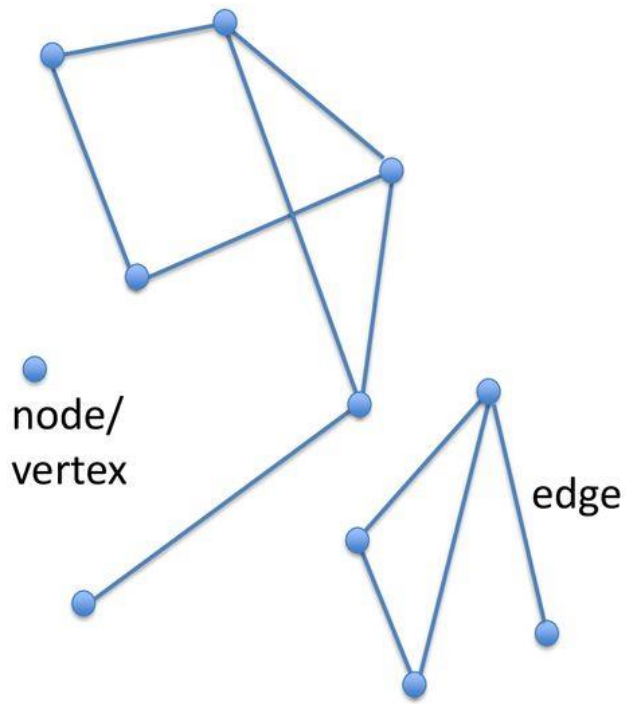
Social Network Analysis

Metrics and Measures

Saeed Roshani (Ph.D.)

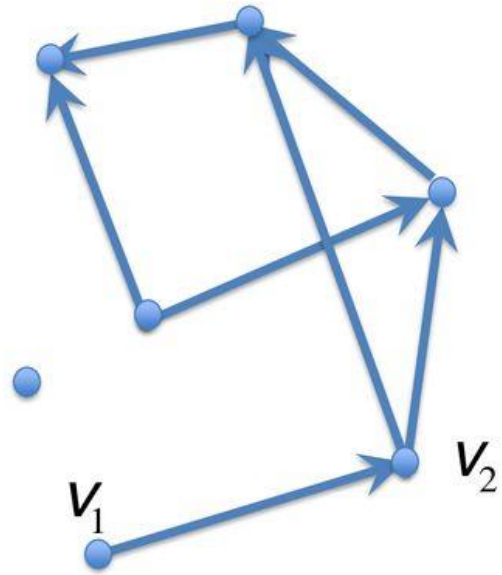
Network: definition



A *network/graph*
consists of
nodes/vertices
connected by
links/edges

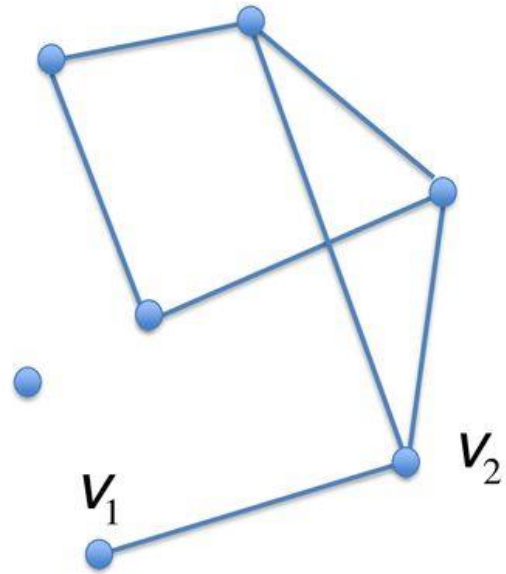
Notation:
$G, V(G), E(G), |G|$

Directed vs. undirected graph



In a directed graph, an edge has a beginning and an end, i.e. (v_1, v_2) (v_2, v_1)

Asymmetric relationship

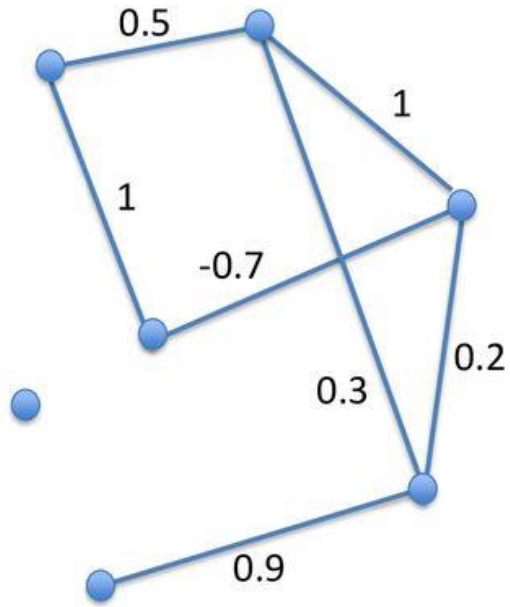


In a simple graph, an edge is an unordered pair of vertices, i.e. $(v_1, v_2) = (v_2, v_1)$

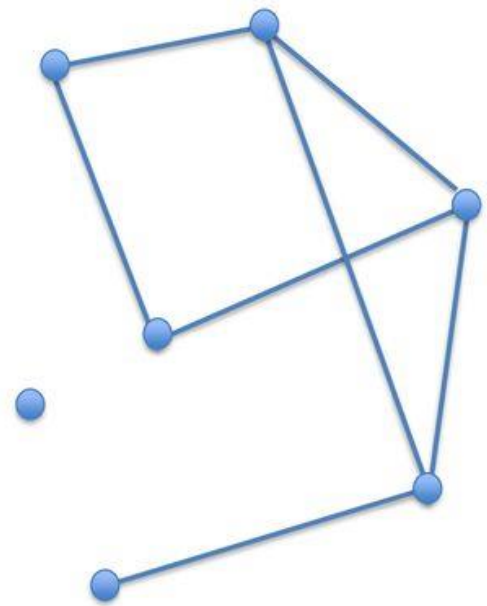
Symmetric relationship

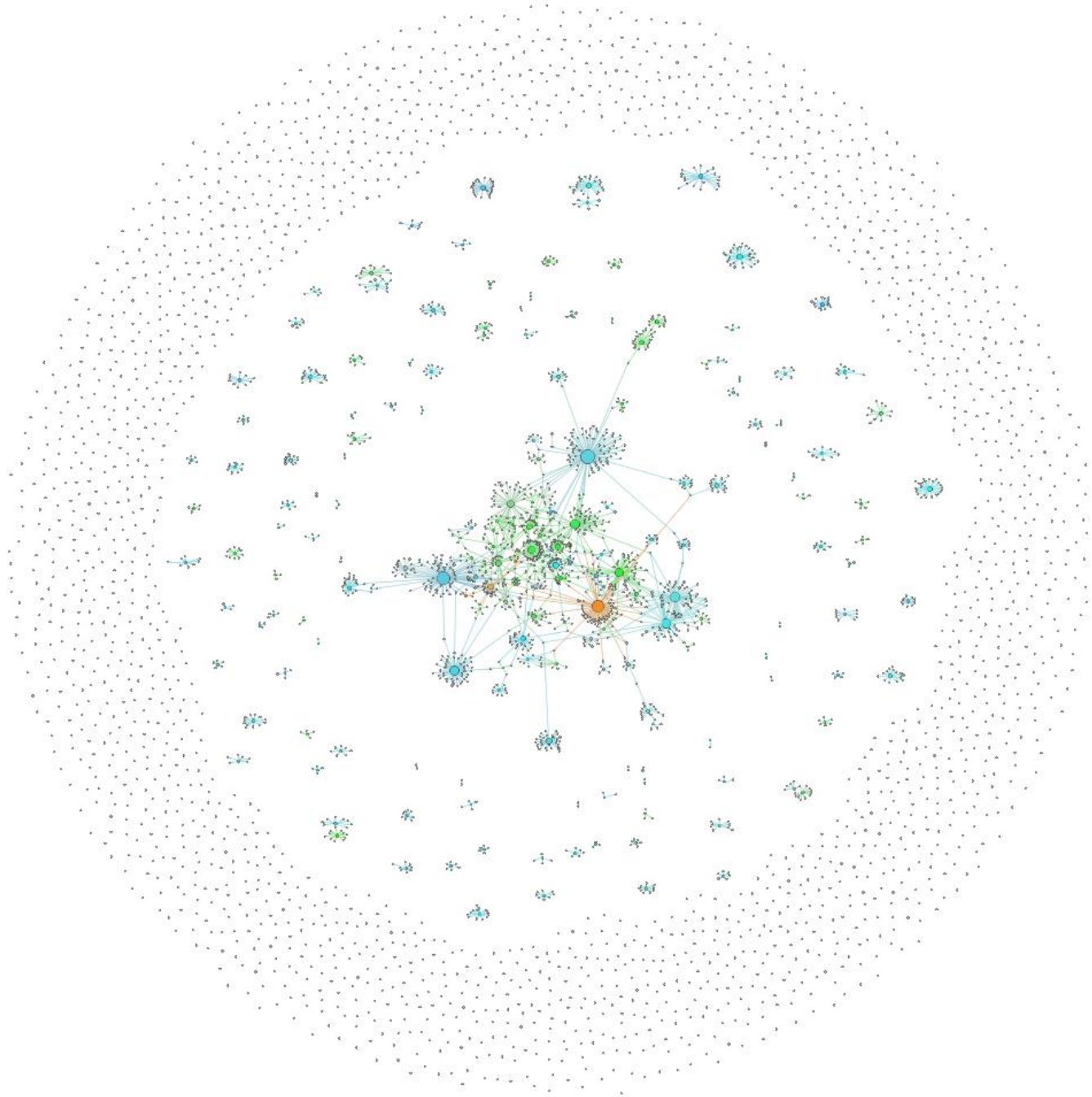
Weighted vs. unweighted graph

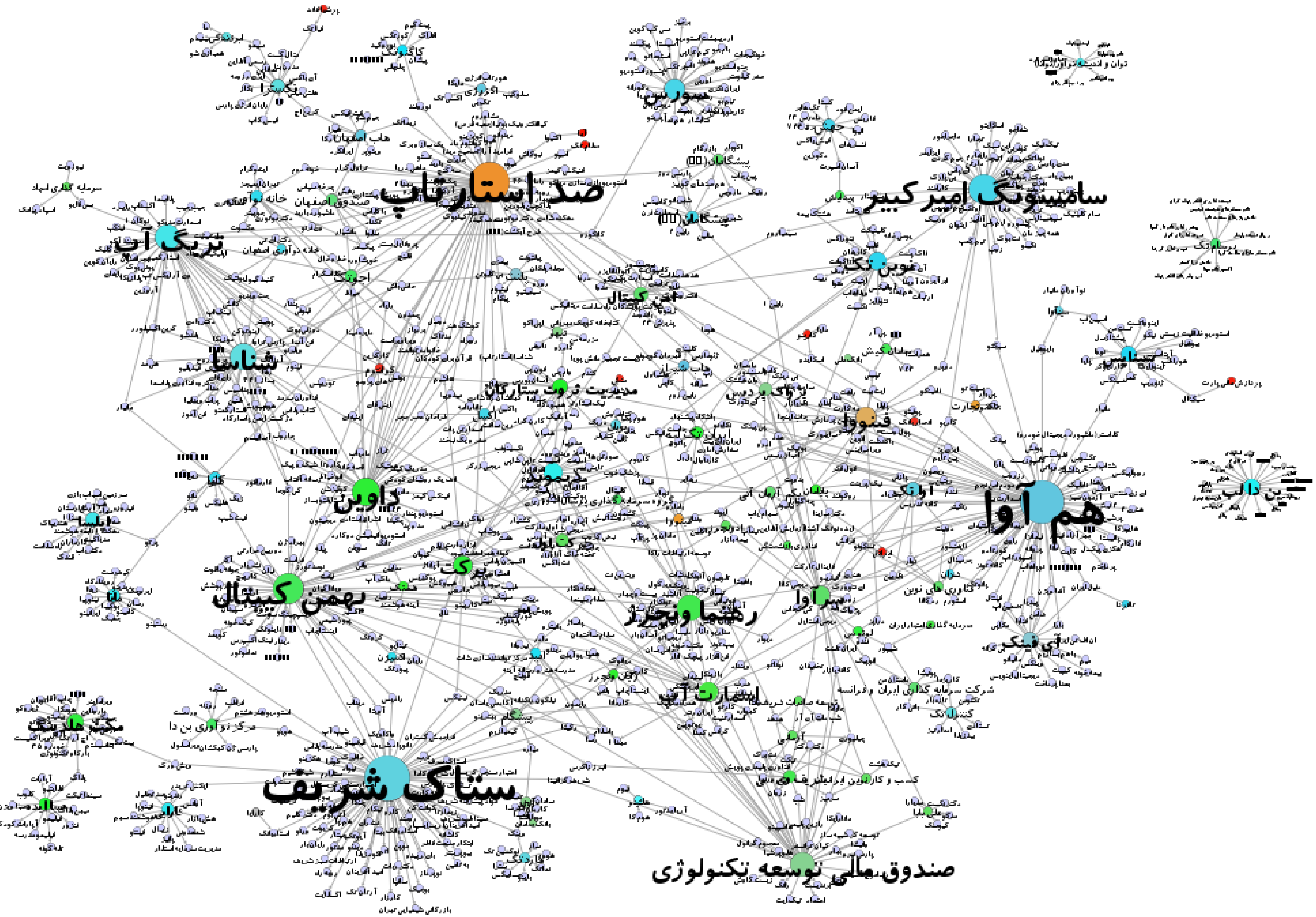
Each edge has an associated ‘strength’:



Carries only information about presence / absence of an edge:



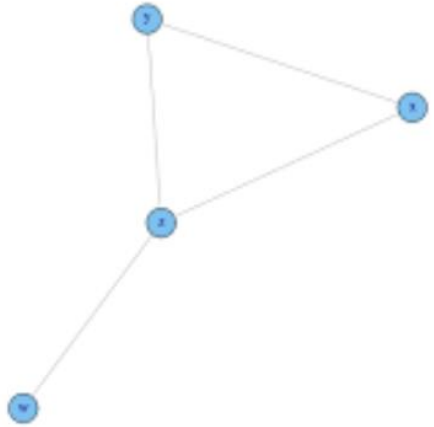




صد استارتاپ
سامسونگ امیرکبیر
هم آوا
ستاک شریف
بهمن کپیتال
رهنما ویچرز
صندوق مالی توسعه تکنولوژی
شناسا
ترگ آپ
داوین
سورس

Network Analysis (Macro Analysis)

What are the most important measures?



Transitivity

$$T = \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triples of nodes in the network}}.$$

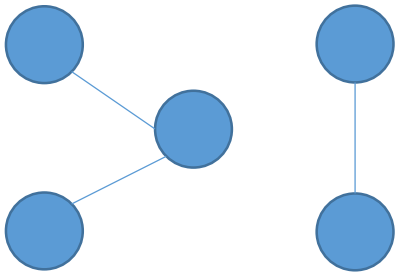
The transitivity T of a graph is based on the relative number of triangles in the graph, compared to total number of connected triples of nodes.

The factor of three in the number accounts for the fact that each triangle contributes to three different connected triples in the graph, one centered at each node of the triangle. With this definition, $0 \leq T \leq 1$, and $T=1$ if the network contains all possible edges.
The transitivity of a graph is closely related to the **clustering coefficient** of a graph, as both measure the relative frequency of triangles.

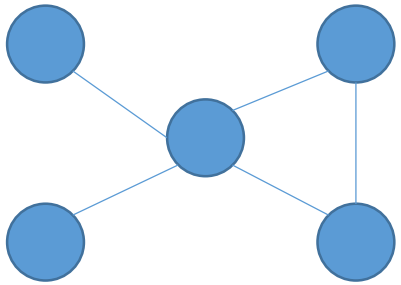
Alpha Index

$$\alpha = \frac{u = e - v + p}{2v - 5}$$

- A measure of connectivity which evaluates the number of cycles in a graph in comparison with the maximum number of cycles.
- The higher the alpha index, the more a network is connected.
- Number of cycles: The maximum number of independent cycles in a graph.
- This number (u) is estimated through the number of nodes (v), edges (e), and sub-graphs (p).



A



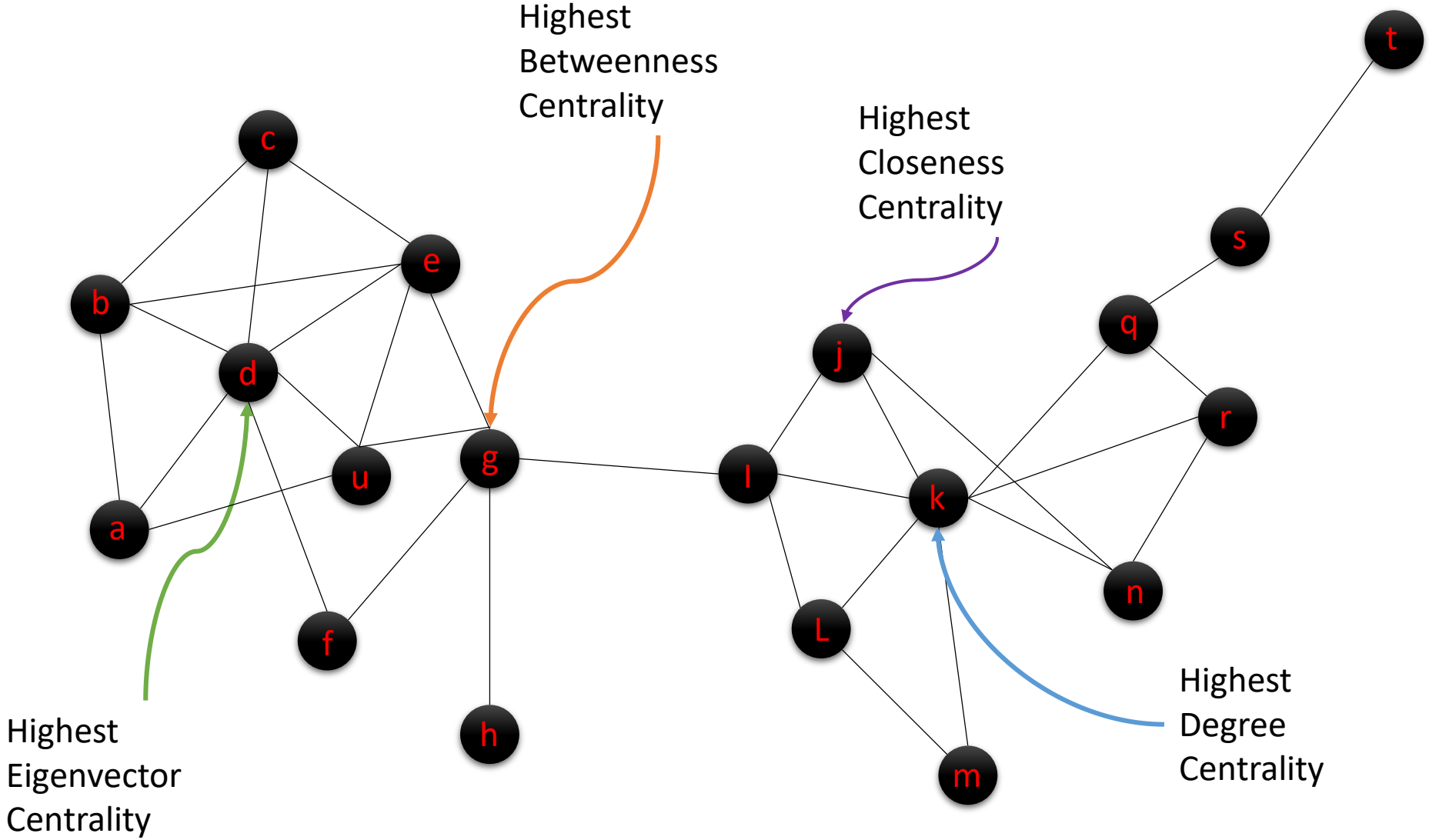
B

Value	Description	Formula	indicator	Criteria
0 <α< 1	سنجش سطح اتصال گره‌های شبکه مبتنی بر تعداد سیکل-های درون شبکه	$\alpha = \frac{e - v + p}{2(v) - 5}$	α	**Alpha**
β> 0	سنجش سطح اتصال شبکه مبتنی بر تعداد اتصال‌های موجود بر تعداد گره‌ها	$\beta = \frac{e}{v}$	β	**Beta**
$0 < \varphi < 1$	سنجش اتصال شبکه از طریق تعداد اتصال‌های موجود بر اتصال‌های ممکن	$\varphi = \frac{e}{\frac{v(v-1)}{2}}$	φ	**Gamma**
0 <T< 1	سنجش میزان اتصال شبکه از طریق اندازه‌گیری تمایل گره‌ها در ایجاد اتصال به یکدیگر	$T(G) = \frac{3\delta(G)}{\tau(G)}$	T	**Transitivity or Clustering co-efficient**

$\delta(G) = 1/3\sum\delta(v)$نشان‌دهنده تعداد مثلث‌های گراف G

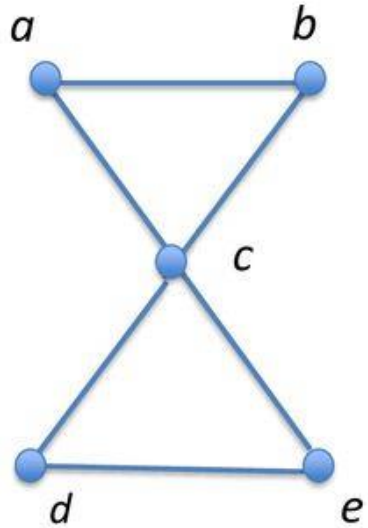
Network Analysis (Micro Analysis)

What are the most important measures?


Highest
Betweenness
Centrality
Highest
Closeness
Centrality
Highest
Eigenvector
Centrality
Highest
Degree
Centrality
c
e
b
d
u
g
a
f
h
j
l
k
L
m
n
q
r
s
t

1.a Connectivity: ‘Can I get there?’

- *Edge/vertex connectivity* = the minimum number of edges/vertices that need to be removed to disconnect a graph

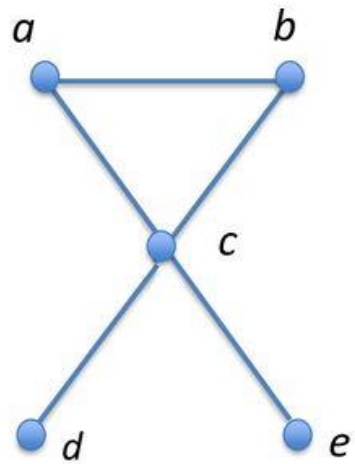


Edge connectivity: 2

Vertex connectivity: 1

1.b Shortest paths: ‘How far is it?’

- *Diameter* = the length of a longest shortest path between a pair of vertices in a graph
  - ‘at most how far are two vertices from each other?’
- *Average shortest path length* = average of the lengths of shortest paths between all pairs of vertices

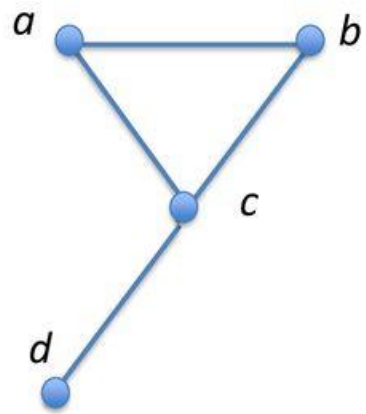


2.a Degree centrality

For every vertex *v* in a graph,

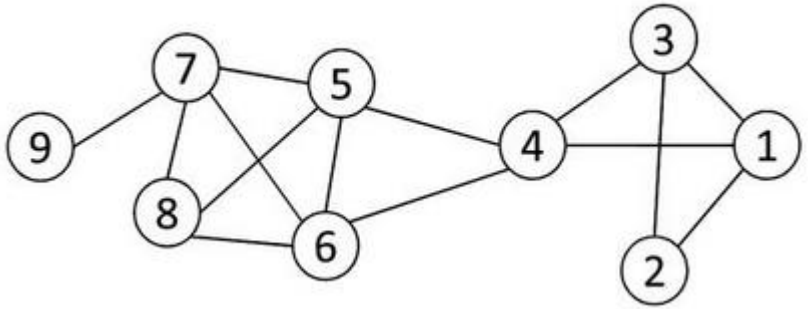
- *Degree centrality (v)* = deg(*v*)

'The most important in the graph are the vertices with highest degrees'



$a \rightarrow 2$
$b \rightarrow 2$
$c \rightarrow 3$
$d \rightarrow 1$

2.b Betweenness Centrality

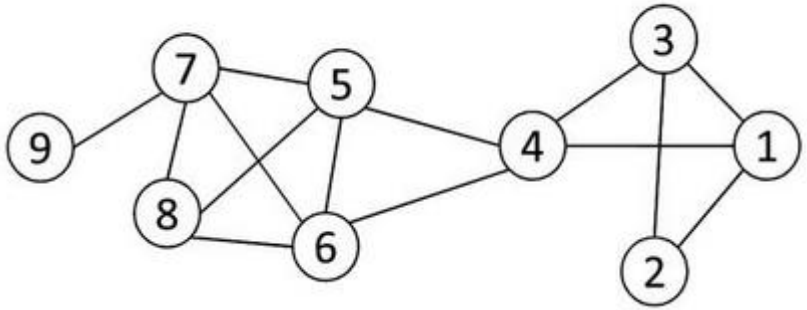


$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

σ_{st} : The number of shortest paths between s and t

$\sigma_{st}(v_i)$: The number of shortest paths between s and t that pass v_i

2.c Closeness Centrality

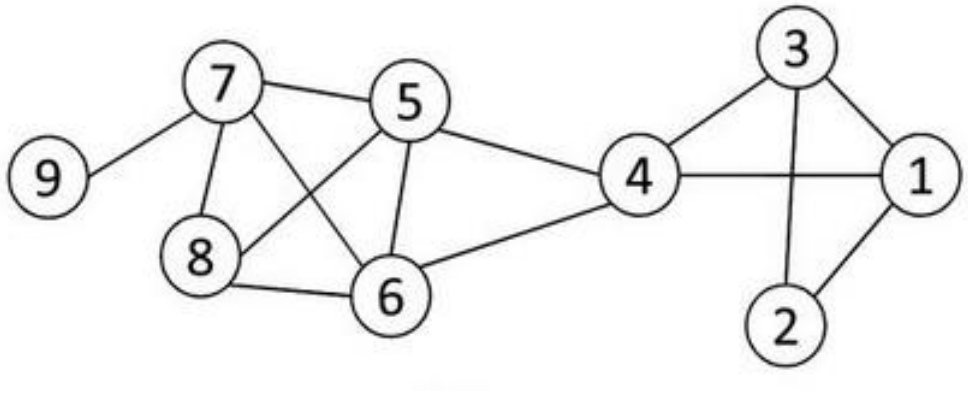


Closeness is based on the average distance between a vertex *i* and all vertices in the graph (consider vertices in the same component only):

$$C_i = 1 \Big/ \sum_{j=1}^{n} d(i,j)$$

depends on inverse distance to other vertices

Closeness centrality can be viewed as the efficiency of a vertex in spreading information to all other vertices.



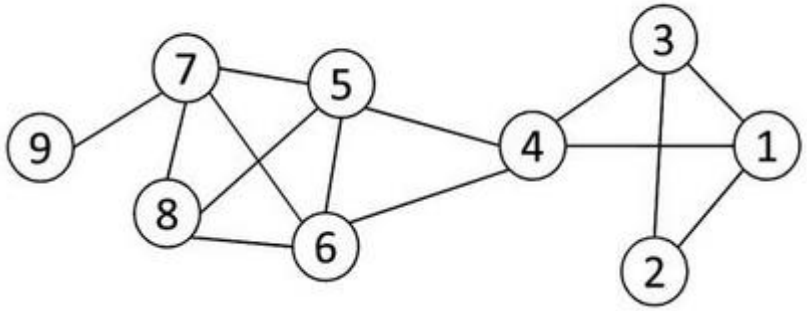
Node	1	2	3	4	5	6	7	8	9
1	0	1	1	1	2	2	3	3	4
2	1	0	1	2	3	3	4	4	5
3	1	1	0	1	2	2	3	3	4
4	1	2	1	0	1	1	2	2	3
5	2	3	2	1	0	1	1	1	2
6	2	3	2	1	1	0	1	1	2
7	3	4	3	2	1	1	0	1	1
8	3	4	3	2	1	1	1	0	2
9	4	5	4	3	2	2	1	2	0

Is node 4 is more central than node 3?

$$C_c(3) = \frac{1}{1+1+1+2+2+3+3+4} = 1/17 = 0.058$$

$$C_c(4) = \frac{1}{1+2+1+1+1+2+2+3} = 1/13 = 0.076$$

2.d Eigenvector Centrality



One's importance is determined by his friends.
If one has many important friends, he should be important as well.

$$C_E(v_i) \propto \sum_{v_j \in N_i} A_{ij} C_E(v_j)$$

A variant of this **eigenvector centrality** is the PageRank score