

Clustering: An Example

- 6 genes (G1..G6)
- 2 conditions (hybridizations, x-dim,y-dim)
- Calculate distance matrix using Manhattan distance
- Build gene tree using average linkage

Clustering: An Example

- 6 gene Vectors

G1: (11,4)

G2: (13,7)

G2: (6,10)

G4: (14,13)

G5: (10,12)

G6: (5,14)

Clustering: An Example

- 6 gene Vecotors

G1: (11,4)

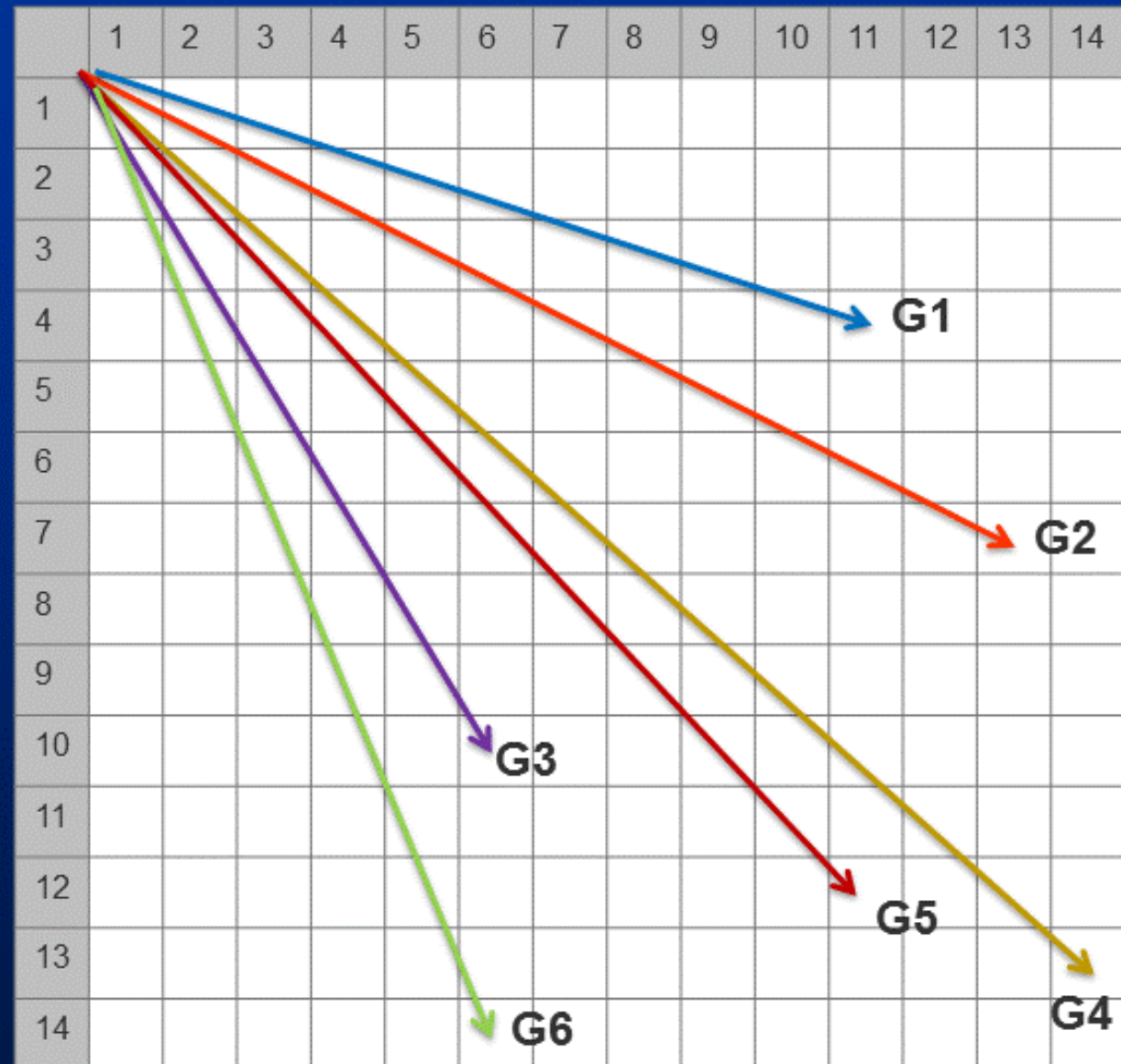
G2: (13,7)

G2: (6,10)

G4: (14,13)

G5: (10,12)

G6: (5,14)



Clustering: Compute distance matrix

6 genes x 6 genes

	G1	G2	G3	G4	G5	G6
G1						
G2						
G3						
G4						
G5						
G6						

Clustering: Compute distance matrix

6 genes x 6 genes

Self $\leftrightarrow$ Self distances = 0 (e.g. G1 $\leftrightarrow$ G1)

=> All main diagonal elements = 0

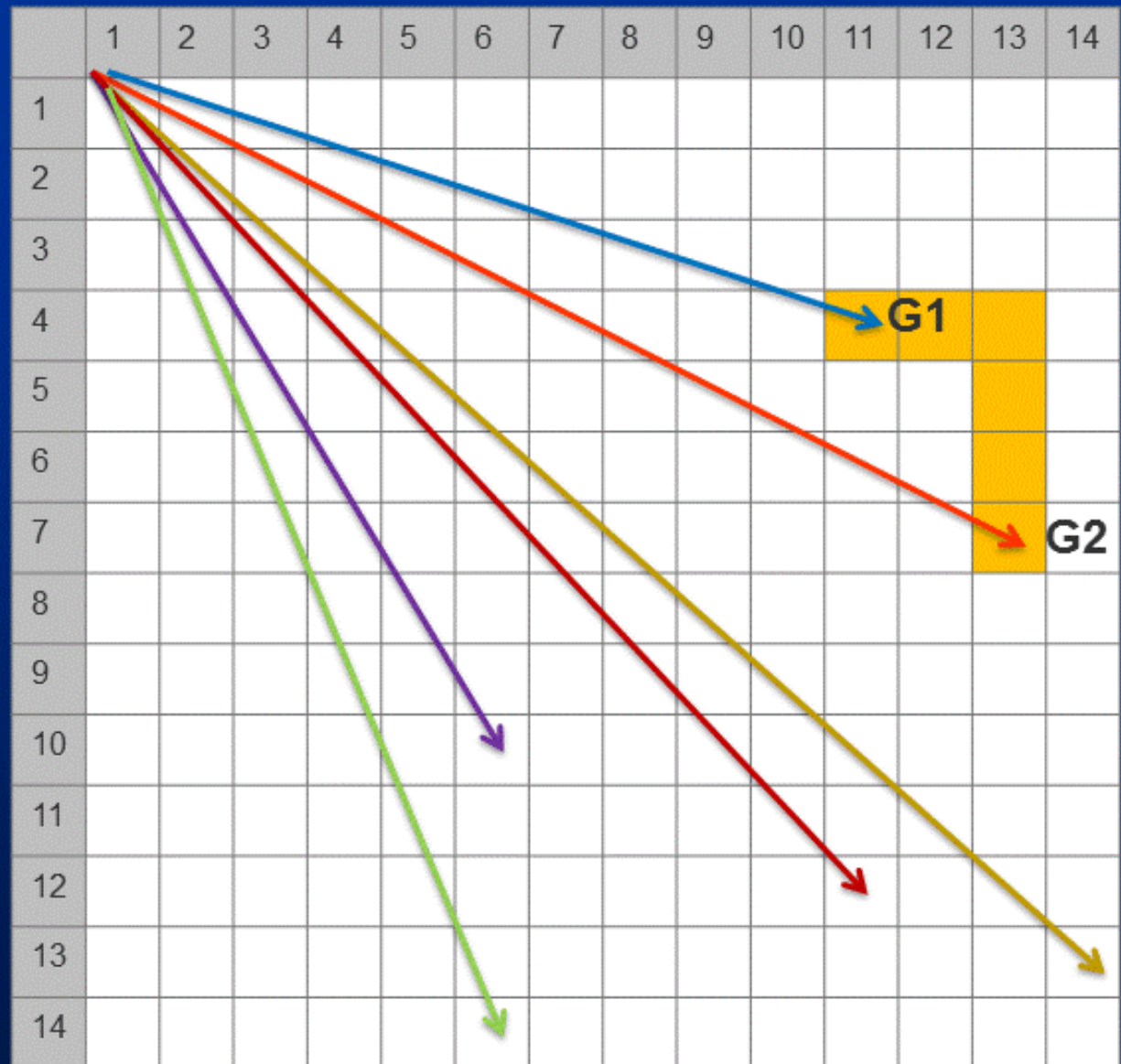
	G1	G2	G3	G4	G5	G6
G1	0					
G2		0				
G3			0			
G4				0		
G5					0	
G6						0

Clustering: Compute all other distances

$$G1 \leftrightarrow G2 = |13-11| + |7-4| \\ = 2 + 3 = 5$$

$$G2 \leftrightarrow G1 = G1 \leftrightarrow G2 = 5$$

	G1	G2	G3	G4	G5	G6
G1	0	5				
G2	5	0				
G3			0			
G4				0		
G5					0	
G6						0

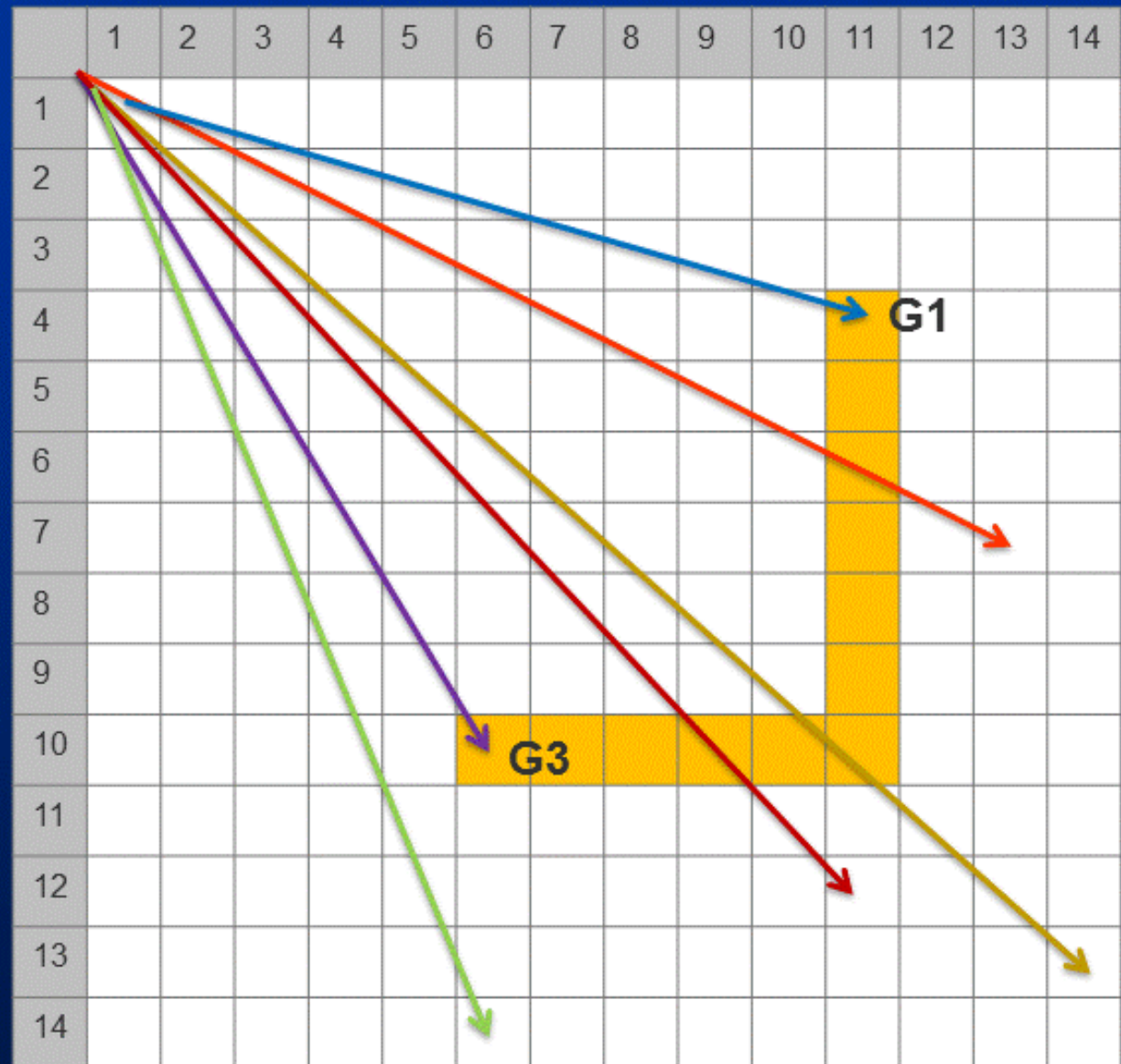


Clustering: Compute all other distances

$$G1 \leftrightarrow G3 = |11-6| + |10-4| \\ = 5 + 6 = 11$$

$$G3 \leftrightarrow G1 = G1 \leftrightarrow G3 = 11$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11			
G2	5	0				
G3	11		0			
G4				0		
G5					0	
G6						0

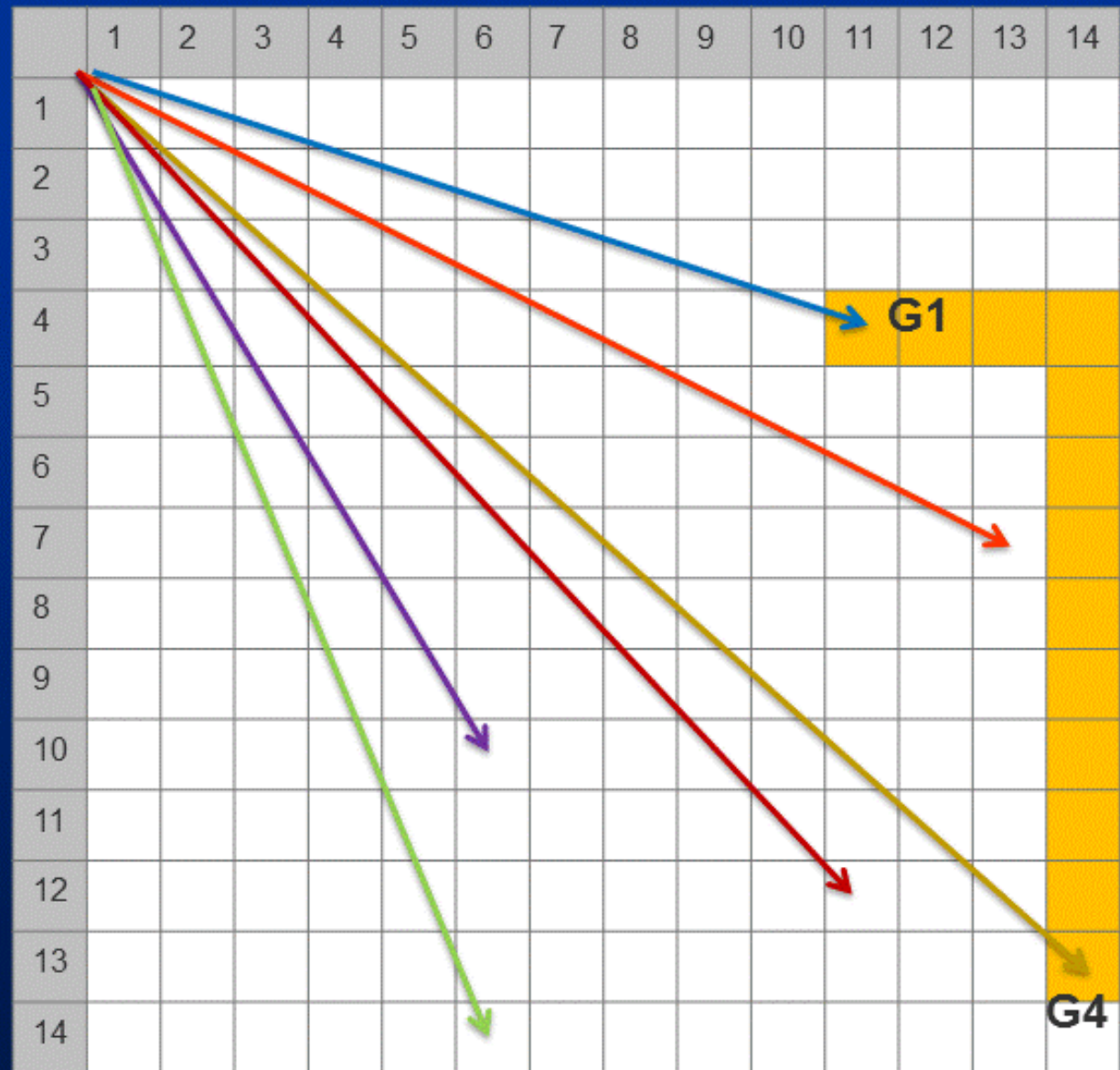


Clustering: Compute all other distances

$$G1 \Leftrightarrow G4 = |13-4| + |14-11| \\ = 9 + 3 = 12$$

$$G4 \Leftrightarrow G1 = G1 \Leftrightarrow G4 = 12$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12		
G2	5	0				
G3	11		0			
G4	12			0		
G5					0	
G6						0

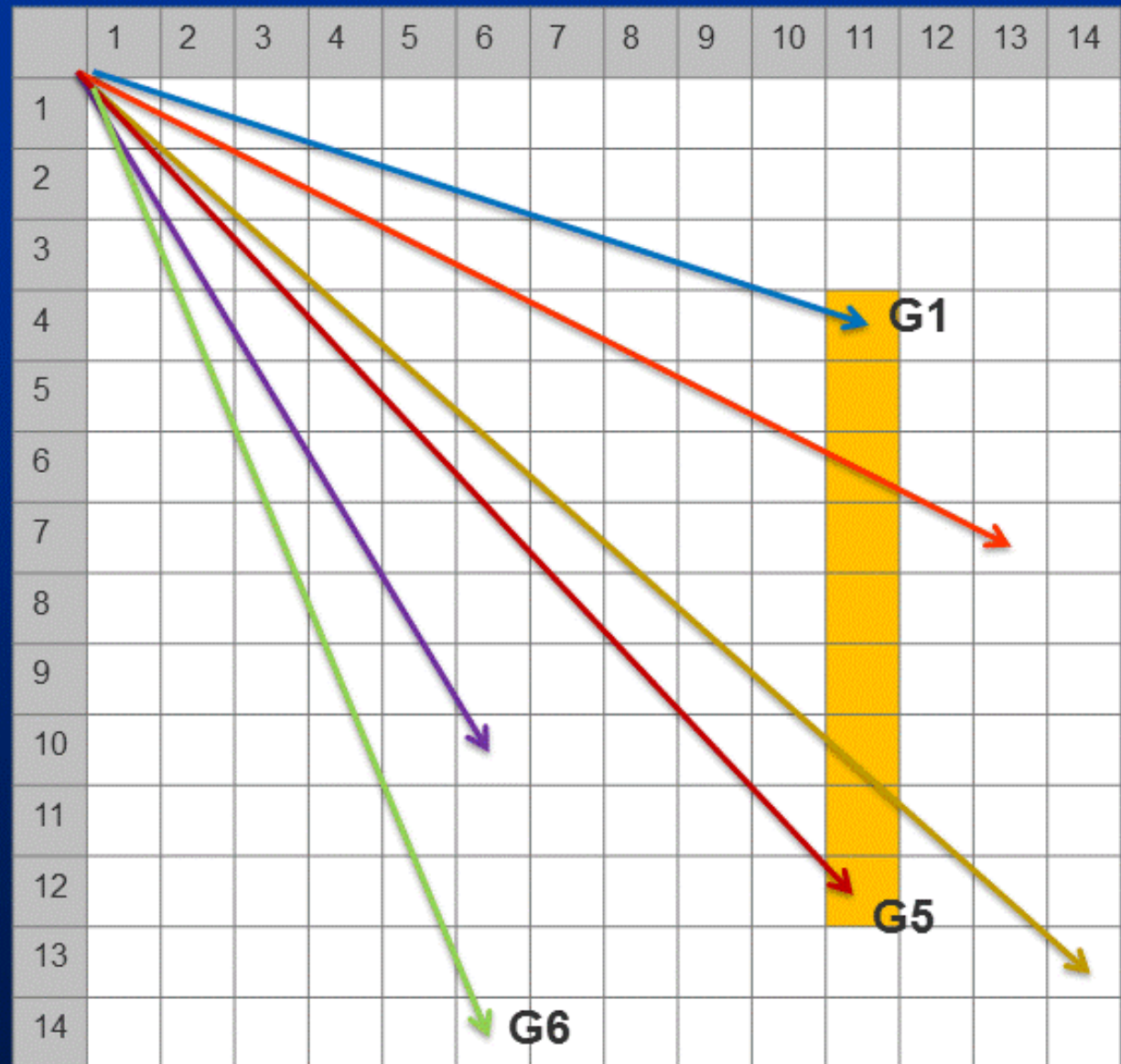


Clustering: Compute all other distances

$$G1 \leftrightarrow G5 = |12-4| + |11-11| \\ = 8 + 0 = 8$$

$$G5 \leftrightarrow G1 = G1 \leftrightarrow G5 = 8$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	
G2	5	0				
G3	11		0			
G4	12			0		
G5	8				0	
G6						0

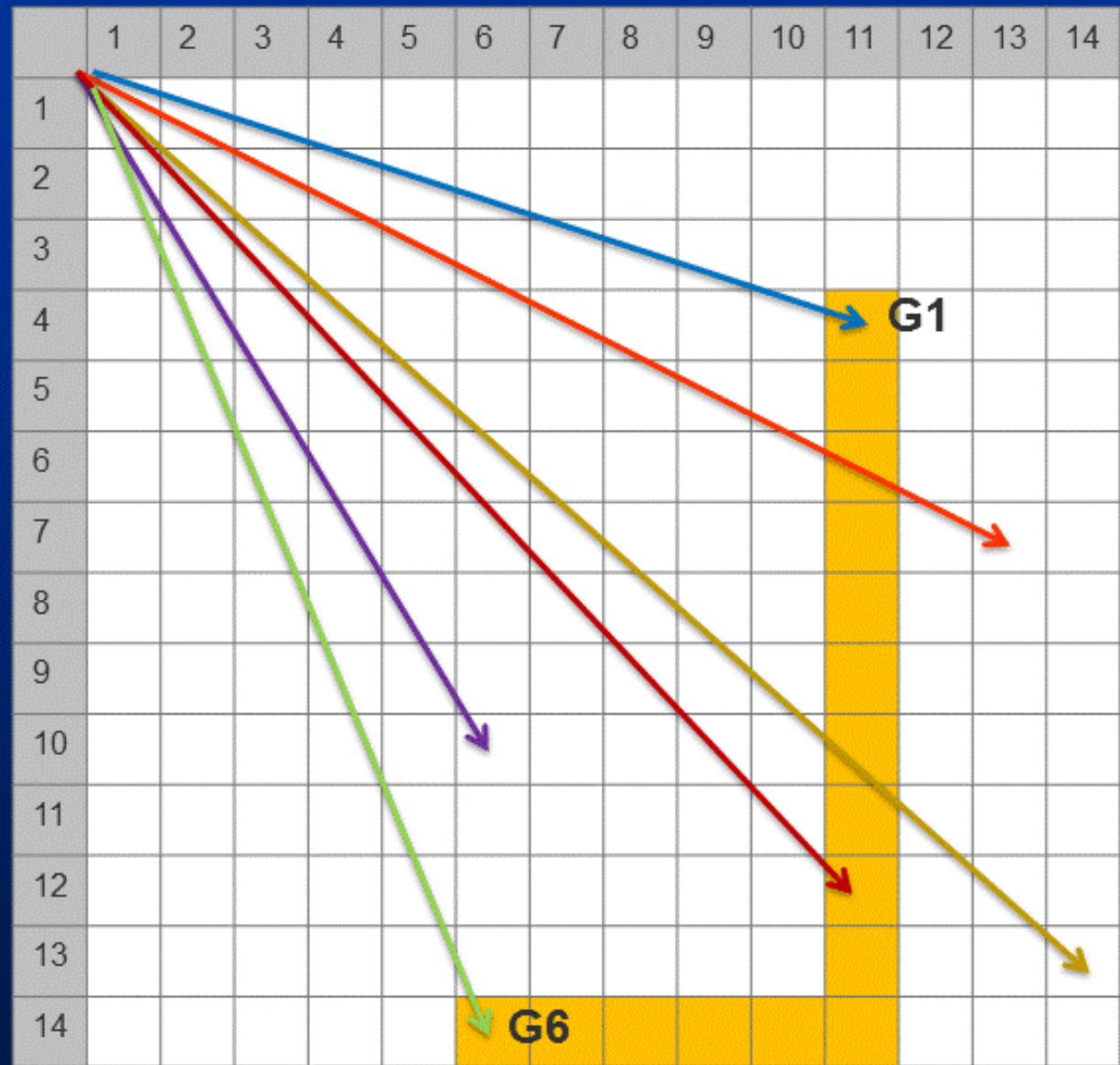


Clustering: Compute all other distances

$$G1 \leftrightarrow G6 = |11-6| + |14-4| \\ = 5 + 10$$

$$G6 \leftrightarrow G1 = G1 \leftrightarrow G6 = 15$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0				
G3	11		0			
G4	12			0		
G5	8				0	
G6	15					0

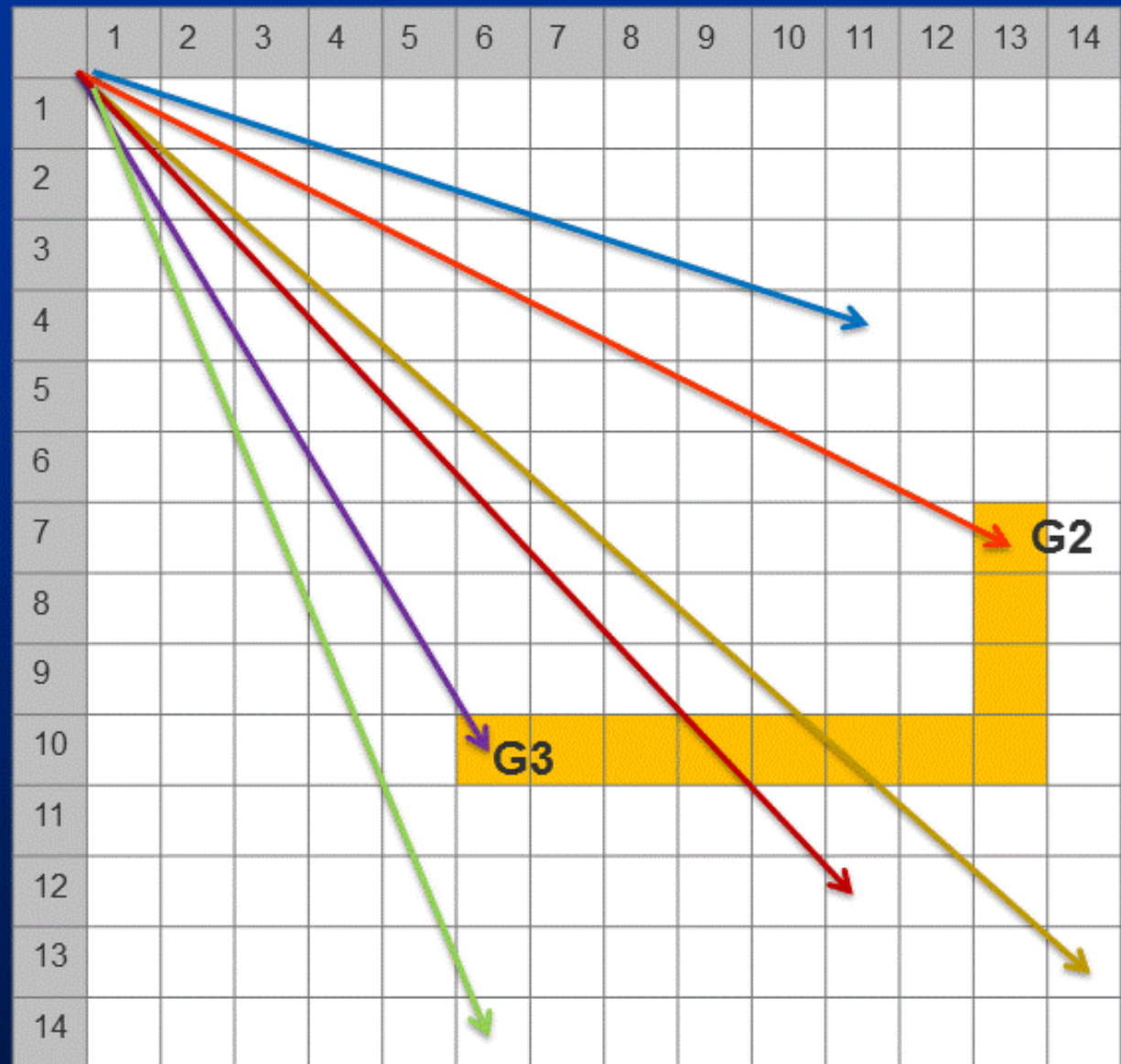


Clustering: Compute all other distances

$$G2 \Leftrightarrow G3 = |13-6| + |10-7| \\ = 7 + 3 = 10$$

$$G3 \Leftrightarrow G2 = G2 \Leftrightarrow G3 = 10$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10			
G3	11	10	0			
G4	12			0		
G5	8				0	
G6	15					0

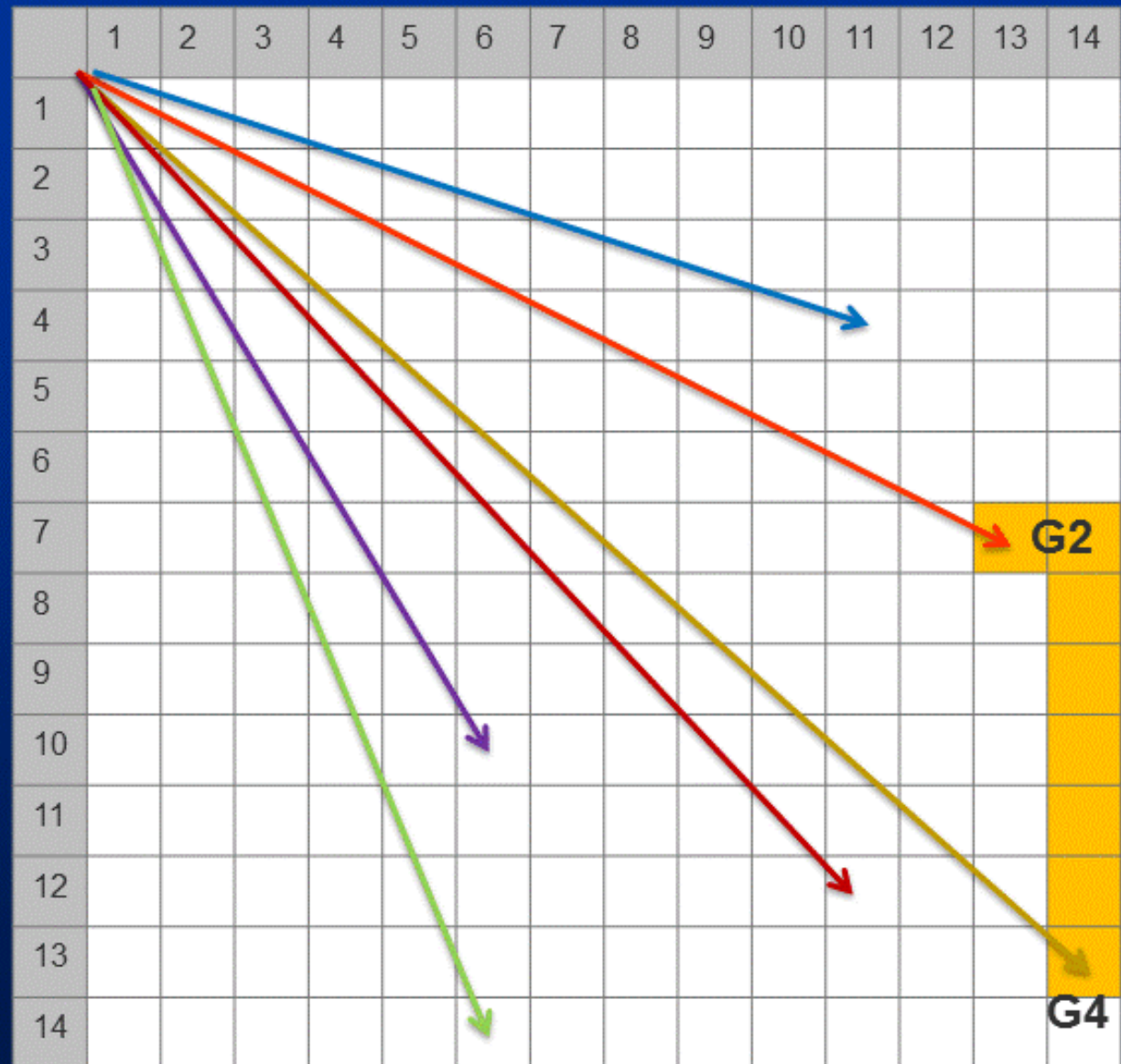


Clustering: Compute all other distances

$$G2 \Leftrightarrow G4 = |14-13| + |13-7| \\ = 1 + 6 = 7$$

$$G4 \Leftrightarrow G2 = G2 \Leftrightarrow G4 = 7$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7		
G3	11	11	0			
G4	12	7		0		
G5	8				0	
G6	15					0

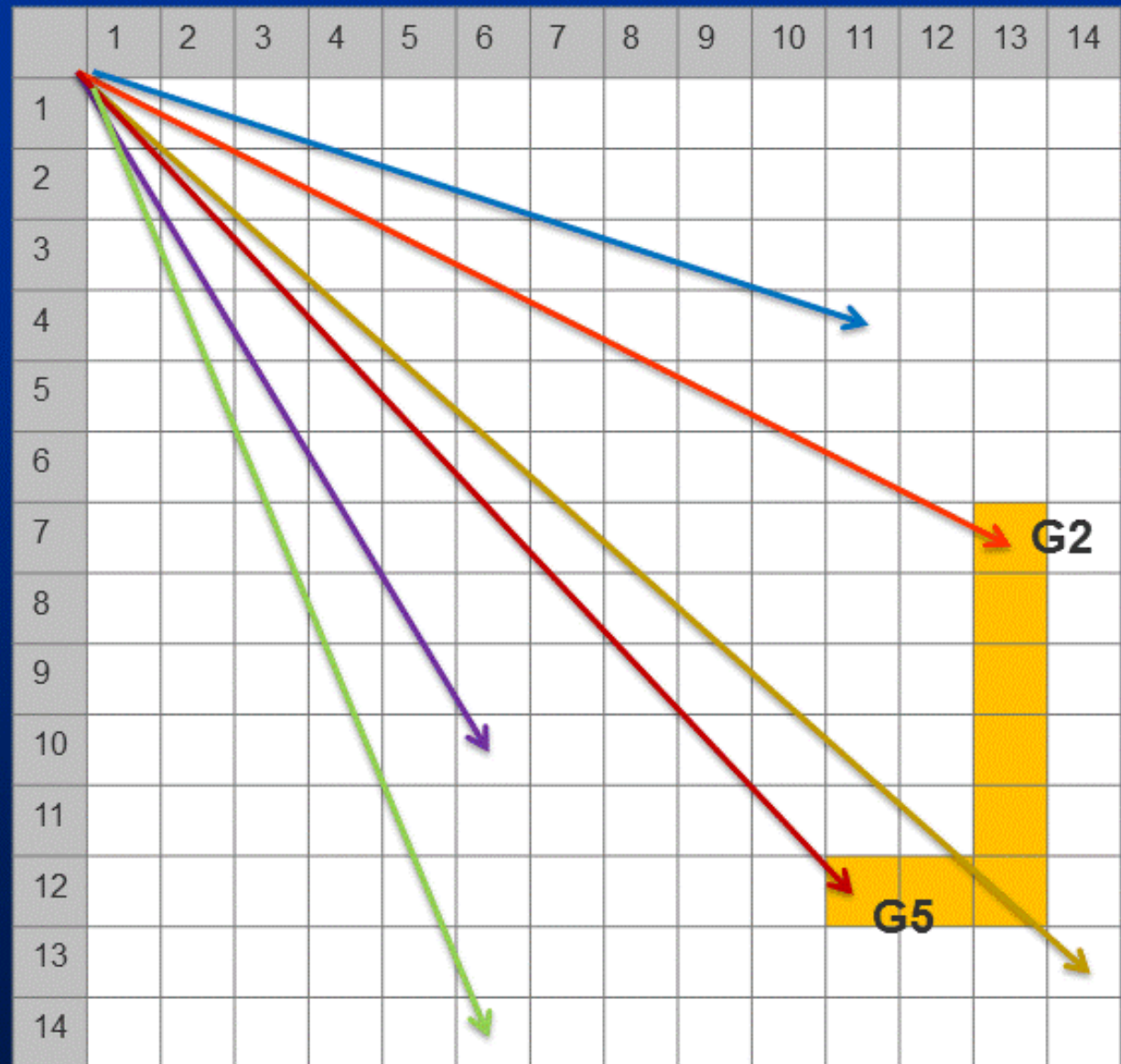


Clustering: Compute all other distances

$$G2 \leftrightarrow G5 = |13-11| + |12-7| \\ = 2 + 5 = 7$$

$$G5 \leftrightarrow G2 = G2 \leftrightarrow G5 = 7$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	
G3	11	10	0			
G4	12	7		0		
G5	8	7			0	
G6	15					0

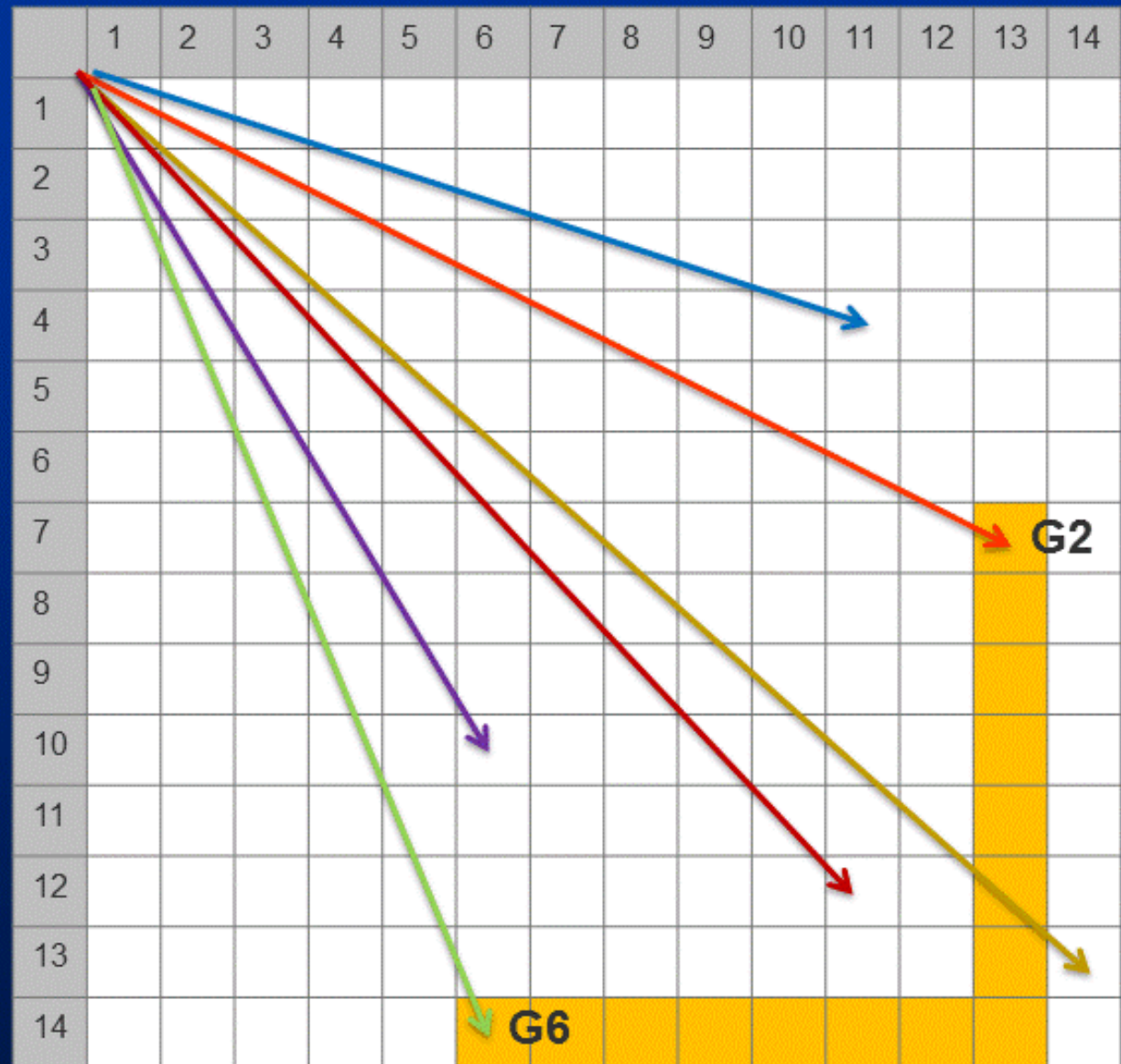


Clustering: Compute all other distances

$$G2 \leftrightarrow G6 = |13-6| + |14-7| \\ = 7 + 7 = 14$$

$$G6 \leftrightarrow G2 = G2 \leftrightarrow G6 = 14$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0			
G4	12	7		0		
G5	8	7			0	
G6	15	14				0

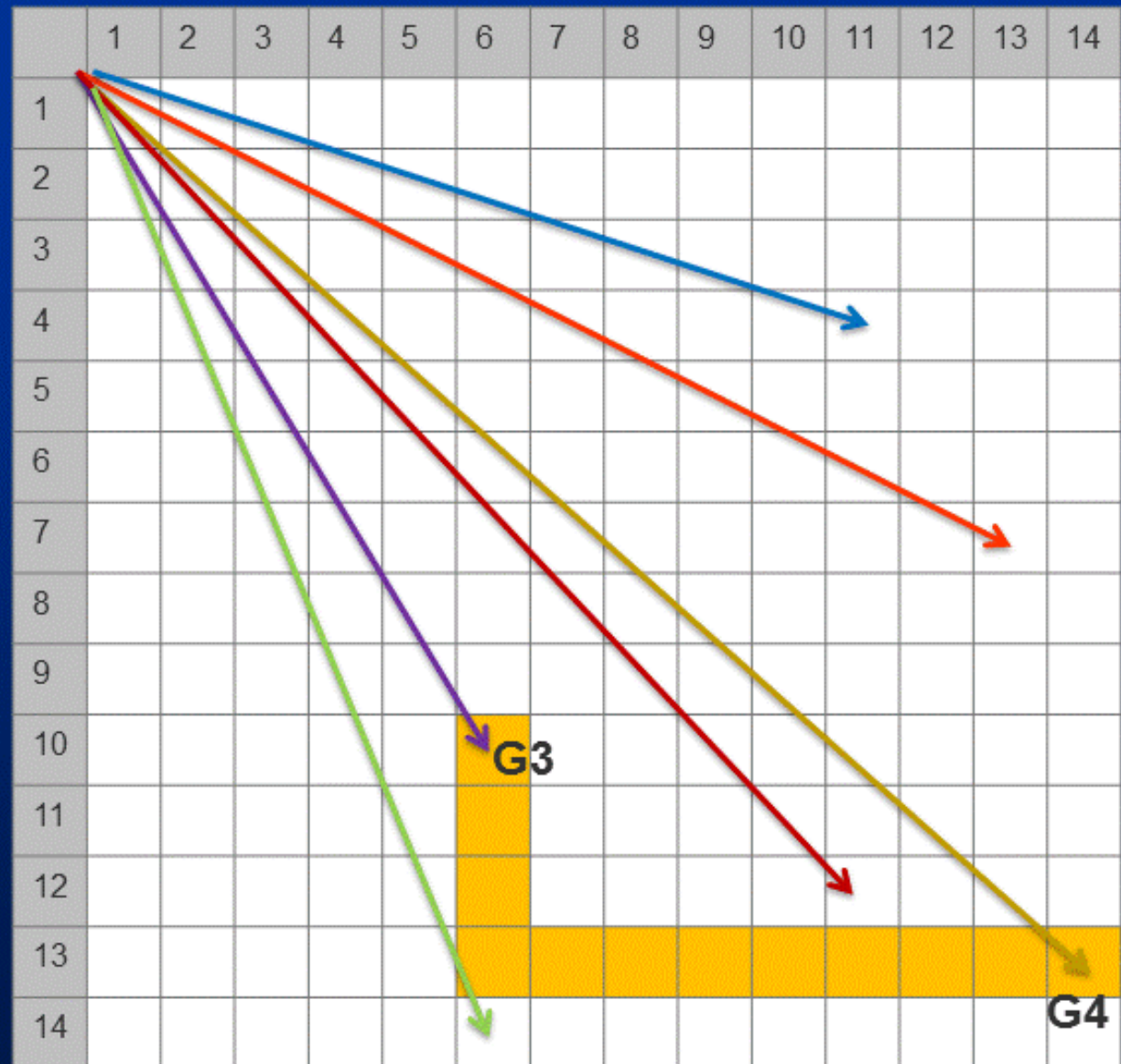


Clustering: Compute all other distances

$$G3 \leftrightarrow G4 = |14-6| + |13-10| \\ = 8 + 3 = 11$$

$$G4 \leftrightarrow G3 = G3 \leftrightarrow G4 = 11$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	18	0	11		
G4	12	7	11	0		
G5	8	7			0	
G6	14	14				0

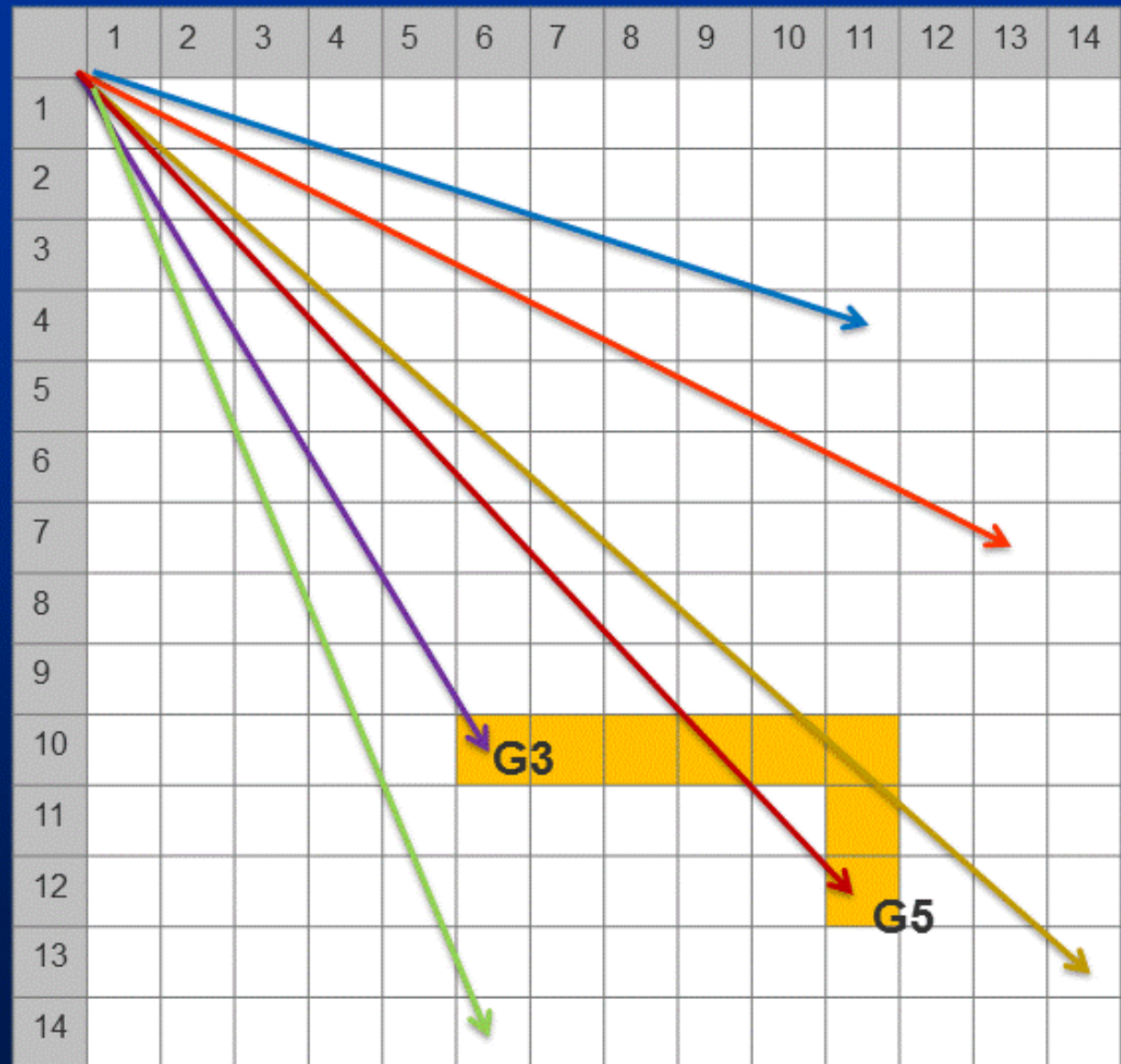


Clustering: Compute all other distances

$$G3 \leftrightarrow G5 = |11-6| + |12-10| \\ = 5 - 2 = 7$$

$$G5 \leftrightarrow G3 = G3 \leftrightarrow G5 = 7$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	11	7	
G4	12	7	11	0		
G5	8	7	7		0	
G6	15	14				0

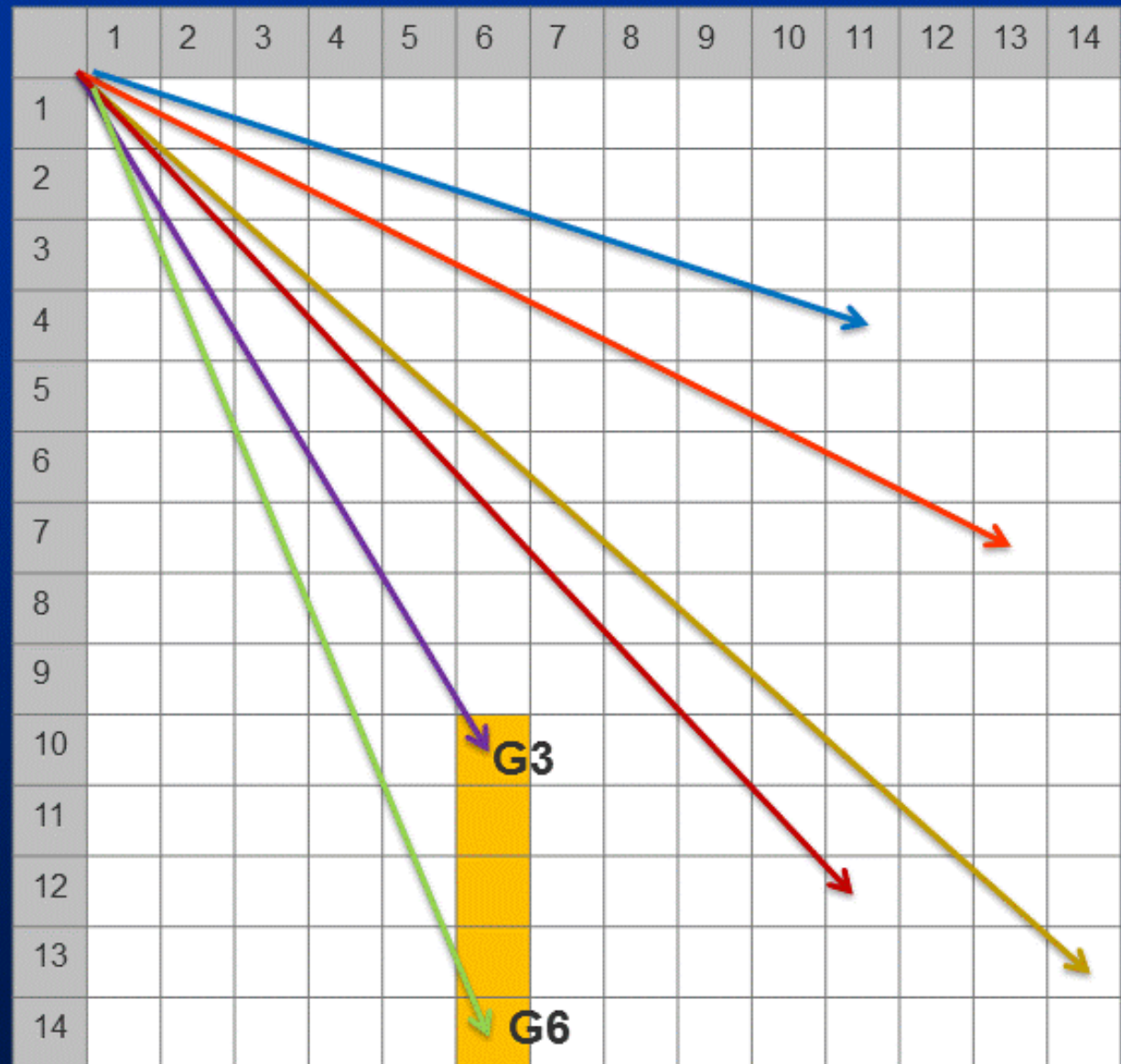


Clustering: Compute all other distances

$$G3 \leftrightarrow G6 = |6-6| + |14-10| = 0 + 4 = 4$$

$$G6 \leftrightarrow G3 = G6 \leftrightarrow G4 = 4$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	11	7	4
G4	12	7	11	0		
G5	8	7	7		0	
G6	15	14	4			0

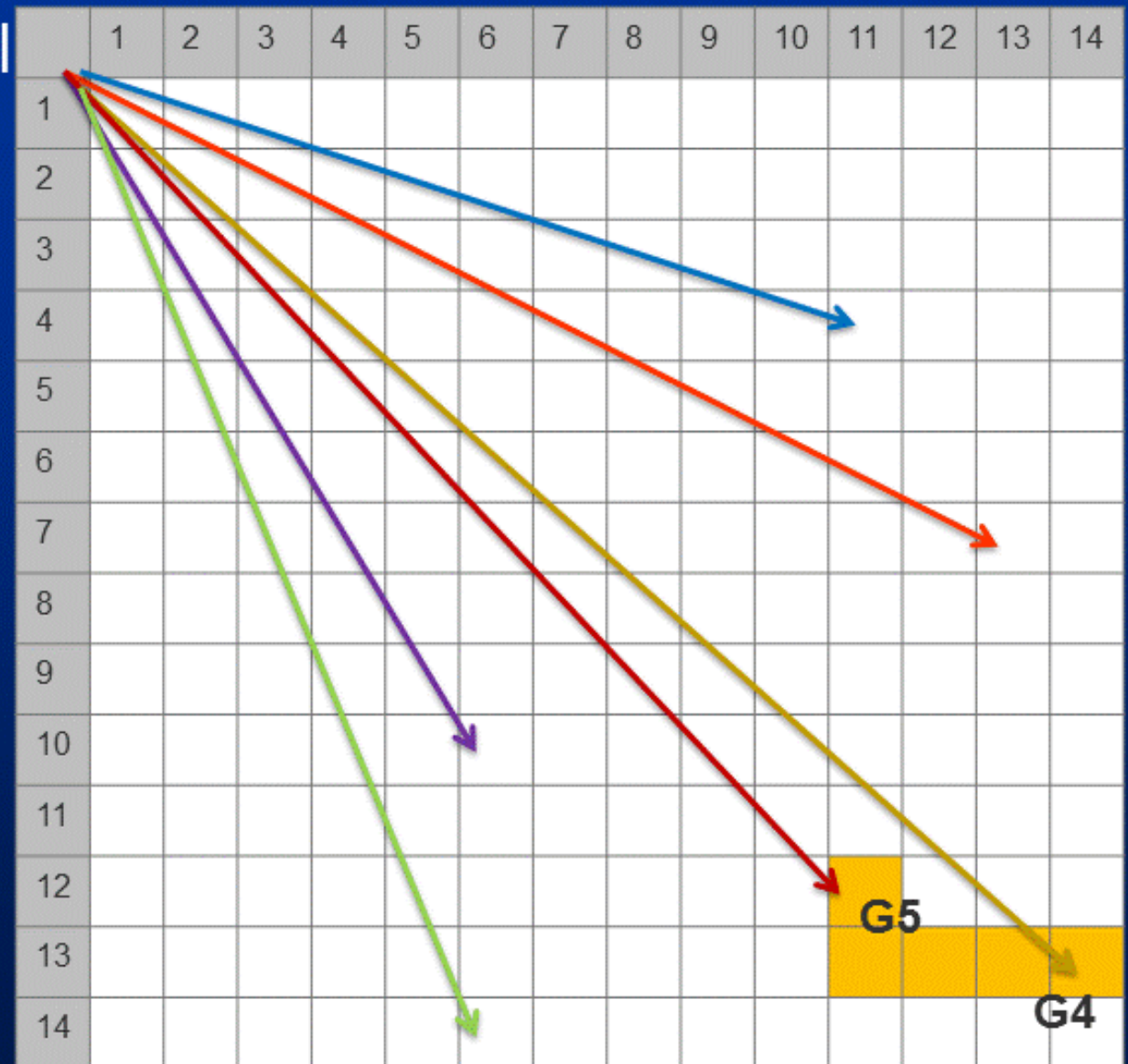


Clustering: Compute all other distances

$$G4 \leftrightarrow G5 = |14-11| + |13-12| \\ = 3 + 1 = 4$$

$$G5 \leftrightarrow G4 = G4 \leftrightarrow G5 = 4$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	10	7	4
G4	12	7	11	0	4	
G5	8	7	7	4	0	
G6	15	14	4			0

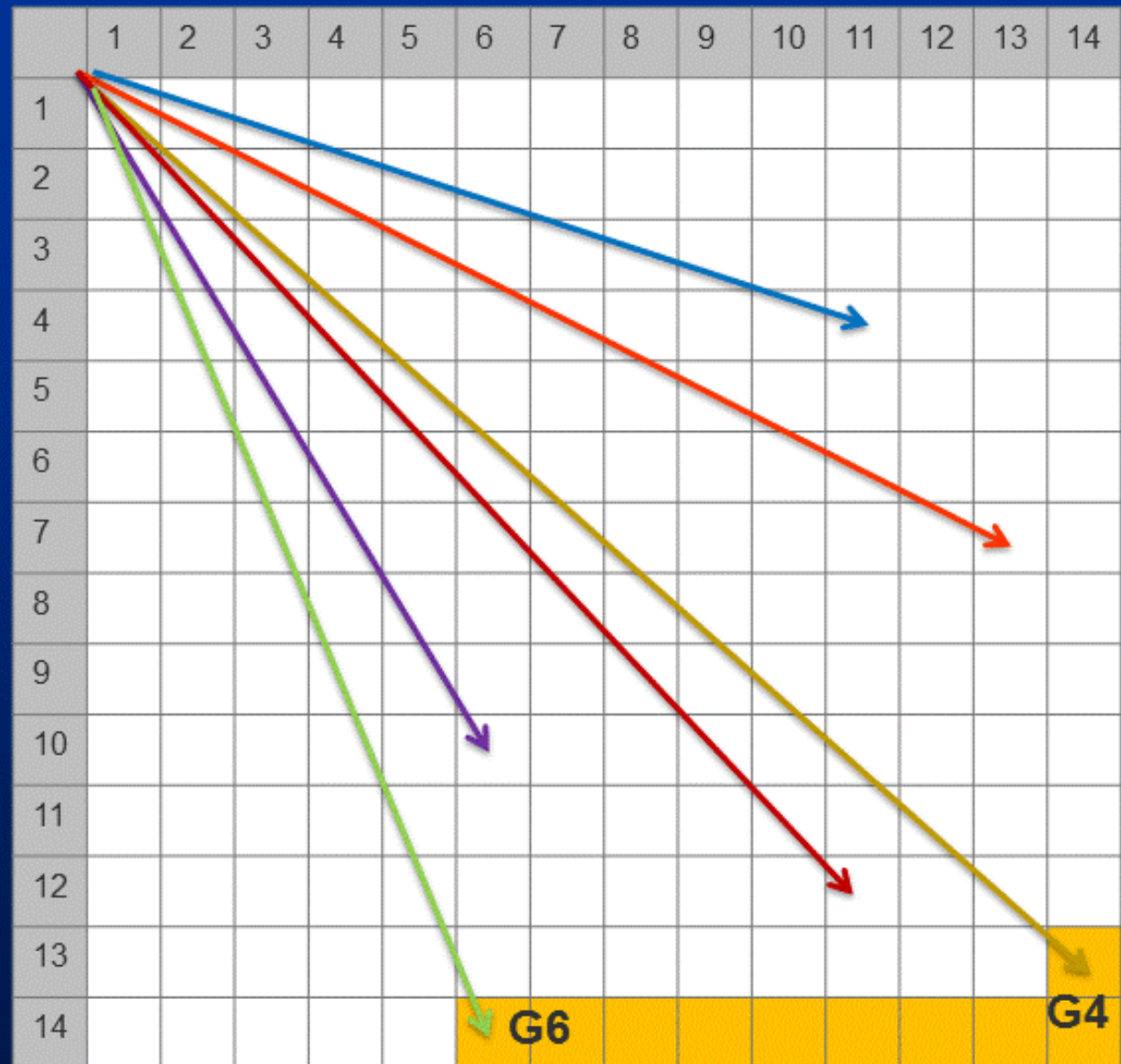


Clustering: Compute all other distances

$$G4 \leftrightarrow G6 = |14-6| + |14-13| \\ = 8 + 1 = 9$$

$$G6 \leftrightarrow G4 = G4 \leftrightarrow G6 = 9$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	11	7	5
G4	12	7	11	0	4	9
G5	8	7	7	4	0	
G6	15	14	4	9		0

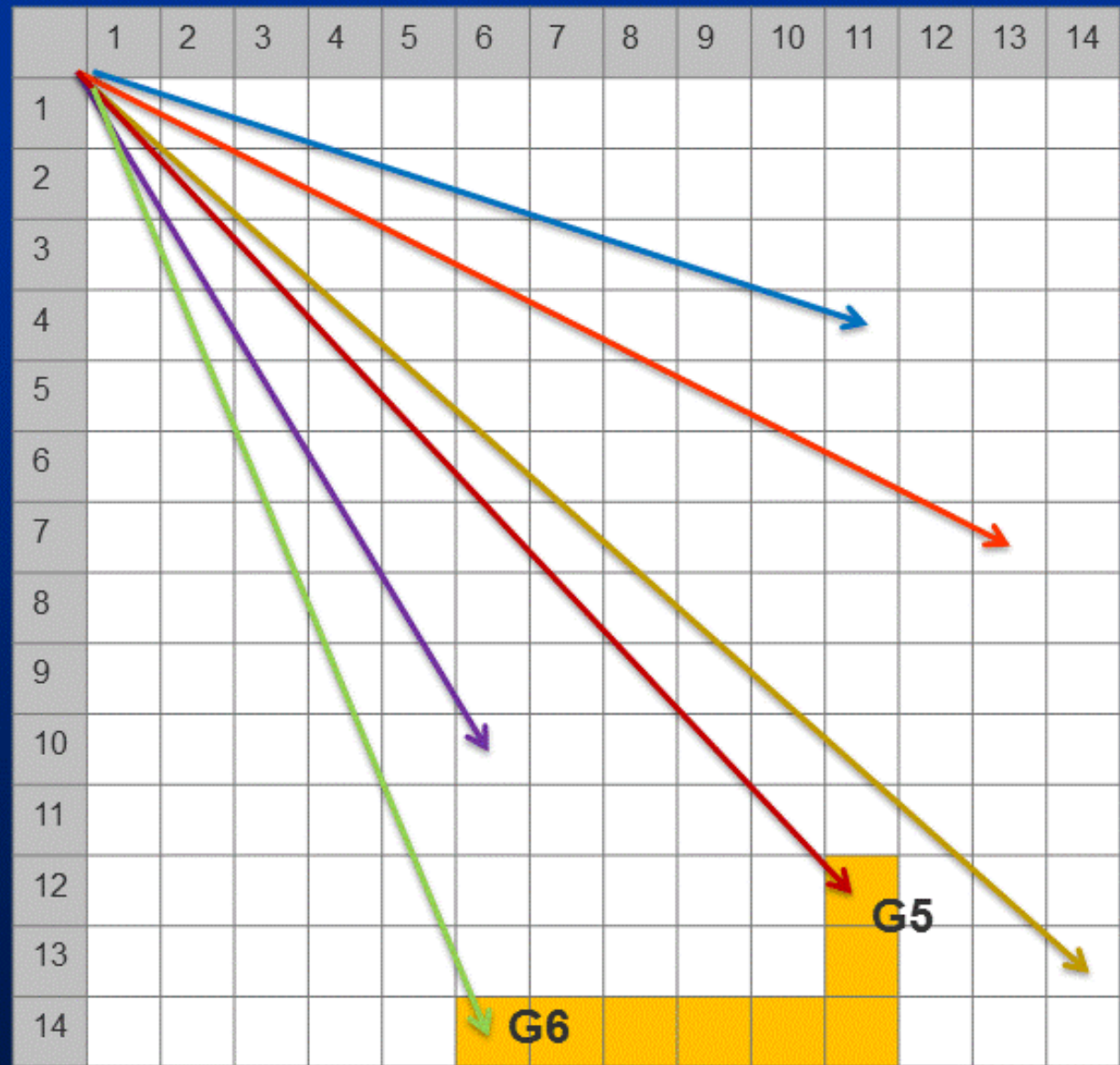


Clustering: Compute all other distances

$$G5 \leftrightarrow G6 = |11-6| + |14-12| \\ = 5 + 2 = 7$$

$$G6 \leftrightarrow G5 = G5 \leftrightarrow G6 = 7$$

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	11	7	4
G4	12	7	11	0	4	9
G5	8	7	7	4	0	7
G6	15	14	4	9	7	0



Clustering: Start clustering

Each gene a single unconnected node:



Clustering: Start clustering

Each gene a single unconnected node:



Search minimal distance in distance matrix.

Keep in mind:

- Matrix is symmetric (only search e.g. Upper triangular matrix)
- Main diagonal is always 0 - ignore

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	11	7	4
G4	12	7	11	0	4	9
G5	8	7	7	4	0	7
G6	15	14	4	9	7	0

Clustering: minimal distance

Two matrix elements show minimal distance:

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	11	7	4
G4	12	7	11	0	4	9
G5	8	7	7	4	0	7
G6	15	14	4	9	7	0

Place G3/G6 and G4/G5 into two independent cluster:



Clustering: Rebuild distance matrix

Compute distances for newly built cluster to rest of matrix

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	11	7	4
G4	12	7	11	0	4	9
G5	8	7	7	4	0	7
G6	15	14	4	9	7	0



**Average
linkage**

	G1	G2	G3/G6	G4/G5
G1	0	5		
G2	5	0		
G3/G6			0	
G4/G5				0



Clustering: Average linkage

Mean (Distance G1-G3, G1-G6) into new matrix:
 Mean (11,15) = 13

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	11	7	4
G4	12	7	11	0	4	9
G5	8	7	7	4	0	7
G6	15	14	4	9	7	0



	G1	G2	G3/G6	G4/G5
G1	0	5	13	
G2	5	0		
G3/G6	13		0	
G4/G5				0



Clustering: Average linkage

Mean (G1-G4, G1-G5) into new matrix:
 Mean (12,8) = 10

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	11	7	4
G4	12	7	11	0	4	9
G5	8	7	7	4	0	7
G6	15	14	4	9	7	0



	G1	G2	G3/G6	G4/G5
G1	0	5	13	10
G2	5	0		
G3/G6	13		0	
G4/G5	10			0



Clustering: Average linkage

Mean (G2-G6, G2-G6) into new matrix:
Mean (10,14) = 12

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	11	7	4
G4	12	7	11	0	4	9
G5	8	7	7	4	0	7
G6	15	14	4	9	7	0



	G1	G2	G3/G6	G4/G5
G1	0	6	14	11
G2	6	0	12	
G3/G6	14	13	0	
G4/G5	12			0



Clustering: Average linkage

Mean (G2-G4, G2-G5) into new matrix:

Mean (7,7) = 7

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	11	7	4
G4	12	7	11	0	4	9
G5	8	7	7	4	0	7
G6	15	14	4	9	7	0



	G1	G2	G3/G6	G4/G5
G1	0	5	13	10
G2	5	0	12	7
G3/G6	13	12	0	
G4/G5	10	7		0



Clustering: Average linkage

Mean (G3-G4, G3-G5, G6-G4, G6,G5) into new matrix:
Mean (11,7,4,9) ~ 8

	G1	G2	G3	G4	G5	G6
G1	0	5	11	12	8	15
G2	5	0	10	7	7	14
G3	11	10	0	11	7	4
G4	12	7	11	0	4	9
G5	8	7	7	4	0	7
G6	15	14	4	9	7	0



	G1	G2	G3/G6	G4/G5
G1	0	5	13	10
G2	5	0	12	7
G3/G6	13	12	0	8
G4/G5	10	7	8	0



Clustering: 2. round - minimal distance

One matrix elements show minimal distance:

	G1	G2	G3/G6	G4/G5
G1	0	5	13	10
G2	5	0	12	7
G3/G6	13	12	0	8
G4/G5	10	7	8	0

Place G1/G2 into independent cluster:



Clustering: Recompute distance matrix

	G1	G2	G3/G6	G4/G5
G1		5	13	10
G2			12	7
G3/G6				8
G4/G5				



	G1/G2	G3/G6	G4/G5
G1/G2			
G3/G6			
G4/G5			



Clustering: Recompute distance matrix

Mean (G1-G3, G1-G6, G2-G3, G2-G6) into new matrix:

Mean (11,15,10,14) ~ 13

	G1	G2	G3	G4	G5	G6
G1		5	11	12	8	15
G2			10	7	7	14
G3				11	7	4
G4					4	9
G5						7
G6						



	G1/G2	G3/G6	G4/G5
G1/G2		13	
G3/G6			
G4/G5			



Clustering: Recompute distance matrix

Mean (G1-G4, G1-G5, G2-G4, G2-G5) into new matrix:

Mean (12,8,7,7) ~ 9

	G1	G2	G3	G4	G5	G6
G1		5	11	12	8	15
G2			10	7	7	14
G3				11	7	4
G4					4	9
G5						7
G6						



	G1/G2	G3/G6	G4/G5
G1/G2		13	9
G3/G6			
G4/G5			



Clustering: Recompute distance matrix

Mean (G3-G4, G3-G5, G6-G4, G6-G5) into new matrix:

Mean (11,7,9,7) ~ 9

	G1	G2	G3	G4	G5	G6
G1		5	11	12	8	15
G2			10	7	7	14
G3				11	7	4
G4					4	9
G5						7
G6						



	G1/G2	G3/G6	G4/G5
G1/G2		13	9
G3/G6			9
G4/G5			



Clustering: 3. round - minimal distance

One matrix element shows minimal distance:

	G1/G2	G3/G6	G4/G5
G1/G2		13	9
G3/G6			9
G4/G5			

Join cluster **G1/G2** - **G4/G5** :



Clustering: 4. round - minimal distance

After several rounds of clustering we can just join the remaining two.



(If needed we can compute the distance between the two last clusters)

Clustering: Final result

