluster 3: {p8}

Cluster 4: {p9}

The End