Introduction to Graph Theory

Dokuz Eylül University
Engineering Faculty
Computer Engineering Department
14 March – 26 May, 2000

□ Contents:

□ Basic Principles

Definitions:

- vertex (*pl. vertices*) & edge
  (node & arc)

- multiple edges & loop
- simple graph

- connected/disconnected graph

- subgraph

- degree of a vertex
- regular graph
  – all vertices have a same degree

- isomorphic

- labeled/unlabeled

□ Basic Principles cont'd

- Handshaking Lemma:

$$\sum_{i=1}^{N} (\deg Q_i) = 2N.$$

- Corollary 1:

  The sum of degree of all vertices is always even.

- Corollary 2:

  The number of edges with odd degree is even.

- Corollary 3:

  The number of arcs in a regular graph whose degree is r and number of vertices is n

  $\Downarrow$ is

  $nr/2$

□ Basic Principles cont'd

• number of non-isomorphic simple graphs

- unlabeled

number of vertices	1	2	3	4	$\cdots$
number of graphs	1	2	8	64	$\cdots$

* The number of all the possible arcs is $n(n-1)/2$
$\Rightarrow$ exist or not $\Rightarrow 2^{n(n-1)/2}$

- labeled

number of vertices	1	2	3	4	5	$\cdots$
number of graphs	1	2	4	11	34	$\cdots$
number of connected ones	1	1	2	6	21	$\cdots$
number of regular ones	1	2	2	4	3	$\cdots$

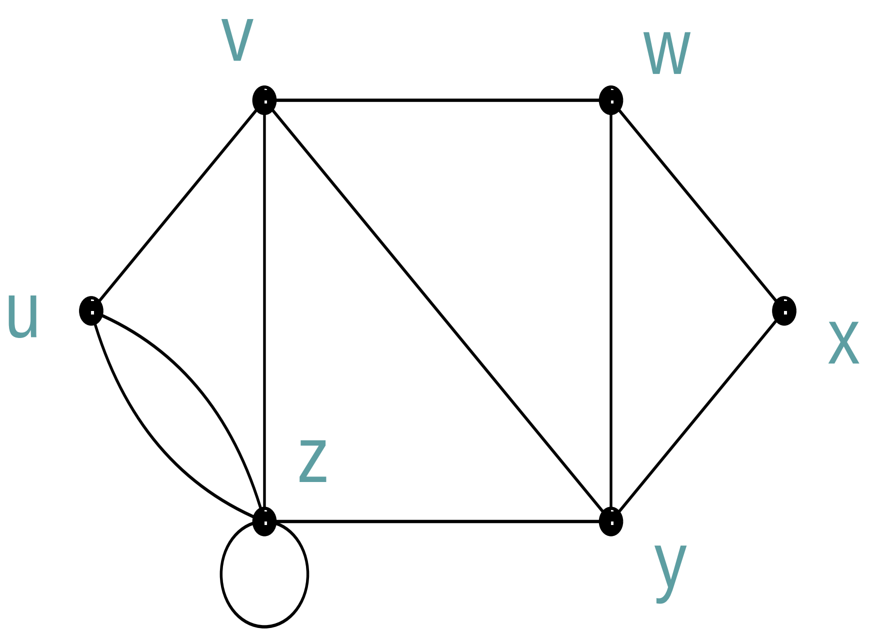
* No such formula exist.

□ Euler & Hamilton Graph

- Definitions:
  - walk of length k
  - walk between P and Q

  - trail
    - a walk made up of different arcs (but may include the identical nodes.)
  - path
    - A walk made up of different arcs and different nodes.

  - closed walk
  - closed trail
  - closed path ⇒ called "cycle"



walk of length 3: uv,vw,wy → uvwy

walk between u and z: uvwyz

trail: uvywx

path: uvywvz

□ Hamilton and Euler Graph

- **Travelers Problem**
  - Travelers are interested in visiting cities as many as possible.
  - They must visit every city once and only once but may pass some arcs.

  So the problem is to search for

  - *cycle(s) that includes all nodes in the graph.*

    $\Downarrow$ if such cycle(s) exists, it is called

    Hamilton Graph or **Hamiltonian**,
    and the cycle is called Hamilton cycle.

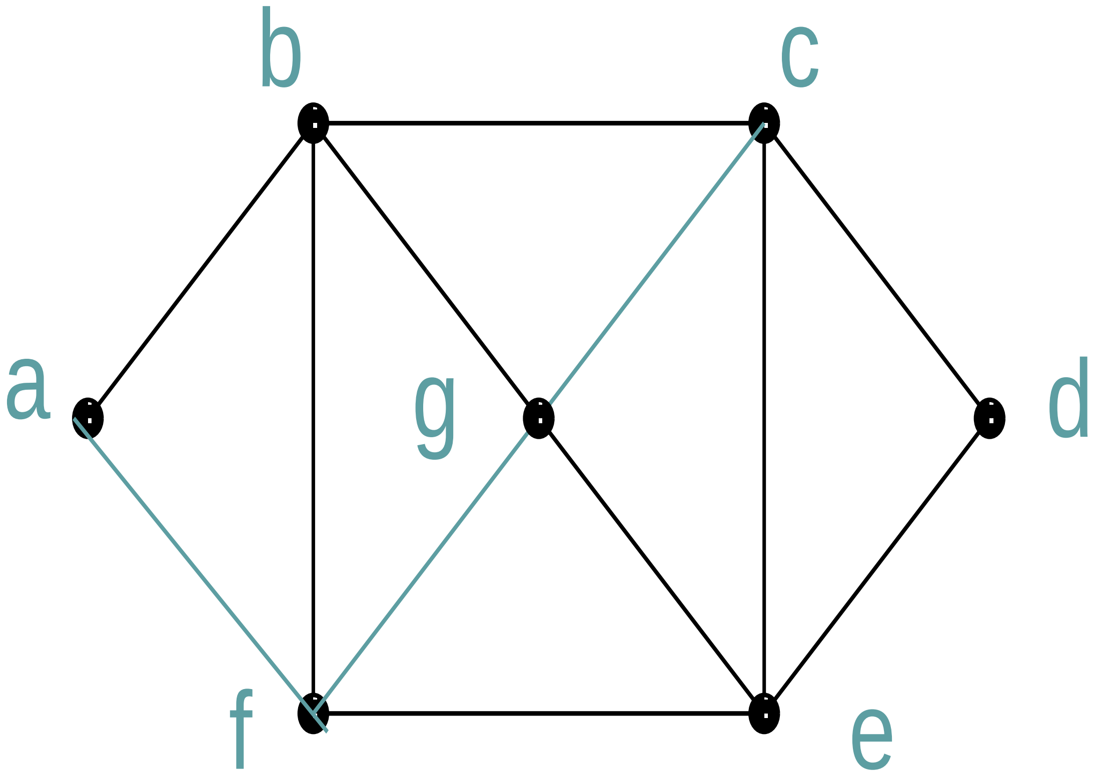
- Explorers Problem
  - Explorers are interested in trails instead.
  - They may visit same city multiple time but only once for one arc.

  So the problem is to search for

  - *closed trail(s) that includes all arcs in the graph.*

    $\Downarrow$ if such trail(s) exists, it is called

    Euler Graph or **Eulerian**.



Hamilton cycle: abcdegfa

Eulerian trail: abcdefbgcegfa

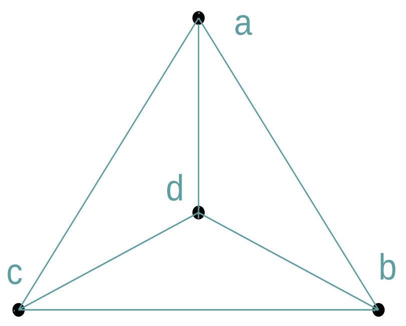
□ Theorem:

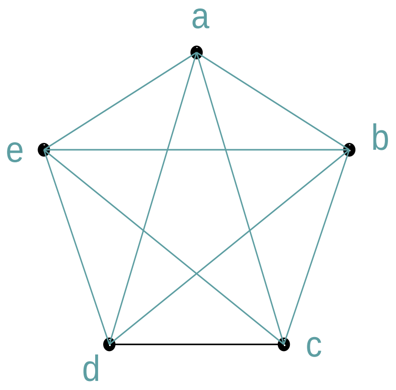
- Connected graph G is Eulerian iff (if and only if) every node has an even order.

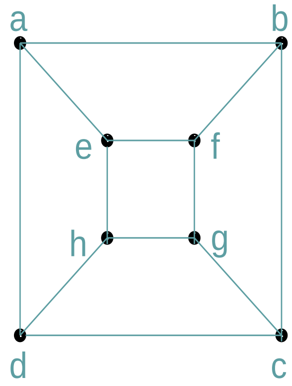
Note:
There's no such theorem for Hamiltonian.

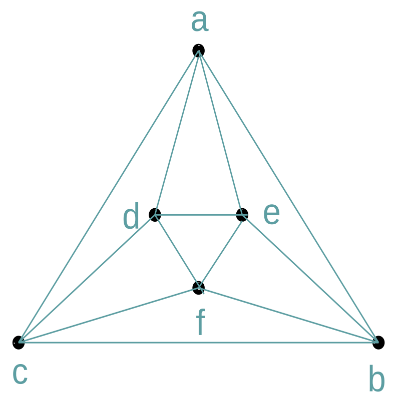
□ Exercise:

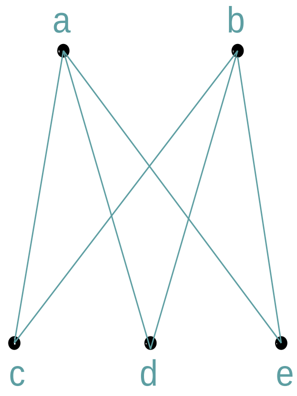
- Which graphs of the follows
  are Hamiltonian and which one are Eulerian?

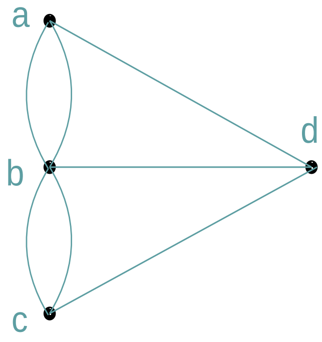

a
d
c
b


a
e
b
d
c


a
b
e
f
h
g
d
c


a
d
e
f
c
b


a
b
c
d
e


a
d
b
c

□ Chinese Postman Problem:

Since this was proposed by a Chinese mathematician named Meign Guan (1962), this is called Chinese ⋯. not a problem regarding ḫinese postman”

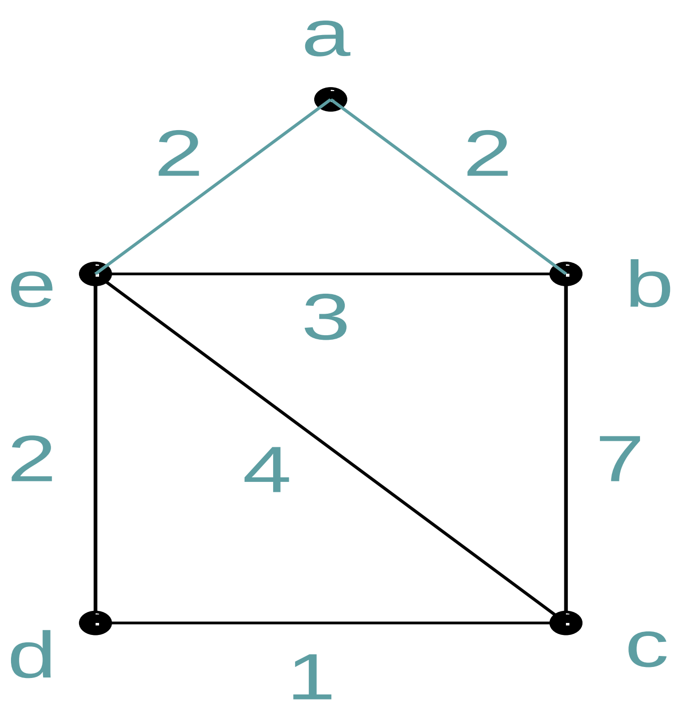
- A postman starts his post office to deliver letters passing every street, and return to the office with the total distance being minimum.

□ Traveler’s Salesman Problem (TSP):

- A traveler’s salesman starts a city visiting every city once and only once with the total distance being minimum.

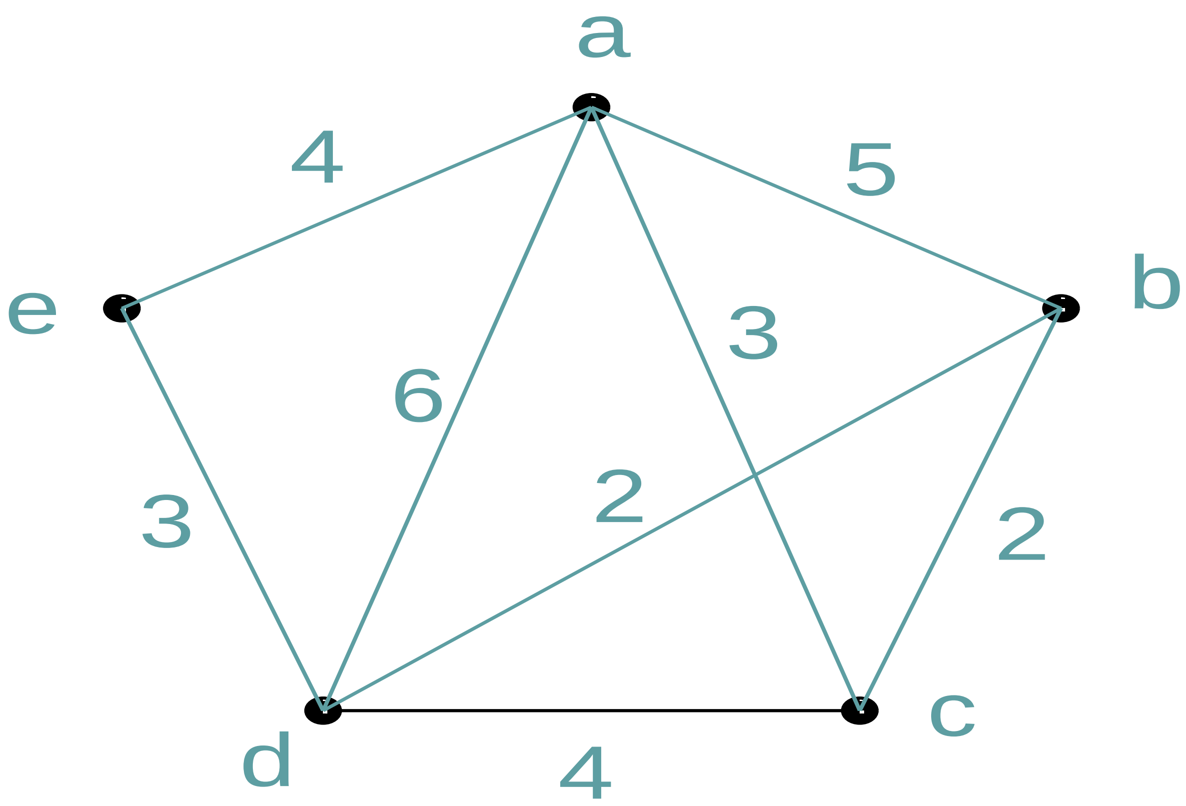
□ The Shortest Path Problem:

□ The Longest Path Problem:



a postman problem's solution:

abcdedcebea



a TSP's solution: acbdea

□ The Shortest Path Algorithm.

Notation:
S: the node to start with, T: The goal;
$E(U)$: energy of node U (the shortest distance from S)
$e(U)$: energy of node U

Step 1

- Assign energy zero to S: $E(S) = 0$.
- Assign label ∞ to all nodes other than S: $e(S) = \infty$.

Step 2

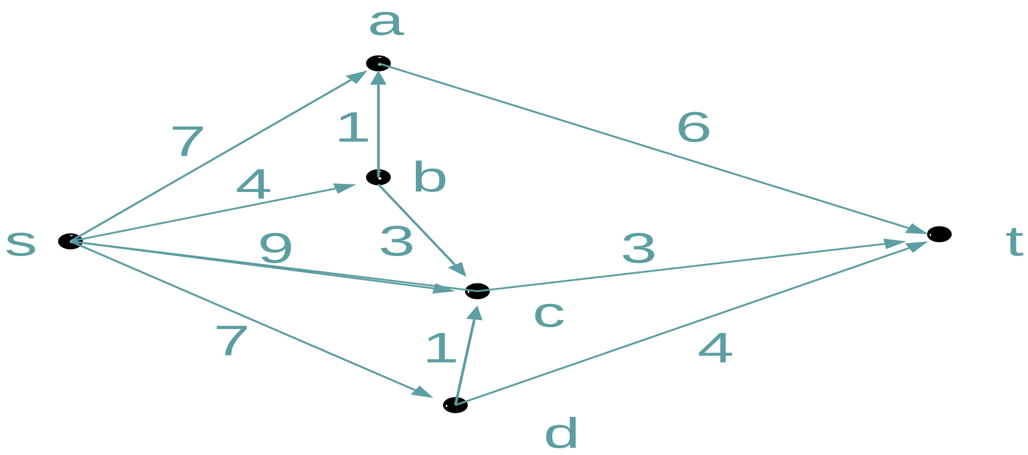
- FOR each of the nodes U_i that was already assigned an energy
  - FOR each V_j that are directly reachable from U_i
    - IF $E(U_i) + r(U_i, V_j) < e(W_j)$
      THEN $e(V_j) = E(U_i) + r(U_i, V_j)$
  - Find a minimum value of labels among nodes that were not assigned an energy yet, and assign energy of that value to the point(s).

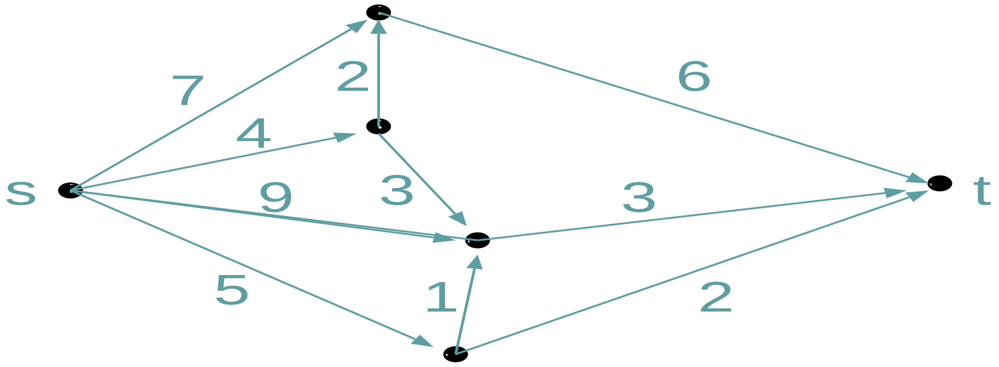
Step 3

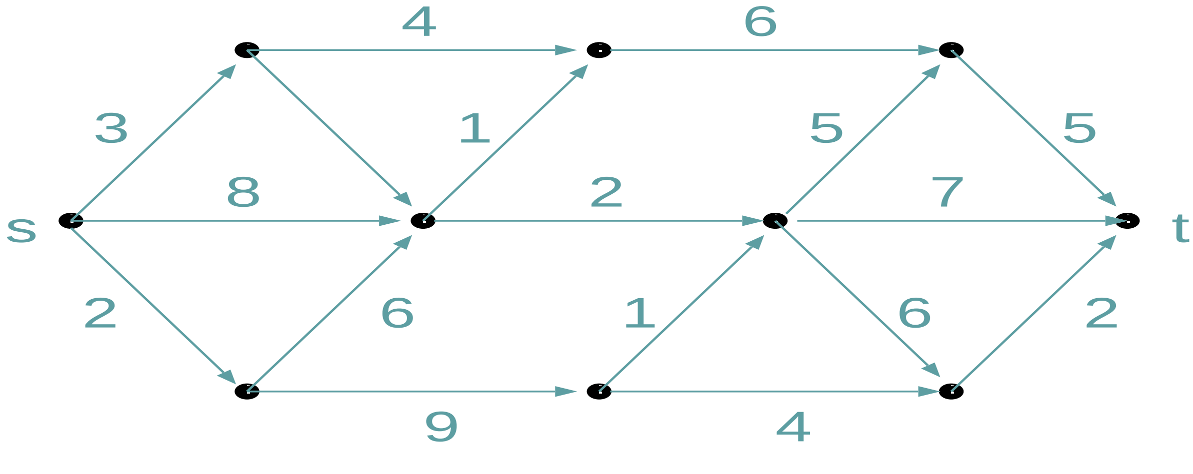
- Repeat step 2 until all nodes are assigned an energy.

Step 4

- Pick up arcs when $E(A) - E(B) = r(A, B)$


a
7
1
6
4
b
s
9
3
3
t
c
7
1
4
d


7
2
6
4
s
9
3
3
t
5
1
2


4
6
3
1
5
5
8
2
7
s
t
2
6
1
6
2
9
4

Chapter 2

Graph Theory for Networking.

□ Network Flow Problem

- Traveler Problem
- Explorer Problem

  ⇒ arcs have nothing to do with values.

- (Chinese) Postman Problem
- Traveler's Salesman Problem
- The Shortest Path Algorithm
- The Longest Path Algorithm

  ⇒ each arcs is assigned a value corresponding to **distance**.

- Network Flow Problem

  ⇒ each arc has a value corresponding to **capacity**.

□ Definitions:

- Bipartite graph:
  - If every node belongs to either A or B.
- Complete bipartite graph:
  - If every node that belongs to A is connected all nodes belongs to B with one and only one arc.
- Complete matching:
  - If *one-to-one-mapping* from A to B exists.

□ What condition is needed for complete matching to exist?

$\Downarrow$

—— Marriage Problem ——

Each boy is a friend of some girls.
What condition is needed
for every boy to be able to get married
with his girl friend?

$\Downarrow$ e.g.,

b_1 knows g_1, g_4 and g_5
b_2 knows g_1
b_3 knows g_2, g_3 and g_4
b_4 knows g_2, g_4

□ Theorem of Marriage (P.Hall 1935)

The condition for the Marriage Problem to exist is that k boys have a total of more than k girl friends for any k that is $1 \leq k \leq m$, where m is the number of boys.

- Example 2:

When

- a knows w, y and z
- b knows x and z
- c knows x and y

answer the following questions.

(1) Draw the bipartite graph.

(2) Find solutions as much as possible, say 5.

(3) Apply the Marriage Theorem.

- Example 3:
  - A university looks for 4 professors of Mathematics Physics, Software Engineering, and Graph Theory. There are 5 candidates who can teach
    * the 1st $\Rightarrow$ Math.
    * the 2nd $\Rightarrow$ Phys.
    * the 3rd $\Rightarrow$ Software Eng.
    * the 2nd $\Rightarrow$ Graph Th.

    Then, answer the following questions.

    (1) Draw the bipartite graph.
    (2) Find solutions.
    (3) Apply the Marriage Theorem.

- Example 4:
  - How about the follows?
    * a knows y and u
    * b knows x, z and v
    * c knows y and u
    * d knows y and u

□ Transversal

The sets of girls known by the four boys:

$$\{g_1, g_4, g_5\}, \{g_1\}, \{g_2, g_3, g_4\}, \{g_2, g_4\}$$

$\Downarrow$ solution is to find

four distinct g_i, one from each of these sets of girls.

$\Downarrow$ In general,

If E is a non-empty finite set, and if $\xi = (S_1, S_2, \cdots, S_m)$ is a family of non-empty subsets of E (not necessarily distinct.) Then, a transversal of ξ is a set of m distinct elements of E, one chosen from each set S_i.

□ Example

$E = \{1, 2, 3, 4, 5, 6\}$
$S_1 = S_2 = \{1, 2\}, S_3 = S_4 = \{2, 3\}, S_5 = \{1, 4, 5, 6\}$

□ Theorem:

- Let E be a non-empty finite set, and let $\xi = (S_1, S_2, \cdots, S_m)$ be a family of nin-empty subsets of E. Then ξ has a transversal if and only if the union of any k of the subsets S_i contains at least k elements $(1 \leq k \leq m)$

□ Partial Transversal

In the previous example:

$E = \{1, 2, 3, 4, 5, 6\}$
$S_1 = S_2 = \{1, 2\}, S_3 = S_4 = \{2, 3\}, S_5 = \{1, 4, 5, 6\}$

ξ has no transversal.

$\Downarrow$ However

Subfamily (S_1, S_2, S_3, S_5) has a transversal, e.g., $\{1,\ 2,\ 3,\ 4\}$.

$\Downarrow$ A transversal of a subfamily of ξ is calld a

partial transversal of ξ

$\Downarrow$ Theorem

- ξ has a partial transversal of size t iff the union of any k of the subsets S_i contains at least $k + t - m$ elements.

$\Downarrow$

A condition under which at least t boys can marry girls known to them.

□ Toward Menger's Theorem:

- definitions
  - * **edge-disjoint path**

    paths between two nodes no two of which have an edge in common.

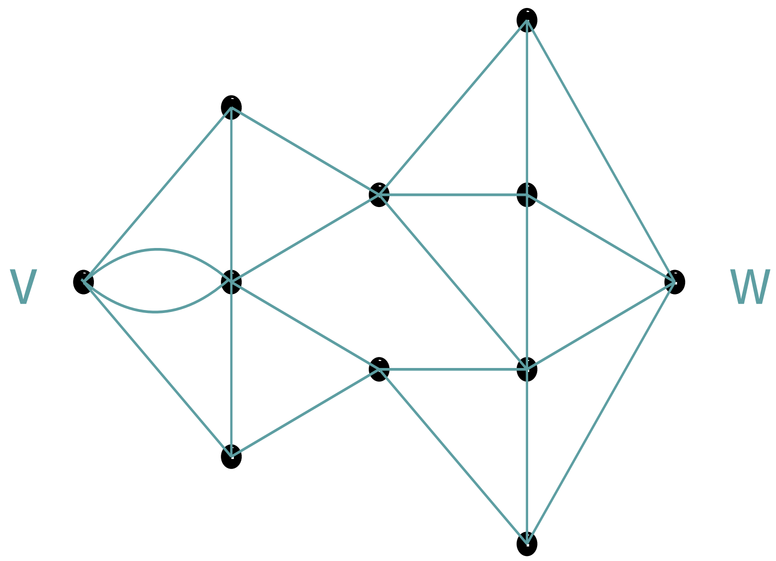
  - * **vertex-disjoint path**

    no two of which have an vertex in common.

- What is the maximum number of these paths?

Find edge/vertex-disjoint-path

as many as possible.



- definitions 2
  - * vw-**disconnecting set of** G

    a set E of edges of G such that each path from v to w includes an edge of E

  - * vw-**separating set of** G

    a set of S of vertices (other than v and w) such that each path from v to w passes through a vertex of S

□ Two versions of Menger's Theorem:

- Edge form:
  - The maximum number of edge-disjoint paths connecting two distinct vertices v and w of a connected graph is equal to the minimum number of edges in a vw-disconnecting set.

- Vertex form:
  - The maximum number of vertex-disjoint paths connecting two distinct vertices v and w of a connected two distinct non-adjacent vertices v and w of a graph is equal to the minimum number of vertices in a vw-separating set.

- definitions 3
  - * **k-edge-connected**

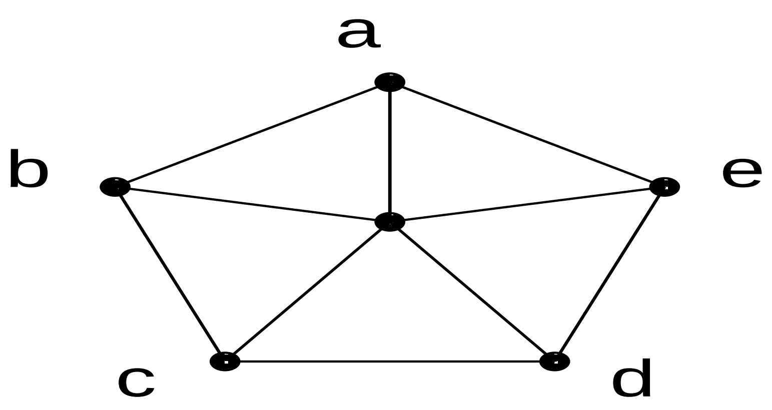
    If the minimum number of edges that we need to delete in order to disconnect $G \geq k$, G is calld k-edge-connected

- Corollary 1
  - * A graph G is k-edge-connected iff any two distinct vertices of G are connected by at least k edge-disjoint pahts.

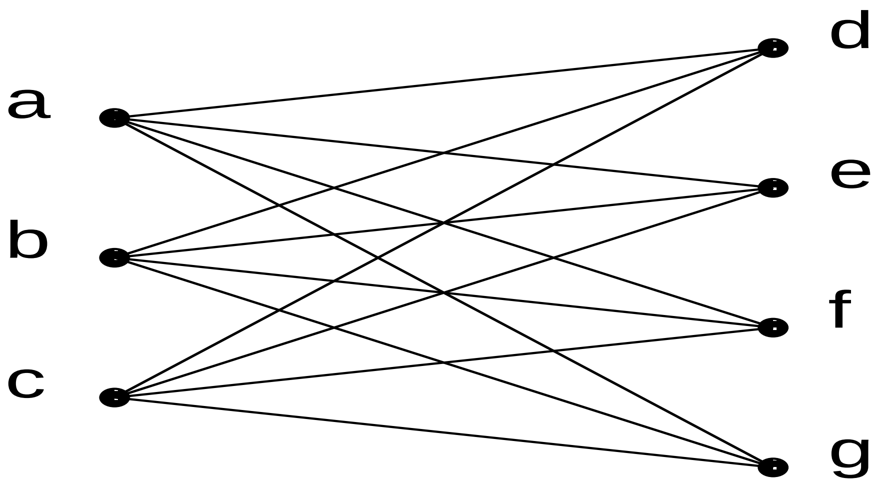
- Corollary 2
  - * A graph G with at least $k+1$ vertoces is k-edge-connected iff any two distinct vertices of G are connected by at least k vertices-disjoint pahts.

Ascertain the corollary 1

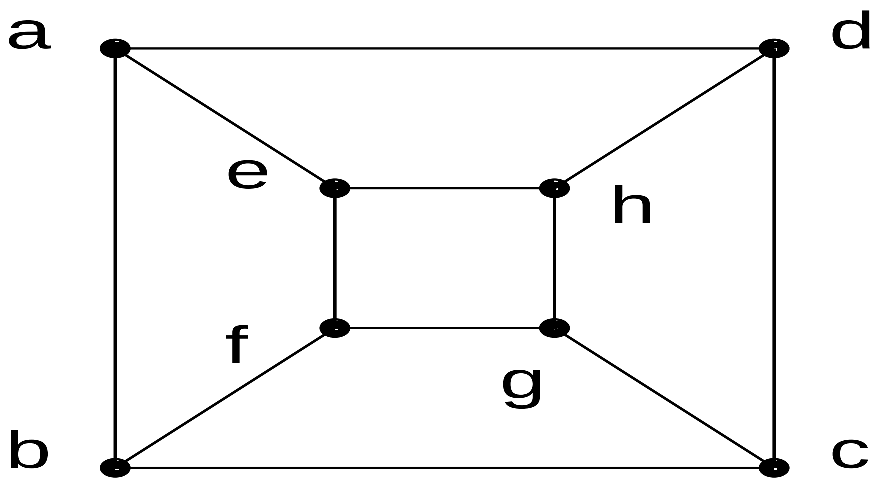
1)



2)

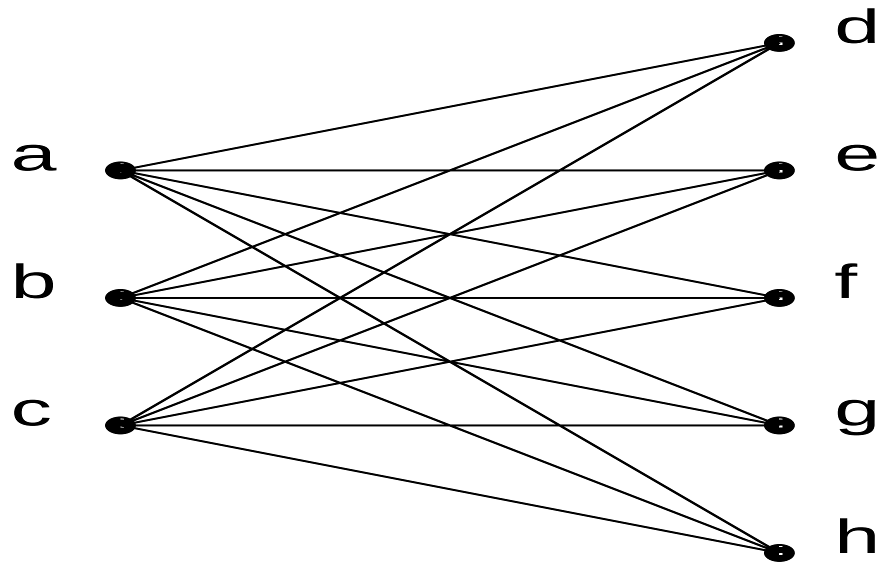


3)

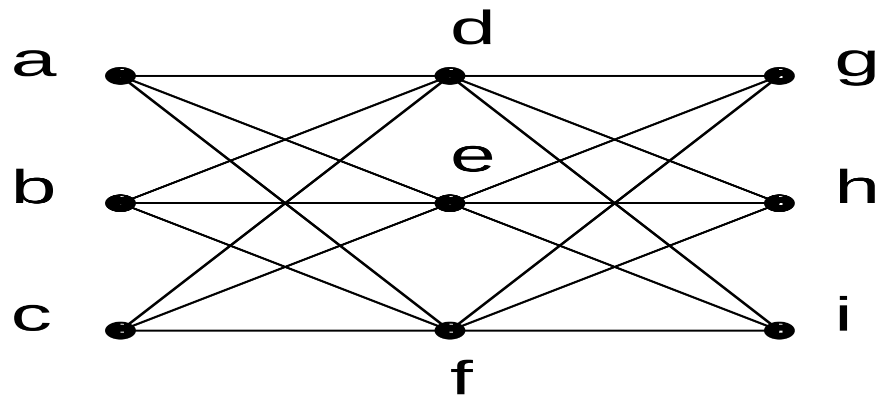


Ascertain the corollary 2

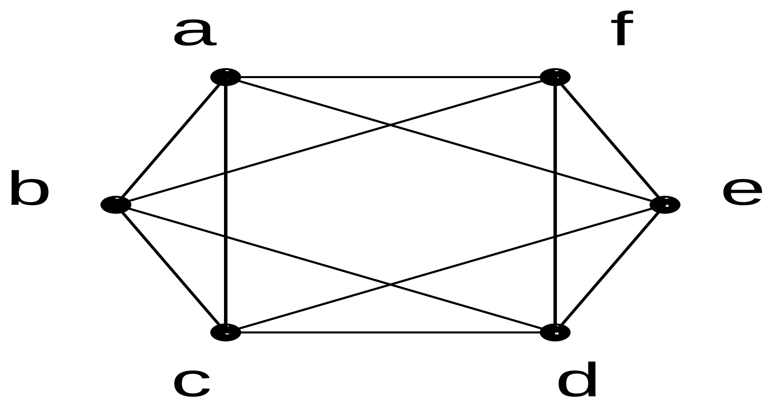
1)



2)



3)



- Theorem
  - * The maximum number of arc-disjoint path from a vertex v to a vertex w in a directed graph is equal to the minimum number of arcs in a vw-desconnecting set.

- Theorem
  - * Menger's theorem implies Hall's theorem.

$\Downarrow$ Hints for proof:

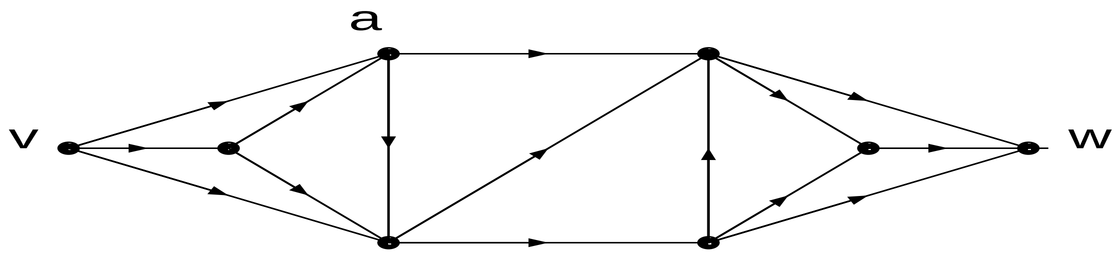
Hall's Theorem $\Rightarrow$ bipartite graph
while
Menger's Theorem $\Rightarrow$ general graph

$\Downarrow$ so

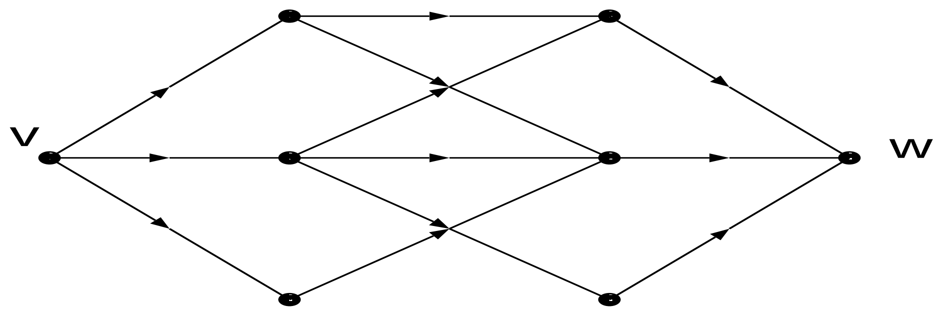
1. Add two nodes v, w to the bipartite graph.
2. Think what does it mean by *edge disjoint path.*

Ascertain the Menger's theorem

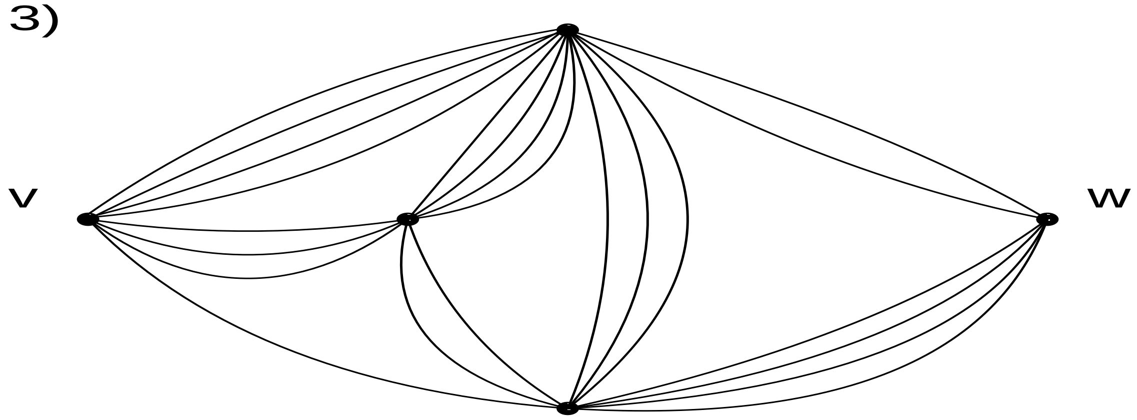
1)

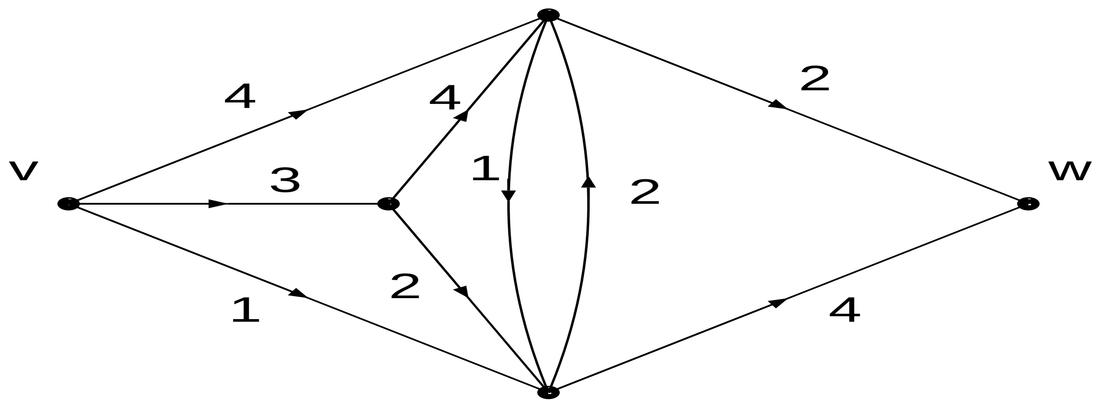


2)



3)





□ Network Flows:

Definitions:

- network
  - a directed graph to each arc of which is assigned a non-negative real number $\Pi(a)$, called its capacity.
- out-degree
  - out-degree(x) of a node x is the sum of the arcs of the form xz, i.e.,
    * out-degree$(x) = \Sigma_z\, xz$
- in-degree
  - similarly,
    * in-degree$(x) = \Sigma_z\, zx$
- flow
  - a flow in a network is a function π that assigns to each arc p a non-negative real number $\pi(p)$ in such a way that

    (1) for each arc p
    * $\pi(p) \leq \Pi(p)$

    (2) in-degree and out-degree of each node (except for v and w) are equal
- cut
  - a set A of arcs such that each path from v to w includes an arc in A

□ Network Flows: cont'd

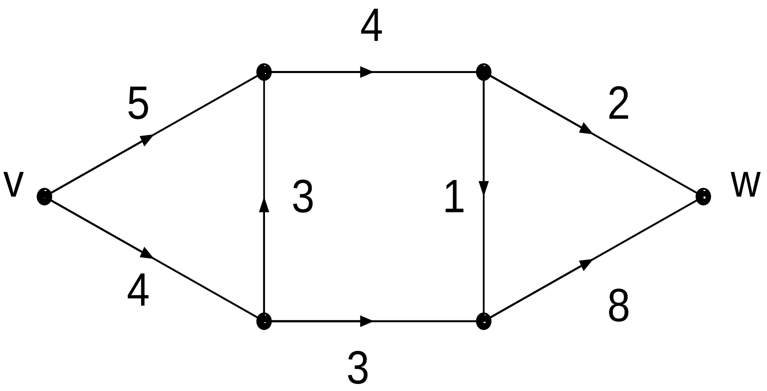
Definitions:

- zero flow
  - the flow in which every arc is zero
- saturated
  - an arc a for which $\pi(a) = \Pi(a)$ is called saturaed
- maximum flows
  - flows whose value is as large as possible
    - although a network can have several different maximum flows, their values must be equal.
- minimum cut
  - a cut in a network is a vw-disconnecting set in the corresponding directed graph
  - the sum of the capacities of the arcs in the cut is called capacity of a cut
  - those cuts whose capacity is as small as possible

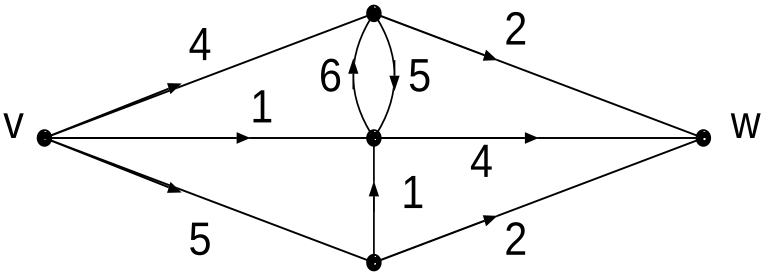
    $\Downarrow$ called
  - minimum cuts

- List all the cuts, and find a minimum one.
- Find a maximul flow and verify the max-flow min-cut theorem.

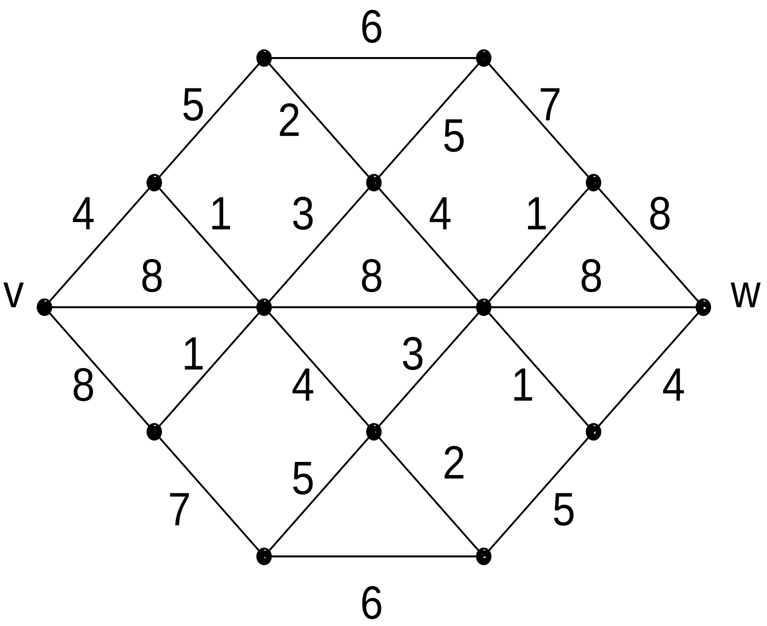
1)



2)



3)



□ An Algorithm for the Muximum Flows:

1. Choose a path, regardless of direction of arcs,

$$\sum_{i=v}^{w} (e_{ij})$$

where each e_{ij} has a value f_{ij}.

This flow is called *flow-augmenting* iff

- when e_{ij} is a same direction as the arc
  - ⋆ $f_{ij} < c_{ij} \quad \Rightarrow \quad \epsilon_{ij} = c_{ij} - f_{ij}$
- when opposite direction
  - ⋆ $f_{ij} > 0 \quad \Rightarrow \quad \epsilon_{ij} = f_{ij}$

---

2. Obtain

$$\epsilon_{\text{min}} = \min\{\epsilon_{ij}\}$$

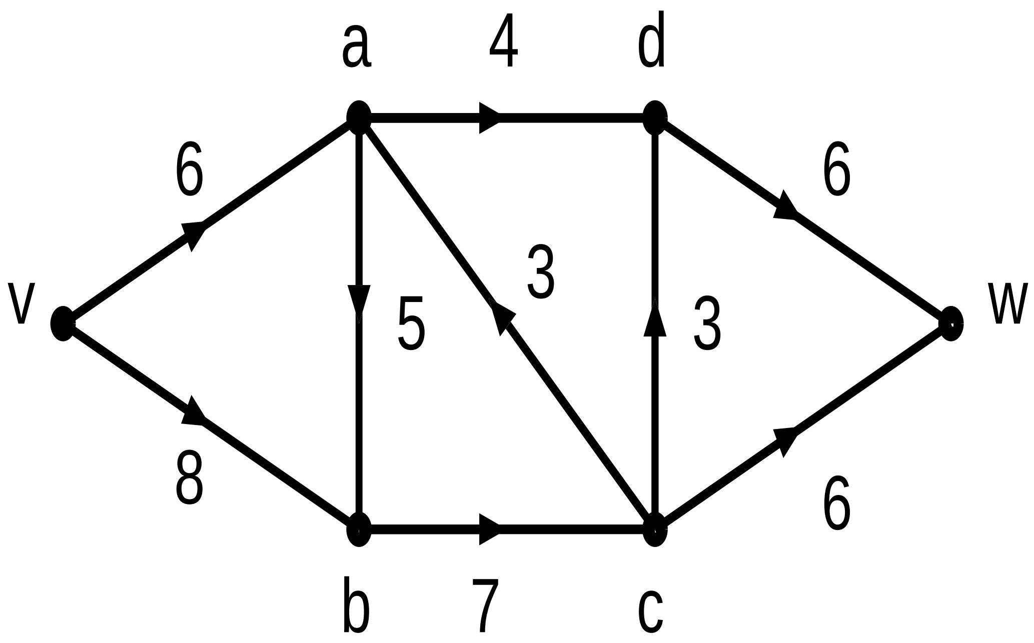
---

3. renew f_{ij} as:

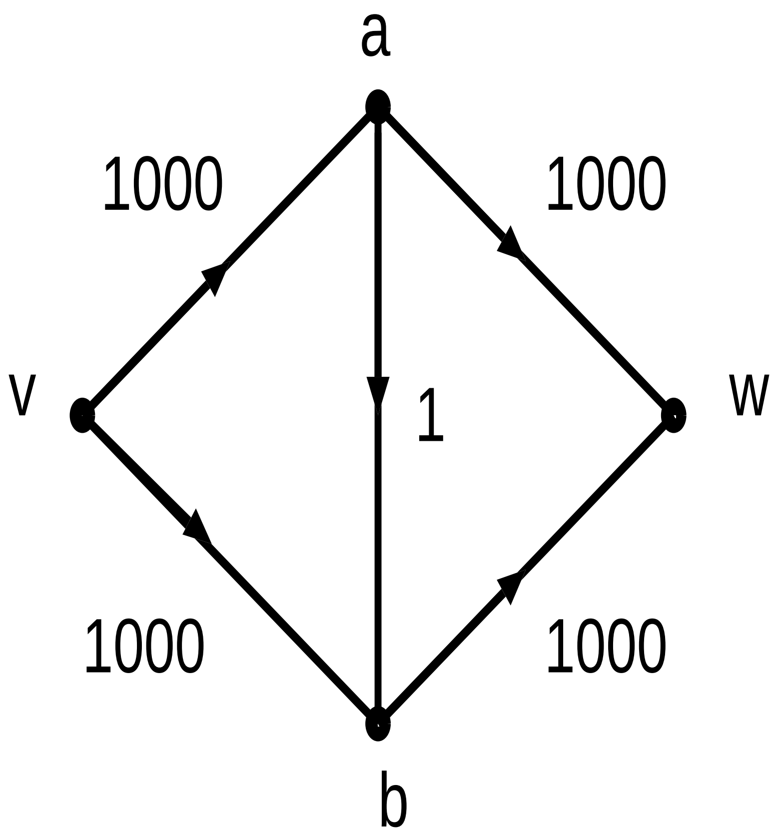
- when e_{ij} is a same direction as the arc
  - ⋆ $f_{ij} = f_{ij} + \epsilon_{\text{min}}$
- when opposite direction
  - ⋆ $f_{ij} = f_{ij} - \epsilon_{\text{min}}$

- Apply the algorithm to find a maximum flow.

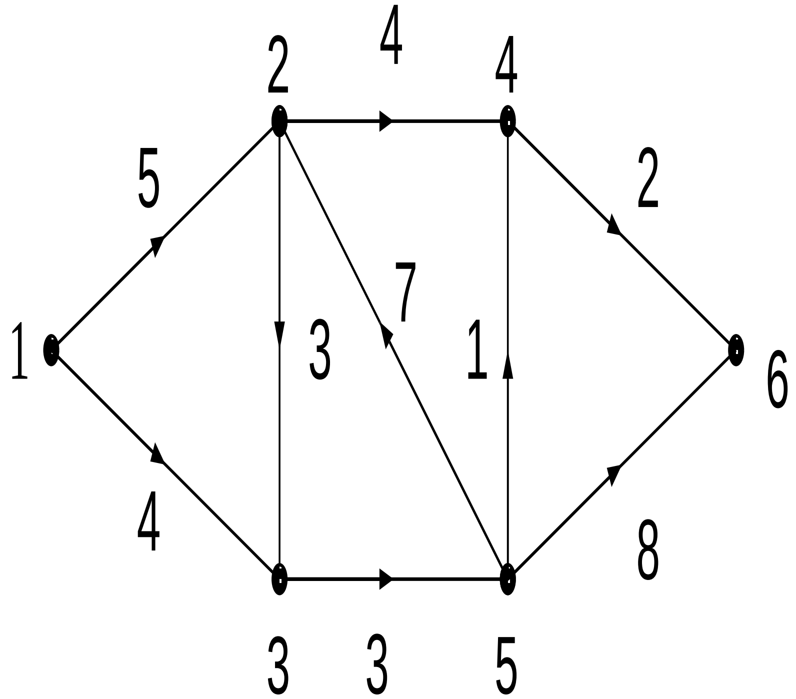
1)



2)



How we express a network?



example
of
the input

1 2 5
1 3 4
2 3 3
2 4 4
3 5 3
4 6 2
5 2 7
5 4 1
5 6 8

□ Homework:

- Create a program in C to find a max-flow of **any** network whose number of nodes and arcs are not exceeding 40.
- Input, which is given as a set of connection and capacities, should be fed by "*redirection*".
  - Note that I will check the code under LINUX with using some of my test cases.
- Output should be a max-flow.
- E-mail to akira@cs.deu.edu.tr with attaching two files:
  - source code (in ASCII text)
  - description in POST-SCRIPT

  by the due-date of May 26 (Friday)

□ Toward a Concept of Network Reliability:

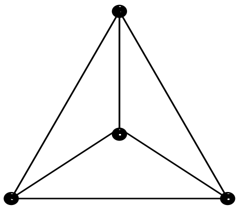
Definitions:

- Complete graph $\Rightarrow K_n$
  - When all possible two distinct nodes are connected with each other.
  - Complete graph with n nodes is denoted as K_n
  - K_n has $\{n(n-1)/2\}$ arcs.
- Bipartite graph $\Rightarrow K_{n,m}$
- Edges/Vertices deletion $\Rightarrow G - e/G - v$
- Contraction $\Rightarrow G/e$
  - Remove an edge and identify its both ends.
- Plane graph
  - A graph whose arcs never cross.
- Planar graph
  - A graph isomorphic to a plane graph.

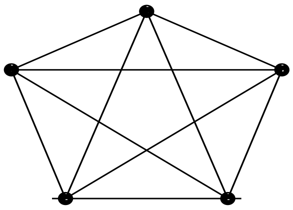
Note:

In the 3-D space, every graph can be drawn in such a way.

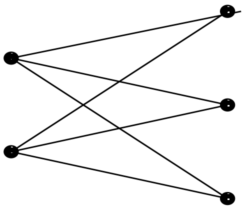
● K_n and $K_{i,j}$



K_4

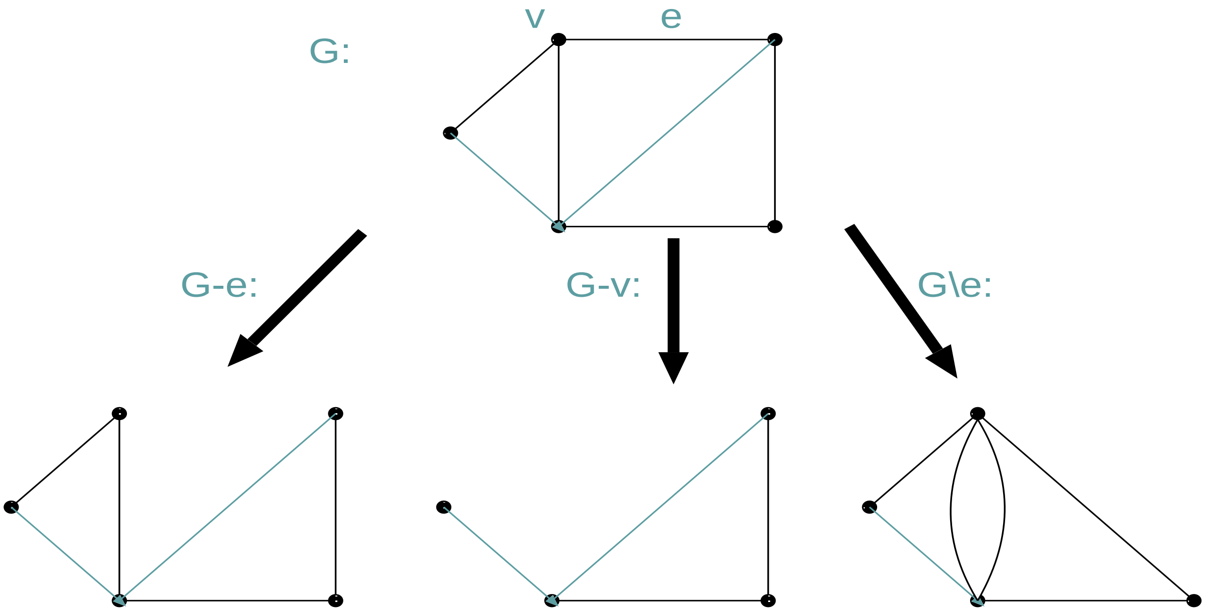


K_5



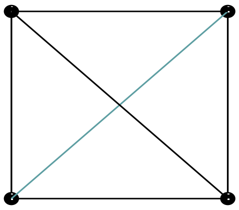
$K_{2,3}$

● deletion and contraction

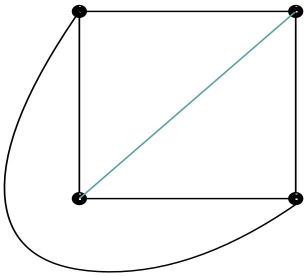


● planar and plane graph

·planar



·plane



Theorems:

- $K_{3,3}$ and K_5 are not planar.

- A graph is planar iff it cantains no subgraph isomorphic to $K_{3,3}$ and K_5 (Kuratowski).

- A graph is planar iff it contains no subgraph contractible to $K_{3,3}$ and K_5.

Connectivity revisit:

- Vertex connectivity $\Rightarrow \kappa(G)$

  (for $\forall k < \kappa(G) : G$ is k-vertex connected)

  - communication is guaranteed even if terminals less than $\kappa(G)$ are out-of-order.
  - The more the arcs, the more reliable the connectivity

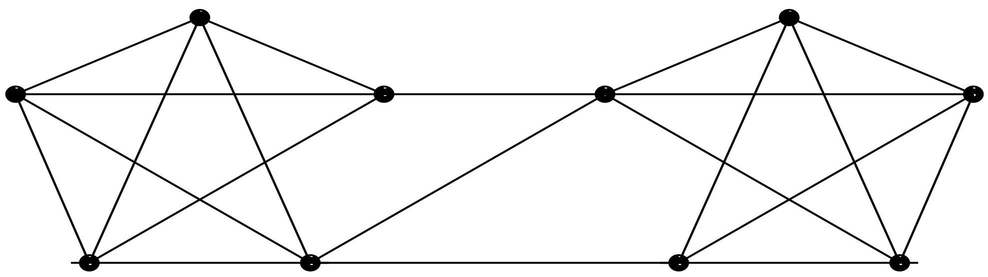
    $\Downarrow$ But from financial reason, we want

    a network as reliable as possible

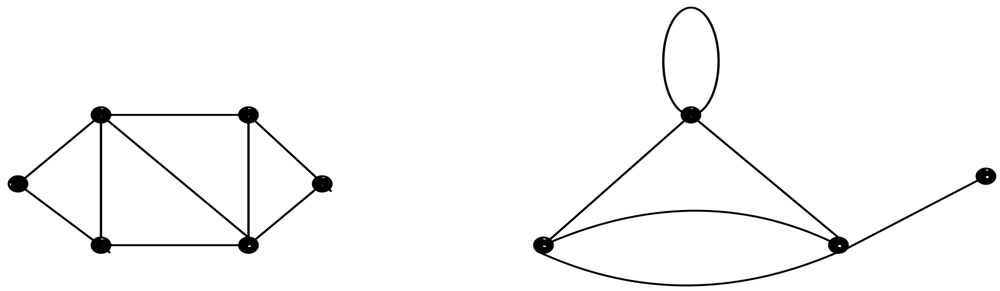
    with arcs as few as possible.

- Edge connectivity $\Rightarrow \lambda(G)$
  - same as above

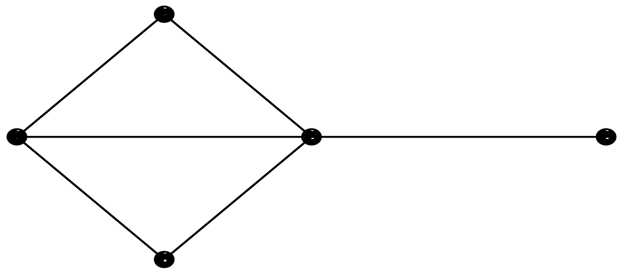
- Which value is larger?



- Euler's Formula



- dual graph



- Which network is more reliable?

Which Connectivity is larger?

(Compare also with minimum degree $\delta(G)$ and average degree $a(G) = \{\text{total-degree}\}/\{\text{number of nodes}\} = 2m/n$).

- Theorem

$$\kappa(G) \le \lambda(G) \le \delta(G) \le a(G).$$

Edge-connectivity of plane graph

- Euler's Formula
$$n - m + f = 2.$$
- Graph may include loop, multiple edges.

- What is the maximum number of f for fixed n?

$$\text{e.g. when } n = 4.$$

- When all the *faces* are bounded with 3 arcs:

$$m = \frac{3f}{2} \rightarrow f = \frac{2m}{3} \rightarrow n - m + \frac{2m}{3} = 2 \rightarrow m = 3n - 6.$$

- In general

$$m \leq 3n - 6$$

- So,

$$a(G) = \frac{2m}{n} \leq \frac{6n - 12}{n} = 6 - \frac{12}{n},$$

- and hence,

$$a(G) < 6.$$

- Therefore

$$\kappa(G), \lambda(G) \leq 5.$$

- Dual graph of G $\Rightarrow G^*$

1. Choose a point v^* inside each face of G.

   $\Downarrow$

   These points will be nodes of G^*

2. Draw a line e^* so that it cross each edge e of G, and joins the node v^* in the face f adjoining e.

- Edge connectivity of plane graph

if	G^* has a loop	$\lambda = 1$
else if	G^* has a multiple arc	$\lambda = 2$
else if	G^* has a face surrounded by 3 edges	$\lambda = 3$
else if	G^* has a face surrounded by 4 edges	$\lambda = 4$
else		$\lambda = 5$

□ Diameter, Radius, Center and Periphery of a Graph

· From now on, we consider undirected graph.

· What we imagine is a trafic network with weight being capacity of the road, but taking no notice of weights for the time being.

- **Distance** $d(x, y)$:
  - the minimumu number of arcs from node x to node y.
    - this fulfill the axiom of distance, i.e.,
      - · $d(x, x) = 0$
      - · $d(x, y) = 0 \Rightarrow x = y$
      - · $d(x, y) = d(y, x)$
      - · $d(x, y) + d(y, z) > d(x, z)$

- **Eccentricity** (Longest distance)
  - for $\forall x$, $\forall y$, $e(x) = \max d(x, y)$ is called *eccentricity*

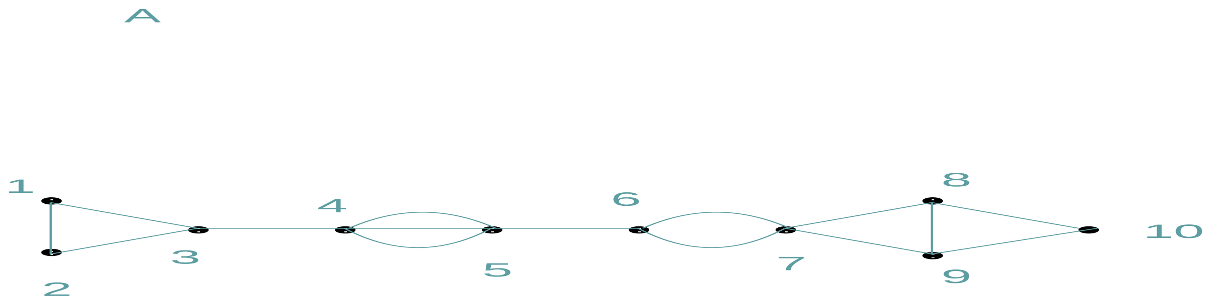
- **Radius** $R(G)$
  - for $\forall x$ $\min(e(x))$

- **Diameter** $D(G)$
  - for $\forall x$ $\max(e(x))$

- **Center**
  - Points where $e(x) = R(G)$ are called *center*.

- **Periphery**
  - Points where $e(x) = D(G)$ are called *periphery*.

• An example

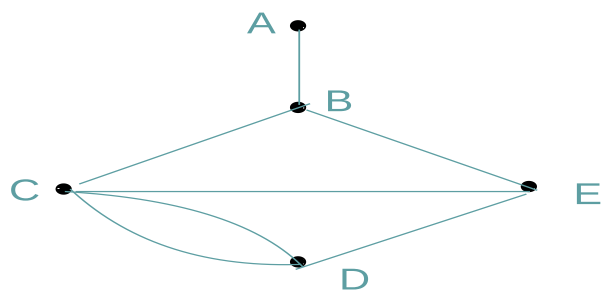


σ	1	2	3	4	5	6	7	8	9	10
$e(v)$	7	7	6	5	4	4	5	6	6	7

$$\Downarrow$$

$$R(G) = 4$$
$$D(G) = 7$$
$$\text{center} \Rightarrow 5,\ 6$$
$$\text{periphery} \Rightarrow 1,\ 2,\ 10$$

- Adjacency matrix A



- What does $(A^n)_{ij}$ mean?
  - the number of path from node i to node j whose distance is n.
  - example

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 1 & 1 & 2 & 0 \end{pmatrix}$$

$A^2 = ?$

$A^3 = ?$

$(A^3)_{2,3} = 7 \quad \Rightarrow \quad$ the number of path from 2 to 3 whose distance is 3.

- $S(r) = A + A^2 + A^3 + \cdots + A^r$.
  - the number of path from i to j whose distance is less than or equal to r.

□ How to obtain $D(G)$?

- Step 1
  - Calculate A^r and $S(r)$ starting with $r = 1$ by incrementing r.
- Step 2
  - When all the non-diagonal element are zero at first time, then r is $D(G)$.

□ Terminal Capacity

A path whose length is i is worth than a length $i+1$ path.

$\Downarrow$

Devide every element A^i by η^i, where usually $\eta = n$

$\Downarrow$

$$W(r) = (A/\eta) + (A/\eta)^2 + (A/\eta)^3 + \cdots + (A/\eta)^n$$

$\Downarrow$

Summation of the i-th column of $W(r)$:
$(W(r))_{1k} + (W(r))_{2k} + (W(r))_{3k} + \cdots + (W(r))_{nk}$

$\Downarrow$ implies

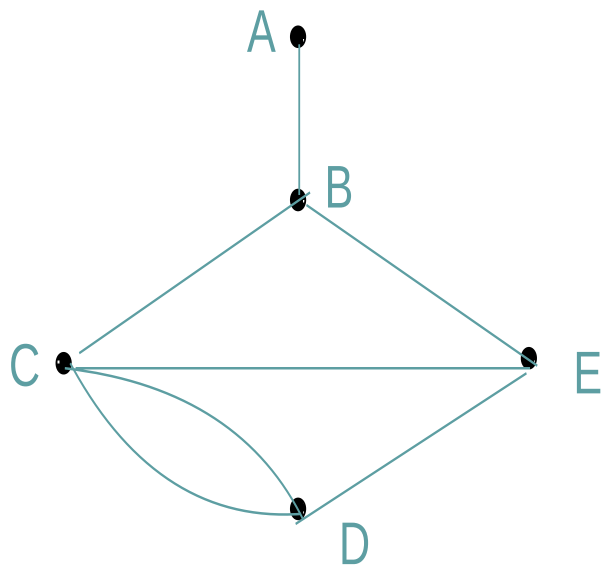
the number of paths to node k whose lenght is less than r from every node of the Graph,

$\Downarrow$ called

Terminal Capacity of Length r

□ An Example of Terminal Capacity

$$W(3) = \frac{1}{125}\begin{pmatrix} 5 & 28 & 6 & 3 & 6 \\ 18 & 17 & 37 & 18 & 40 \\ 6 & 37 & 21 & 44 & 48 \\ 3 & 18 & 44 & 27 & 70 \\ 6 & 40 & 48 & 70 & 36 \end{pmatrix}$$



□ Terminal Capacity for $n \to \infty$

Recall the equation:

$$1 + x + x^2 + x^3 \cdots = \frac{1}{1-x}$$

This corresponds to Matrix relation of:

$$(I - X)(I + X + X^2 + X^3 \cdots) = I$$

So, when $|I - X| \neq 0$

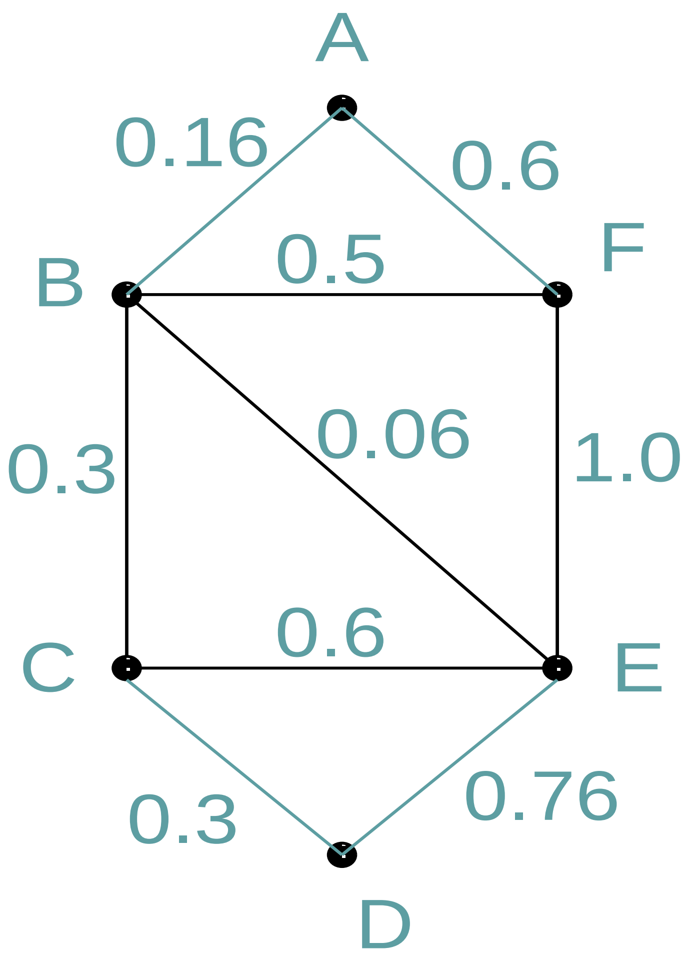
$$X + X^2 + X^3 \cdots = (I - X)^{-1} - I$$

Therefore

$$W(\infty) = \{I - (A/\eta)\}^{-1} - I$$

Note that this does not require calculation of power of matrix.

Which road should be improved?



city	population	capacity
A	3000,000	?
B	4000,000	?
C	6000,000	?
D	8000,000	?
E	1,0000,000	?
F	7000,000	?

$$A = \begin{pmatrix} 0.00 & 0.16 & 0.00 & 0.00 & 0.00 & 0.60 \\ 0.16 & 0.00 & 0.30 & 0.00 & 0.06 & 0.50 \\ 0.00 & 0.30 & 0.00 & 0.30 & 0.60 & 0.00 \\ 0.00 & 0.00 & 0.30 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.06 & 0.60 & 0.00 & 0.00 & 1.00 \\ 0.60 & 0.50 & 0.00 & 0.00 & 1.00 & 0.00 \end{pmatrix}$$

$$W(3) = \begin{pmatrix} 0.0112 & 0.0357 & 0.0034 & 0.0022 & 0.0115 & 0.1068 \\ 0.0357 & 0.0111 & 0.0536 & 0.0062 & 0.0303 & 0.0925 \\ 0.0034 & 0.0536 & 0.0164 & 0.0643 & 0.1134 & 0.0222 \\ 0.0022 & 0.0062 & 0.0643 & 0.0198 & 0.1388 & 0.0223 \\ 0.0115 & 0.0303 & 0.1134 & 0.1388 & 0.0556 & 0.1798 \\ 0.1068 & 0.0925 & 0.0222 & 0.0223 & 0.1798 & 0.0454 \end{pmatrix}$$

$$W(\infty) = \begin{pmatrix} 0.0117 & 0.0363 & 0.0038 & 0.0026 & 0.0190 & 0.1074 \\ 0.0363 & 0.0118 & 0.0541 & 0.0068 & 0.0319 & 0.0933 \\ 0.0038 & 0.0541 & 0.0174 & 0.0653 & 0.1146 & 0.0240 \\ 0.0026 & 0.0068 & 0.0653 & 0.0210 & 0.1400 & 0.0241 \\ 0.0190 & 0.0319 & 0.1146 & 0.1400 & 0.0597 & 0.1812 \\ 0.1074 & 0.0933 & 0.0240 & 0.0241 & 0.1812 & 0.0487 \end{pmatrix}$$

□ Markov Chain

Random Walk

- A drunken walk:
  An example of the 1-diomensional random walk.

  – Denote the place at which he can stop by with
  $$E_1, E_2, \cdots, E_6$$

  – The initial position: E_4

  – Every time he steps to either right, left or stay unmoved with the probability 1/2, 1/3, 1/6, respectively.

  – If he arrives at either E_1 or E_6, he stays there
  $$E_1 \text{ and } E_6 \Rightarrow \text{absorbing state.}$$

  – Then

    * Which of the E_1 and E_6, he is more likely to arrive first?

    * The probabilities of his position
      · after one step $\Rightarrow (0, 0, 1/2, 1/6, 1/3, 0)$
      · after two steps $\Rightarrow (0, 1/4, 1/6, 1/36, 1/9, 1/9)$

- definition

  - initial position vector
  $$\mathbf{x} = (0, 0, 0, 1, 0, 0)$$

  - transition matrix
  $$P = (p_{ij})$$
  where p_{ij}'s are the probability that he moves from E_i to E_j in one step

  In the drunken example above:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/6 & 1/3 & 0 & 0 & 0 \\ 0 & 1/2 & 1/6 & 1/3 & 0 & 0 \\ 0 & 0 & 1/2 & 1/6 & 1/3 & 0 \\ 0 & 0 & 0 & 1/2 & 1/6 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

  - The probability that he is at E_i after k steps is
  $$x \cdot P^k$$

- Markov chain is (P, x) where
  - $P \Rightarrow n \times n$ transition matrix
  - $x \Rightarrow 1 \times n$ row vector

- E_i is called *states* of the chain.

- Our goal here is to classify these states.

- Assume that we are only concerned whether we can get from a given state to another state and if so how long it takes.
  - In the drunken example above:
    - E_4 to E_1 with three steps.
    - E_4 to E_6 with two steps.
    - E_1 to E_4 is impossible.

$\Downarrow$ So,

- our concern is not with the actual value of p_{ij}. but whether p_{ij}'s are non-zero.

$\Downarrow$ So a chain is expressed by a

- directed graph

- We can get from a state E_1 to a state E_j iff there is a path frow i to j, and the least possible time is the length of the shortest such path

$$\Downarrow$$

- A Markov chain in which we can get from any state to any other is called *irreducible chain*.

$\Downarrow$ clearly

- A Markov chain is irreducible iff the corresponding directed graph is strongly connected.

□ Directed Graph Revisit

- Definition
  - underlying graph
    - ∗ when we revome arrows from the directed graph.
  - A digraph is *connected* if its underlying graph is a connected graph.
  - A digraph is *strongly connected* if, for any two nodes v, w, there is a paht from v to j.

  Note that
  - every strong connected digraph is connected but not all conndected digraph is strong connected.

- The distinction between a connected and a strongly connected digraph.

  - Consider the road map of a city, all of whose streets are one-way.

  - If the road map is connected, we can drive from any part of the city to any other, ignoring the direction of the one-way streets.

  - If the map is strongly connected, we can drive from any part of the city to any other, always going the right way down the one-way streets.

- Every one-way system should be strongly connected.

- Then how we can impose a one-way system in such a way that we can drive from any part of the city to any other.
  - If, for example, the city consists of two parts connected only by a bridge, then we cannot impose such a one-way system, since whatever direction we give to the bridge, one part of the city must be cut off.

- Define a graph to be *orientable* if each edge of the graph can be directed so that the resulting digraph is strongly connected.
  - Note that any Eulerian graph is orientable.

- Theorem
  - Let G be a connected graph. Then G is orientable iff each edge of G is contained in at least one cycle.

□ A classification of States - 1

- **Persistent/Transient**
  - States to which we keep on retruning however long we continue.

    ⇒ persistent states
  - States we visit a few times and never return to

    ⇒ transient states

in other words,

- If the probability of returning to E_i at some later stage is 1

  ⇒ persistent states (e.g. E_1, E_6)

- otherwise

  ⇒ transient states (e.g. $E_2 \cdots, E_5$)

Theorem

- E_j is persistent iff the existence of a path from i to j in the corresponding graph implies the existence of a path from j to i.

In the example (2) above:

There is a path from 1 to 4, but no path from 4 to 1,

$\Downarrow$ thus

E_1 is *transient*, as is E_3,
whereas E_2, E_4, E_5 and E_6 are *persistent.*

□ A classification of States - 2

- **Periodicity**
  - A Markov chain is *periodic* of period t ($t \neq 1$) if it is possible to return to E_i only after a period of time that is a multiple of t.
  - If no such t exists $\Rightarrow$ *aperiodic*.

□ A classification of States - 3

- **Ergodicity**
  - A state is **ergodic** if it is both persistent and aperiodic.
  - A Markov chain is **ergodic** if every state is ergodic.

□ Exercise

- In the drunken walk example:
  - suppose E_6 ejects him as soon as he gets there.
    - Write down the transition matrix.
    - Draw the corresponding digraph.
    - Classify the states.
  - How about if both eject him?