Introduction to graph theory

Graphs

Size and order
Degree and degree distribution
Subgraphs
Paths, components
Geodesics
Some special graphs
Centrality and centralisation

Directed graphs

Dyad and triad census
Paths, semipaths, geodesics, strong and weak components
Centrality for directed graphs
Some special directed graphs

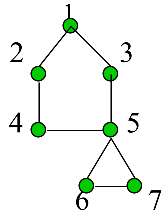
©Department of Psychology, University of Melbourne

Definition of a graph

A ***graph*** *G* comprises a set *V* of vertices and a set *E* of edges

Each ***edge*** in *E* is a pair (*a*,*b*) of vertices in *V*
If (*a*,*b*) is an edge in *E*, we connect *a* and *b* in the ***graph drawing*** of *G*

Example:



V={1,2,3,4,5,6,7}
E={(1,2),(1,3),(2,4).
(4,5),(3,5),(4,5),
(5,6),(6,7)}

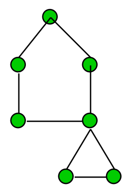
©Department of Psychology, University of Melbourne

Size and order

The ***size*** of G is the number n of vertices in V

The ***order*** of G is the number L of edges in E

Minimum possible order is 0 (***empty*** graph)
Maximum possible order is $n(n-1)/2$ (***complete*** graph)



Size = 7, Order = 8

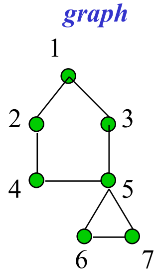
©Department of Psychology, University of Melbourne

Adjacency matrix for a graph

The adjacency matrix $\mathbf{x} = [x_{ab}]$ for G is a matrix with n rows and n colums and entries given by:

x_{ab} = 1 if (a,b) is an edge in G
0 otherwise

Example: ***graph***



adjaceny matrix

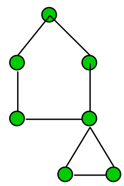
0110000
1001000
1000100
0100100 ***symmetric***
0011011
0000101
0000110

©Department of Psychology, University of Melbourne

Density

The ***density*** of *G* is the ratio of edges in *G* to the maximum possible number of edges

$$\text{Density} = \frac{2L}{n(n-1)}$$



Density = 2×8/(7×6) = 8/21

©Department of Psychology, University of Melbourne

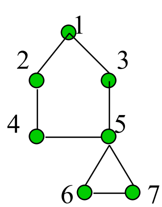
Degrees and degree sequence

The ***degree*** d_a of vertex *a* is the number of vertices to which *a* is linked by an edge

The minimum possible degree is 0

The maximum possible degree is n-1

The ***degree sequence*** for a graph is the vector $(d_1, d_2, \ldots, d_n)$

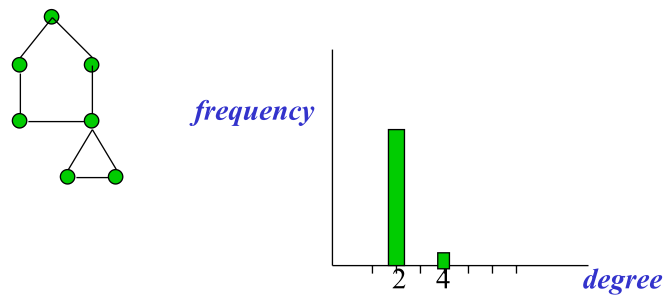


Degree sequence = (2,2,2,2,4,2,2)

©Department of Psychology, University of Melbourne

Degree distribution

The ***degree distribution*** for the graph is (k_0, k_1,…, k_{n-1}), where k_j = the number of nodes with degree j

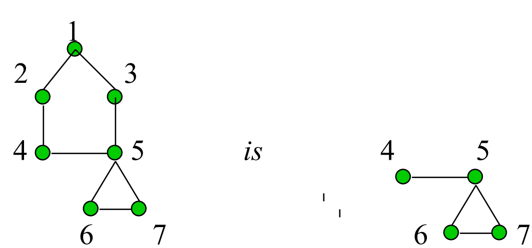


©Department of Psychology, University of Melbourne

Subgraphs

A ***subgraph*** of G=G(V,E) is a subset W of the vertex set V together with all of the edges that connect pairs of vertices in W

Eg if W={4,5,6,7}, the subgraph of



©Department of Psychology, University of Melbourne

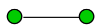
Subgraph counts: the dyad census

The graph G has $n(n\text{-}1)/2$ subgraphs of size 2

Each subgraph of size 2 comprises a pair of vertices, and the edge between them is either present or absent:

subgraph ***count***

$D_0 = n(n\text{-}1)/2 – L$



$D_1 = L$

Dyad census = (D_0,D_1) ***count of the no. of each type of dyad subgraph***

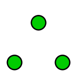
©Department of Psychology, University of Melbourne

Subgraph counts: the triad census

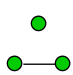
The graph G has $n(n\text{-}1)(n\text{-}2)/6$ subgraphs of size 3

Each subgraph of size 3 comprises a triple of vertices, and the possible forms are:

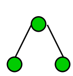
subgraph count



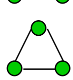
$T_0 = (1/6)\Sigma_{i,j,k}\,(1\text{-}x_{ij})(1\text{-}x_{ik})(1\text{-}x_{jk})$



$T_1 = (1/6)\Sigma_{i,j,k}\,(1\text{-}x_{ij})(1\text{-}x_{ik})x_{jk}$



$T_2 = (1/6)\,\Sigma_{\,i,j,k}(1\text{-}x_{ij})x_{ik}x_{jk}$



$T_3 = (1/6)\,\Sigma_{\,i,j,k}\,x_{ij}x_{ik}x_{jk}$

Triad census = (T_0,T_1,T_2,T_3)

©Department of Psychology, University of Melbourne

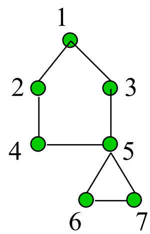
Paths

A ***path*** from vertex a to vertex b is an ordered sequence

$a=v_0, v_1, \ldots, v_m=b$

of distinct vertices in which each adjacent pair (v_{j-1},v_j) is linked by an edge. The ***length*** of the path is m

There is:



a path of length 1 from 1 to 2
a path of length 2 from 1 to 4
a path of length 3 from 1 to 4
a path of length 3 from 1 to 6

©Department of Psychology, University of Melbourne

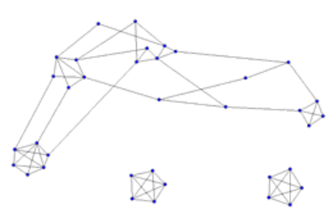
Reachability and connectedness

If there is a path from vertex a to vertex b, a is ***reachable*** from b

If each vertex in G is reachable from each other vertex, then G is ***connected***

A ***component*** of G is a maximal connected subgraph (ie a connected subgraph with vertex set W for which no larger set Z containing W is connected)

A graph with 3 components



©Department of Psychol

Geodesics

A ***geodesic*** from *a* to *b* is a path of minimum length

The ***geodesic distance*** d_{ab} between *a* and *b* is the length of the geodesic

If there is no path from *a* to *b*, the geodesic distance is ***infinite***

For the graph

The geodesic distances are:

$d_{AB} = 1$, $d_{AC} = 1$, $d_{AD} = 1$, $d_{BC} = 1$, $d_{BD} = 2$, $d_{CD} = 2$

©Department of Psychology, University of Melbourne

Cycles

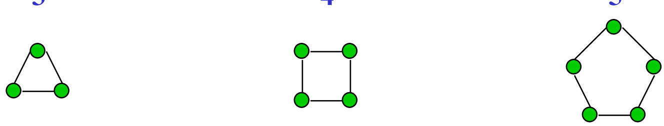
A ***cycle*** is an ordered sequence

$a=v_0, v_1, \ldots, v_m=a$

of vertices in which each adjacent pair (v_{j-1},v_j) of vertices is linked by an edge, and $v_0, v_1, \ldots, v_{m-1}$ are distinct . The ***length*** of the cycle is *m*

Cycles of length

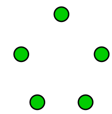
3 **4** **5**



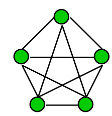
©Department of Psychology, University of Melbourne

Some special graphs: trivial, empty and complete graphs

The ***empty*** graph on 5 vertices (Z_5)



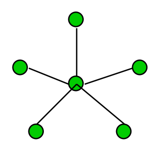
The ***complete*** graph on 5 vertices (K_5)



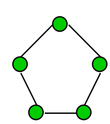
©Department of Psychology, University of Melbourne

Star and cyclic graphs

A ***star*** graph on 6 vertices



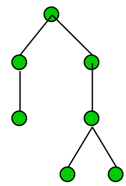
A ***cyclic*** graph on 5 vertices (C_5)



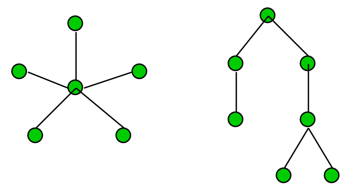
©Department of Psychology, University of Melbourne

Trees and forests

A ***tree*** (a connected acyclic graph)



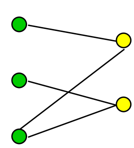
A ***forest*** (a graph with tree components)



©Department of Psychology, University of Melbourne

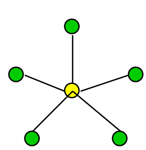
Bipartite graphs

A ***bipartite*** graph (vertex set can be partitioned into 2 subsets, and there are no edges linking vertices in the same set)

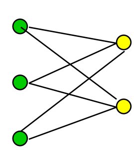


A ***complete*** bipartite graph (all possible edges are present)

$K_{1,5}$



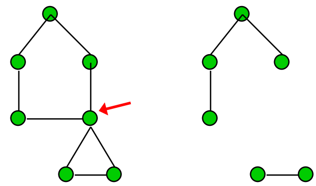
$K_{3,2}$



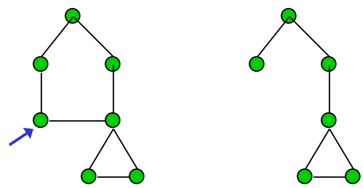
©Department of Psychology, University of Melbourne

Cutpoints

A vertex is a ***cutpoint*** if its removal increases the number of components in the graph



the vertex marked by the red arrow is a cutpoint

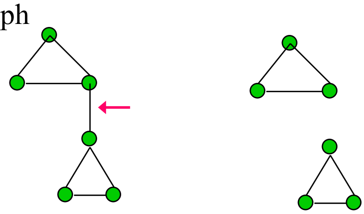


the vertex marked by the blue arrow is not

©Department of Psychology, University of Melbourne

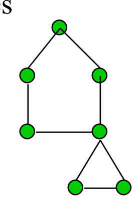
Bridges

An edge is a ***bridge*** if its removal increases the number of components in the graph



the edge marked by the red arrow is a bridge

This graph has no bridges

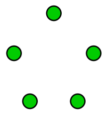
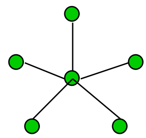
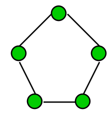
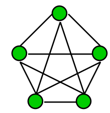


©Department of Psychology, University of Melbourne

Connectivity

The ***connectivity*** κ(*G*) of a connected graph *G* is the minimum number of vertices that need to be removed to disconnect the graph (or make it empty)

A graph with more than one component has connectivity 0

Graph				
Connectivity	0	1	2	4

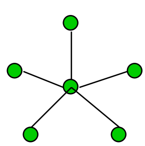
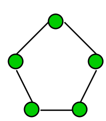
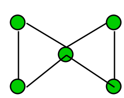
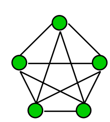
A graph with connectivity *k* is termed ***k-connected***

©Department of Psychology, University of Melbourne

Edge-connectivity

The ***edge-connectivity*** λ(*G*) of a connected graph *G* is the minimum number of ***edges*** that need to be removed to disconnect the graph

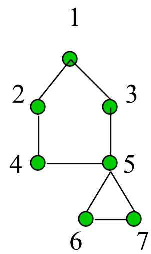
A graph with more than one component has edge-connectivity 0

Graph				
Edge-Connectivity	1	2	2	4
Connectivity	1	2	1	4

©Department of Psychology, University of Melbourne

Independent and edge-independent paths

Two paths from *a* to *b* are ***independent*** if they have no nodes in common apart from *a* and *b*   ***e.g. paths 1-2-4-5 and 1-3-5***



Two paths from *a* to *b* are ***edge-independent*** if they have no edges in common   ***e.g. paths 1-2-4-5-6 and 1-3-5-7-6***

©Department of Psychology, University of Melbourne

Several theorems about connectivity

Whitney's theorem

For any graph *G*, $\kappa(G) \leq \lambda(G) \leq \delta(G)$, where $\delta(G)$ is the minimum degree of any vertex in *G*

Menger's theorem

A graph *G* is *k*-connected if and only if any pair of vertices in *G* are linked by at least *k* independent paths

Menger's theorem

A graph *G* is *k*-edge-connected if and only if any pair of vertices in *G* are linked by at least *k* edge-independent paths

For application, see Harary & White (2001)

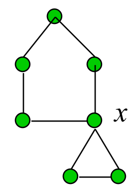
©Department of Psychology, University of Melbourne

Degree Centrality

Freeman (1979) described three measures of ***vertex*** centrality:

Degree centrality (communication potential)

Degree centrality of node *a:* $C_D(a) = d_a$ **degree of node a**



Normalised degree centrality of node a: $d_a/(n\text{-}1)$

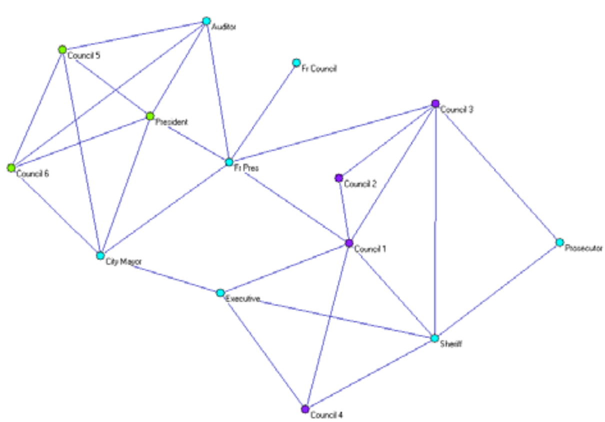
Example

Node *x*: degree centrality = 4

normalised degree centrality = 4/6 = 0.67

©Department of Psychology, University of Melbourne

A political network (Doreian, 1988)



©Department of Psychology, University of Melbourne

Degree centrality in Doreian's (1988) political network

		Degree	NrmDegree	Share
4	***Council 1***	***6.000***	***46.154***	***0.107***
12	***Fr Pres***	***6.000***	***46.154***	***0.107***
3	Sheriff	5.000	38.462	0.089
8	President	5.000	38.462	0.089
6	Council 3	5.000	38.462	0.089
13	City Mayor	5.000	38.462	0.089
2	Auditor	4.000	30.769	0.071
1	Executive	4.000	30.769	0.071
9	Council 5	4.000	30.769	0.071
10	Council 6	4.000	30.769	0.071
7	Council 4	3.000	23.077	0.054
14	Prosecutor	2.000	15.385	0.036
5	Council 2	2.000	15.385	0.036
11	Fr Council	1.000	7.692	0.018

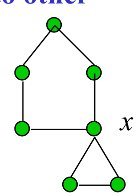
©Department of Psychology, University of Melbourne

Closeness Centrality

Closeness centrality (potential for independent communication)

Closeness centrality of node *a*: $C_D(a) = 1/\sum_b d_{ab}$ **inverse sum of distances to other nodes b**

Normalised closeness centrality of node a: $(n\text{-}1)/\sum_b d_{ab}$



Example

Node *x*: closeness centrality = 1/[1+1+1+1+2+2]=1/8 = 0.125

normalised closeness centrality = 6/8=0.75

©Department of Psychology, University of Melbourne

Closeness centrality in Doreian’s political network

		Farness	nCloseness
12	***Fr Pres***	***20.000***	***65.000***
4	Council 1	22.000	59.091
6	Council 3	23.000	56.522
13	City Mayor	23.000	56.522
1	Executive	24.000	54.167
8	President	25.000	52.000
3	Sheriff	26.000	50.000
2	Auditor	27.000	48.148
7	Council 4	28.000	46.429
9	Council 5	31.000	41.935
10	Council 6	31.000	41.935
11	Fr Council	32.000	40.625
5	Council 2	32.000	40.625
14	Prosecutor	32.000	40.625

©Department of Psychology, University of Melbourne

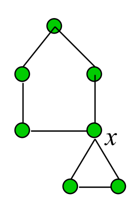
Betweeness centrality

Betweeness centrality (Potential for control of communication)

Betweenness centrality of node a: $C_D(a) = \sum_{b<c}[g_{bc(a)}/g_{bc}]$

Where g_{bc} is the number of geodesics between b and c, and $g_{bc(a)}$ is the number of geodesics between b and c that contain a

sum over all pairs (*b*,*c*) of the proportion of geodesics linking the pair that contain node *a*



Normalised betweeness centrality of node a:
$2\sum_{b<c}[g_{bc(a)}/g_{bc}]/[n^2 - 3n + 2]$

Example

Node x: betweeness centrality = 14

normalised betweeness centrality = 14/[49-21+2]=7/15

©Department of Psychology, University of Melbourne

Betweeness centrality in Doreian's political network

		Betweenness	nBetweenness
12	Fr Pres	33.198	42.561
13	City Mayor	14.490	18.578
4	Council 1	13.843	17.747
6	Council 3	13.024	16.697
1	Executive	9.452	12.118
3	Sheriff	4.500	5.769
8	President	4.021	5.156
2	Auditor	3.571	4.579
9	Council 5	0.450	0.577
10	Council 6	0.450	0.577
11	Fr Council	0.000	0.000
5	Council 2	0.000	0.000
7	Council 4	0.000	0.000
14	Prosecutor	0.000	0.000

©Department of Psychology, University of Melbourne

Maximum centrality

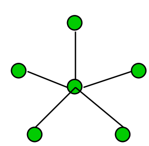
Centrality is at a maximum on all measures for the ***central node in a star*** configuration:

Normalised measures

Degree centrality: 1

Closeness centrality: 1

Betweeness centrality: 1



©Department of Psychology, University of Melbourne

Centralisation

Graph-level measure of ***centralisation***:

Degree to which the centrality of the most central vertex exceeds the centrality of all other vertices, compared to the maximum possible discrepancy

Index has the form:

Sum over nodes a of (max centrality in G – centrality of node a)
--
Max value of the sum in all graphs on the same number of vertices

Can be used with any centrality measure

©Department of Psychology, University of Melbourne

Eigenvector centrality (Bonacich, 1972)

The ***eigenvector centrality*** of vertex *a* is the sum of its connections to other nodes, weighted by ***their*** centrality

It is hence given by the solution for c_a of the equation
$c_a = (1/\lambda) \sum_b x_{ab} c_b$ where λ is a constant

[***Mathematical note***: this is equivalent to:
$\lambda\mathbf{c} = \mathbf{x}\mathbf{c}$
where **x** is the adjacency matrix and **c** is the vector of centrality measures; hence **c** is an eigenvector of **x**, usually taken to be the one associated with the largest eigenvalue λ]

Use with connected graphs only

©Department of Psychology, University of Melbourne

Eigenvector centrality for Doreian network

		Eigenvec	nEigenvec
1	Executive	0.249	35.207
2	Auditor	0.287	40.585
3	Sheriff	0.257	36.368
4	***Council 1***	***0.328***	***46.358***
5	Council 2	0.134	18.887
6	Council 3	0.270	38.154
7	Council 4	0.186	26.357
8	***President***	***0.350***	***49.435***
9	Council 5	0.282	39.854
10	Council 6	0.282	39.854
11	Fr Council	0.083	11.723
12	***Fr Pres***	***0.371***	***52.454***
13	***City Mayor***	***0.343***	***48.453***
14	Prosecutor	0.118	16.655

©Department of Psychology, University of Melbourne

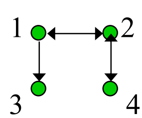
Directed graph

A ***directed graph*** *G* comprises a set *V* of vertices and a set *E* of ***arcs***

Each ***arc*** in *E* is an ***ordered pair*** (*a*,*b*) of vertices in *V*
If (*a*,*b*) is an arc in *E*, we draw an arc ***from a to b*** in the ***graph drawing*** of *G*

Example:

V={1,2,3,4}
E={(1,2),(2,1),(2,4),(1,3),(4,2)}



©Department of Psychology, University of Melbourne

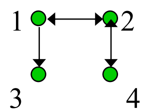
Adjacency matrix for a directed graph

The adjacency matrix **x** =[x_{ab}] for G is a matrix with n rows and n colums and entries given by:

x_{ab} = 1 if (a,b) is an arc in G

0 otherwise

Example



adjacency matrix

0110

1001

0000

0100

not necessarily symmetric

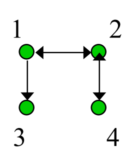
©Department of Psychology, University of Melbourne

Density for a directed graph

The ***density*** of G is the ratio of arcs in G to the maximum possible number of arcs

$$\text{Density} = \frac{L}{n(n-1)}$$

Example



density = 5/12

©Department of Psychology, University of Melbourne

Indegrees and outdegrees

The ***indegree*** i_a of vertex *a* is the number of vertices to which *a* is linked by an arc

The ***outdegree*** o_a of vertex *a* is the number of vertices linked to *a* by an arc

The minimum possible indegree (or outdegree) is 0

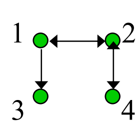
The maximum possible indegree(or outdegree) is n-1

©Department of Psychology, University of Melbourne

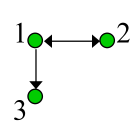
Directed subgraphs

A ***subgraph*** of G=$G(V,E)$ is a subset W of the vertex set V together with all of the arcs that connect pairs of vertices in W

Eg if W={1,2,3}, the subgraph of



is



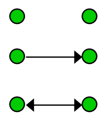
©Department of Psychology, University of Melbourne

Directed subgraph counts: the dyad census

The directed graph G has $n(n-1)/2$ subgraphs of size 2

Each subgraph of size 2 comprises a pair of vertices, and there are either 0, 1 or 2 arcs linking them:

subgraph ***count***



N = number of null dyads

A = number of asymmetric arcs

M = number of mutual arcs

Dyad census = (*M*,*A*,*N*) ***count of the no. of each type of dyadic subgraph***

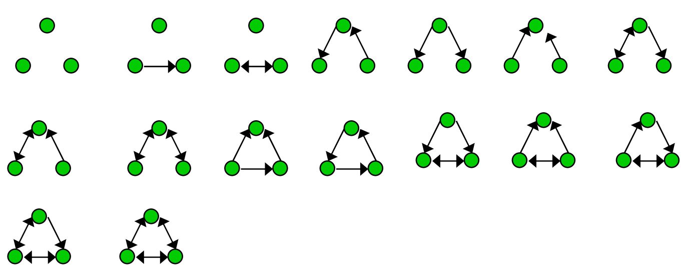
©Department of Psychology, University of Melbourne

Directed subgraph counts: the triad census

The graph G has $n(n-1)(n-2)/6$ subgraphs of size 3

Each subgraph of size 3 comprises a triple of vertices, and there are 16 possible forms:

subgraphs



Triad census: counts of each of the 16 forms across all subgraphs of G

©Department of Psychology, University of Melbourne

Paths and semipaths

A ***path*** from vertex *a* to vertex *b* is an ordered sequence

$a=v_0, v_1, \ldots, v_m=b$

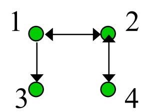
of distinct vertices in which each adjacent pair (v_{j-1},v_j) is linked by an ***arc***. The ***length*** of the path is *m*

A ***semipath*** from vertex *a* to vertex *b* is an ordered sequence

$a=v_0, v_1, \ldots, v_m=b$

of distinct vertices in which either (v_{j-1},v_j) and/or (v_j,v_{j-1}) is linked by an ***arc***. The ***length*** of the semipath is *m*

e.g.



there is a path from 2 to 3 of length 2
there is no path from 3 to 4
there s a semipath from 3 to 4 of length 3

©Department of Psychology, University of Melbourne

Strong and weak connectedness; strong and weak components

If there is a path from vertex *a* to vertex *b*, *b* is ***reachable*** from *a*

If each vertex in *G* is reachable from each other vertex, then *G* is ***strongly connected***

If there is a semipath from each vertex in *G* to each other vertex, then *G* is ***weakly connected***

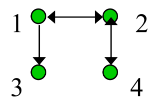
A ***strong (weak) component*** of *G* is a maximal strongly (weakly) connected subgraph (ie a strongly (weakly) connected subgraph with vertex set *W* for which no larger set *Z* containing *W* is strongly (weakly) connected)

©Department of Psychology, University of Melbourne

Geodesics

A ***geodesic*** from *a* to *b* is a path of minimum length
The ***geodesic distance*** d_{ab} between *a* and *b* is the length of the geodesic
If there is no path from *a* to *b*, the geodesic distance is ***infinite***

For the directed graph



The geodesic distances are:

$d_{12} = 1$, $d_{13} = 1$, $d_{14} = 2$, $d_{21} = 2$ $d_{23} = 2$, $d_{24} = 1$,
d_{31} = infinite, d_{32} = infinite, d_{34} = infinite, $d_{41} = 2$, $d_{42} = 1$, $d_{43} = 3$

©Department of Psychology, University of Melbourne

Centrality in directed graphs

As for graphs, but note that:

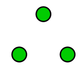
Indegree and ***outdegree*** centrality replace degree centrality

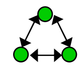
Eigenvector centrality is only computed by UCINET for graphs

©Department of Psychology, University of Melbourne

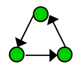
Some special directed graphs

Empty and ***complete*** directed graphs

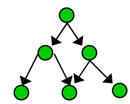




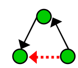
Cycle



Acyclic directed graph: a directed graph with no cycles



Transitive directed graph: every two path is accompanied by a direct path



©Department of Psychology, University of Melbourne