PART-C

11) Complexities in data collection:

1. Dynamic nature of data demands collection at uniform intervals of time.
2. Accuracy & Reliability of data collection.
3. Privacy concerns of data.
4. Takes a lot of effort - 50% of effort wasted in data collection.
5. Since the collection process is difficult, collected data is reused often.

Alternatives

1. Online e-data can be collected.
The web is vast and diverse. The information from web can be collected and used for the purpose of SNA.
For eg: Information Dynamics Lab at HP investigated the presence of ties based off of corporate archive of mail. If enough mails were sent, a tie was assumed.

However, data privacy issues are a major concern. ~~According to HP~~ since private, sensitive data cannot be easily shared.

Instead, data from ~~mail~~ open-source archives can be used. Data Visualisation & diffusion analysis can also be performed.

2. Data not meant for SNA can also be used:
   → archival data (minute of meeting/court records, etc)
   → web-data through web-mining.

3. Blogs:
   They provide a platform for people to interact through comments and links. They're mostly used for marketing trend analysis to predict consumer choices.
   → Links are used since they're representative of relationships & are authoritative.
   → Co-occurences are also used as an indicator of ties.
   Referral Web first used web-mining for referral chaining - to find an expert of a particular niche closest to the user.
   Jaccard Coefficient: $\frac{\text{no. of cooccurences}}{\text{no. of indiv. occurences}}$
   To extract ties, we need an efficient commercial search engine.
   First, the names are extracted using NLP & further used to extract newer names [snow-balling]

4. Social media platforms.

→ Better way of social networking
→ Extra features like photo sharing, messaging, etc
→ To analyze changes in memberships communities & effect of structure
→ Guarded data with security measure.
→ Alternatively, use FOAF networkes which store user profiles on their personal blogs, connected by links
→ Less support for maintaining & extraction of info.

Other disadvantages

- Technological limits to accuracy.
- Can be costly.

13) a) Eccentricity is the longest shortest path seen by node 3

Path	length
(3,1)	2
(3,2)	1
(3,3)	0
(3,4)	2
(3,5)	3
(3,6)	4
(3,7)	5
(3,8)	4
(3,9)	3
(3,10)	4
(3,11)	4

Eccentricity - 5

b) A walk is a ~~pat~~ sequence of nodes & lines in the graph w possible repetitions

Walk - $n0, n4, n9, n11, n10, n9, n11$

c) A trial is one walks where the lines are unique and not repeated

Trial - $n_1, n_2, n_3, n0$.

d) A tour is a closed walk in which all lines occur atleast once.

n_0, n_3, n_2, n_1, n_0 ~~-~~ $- n_4 - n_5 - n_8 - n_7 - n_6 - n_5 - n_4 - n_9 - n_{11} - n_{10} - n_9 - n_4 - n_0$.

e) Nodal degree - no. of lines incident

$n0 - 3$ $n5 - 3$

$n1 - 2$ $n6 - 2$

$n2 - 2$. $n7 - 2$. $n11$

$n3 - 2$, $n8 - 2$.

$n4 - 3$ $n9 - 3$.

$n10 - 2$.

f) mean nodal degree.

$$\bar{d} = \frac{\sum_{i=0}^{g} d(i)}{g} = \frac{3+2+2+}{}$$

$$= \frac{28}{12} = 2.33$$

g) Variance.

$$s^2 = \frac{\Sigma(d_i - \bar{d})^2}{g}$$

$$= \frac{(3-2.33)^2 + (2-2.33)^2 + \dots}{12.}$$

$$= \frac{4(3-2.33)^2 + 8(2-2.33)^2}{12} = 1.7$$

$$= \frac{2.6668}{12} = 0.222$$ is the varian

e) Nodal degree - no. of lines incident to it.

$n_0 - 3$ $n_5 - 3$

$n_1 - 2$ $n_6 - 2$

$n_2 - 2$ $n_7 - 2$ $n_{11} - 2$

$n_3 - 2$ $n_8 - 2$

$n_4 - 3$ $n_9 - 3$

$n_{10} - 2$

f) mean nodal degree.

$$\bar{d} = \frac{\sum_{i=0}^{g} d(i)}{g} = \frac{3+2+2+2+3+3+2+2+2+3+2+}{12}$$

$$= \frac{28}{12} = 2.33 \text{ degrees/node}$$

g) Variance.

$$S^2 = \frac{\Sigma (d_i - \bar{d})^2}{g}$$

$$= \frac{(3-2.33)^2 + (2-2.33)^2 + \ldots}{12}$$

$$= \frac{4(3-2.33)^2 + 8(2-2.33)^2}{12} = \frac{1.7956 + 0.8712}{12}$$

$$= \frac{2.6668}{12} = 0.222 \text{ is the variance}$$

h) Density of graph.

$$= \frac{d}{(g-1)} = \frac{2L}{g(g-1)}$$

$L = 14$

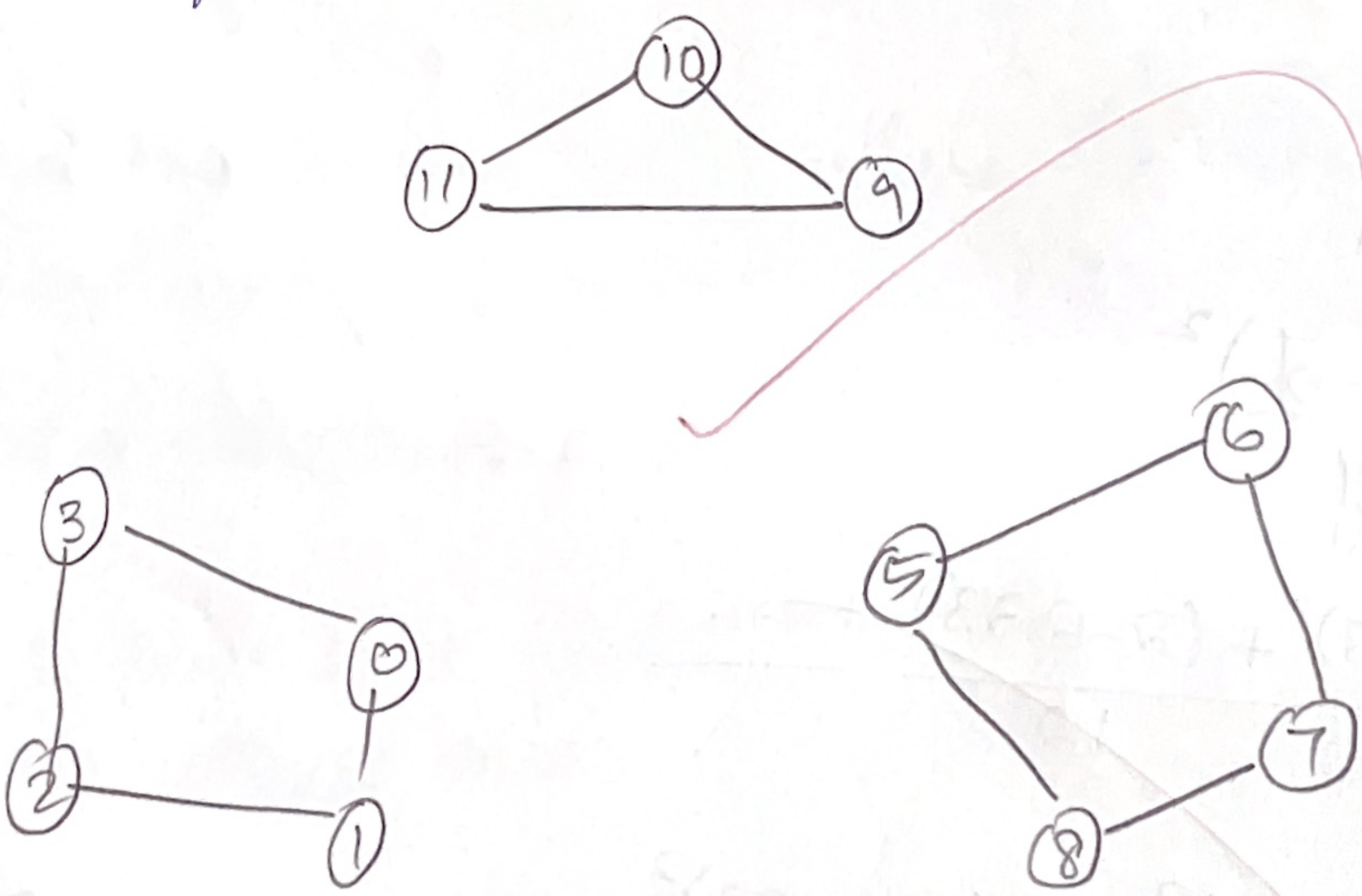
$g = 12$

Density $\Delta = \frac{2 \times 14}{12 \times 11} = \frac{28}{12 \times 11} = 0.21\overline{21}\ldots$

↳ the density

(i) cut-point/node connectivity is the set of nodes for which the graph will become disconnected if removed.

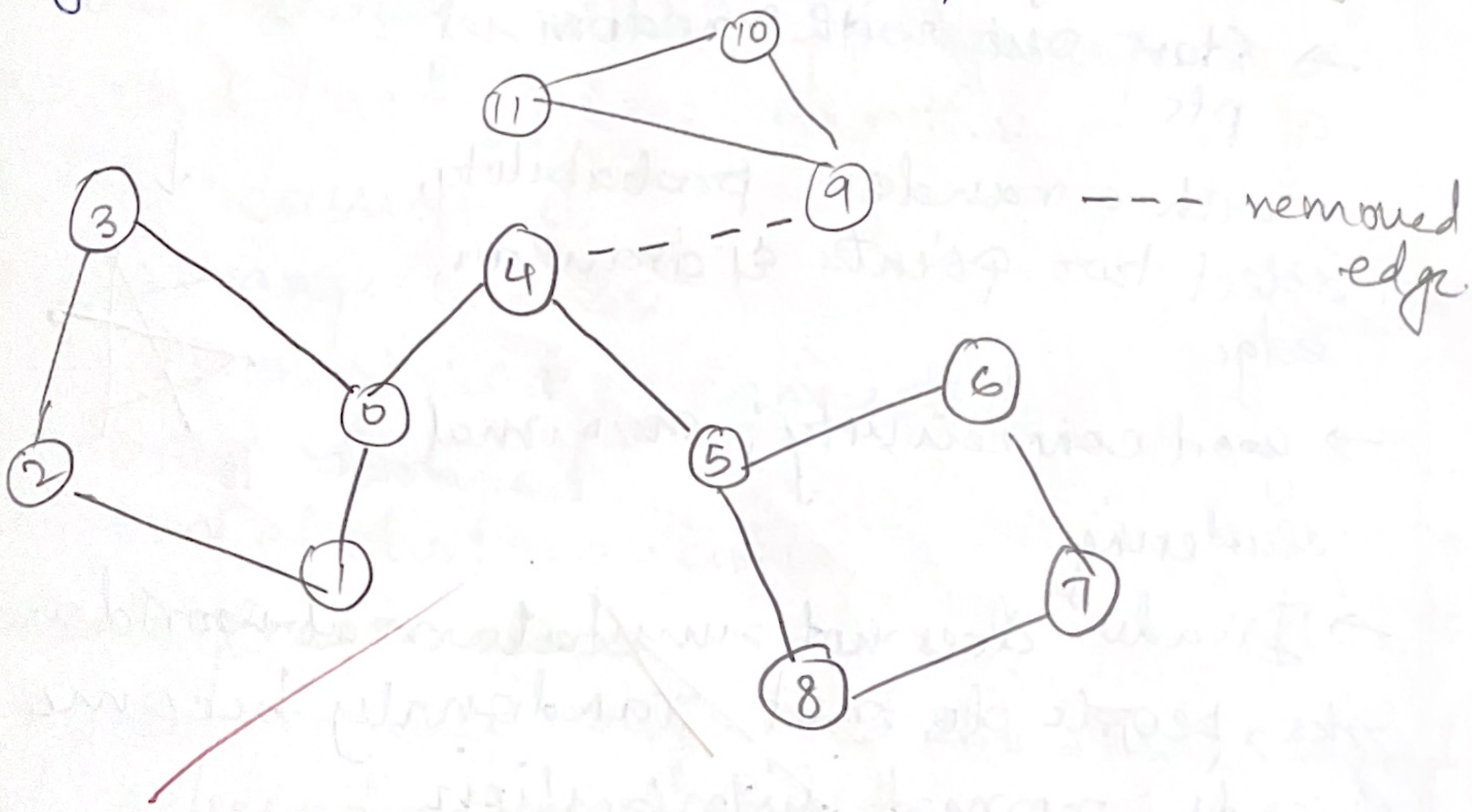
cut-point = $\{n4\}$ 1 → point connectivity



disconnected

(j) Edge connectivity - min no. of edges that should be removed to disconnect the graph. SSN 11

Edge {n4-n9} Value - 1



PART-B

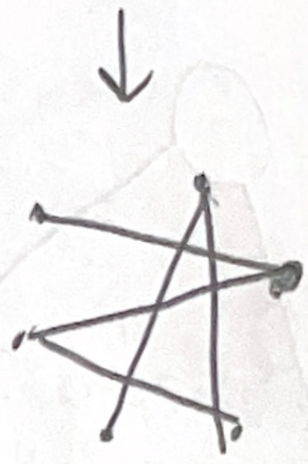
SSN

1) Different models for graph networks

1) Random graph model

→ Start out with random set of pts

→ with a random probability, select two points & draw an edge



→ good connectivity; minimal clustering

→ Disadv: does not simulate real-world networks, people do not randomly become friends; normal distribution

→ can use to simulate complex structures

2) Alpha model

→ proposed in seminal paper.

→ The probability of having a connection between 2 nodes proportional to the no. of shared mutual neighbors.

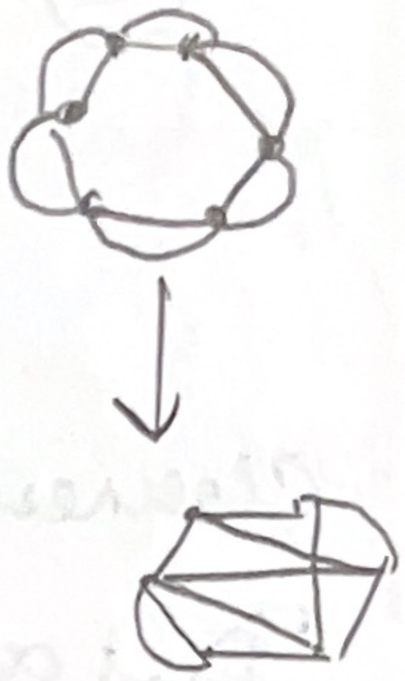
→ α parameter used to control the influence of mutual friends on connection

→ highly connected; Great clustering

→ However, normally distributed

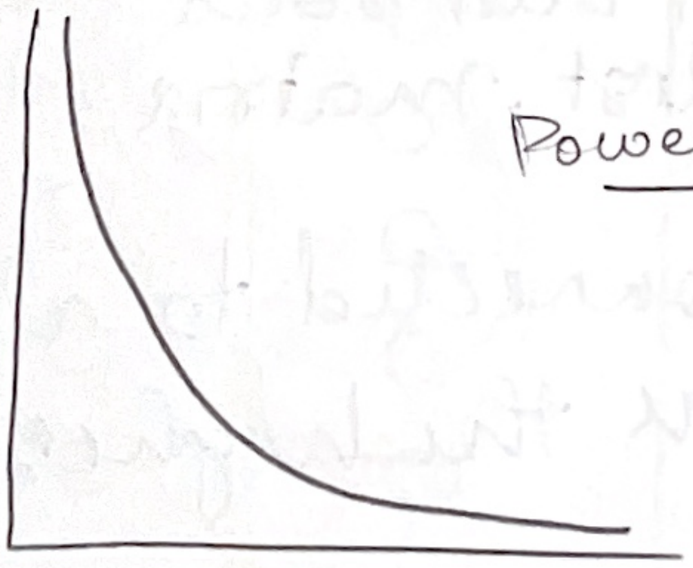
3) Beta model

→ Starts with a toroidal lattice,
→ Nodes connected to neighbours & neighbors of neighbors.
→ with random probability β, keeping one end fixed, select another end & reassign it.
→ Rearrange like so.
→ Well clustered & connected networks formed
→ Normal distribution

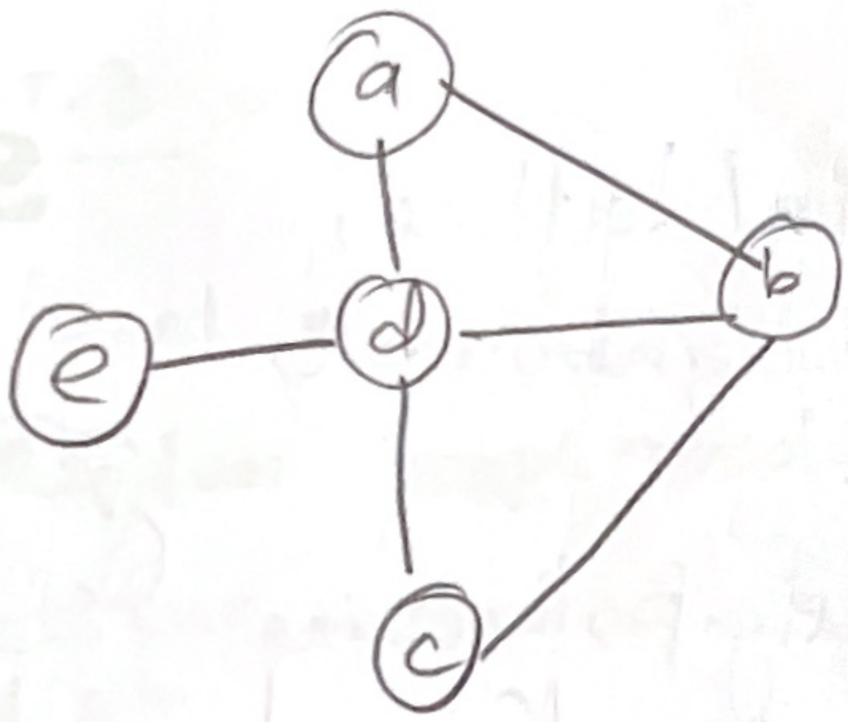


4) Scale-free models

→ prev. models only cared abt small world pro not scale-free characteristic.
→ nodes with large no. of degrees lesser than ones with less no. of degrees
→ Rich get richer idea.
→ Distribution follows power law d^{-k} for $k>0$ -



8)



Closeness centrality.

Distance matrix.

	A	B	C	D	E	Sum.	closeness centrality
A	0	1	2	1	2	6	0.1666
B	1	0	1	1	2	5	0.2
C	2	1	0	1	2	6	0.1666
D	1	1	1	0	1	4	0.25
E	2	2	2	1	0	7	0.143

where Dij = shortest path between node i and node j

$$\text{closeness centrality} = \frac{1}{\text{sum over rows of dist. matrix}}$$

Since D is most closely connected to all other nodes, its value is the highest, with 0.25

Betweeness centrality = $\frac{\sigma_{uv}(w)}{\sigma_{uv}}$

σ_{uv} = shortest path between u & v

$\sigma_{uv}(w)$ = shortest path btw u & v, including w.

w = A.

Path	$\sigma_{uv}(w)$	σ_{uv}	BC
BC	0	1	0
BD	1	2	0.5
BE	1	2	0.5
CD	0	1	0
CE	0	1	0
DE	0	1	0
			1

w = B

Path	$\sigma_{uv}(w)$	σ_{uv}	BC
AC	1	2	0.5
AD	0	1	0
AE	0	1	0
CD	0	1	0
CE	0	1	0
DE	0	1	0
			0.5

W = C

Path	σ_{ij}	$\sigma_{ij}(w)$	BC
AB	1	0	0
AD	1	0	0
AE	1	0	0
BD	2	1	0.5
BE	2	1	0.5
DE	0	1	0
			1

W = D

Path	σ_{ij}	$\sigma_{ij}(w)$	BC
AB	1	0	0
AC	2	1	0.5
AE	1	1	1
BC	1	0	0
BE	2	2	1
CE	1	1	1
			3.5

W = E

Path	σ_{ij}	$\sigma_{ij}(w)$	BC
AB	1	0	0
AC	2	0	0
AD	1	0	0
BC	1	0	0
BD	2	0	0
CD	1	0	0
			0

highest for node D = 3.5

a) 2-mode matrix.

Let people be

Allison : P1

Drew : P2

Elliot : P3

Keith : P4

Rees : P5

Sarah : P6

Events / Affiliation

Party 1 = a1

Party 2 = a2

Party 3 = a3

	P1	P2	P3	P4	P5	P6	a1	a2	a3
P1	0	0	0	0	0	0	1	0	1
P2	0	0	0	0	0	0	0	1	0
P3	0	0	0	0	0	0	0	1	1
P4	0	0	0	0	0	0	0	0	1
P5	0	0	0	0	0	0	1	1	1
P6	0	0	0	0	0	0	1	1	0
a1	1	0	0	0	1	1	0	0	0
a2	0	1	1	0	1	1	0	0	0
a3	1	0	1	1	1	0	0	0	0

one-mode

SSN

	P_1	P_2	P_3	P_4	P_5	P_6
P_1	2	0	1	1	2	1
P_2	0	1	1	0	1	1
P_3	1	1	2	1	2	1
P_4	1	0	1	1	1	0
P_5	2	1	2	1	3	2
P_6	1	1	1	0	2	2

PART-A

1) Social network analysis is the conceptualisation of social networks & their representation in various forms like graphs / sociomatrices. It also involves the analysis of these structures to form general inferences on the nature of social groups

2) → Highly connected subgroups with sparse intergroup connection.

→ Two mode - affiliation network rep by bipartite graph.

→ core-periphery networks: Core has highly connected nodes; peripheral nodes are not connected directly, but with core; minimises error

Global Views of social structure

5) powerful, but not central:

~~[redacted] : has a lot of connections~~

A french actor, in a web of actors might be powerful within her own community, but not central, since they do not act as a connection to the overall web.

Similarly, in a film network, a director might be central, in knowing/connecting other people. However, they might not be powerful themselves

6) Three schemes:

→ Sociometric Analysis:
- ~~[redacted]~~ equivalence of blocks.

→ ~~[redacted]~~ Relational Algebra:
- Role & positional Analysis

→ Graph Theoretic
- for dyad / triad / closeness analysis.