INTRODUCTION TO GRAPH THEORY AND APPLICATIONS

Andrea Scozzari

University Niccolò Cusano
Dept. of Economics, Psychological and Communication Science.

andrea.scozzari@unicusano.it

MAXIMUM FLOW

Given a directed graph $G=(V,A)$, assume that there exists a source node $s \in V$ and a sink node $t \in V$. All other nodes are transit nodes.

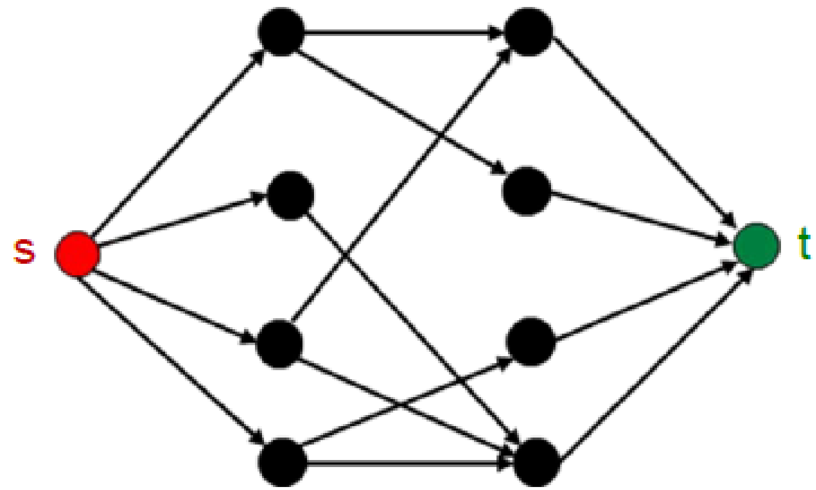
Upper Capacities U_{ij} for each $(i,j) \in A$

Lower Capacities L_{ij} for each $(i,j) \in A$ and such that $L_{ij} \leq U_{ij}$.

We may assume $L_{ij} = 0$ for each $(i,j) \in A$

MAXIMUM FLOW

Given a directed graph G=(V,A) with |V| = n, let 1 and n be the source and sink nodes, respectively. Let F be the (unknown) flow to be sent from node 1 to node n



MAXIMUM FLOW

Given a directed graph $G=(V,A)$ with $|V| = n$, let 1 and n be the source and sink nodes, respectively. Let F be the (unknown) flow to be sent from node 1 to node n

Flow-Conservation constraints

$$\left\{\begin{array}{ll} \displaystyle\sum_{j \in S(1)} x_{1j} = F & \\ \displaystyle\sum_{j \in P(i)} x_{ji} - \sum_{j \in S(i)} x_{ij} = 0 & \forall i \in V \setminus \{1, n\} \\ \displaystyle -\sum_{j \in P(n)} x_{jt} = -F & \end{array}\right.$$

Capacity constraints

$$0 \le x_{ij} \le u_{ij} \qquad \forall (i,j) \in A$$

$$F \in \mathbb{R}_+$$

MAXIMUM FLOW

Objective function $\text{Max}\ \boldsymbol{F}$

Flow-Conservation constraints

$$\sum_{j \in S(1)} x_{1j} = F$$

$$\sum_{j \in P(i)} x_{ji} - \sum_{j \in S(i)} x_{ij} = 0 \qquad \forall i \in V \backslash \{1, n\}$$

$$-\sum_{j \in P(n)} x_{jt} = -F$$

Capacity constraints $0 \leq x_{ij} \leq u_{ij} \qquad \forall (i, j) \in A$

$$F \in \mathbb{R}_+$$

CENTRALITY INDEXES

Centrality Indexes measure the importance of a node in a network G=(V,A)

Betweeness centrality: measures the centrality of a node k in the network computed as the number of Shortest Paths in the network that pass through node k, i.e. paths that use node k as an intermediary node.

Closeness centrality: measures the ability of a node to communicate with many other nodes in the network using the minimum number of intermediary nodes.

CENTRALITY INDEXES

Centrality Indexes measure the importance of a node in a network G=(V,A)

Betweeness centrality: measures the centrality of a node k in the network computed as the number of Shortest Paths in the network that pass through node k, i.e. paths that use node k as an intermediary node.

Closeness centrality: measures the ability of a node to communicate with many other nodes in the network using the minimum number of intermediary nodes.

CENTRALITY INDEXES

Classical centrality indices compute, in practice, how much information passes through nodes.

They assume that information passes through the Shortest Paths between pairs of nodes in the network G.

However, classical centrality indices do not take into account how much information passes through nodes!!!

In fact, there may be “central” nodes but through which a low flow of information passes.

CENTRALITY INDEXES

Flow Betweenness (FB) (Freeman et al. 1991)

It measures the level of maximum flow between all pairs of nodes in a network w.r.t a specific node k, i.e., it represents the amount of information that uses node k as an intermediary node.

The higher the index value, the greater the centrality of a node.

CENTRALITY INDEXES

Flow Betweenness (FB) (Freeman et al. 1991)

$$FB_k = \frac{\sum_{\{ij\} \in C} F_{ij}^k}{\sum_{\{ij\} \in C | i \neq j} F_{ij}}$$

Where

FB_k is the flow betweenness of node *k*.

F_{ij}^k is the Maximum Flow between *i* and *j* passing through *k*.

F_{ij} is the Maximum Flow between *i* and *j*.

C is the set of all pairs of nodes of the graph $\{i, j\} \in V$

CENTRALITY INDEXES

Flow Closeness (FC) (Freeman et al. 1991)

Measures the propagation strength of an information flow from one node k in the network to all other nodes

The higher the index value, the greater the *propagation strength* of a node.

CENTRALITY INDEXES

Flow Closeness (FC) (Freeman et al. 1991)

$$FC_k = \sum_{j \in V | j \neq k} F_{kj}$$

where

F_{kj} is the Maximum Flow between k and j in the network.

CENTRALITY INDEXES

Influence between Groups (IG) (Renfro, 2001)

Given two sets T and R of nodes in the network, with $T \cap R = \emptyset$, IG measures the influence of the set of nodes in T on the set of nodes in R.

The higher the index value, the greater the *influence between the groups* of nodes.

Note: when T and R are just two nodes in the network, IG reduces to computing the maximum flow between the two nodes.

CENTRALITY INDEXES

Influence between Groups (IG) (Renfro, 2001)

$$IG_{TR} = \sum_{i \in T} \sum_{j \in R} F_{ij}$$

where

F_{ij} is the Maximum Flow between i and j with i that belongs to the set T and j to the set R.

NETWORKS MEASURES

In many real-life applications based on network analysis, one is interested in understanding whether there are particular types of "concentration" of some activities within a business system.

Through concentration analysis, one wants to understand whether an activity being analyzed is homogeneously located within a specific geographical area or within a set of other activities.

NETWORKS MEASURES

Concentration Indexes.

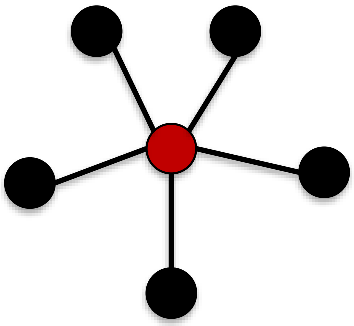
- Freeman concentration index

It takes into account the structure of a system and measures the 'shape' of the network in the sense, that it measure the dishomogeneity of the network with respect to an ideal case.

NETWORKS MEASURES

It takes into account the structure of a system and measures the 'shape' of the network in the sense, that it measure the dishomogeneity of the network with respect to an ideal case.

Star



Freeman's concentration index measures the similarity between the topology of the existing network and the star. Therefore, the higher its value, the more similar the network under observation is to the ideal case.

NETWORKS MEASURES

- Freeman concentration index

Given an undirected graph G=(V,E)

$$F_B = \frac{\sum_{i=1}^{n} [F_B(x^*) - F_B(x_i)]}{n^3 - 4n^2 + 5n - 2}$$

Dove:

- n is the cardinality of V (e.g., the set of firms)
- $F_B(x_i)$ is the *betweenness centrality* of vertex i
- $F_B(x^*)$ is the maximum among the *betweenness centrality* of the vertices in G

NETWORKS MEASURES

Concentration Indexes.

- *Entropy*

It measures the spatial dis-homogeneity, i.e. heterogeneity, in an economic network.

It is used to understand whether a given spatial configurations of companies in a territory are completely arbitrary or whether these configurations reflect a kind of “organization” of the nodes of the network under observation.

NETWORKS MEASURES

- Entropy

Given an undirected graph G=(V,E)

$$E = -\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}\ln(p_{ij})$$

where:

- n is the cardinality of V (e.g., the set of firms)
- $p_{ij} = f(a_{ij}, d_i)$ is the *probability* (i.e., Probability distribution) that there exists a link between vertcices i and j; d_i the degree of i.

The higher the value of E, the higher the network *diversification*.

NETWORKS MEASURES

The concepts of centrality (referring to a single node) and concentration (referring to an entire network) are, of course, strongly linked.

Property 1

A network is "centralized" when one node, or a group of nodes, are able to control the flows within a network. In these cases, the centrality indexes of such nodes (or groups of nodes) are very high.

Property 2

A "centralized" network is always concentrated, but the reverse is not necessarily true

CORPORATE FINANCE APPLICATION

Asset Liability Management

Consider the financial management of a (small) firms that receives payments from customers in certain periods of time and must, in turn, pay suppliers in certain periods of time. The company can invest its excess cash in securities and short-term deposits and can borrow short-term from banks.

CORPORATE FINANCE APPLICATION

Investments: payments, with interest, made to the company in the future.
Loans: payments, with interest, made to the company in the present minus the interest rate.

Suppose we know the term structure of interest rates (assets and liabilities) over the next n years and we know the expected cash flows over those periods.

Problem: Determine an investment and liability plan in line with the payments to be made and that maximises cash availability (positive or negative) at the end of year n.

FLOW WITH GAINS OR LOSSES

In some problems it may happen that the flow x_{ij} passing through an arc (i,j) does not remain constant from i to j but the flow increases or decreases along the arc. For example, a flow of water along leaky pipes, a flow of money between two periods in the presence of interest income or expense.

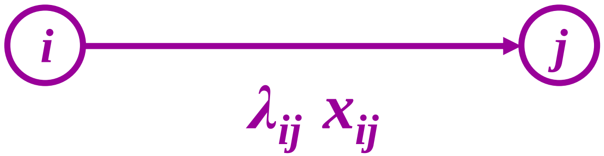
Consider:

x_{ij} flow passing through arc (i,j) and entering in j.

FLOW WITH GAINS OR LOSSES

Proportionality assumption:

the losses or gains in flow along an arc are proportional to the flow through that arc



λ_{ij} increase or decrease of flow from node i to node j.

$\lambda_{ij} < 1$ *decrease*

$\lambda_{ij} > 1$ *increase*

FLOW WITH GAINS OR LOSSES

The new flow-conservation constraint

$$\sum_{j \in P(i)} \lambda_{ji} x_{ji} - \sum_{j \in S(i)} x_{ij} = b_i \quad \forall\, i \in V$$

FLOW WITH GAINS OR LOSSES

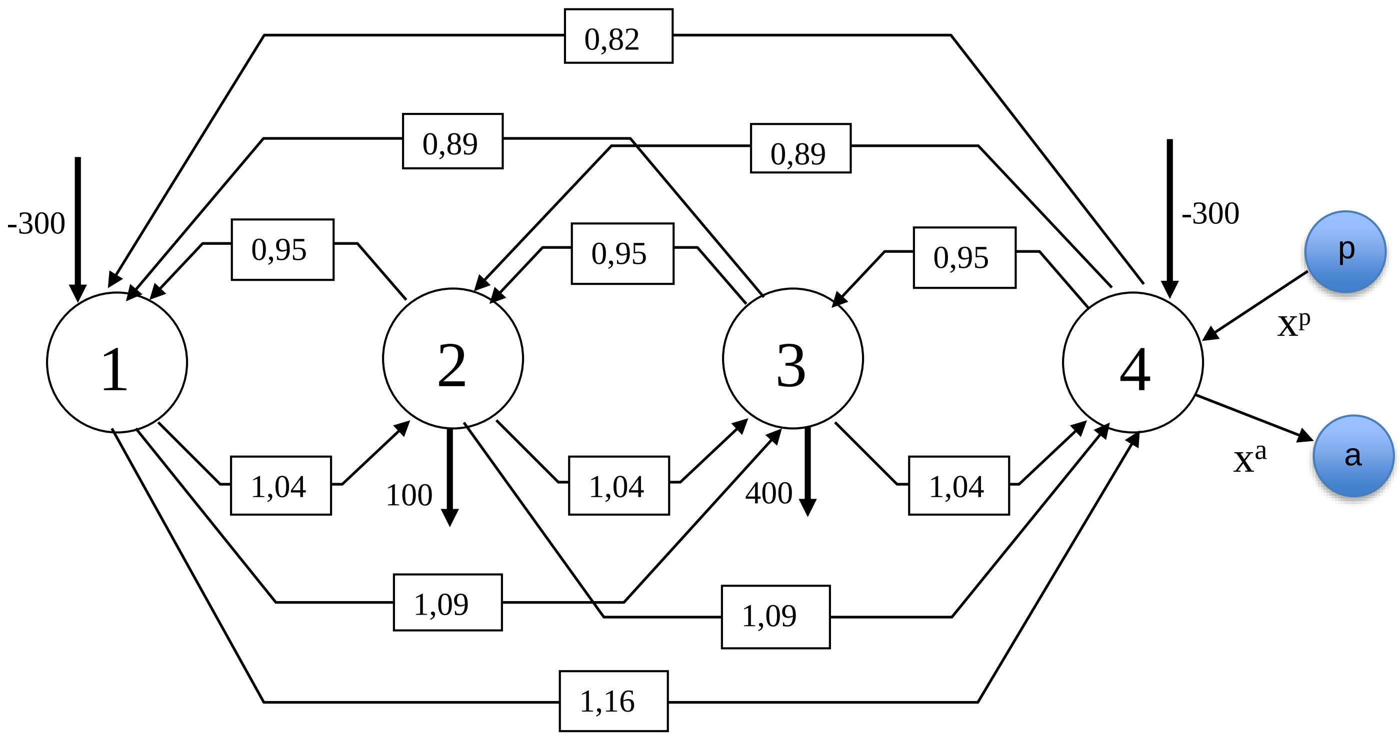
The model

$$\min \sum_{(i,j)\in A} c_{ij} x_{ij}$$

$$\sum_{j\in P(i)} \lambda_{ji} x_{ji} - \sum_{j\in S(i)} x_{ij} = b_i \qquad \forall\, i \in V$$

$$l_{ij} \le x_{ij} \le u_{ij} \qquad \forall\, (i,j) \in A$$

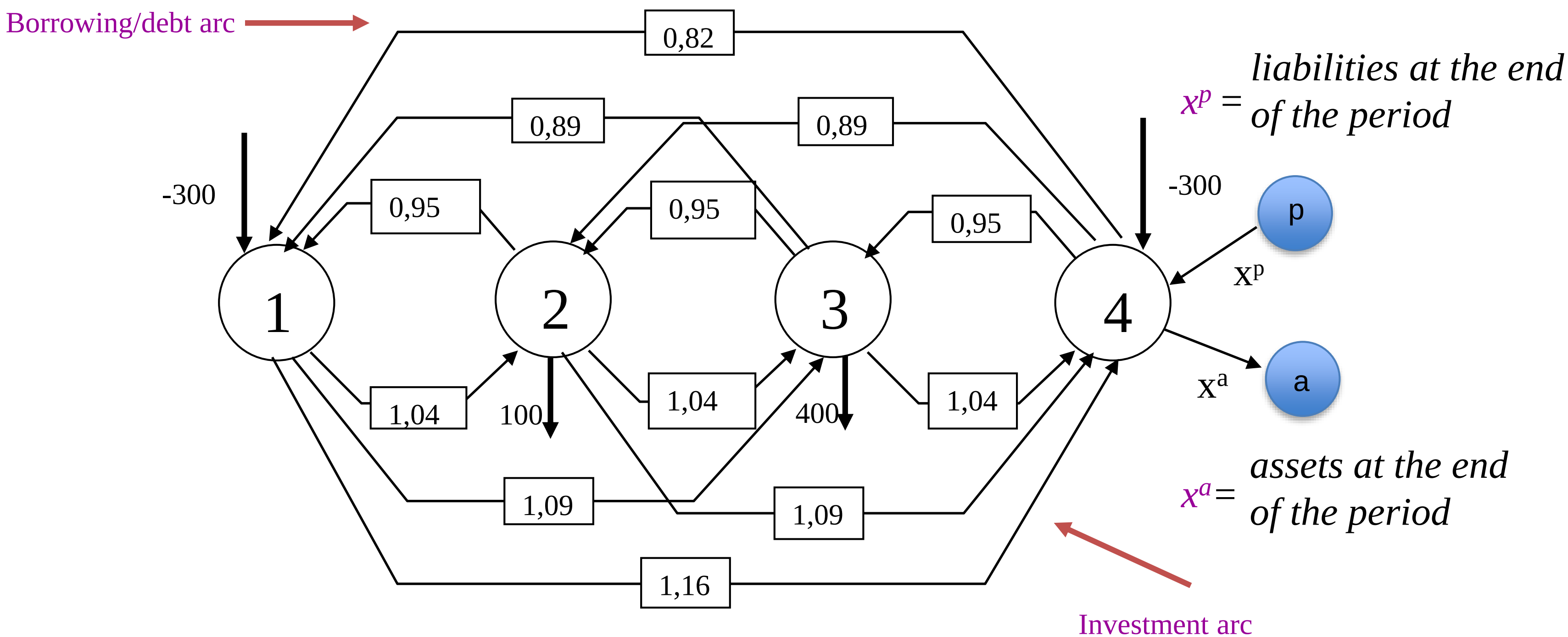
FLOW WITH GAINS OR LOSSES: Example



Let b_i be the cash flow (estimated) at each period i=1,2,3,4. if:

$b_i < 0$ the company expects incoming payments (i=1,4);
$b_i > 0$ the company must make payments (i=2,3).

FLOW WITH GAINS OR LOSSES: Example



r_{ij} = if $i > j$ it is an active rate of return from period i to period j, i.e. for every euro invested in i you collect $(1+r_{ij})$ euro in j

r_{ij} = if $i < j$ is an interest rate payable from period i to period j, i.e. for each euro to be repaid in j you collect $(1-r_{ij})$ euro in i.

x_{ij} = Amount of investments or loans between periods i and j

FLOW WITH GAINS OR LOSSES: Example

Budget constraint: given a node j=1,...,n-1

$$\underbrace{\sum_{i=1}^{j-1}(1+r_{ij})x_{ij}+\sum_{i=j+1}^{n}(1-r_{ij})x_{ij}}_{\substack{\text{Set of arcs of investment and}\\ \text{debt \textbf{entering} the node } j}}-\underbrace{\sum_{i=1}^{j-1}x_{ji}-\sum_{i=j+1}^{n}x_{ji}}_{\substack{\text{Set of arcs of investment and}\\ \text{debt \textbf{coming out} of the node } j}}=b_j \qquad \forall\ j=1,\dots,n-1$$

$$\underbrace{\sum_{i=1}^{n-1}(1+r_{in})x_{in}-\sum_{i=1}^{n-1}x_{ni}}_{\substack{\text{Set of arcs of Investment and}\\ \text{debt \textbf{incidents} in the node } n}}\underbrace{-x^{a}+x^{p}}_{\substack{\text{Cash flow at the end}\\ \text{of the period}}}=b_n$$

FLOW WITH GAINS OR LOSSES: Example

The aim is to maximize (x^a - x^p) subject to the budget constraints at each period:

$$\max \ x^a - x^p$$

$$\sum_{i=1}^{j-1} (1+r_{ij})x_{ij} + \sum_{i=j+1}^{n} (1-r_{ij})x_{ij} - \sum_{i=1}^{j-1} x_{ji} - \sum_{i=j+1}^{n} x_{ji} = b_j \qquad \forall \ j = 1,\ldots,n-1$$

$$\sum_{i=1}^{n-1} (1+r_{in})x_{in} - \sum_{i=1}^{n-1} x_{ni} - x^a + x^p = b_n$$

$$x_{ij} \geq 0 \quad \forall \, (i, j) \in A$$

$$x^a, x^p \geq 0$$

($1+r_{ij}$) and ($1-r_{ij}$) correspond to the losses or gains λ_{ij}

PROJECT SCHEDULING

Project Definition

A project is a set of elementary activities that must be carried out in a given order.

It is convenient to represent any project by means of a graph where the individual activities and precedence relationships between them are represented.

PROJECT SCHEDULING

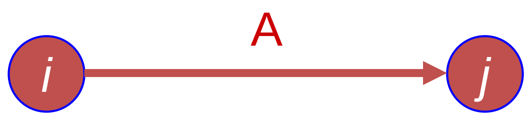
The Critical Path Method (CPM) is a methodology that enables the planning and management of a complex project, i.e. a project consisting of several activities that must be carried out, in such a way as to keep the duration, and thus the overall cost, of the project under control.

CPM benefits

1. It allows you to determine which activities must be carried out first;
2. it finds the critical activities, i.e. those activities whose delay causes a general delay of the whole project;
3. it finds the non-critical activities and how much delay they are allowed;
4. How to plan a project having the lowest total cost and the best duration.

PROJECT SCHEDULING

The activities correspond to arcs of a given graph G=(V,A)



(*i*,*j*) = arc representing activity A

i = event node, i.e., start (tail) of activity A

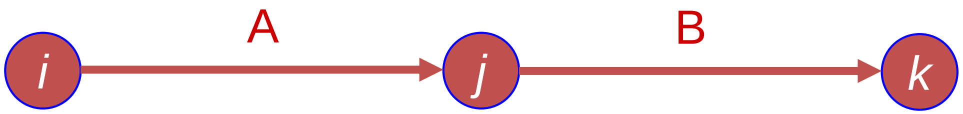
j = event node, i.e., end (head) of activity A

event: is a time instant defining the start or end time of an activity.

PROJECT SCHEDULING

The activities correspond to arcs of a given graph G=(V,A)

arc: the direction of the arc indicates a progression in time, so all the arcs of an activity graph are always directed from left to right



An activity MUST always be finished before the next one can start and the final event of a previous activity coincides with the initial event of the following activity

PROJECT SCHEDULING

In a graph G representing the set of activities of a project, two particular events are distinguished, namely the starting event and the final event of the project. The starting event has NO activity preceding it, and the final event has NO activity following it. They only indicate the project start time and the project end time.

Example:

Assume a project with 5 activities A, B, C, D, E, where:

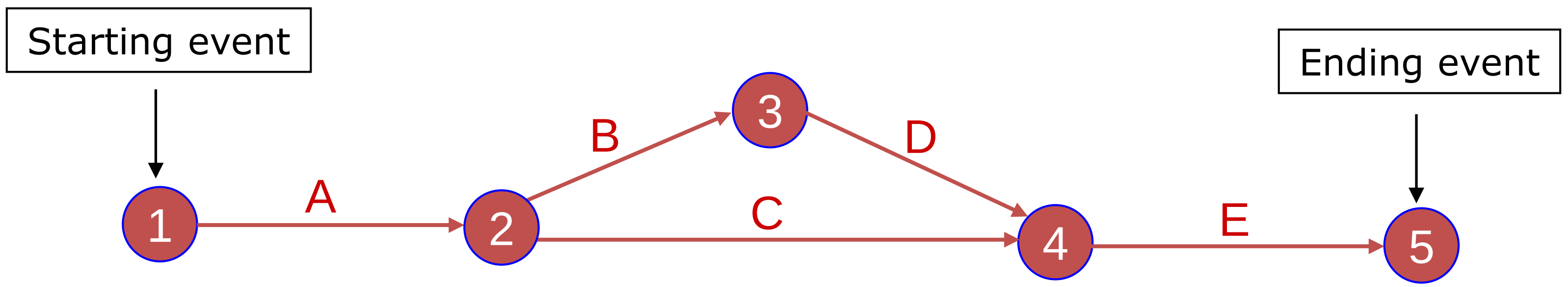
1) Activity A precedes B and C;

2) Activity B precedes D;

3) Activities C and D precede E.

PROJECT SCHEDULING

The corresponding activity graph G is:



PROJECT SCHEDULING

The following rules must hold when representing a project (and all its activities) by a graph G:

Each node must have a numbering, with node 1 representing the start of the project

An additional node representing the end of the project must be included in the graph

All nodes are numbered in such a way that if there is an arc from node i to node j then $i < j$ (Topological sorting).

There are only a starting and ending nodes

Two nodes cannot be connected by more than one arc.

PROJECT SCHEDULING

Exercise 1

Consider a project consisting of organising a sports tournament. Let the set of activities be listed in the table

Activity	Precedences
A) Choice of location	
B) Participants selection	
C) Reservation of facilities	
D) Booking accommodation	
E) Program and ticket printing	
F) Ticket sales	
G) Organisation of side activities	
H) Assignment of grounds for matches	
I) Assignment of training grounds	
L) Running of the Tournament	

PROJECT SCHEDULING

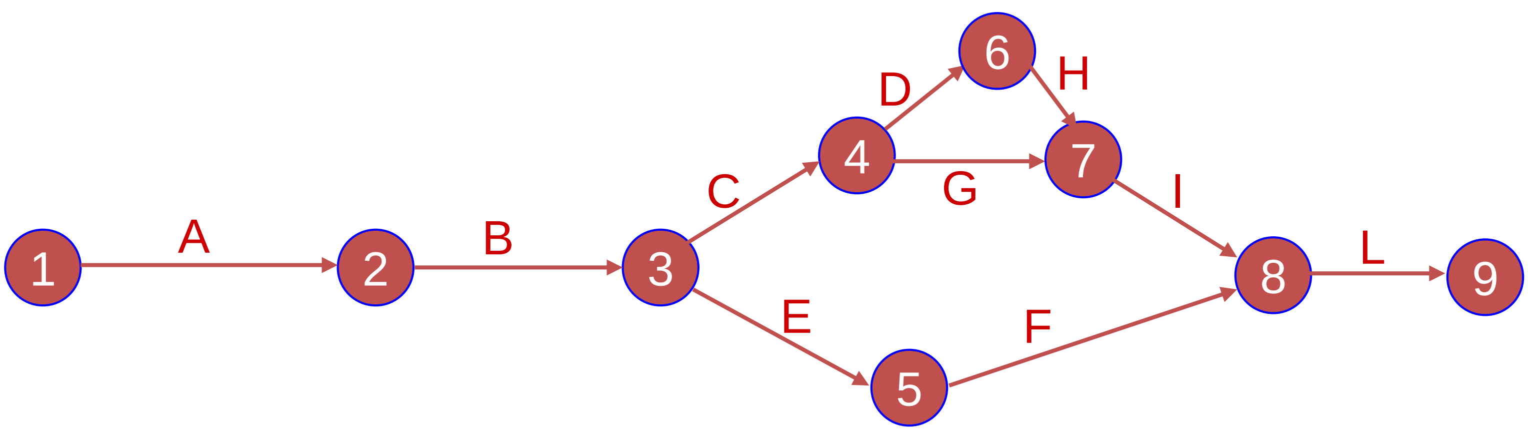
Exercise 1

Consider a project consisting of organising a sports tournament. Let the set of activities be listed in the table

Activity	Precedences
A) Choice of location	-
B) Participants selection	A
C) Reservation of facilities	B
D) Booking accommodation	C
E) Program and ticket printing	B
F) Ticket sales	E
G) Organisation of side activities	C
H) Assignment of grounds for matches	D
I) Assignment of training grounds	G,H
L) Running of the Tournament	F,I

PROJECT SCHEDULING

Activity	Precedences
A) Choice of location	-
B) Participants selection	A
C) Reservation of facilities	B
D) Booking accommodation	C
E) Program and ticket printing	B
F) Ticket sales	E
G) Organisation of side activities	C
H) Assignment of grounds for matches	D
I) Assignment of training grounds	G,H
L) Running of the Tournament	F,I



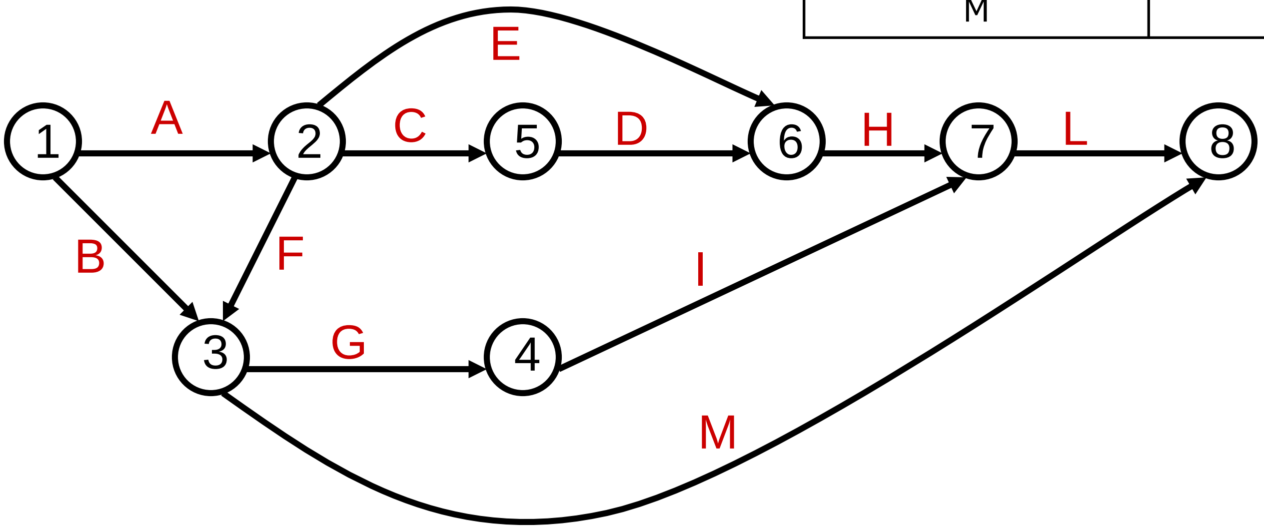
PROJECT SCHEDULING

Find the graph of the following activities

Activities	Predecessors
A	-
B	-
C	A
D	A,C
E	A
F	A
G	B,F
H	D,E
I	G
L	H
M	B,F

PROJECT SCHEDULING

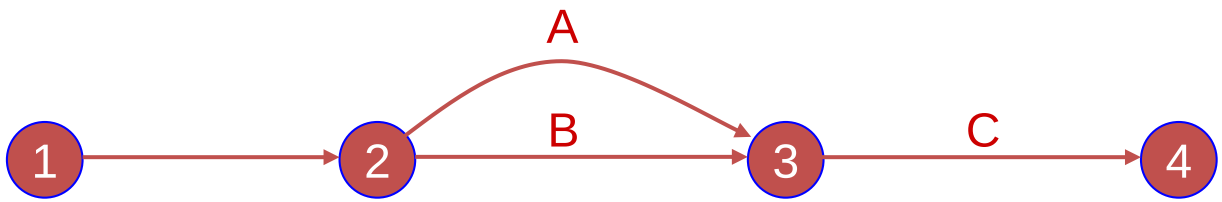
Activities	Predecessors
A	-
B	-
C	A
D	A,C
E	A
F	A
G	B,F
H	D,E
I	G
L	H
M	B,F

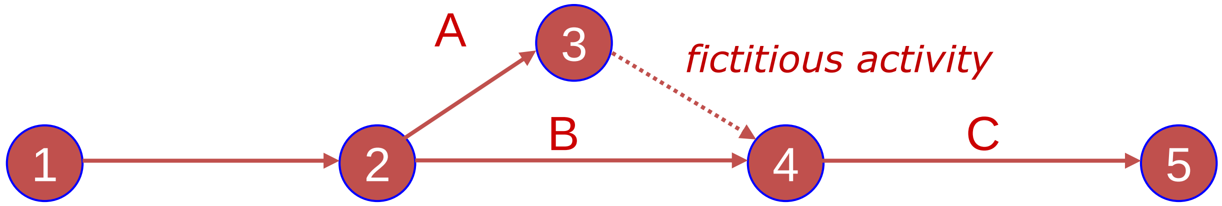


PROJECT SCHEDULING

In the graph representation of a project, it is often necessary to introduce fictitious activities to define the precedence relationship between two activities A and B with respect to an activity C, without this implying precedence between A and B.

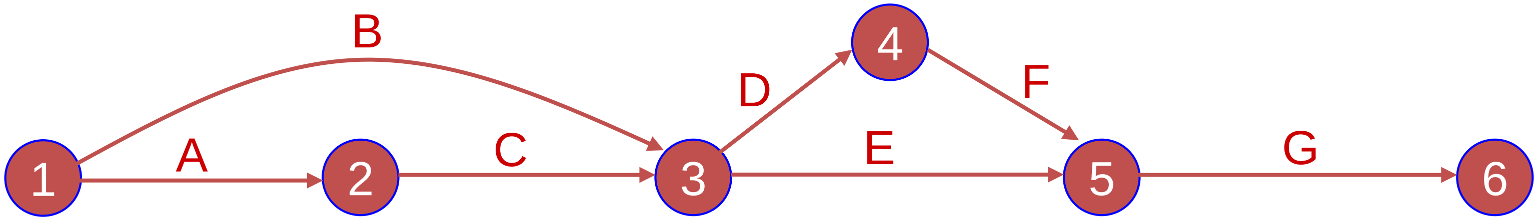
Example:

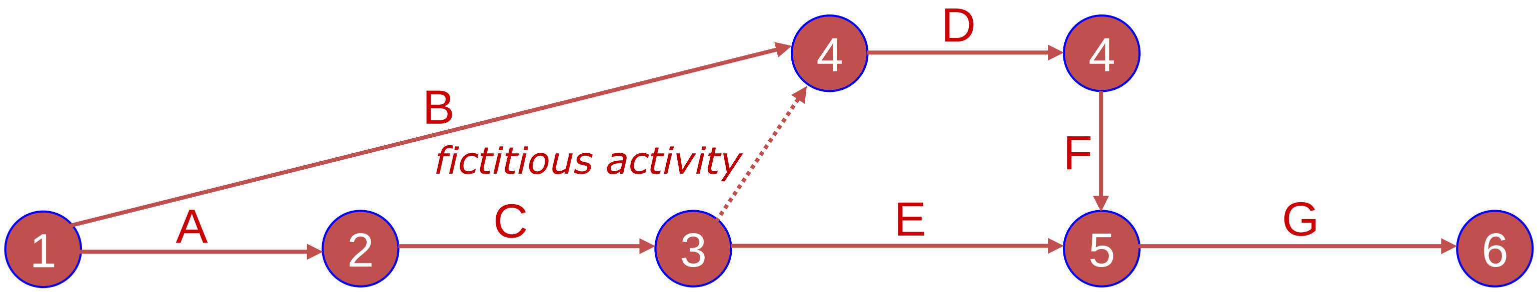




PROJECT SCHEDULING

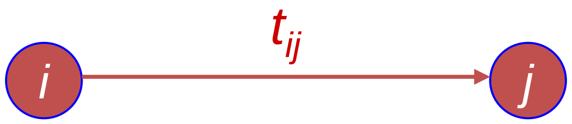
Example: Activities *B* and *C* end at event node 3, but only the activity *B* is needed for strating activity *D*





PROJECT SCHEDULING

Assume that, for each activity (i,j), we knwow its lifetime t_{ij}



Activities	Predecessors	times
A	-	3
B	-	5
C	A	14
D	A,C	5
E	A	7
F	A	10
G	B,F	9
H	D,E	12
I	G	2
L	H	3
M	B,F	9

PROJECT SCHEDULING

In any project, some activities are flexible w.r.t. their starting and ending time, while others are not. The delay of any one of the latter will delay the entire project. These are the critical activities

The (length of the) Critical Path:

THE LONGEST TIMELIFE OF ACTIVITIES THROUGH THE PROJECT

In a project there is always a critical path and sometimes there is more than one.

PROJECT SCHEDULING

The problem of determining the critical path can also be formulated as a linear programming problem in the following way:

x_j = time when we arrive at node j during the project

F = node representing the end of the project

$$\min \; x_F - x_1$$
$$x_j - x_i \geq t(i,j) \qquad \forall (\mathrm{i,j}) \in \mathrm{A}$$
$$x_j \geq 0$$