



Social Networks Analysis

Maciej Medyk

What is Social Network?

- A social structure made of individuals and organizations represented as nodes
- Each are connected by one or more specific types of interdependency such as friendship, beliefs, knowledge, or prestige
- Has a structure of undirected graph
- Analysis often evaluates density, network diameter, betweenness centrality, closeness centrality, edge density, clustering coefficient of nodes and network, and degree distribution of the network.

Social Network Graph Representation

- Social Network is made of interconnected nodes
- It can be stored as adjacency list or adjacency matrix
- Network connection can be measured by network diameter that shows how far nodes are from each other
- Network can be loosely or closely connected through node degree measurements

List of nodes: $V=\{A, B, C, D, E\}$

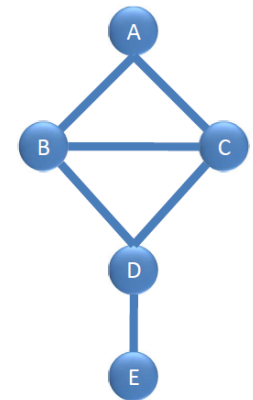
List of edges: $E=\{(A,B), (A,C), (B,C), (B,D), (C,D), (D,E)\}$

Adjacency list:

A: {B, C}; B: {A, C, D}; C: {A, B, D}; D: {B, C, E}; E: {D}

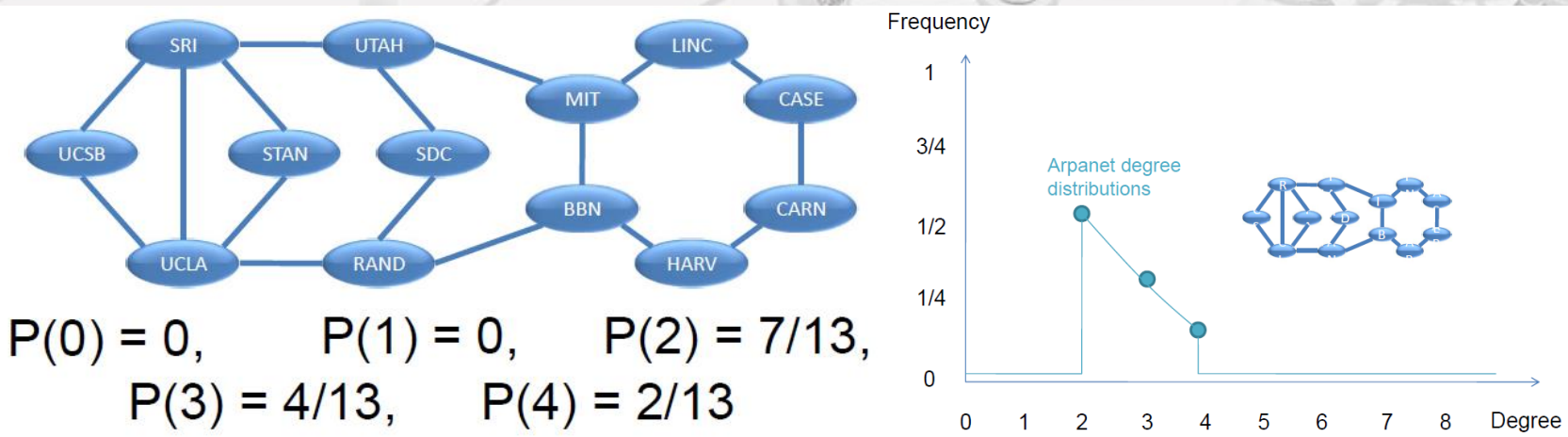
Adjacency matrix: $A_{ij} = 1$ if (i,j) is an edge, else = 0

	A	B	C	D	E
A	0	1	1	0	0
B	1	0	1	1	0
C	1	1	0	1	0
D	0	1	1	0	1
E	0	0	0	1	0



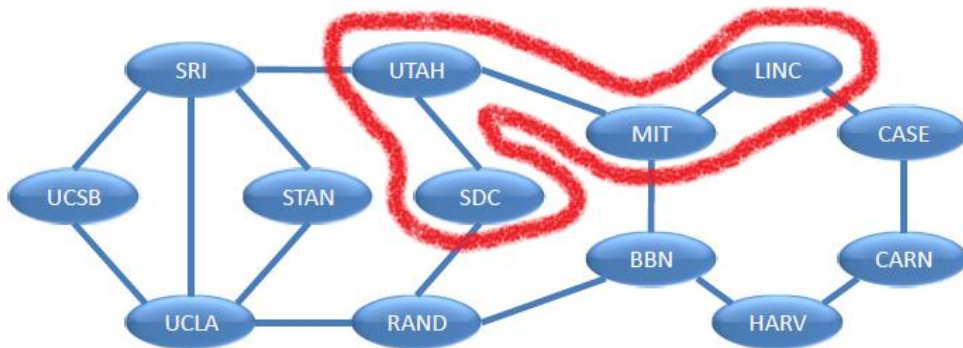
Node Edges and Degree of Distribution

- Connection between nodes is called edge
- For node to be connected it has to have at least one edge
- Usually there is small amount of nodes with high amount of edges and large amount of nodes with low amount of edges

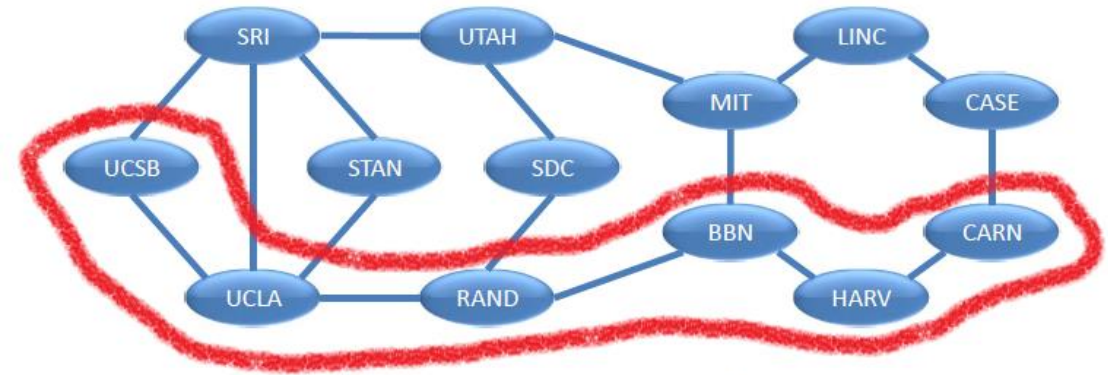


Shortest Path and Network Diameter

- Shortest path is the least amount of hops between two nodes
- Network diameter is the biggest shortest path which identifies how closely the network is connected



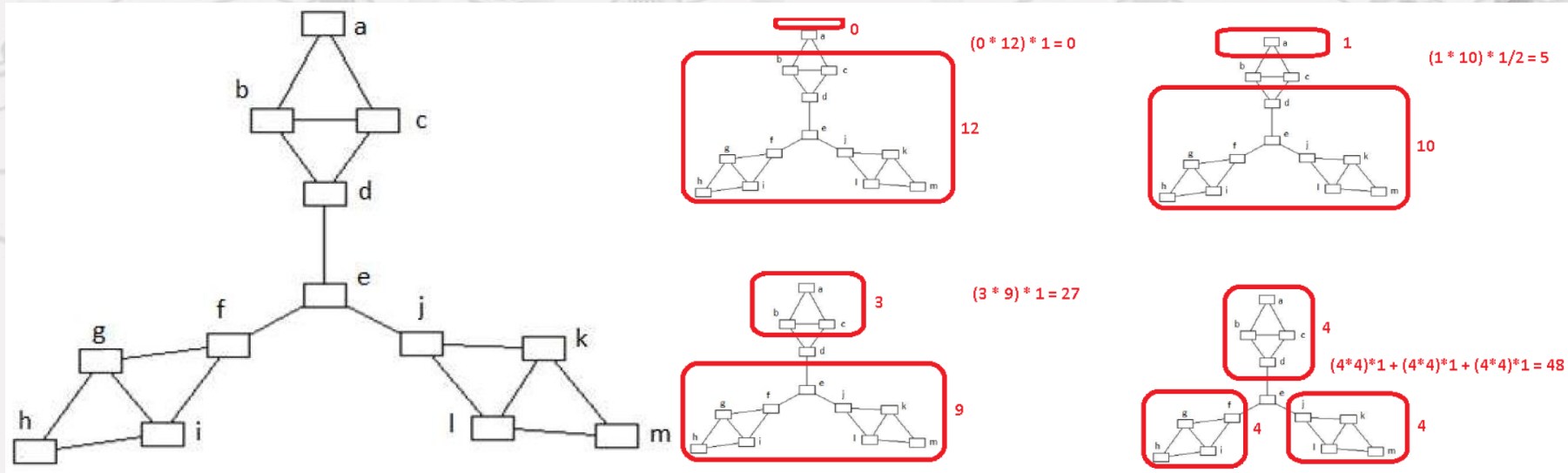
Distance between LINC and SDC is 3.



Diameter is 5.

Betweenness Centrality

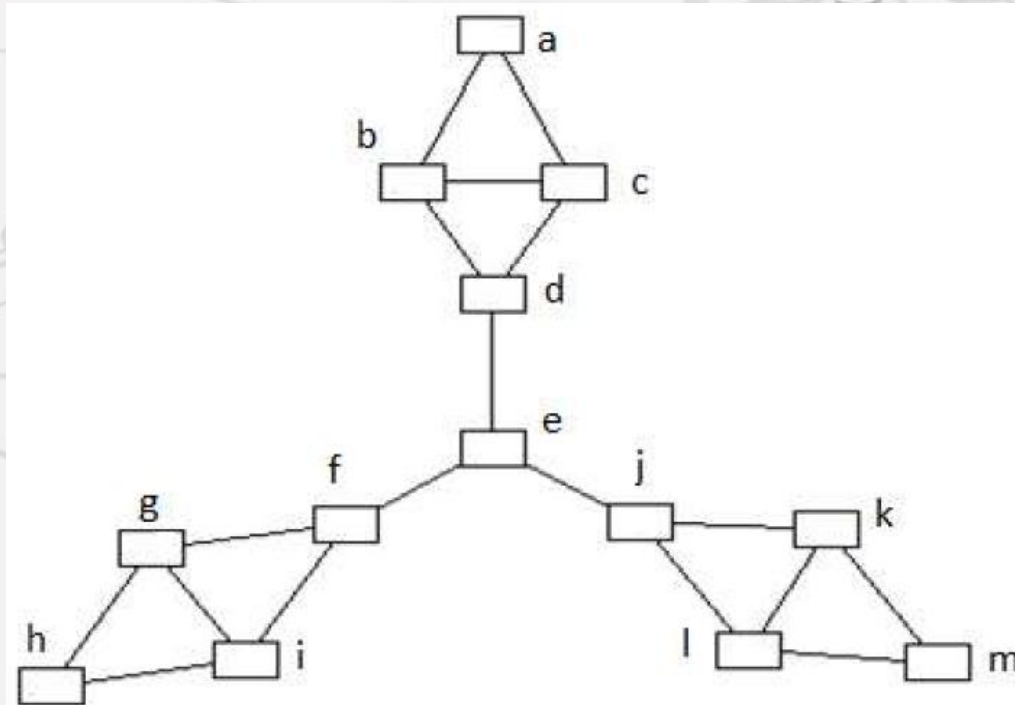
- Betweenness Centrality defines how important node is in connecting other nodes



NODE	SCORE	NORMAL
A	0	0.000
B	5	0.076
C	5	0.076
D	27	0.409
E	48	0.727
F	27	0.409
G	5	0.076
H	0	0.000
I	5	0.076
J	27	0.409
K	5	0.076
L	5	0.076
M	0	0.000

Closeness Centrality

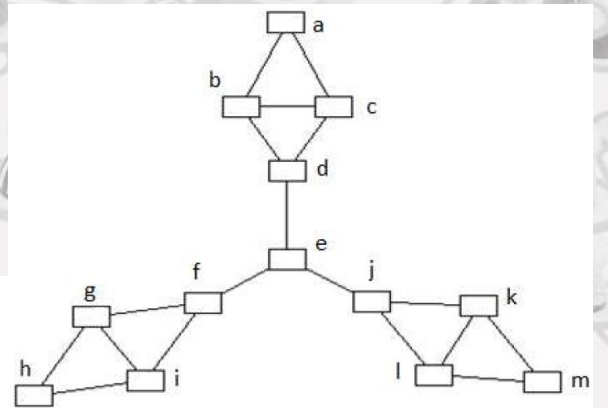
- Closeness Centrality defines how easy it can reach other nodes in the network



	A	B	C	D	E	F	G	H	I	J	K	L	M	
A	0	1	1	2	3	4	5	6	5	4	5	5	6	
B	1	0	1	1	2	3	4	5	4	3	4	4	5	
C	1	1	0	1	2	3	4	5	4	3	4	4	5	
D	2	1	1	0	1	2	3	4	3	2	3	3	4	
E	3	2	2	1	0	1	2	3	2	1	2	2	3	
F	4	3	3	2	1	0	1	2	1	2	3	3	4	
G	5	4	4	3	2	1	0	1	1	3	4	4	5	
H	6	5	5	4	3	2	1	0	1	4	5	5	6	
I	5	4	4	3	2	1	1	1	0	3	4	4	5	
J	4	3	3	2	1	2	3	4	3	0	1	1	2	
K	5	4	4	3	2	3	4	5	4	1	0	1	1	
L	5	4	4	3	2	3	4	5	4	1	1	0	1	
M	6	5	5	4	3	4	5	6	5	2	1	1	0	
SUM	47	37	37	29	24	29	37	47	37	29	37	37	47	SUM(A:M)
SCORE	3.917	3.083	3.083	2.417	2.000	2.417	3.083	3.917	3.083	2.417	3.083	3.083	3.917	SUM(A:M) / N-1
NORMAL	0.255	0.324	0.324	0.414	0.500	0.414	0.324	0.255	0.324	0.414	0.324	0.324	0.255	N-1 / SUM(A:M)

Graph Edge Density

- Edge density describes the portion of the potential connections in a network that are actual connections. It shows how much percentage of complete graph is achieved
- Graph edge density = $\# \text{ of edges} / (\# \text{ of nodes} * ((\# \text{ of nodes} - 1) / 2))$

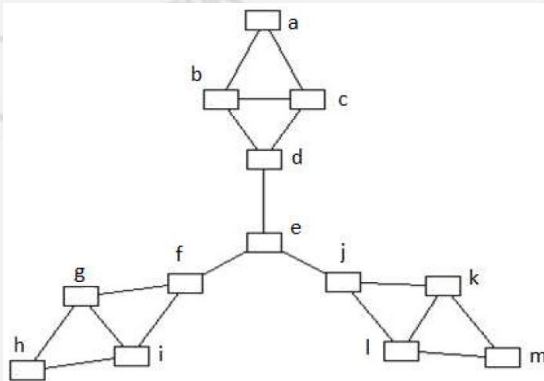


ENTIRE NETWORK	
EDGE DENSITY	$18 / (13 * ((13 - 1) / 2)) = 18 / (13 * (12 / 2)) = 18 / (13 * 6) = 18 / 78 = 0.230769$

Clustering Coefficient

- Clustering coefficient of a node is the fraction of its neighbors that are connected and create clusters (triangles)
- Clustering coefficient of the network is average clustering coefficient of all of its nodes

$$C_i = \begin{cases} \frac{k_i}{d_i \times (d_i - 1) / 2} & d_i > 1 \\ 0 & d_i = 0 \text{ or } 1 \end{cases}$$



CLUSTERING COEFFICIENT OF NODES													
	A	B	C	D	E	F	G	H	I	J	K	L	M
DISTRIBUTION	2	3	3	3	3	3	3	2	3	3	3	3	2
CLUSTERS	1	2	2	1	0	1	2	1	2	1	2	2	1
(DISTRIBUTION - 1) / 2	0.5	1	1	1	1	1	1	0.5	1	1	1	1	0.5
COEFFICIENT	1	0.6667	0.6667	0.3333	0	0.3333	0.6667	1	0.6667	0.3333	0.6667	0.6667	1

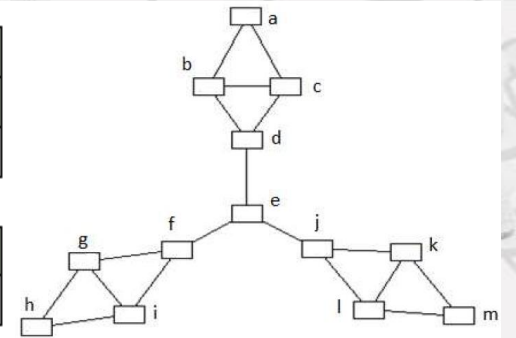
CLUSTERING COEFFICIENT OF NODES	
COEFFICIENT	(1 + 0.6667 + 0.6667 + 0.3333 + 0.3333 + 0.6667 + 1 + 0.6667 + 0.3333 + 0.6667 + 0.6667 + 1) / 13 = 0.615385

Degree Distribution of Network

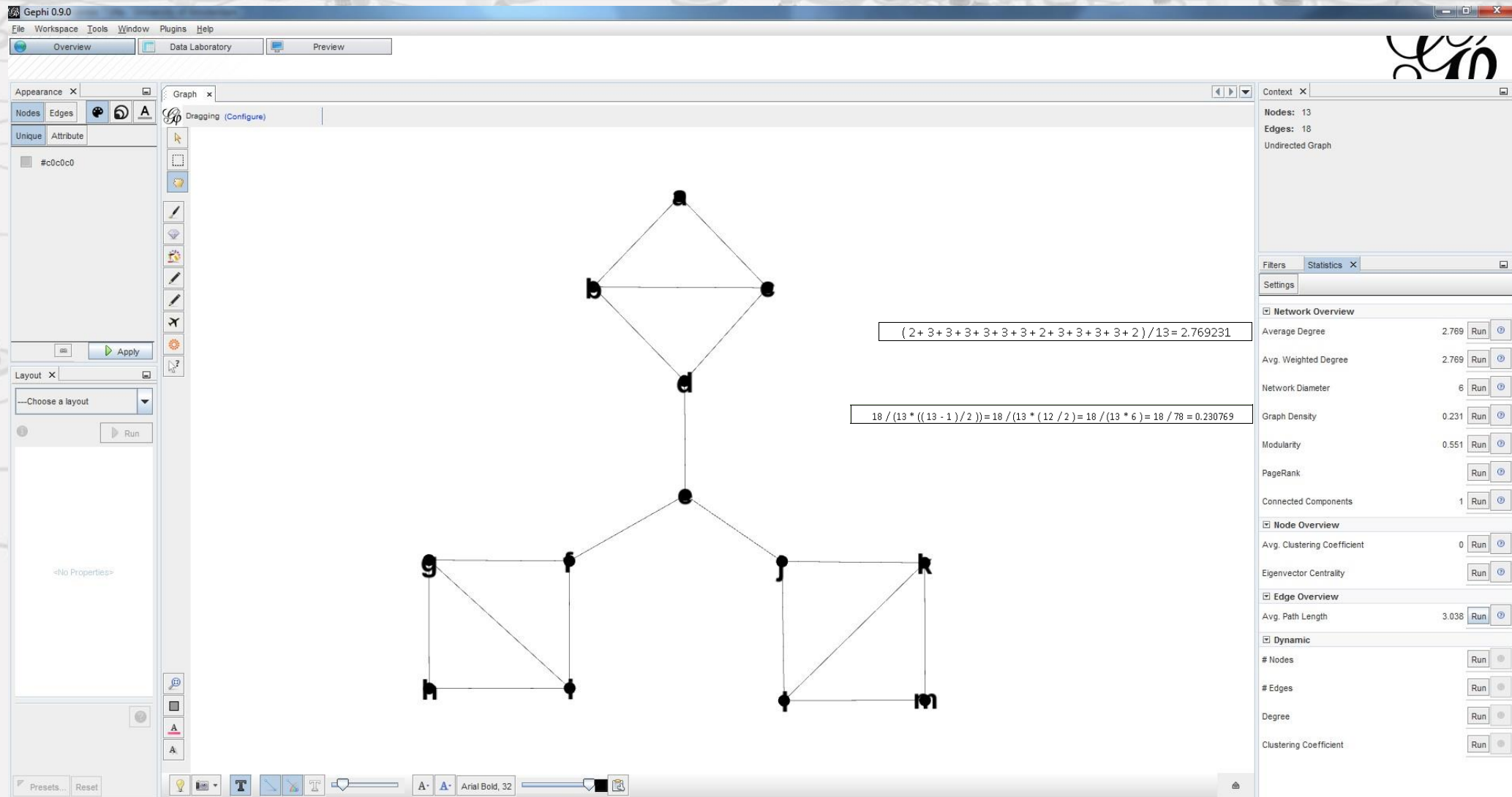
- Degree of a node is the number of the connection it has to other nodes
- Degree distribution is a probability of distribution of these degrees over entire network

INDIVIDUAL NODES													
	A	B	C	D	E	F	G	H	I	J	K	L	M
DISTRIBUTION	2	3	3	3	3	3	3	2	3	3	3	3	2

ENTIRE NETWORK	
DISTRIBUTION	$(2 + 3 + 3 + 3 + 3 + 3 + 3 + 2 + 3 + 3 + 3 + 3 + 2) / 13 = 2.769231$



Gephi Social Network Analysis Software 1/4



Gephi Social Network Analysis Software 2/4

Gephi 0.9.0

File Workspace Tools Window Plugins Help

Overview Data Laboratory Preview

Data Table

Nodes Edges Configuration Add node Add edge Search/Replace Import Spreadsheet Export table More actions

Filter: Id

Id	Label	Timestamp	Degree	Weighted Degree	Eccentricity	Closeness Centrality	Betweenness Centrality	Modularity Class	PageRank	Component ID	Clustering Coefficient	Number of triangles	Eigenvector Centrality
0	a		2	2.0	6.0	3.916667	0.0	0	0.069592	0	0.0	0	0.424
1	b		3	3.0	5.0	3.083333	5.0	0	0.068355	0	0.0	0	0.51221
2	c		3	3.0	5.0	3.083333	5.0	0	0.068355	0	0.0	0	0.51221
3	d		3	3.0	4.0	2.416667	27.0	0	0.09597	0	0.0	0	0.823579
4	e		3	3.0	3.0	2.0	48.0	2	0.093185	0	0.0	0	1.0
5	f		3	3.0	4.0	2.416667	27.0	1	0.09597	0	0.0	0	0.823579
6	g		3	3.0	5.0	3.083333	5.0	1	0.068355	0	0.0	0	0.51221
7	h		2	2.0	6.0	3.916667	0.0	1	0.069592	0	0.0	0	0.424
8	i		3	3.0	5.0	3.083333	5.0	1	0.068355	0	0.0	0	0.51221
9	j		3	3.0	4.0	2.416667	27.0	2	0.09597	0	0.0	0	0.823579
10	k		3	3.0	5.0	3.083333	5.0	2	0.068355	0	0.0	0	0.51221
11	l		3	3.0	5.0	3.083333	5.0	2	0.068355	0	0.0	0	0.51221
12	m		2	2.0	6.0	3.916667	0.0	2	0.069592	0	0.0	0	0.424

SUM	47	37	37	29	24	29	37	47	37	29	37	37	47	SUM(A:M)
SCORE	3.917	3.083	3.083	2.417	2.000	2.417	3.083	3.917	3.083	2.417	3.083	3.083	3.917	SUM(A:M) / N-1
NORMAL	0.255	0.324	0.324	0.414	0.500	0.414	0.324	0.255	0.324	0.414	0.324	0.324	0.255	N-1 / SUM(A:M)

NODE	SCORE	NORMAL
A	0	0.000
B	5	0.076
C	5	0.076
D	27	0.409
E	48	0.727
F	27	0.409
G	5	0.076
H	0	0.000
I	5	0.076
J	27	0.409
K	5	0.076
L	5	0.076
M	0	0.000

Add column Merge columns Delete column Clear column Copy data to other column Fill column with a value Duplicate column

Create a boolean column from regex match Create column with list of regex matching groups Negate boolean values Convert column to dynamic

Gephi Social Network Analysis Software 3/4

Gephi 0.9.0

File Workspace Tools Window Plugins Help

Overview Data Laboratory Preview

Data Table x

Nodes Edges Configuration Add node Add edge Search/Replace Import Spreadsheet Export table More actions Filter: Id

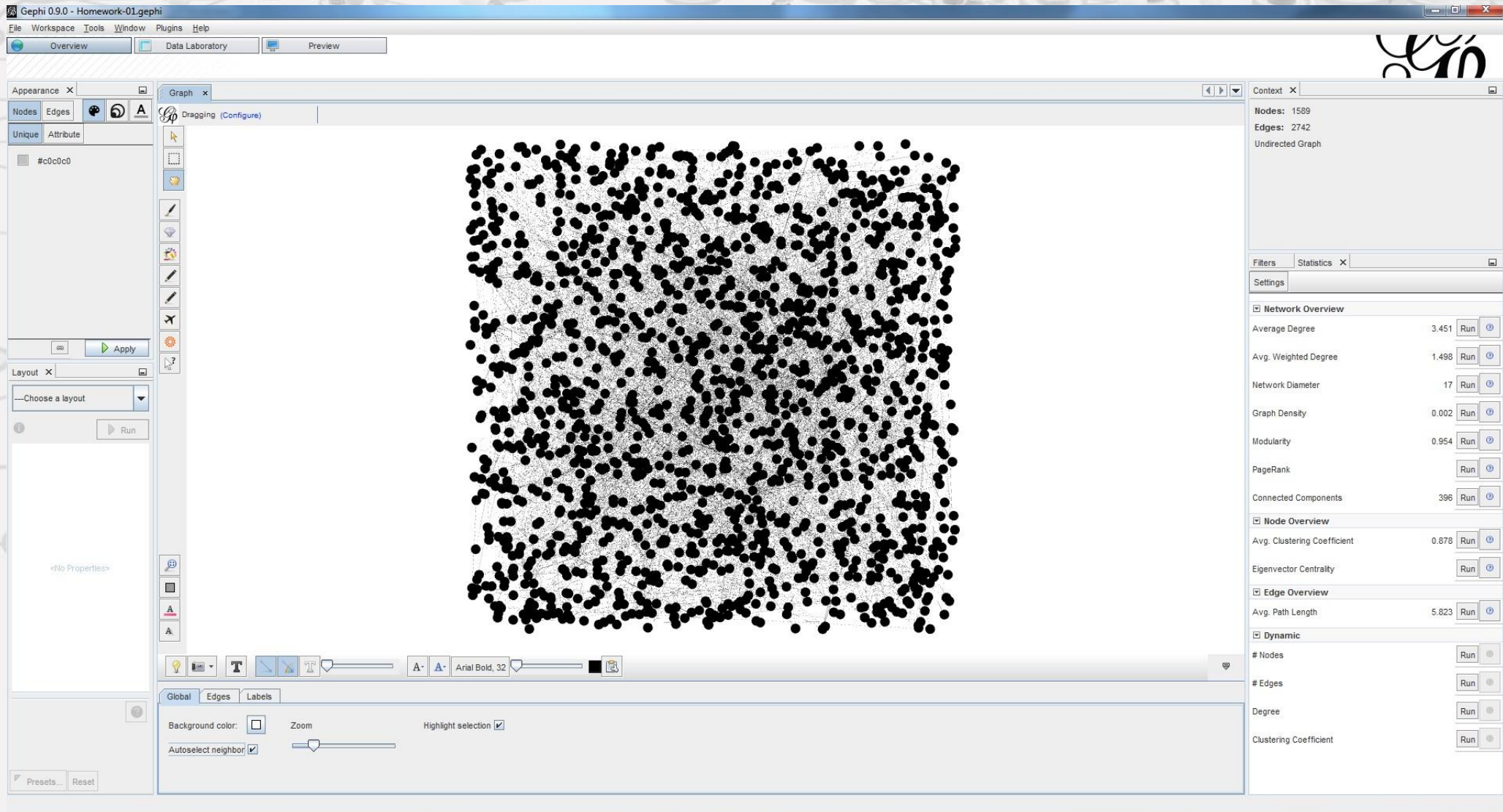
Id	Label	Timestamp	Degree	Weighted Degree	Eccentricity	Closeness Centrality	Betweenness Centrality	Modularity Class	PageRank	Component ID	Clustering Coefficient	Number of triangles	Eigenvector Centrality
0	a		2	2.0	6.0	0.255319	0.0	0	0.069592	0	0.0	0	0.424
1	b		3	3.0	5.0	0.324324	0.075758	0	0.068355	0	0.0	0	0.51221
2	c		3	3.0	5.0	0.324324	0.075758	0	0.068355	0	0.0	0	0.51221
3	d		3	3.0	4.0	0.413793	0.409091	0	0.09597	0	0.0	0	0.823579
4	e		3	3.0	3.0	0.5	0.727273	2	0.093185	0	0.0	0	1.0
5	f		3	3.0	4.0	0.413793	0.409091	1	0.09597	0	0.0	0	0.823579
6	g		3	3.0	5.0	0.324324	0.075758	1	0.068355	0	0.0	0	0.51221
7	h		2	2.0	6.0	0.255319	0.0	1	0.069592	0	0.0	0	0.424
8	i		3	3.0	5.0	0.324324	0.075758	1	0.068355	0	0.0	0	0.51221
9	j		3	3.0	4.0	0.413793	0.409091	2	0.09597	0	0.0	0	0.823579
10	k		3	3.0	5.0	0.324324	0.075758	2	0.068355	0	0.0	0	0.51221
11	l		3	3.0	5.0	0.324324	0.075758	2	0.068355	0	0.0	0	0.51221
12	m		2	2.0	6.0	0.255319	0.0	2	0.069592	0	0.0	0	0.424

SUM	47	37	37	29	24	29	37	47	37	29	37	37	47	SUM(A:M)
SCORE	3.917	3.083	3.083	2.417	2.000	2.417	3.083	3.917	3.083	2.417	3.083	3.083	3.917	SUM(A:M) / N-1
NORMAL	0.255	0.324	0.324	0.414	0.500	0.414	0.324	0.255	0.324	0.414	0.324	0.324	0.255	N-1 / SUM(A:M)

NODE	SCORE	NORMAL
A	0	0.000
B	5	0.076
C	5	0.076
D	27	0.409
E	48	0.727
F	27	0.409
G	5	0.076
H	0	0.000
I	5	0.076
J	27	0.409
K	5	0.076
L	5	0.076
M	0	0.000

Add column Merge columns Delete column Clear column Copy data to other column Fill column with a value Duplicate column Create a boolean column from regex match Create column with list of regex matching groups Negate boolean values Convert column to dynamic

Gephi Social Network Analysis Software 4/4





Questions?