

Markov Clustering

- Overview of algorithm
- Input/Output format
- Clustering example
- Code
- Cluster formation
- Dataset size – Time variation

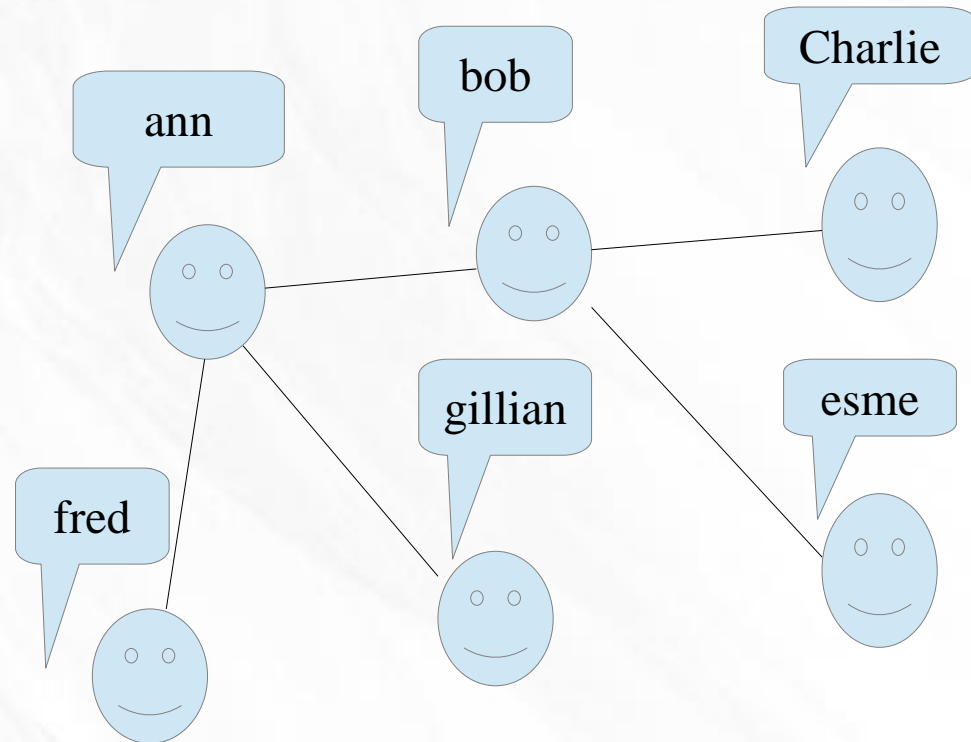
MCL Algorithm

- Input is an undirected graph, power parameter e , and inflation parameter r .
- Create the adjacency matrix.
- Normalize it.
- Expand by taking the e th power of the matrix.
- Inflate by taking inflation of the resulting matrix with parameter r .
- Repeat steps 3, 4 and 5 until a steady state is reached (convergence).
- Interpret resulting matrix to discover clusters.

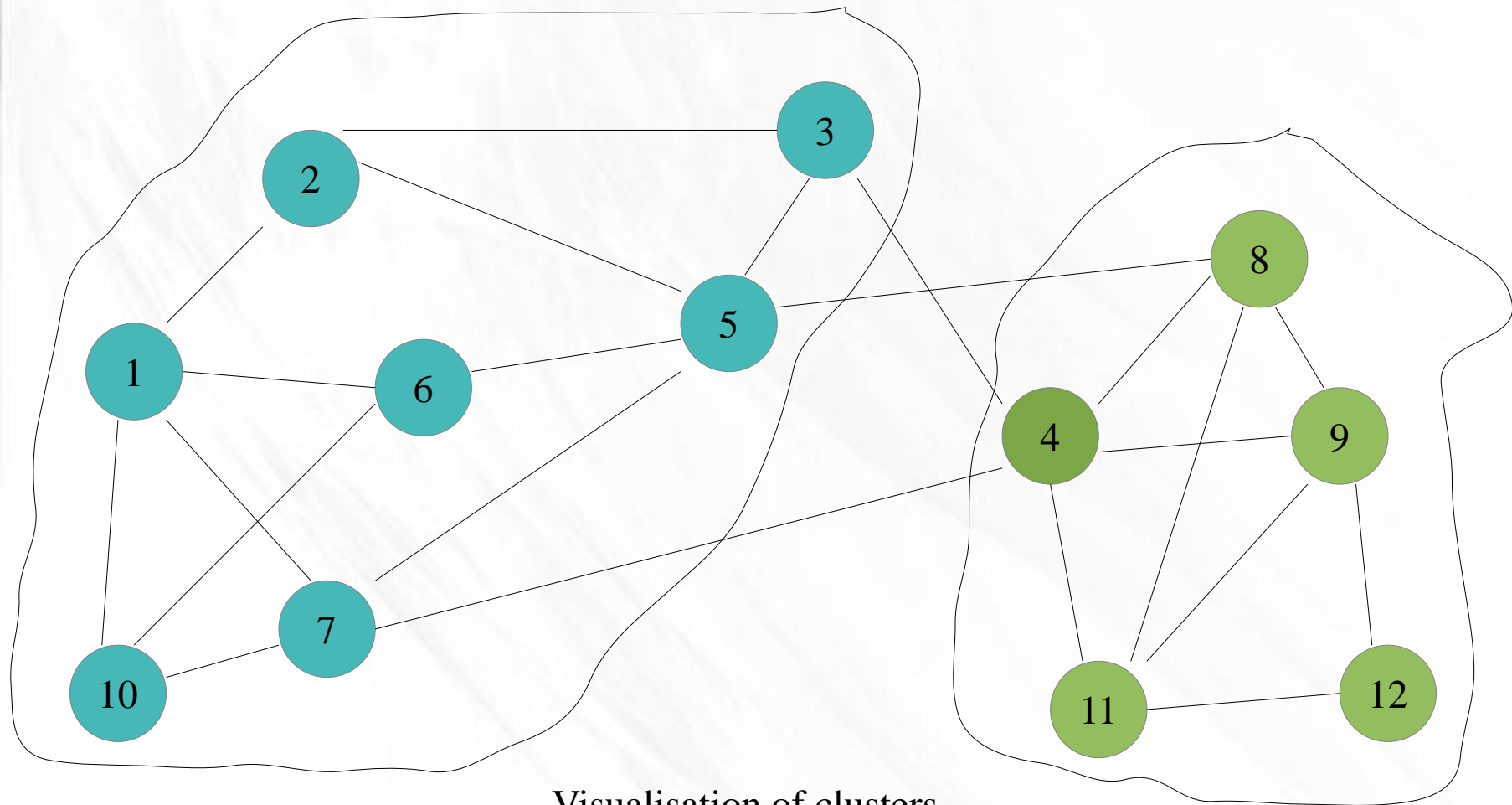
Clustering Example

Data below encodes an acquaintance network

ann	bob
ann	fred
ann	gillian
ann	john
bob	ann
bob	charlie
bob	esme
...	...
...	...
...	...



- Number of people in the dataset – 12
- Each person assigned a unique number from 1 to 12.
- ann – 1, bob – 2 , charlie – 3



Visualisation of clusters

Input / Output Format

•For each value in the adjacency matrix, there is a corresponding line of the form:

•row column 1

•Rows and columns are the numbers or id's assigned to the people and value

•1 is to show an edge between two people.

•Input file -
file -

Output

.1 2 1.000000
0.086955565965730444

[1, 1]

.1 6 1.000000
0.1266169889447161

[11, 1]

.1 7 1.000000
0.08693703269488047

[12, 1]

.1 10 1.000000
0.13568909675430582

[3, 1]

.2 1 1.000000
0.12476118936350429

[4, 1]

.2 3 1.000000
0.12478359398796567

[8, 1]

Interpreting Clusters

•For pruning threshold value – 0.065 and number of iterations – 10, the following results were obtained.

Row Id	Column Id's				
.1	9	11,	12,	4,	8,
.2	9	11,	12,	4,	8,
.3	9	11,	12,	4,	8,
.4	5,	1, 6,	10, 7	2,	3,
.5	5,	1, 6,	10, 7	2,	3,
.6	9	11,	12,	4,	8,
.7	9	11,	12,	4,	8,

Time Comparison

- For data size of 160 MB
- Num Iterations – 1
- Graph size – 105938
- Time taken :
 - Locally – 0:08:28.406041
 - On hadoop – 0.09:35