INTRODUCTION TO GRAPH THEORY AND APPLICATIONS

Andrea Scozzari

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

andrea.scozzari@unicusano.it

CENTRALITY MEASURES

Given an undirected graph G = (V,E) where E is the set of edges of G and V the set of nodes or vertices of G.

Average path length: is the average (geodesic) distance between all pairs of nodes in the graph.

$$L = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i \geq j} d_{ij}$$

where:

d_{ij} = the distance between vertex i and vertex j

n = |V|

The distance d_{ij} between two nodes is defined as the minimum number of edges (cardinality) of a path from node i to node j (and vice versa).

CENTRALITY MEASURES

Given an undirected graph G = (V,E) where E is the set of edges of G and V the set of nodes or vertices of G.

Betweenness Centrality:

$$B(i) = \sum_{j,k \neq i} \frac{P_{jk}(i)}{P_{jk}}$$

where:

P_{jk} is the total number of paths between vertices j and k

$P_{jk}(i)$ is the number of paths between vertices j and k that pass through vertex i

CENTRALITY MEASURES

Given an undirected graph G = (V,E) where E is the set of edges of G and V the set of nodes or vertices of G.

Betweenness Centrality:

$$B(i) = \sum_{j,k \neq i} \frac{P_{jk}(i)}{P_{jk}}$$

Thus, the Betweenness measures the degree of 'attractiveness' of node i with respect to all other nodes in the network.

CENTRALITY MEASURES

Given an undirected graph G = (V,E) where E is the set of edges of G and V the set of nodes or vertices of G.

Eigenvector Centrality: This index measures the importance of a node with respect to its neighbouring nodes, by assuming that the importance of a node i is represented by a parameter x_i, it is defined

$$x_i = \hat{\alpha} \sum_{j \in N_i} x_j$$

where:

N_i = the set of neighbors of vertex i

n = |V|

$\hat{\alpha}$= a parameter of proportionality

CENTRALITY MEASURES

Given an undirected graph G = (V,E) where E is the set of edges of G and V the set of nodes or vertices of G.

Eigenvector Centrality: $x_i = \hat{\alpha} \sum_{j \in N_i} x_j$

Who's $\hat{\alpha}$?

Let $\boldsymbol{X}$ be the (column) vector with n elements where the component x_i represents the importance of vertex i.

Let a_{ij} be the generic element of the adjacency matrix $\boldsymbol{A}$ of G.
For each vertex i, the importance of i with respect to all vertices $j \in N_i$, is:

$$x_i = \hat{\alpha} \sum_{j \in N_i} x_j = \hat{\alpha} \sum_{j=1}^{n} a_{ij} x_j = \hat{\alpha} a_{i\cdot} \boldsymbol{X}$$

CENTRALITY MEASURES

$$x_i = \hat{\alpha} \sum_{j \in N_i} x_j = \hat{\alpha} \sum_{j=1}^{n} a_{ij} x_j = \hat{\alpha} a_{i.} \boldsymbol{X}$$

In a more compact form, for all vertices i=1,…,n, we have:

$$\boldsymbol{AX} - \frac{1}{\hat{\alpha}} \boldsymbol{X} = \boldsymbol{AX} - \alpha \boldsymbol{X} = \boldsymbol{0}$$

That is, the parameter α is the parameter that solves the equation:

$$(\boldsymbol{A} - \alpha \boldsymbol{I})\boldsymbol{X} = \boldsymbol{0}$$

CENTRALITY MEASURES

For the equation:

$$(\boldsymbol{A} - \alpha \boldsymbol{I})\boldsymbol{X} = \boldsymbol{0}$$

There exists a non trivial solution if

$$\mathbf{det}(\boldsymbol{A} - \alpha \boldsymbol{I}) = \boldsymbol{0}$$

That is, α are the eigenvalues of the adjacency matrix $\boldsymbol{A}$.

Let $\alpha_{max} > 0$ be the maximum among the eigenvalues, which exists and it is unique due to the *Perron-Frobenius* Theorem, then:

CENTRALITY MEASURES

Given an undirected graph G = (V,E) where E is the set of edges of G and V the set of nodes or vertices of G.

Eigenvector Centrality: This index measures the importance of a node with respect to its neighbouring nodes, by assuming that the importance of a node i is represented by a parameter x_i, it is:

$$\boldsymbol{X_{max}}$$

The (positive) eigenvector associated to the maximum eigenvalue.

FLOW MODELS AND APPLICATIONS

Graph or network models are special Linear Programming (LP) models that possess peculiar features

1. they allow providing a graphical interpretation of the problem under consideration

2. they can be solved with specific algorithms that are more efficient than those available for LP

Ahuja, Ravindra K., Thomas L. Magnanti, and James B. Orlin: Network flows: Theory, Algorithms and Applications, Pretience Hall, Upper Saddle River, New Jersey, 1988, ISBN 0-13-617459-X.

FLOW MODELS AND APPLICATIONS

In network models, decision variables x_{ij} are associated with each arc (i,j) of the network, which can be interpreted as the amount of flow to be sent from vertex i to vertex j (case $x_{ij} \in \mathbb{R}$), or to state if there is a “relationship” between vertex i and vertex j (case $x_{ij} = 0{,}1$) (Boolean flow).

In economic or financial problems, the nodes or vertices of a network may represent time periods or sites, while the variables x_{ij} can be interpreted as a flow (monetary, of materials, etc.) between one node and another.

FLOW MODELS AND APPLICATIONS

In addition to the decision variables x_{ij} , the arcs (i,j) are often associated with:

1. limitations or maximum capacities u_{ij} and/or minimum capacities l_{ij} for the flows that pass through them

2. travel costs (or profits) c_{ij} for each unit of flow x_{ij}.

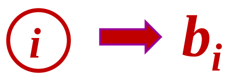
FLOW MODELS AND APPLICATIONS

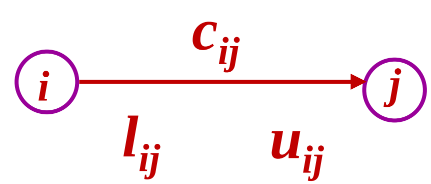
We also assume that there are supplies or demands of flow such as:

1. $b_i < 0$ supply or increase in flow in node i
2. $b_i > 0$ supply or decrease in flow in node i.
3. $b_i = 0$ node i is called a transfer node.

FLOW MODELS AND APPLICATIONS

Let us now consider only directed graphs $G=(V,A)$. Let us assume that the graph is "weighted", i.e. a weight is associated with each arc, that represents either a (generalized) cost or a capacity of the arc. The weights associated to the nodes represent the amount of flow entering or leaving those nodes.





$b_i > 0$ represents the amount of flow leaving the network at node i. Then b_i is the demand at node i, and node i is called the sink.

$b_i < 0$ represents the amount of flow entering the network at node i. Then b_i is the supply of node i, and node i is called the source.

$b_i = 0$ node i ia a transfer node.

$c_{ij} \geq 0$ is the unitary cost of (i,j).

$u_{ij} \geq 0$ capacity upper bound for (i,j).

$l_{ij} \geq 0$ capacity lower bound for (i,j).

FLOW MODELS AND APPLICATIONS

In flow problems, global demand equals global supply, i.e., given D and O, respectively, the set of demand and supply nodes we have:

$$D = \{i \in V \mid b_i > 0\} \quad \text{and} \quad O = \{i \in V \mid b_i < 0\}$$

$$\sum_{i \in D} b_i = -\sum_{i \in O} b_i$$

FLOW MODELS AND APPLICATIONS

In flow problems, all arcs (i,j) are associated with decision variables x_{ij} representing the amount of flow to be sent from vertex i to vertex j. A flow can be interpreted as the amount of goods passing through an arc (i,j) or, in financial applications, represents an amount of money passing between two instants of time i and j.

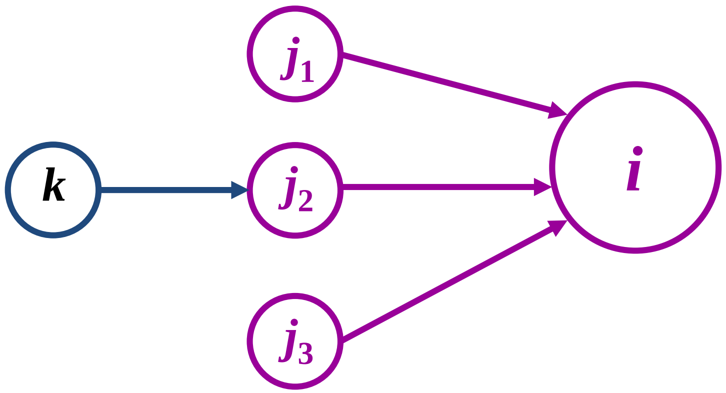
We define:

$P(i) = \{j \in V \mid (j,i) \in A\}$ the set of *predecessor* of node i;

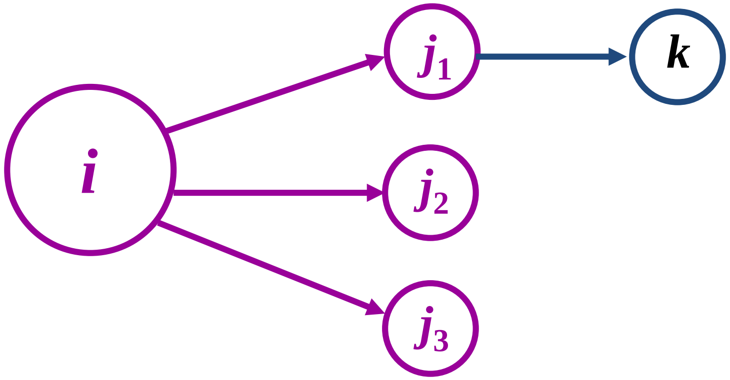
$S(i) = \{j \in V \mid (i,j) \in A\}$ the set of *successor* of node i.

FLOW MODELS AND APPLICATIONS

$P(i) = \{j \in V \mid (j,i) \in A\}$ the set of *predecessor* of node i;



$S(i) = \{j \in V \mid (i,j) \in A\}$ the set of *successor* of node i.



FLOW MODELS AND APPLICATIONS

In flow problems, the objective is to send an amount of flow x_{ij} along the arcs (i,j) of graph G in such a way as to satisfy all the demands of the demand nodes at minimum cost subject to conservation and capacity constraints.

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

MIN COST FLOW MODEL

The model

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

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

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

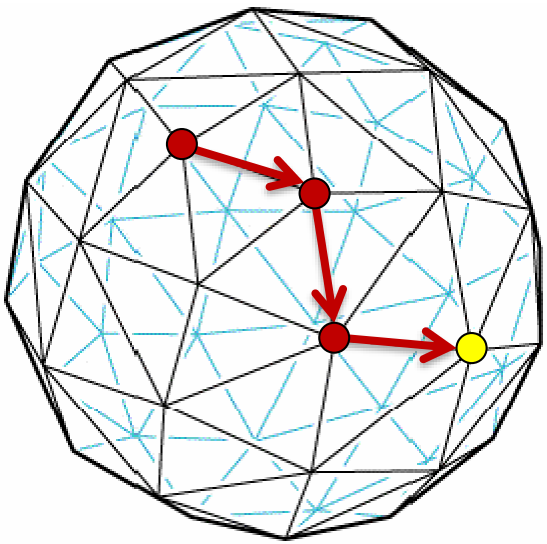
MIN COST FLOW MODEL

Linear Programming (LP)

$$\min \; c^T x$$

$$Ax \geq b$$

$$x \geq 0$$



Fundamental Theorem of LP

A Linear function bounded below on a polyhedron P attains the minimum at a vertex of P

MIN COST FLOW MODEL: Applications

Dynamic Lot Sizing Problem

This is a multi-period production problem where, given a time horizon divided into t periods, in addition to deciding how much to produce in each period t so as to satisfy demand in t, one must also decide how much goods produced must be stored between two successive periods.

MIN COST FLOW MODEL: Applications

Dynamic Lot Sizing Problem

Suppose we want to satisfy the expected demand d_t for a given commodity in K future periods ($t = 1, \ldots, K$). In order to satisfy these demands, one can decide to produce all the required goods in each period t, or one can decide to produce them in larger quantities in t in such a way that the surplus of goods from t to $t+1$ is stored to satisfy demand in the next period, i.e., in $t+1$.

Production must obviously be obtained at minimum cost.

Let c_t be the production cost at time t, with $t = 1,\ldots,K$, and $c_{t,t+1}$ the storage cost form t to $t+1$.

MIN COST FLOW MODEL: Applications

Dynamic Lot Sizing Problem

Construct the following network. Insert a node 0 indicating the time of production at the beginning of each period t, along with the corresponding arc $(0, t)$, $t = 1, \dots, K$. These arcs represent the production process of the goods in period t.

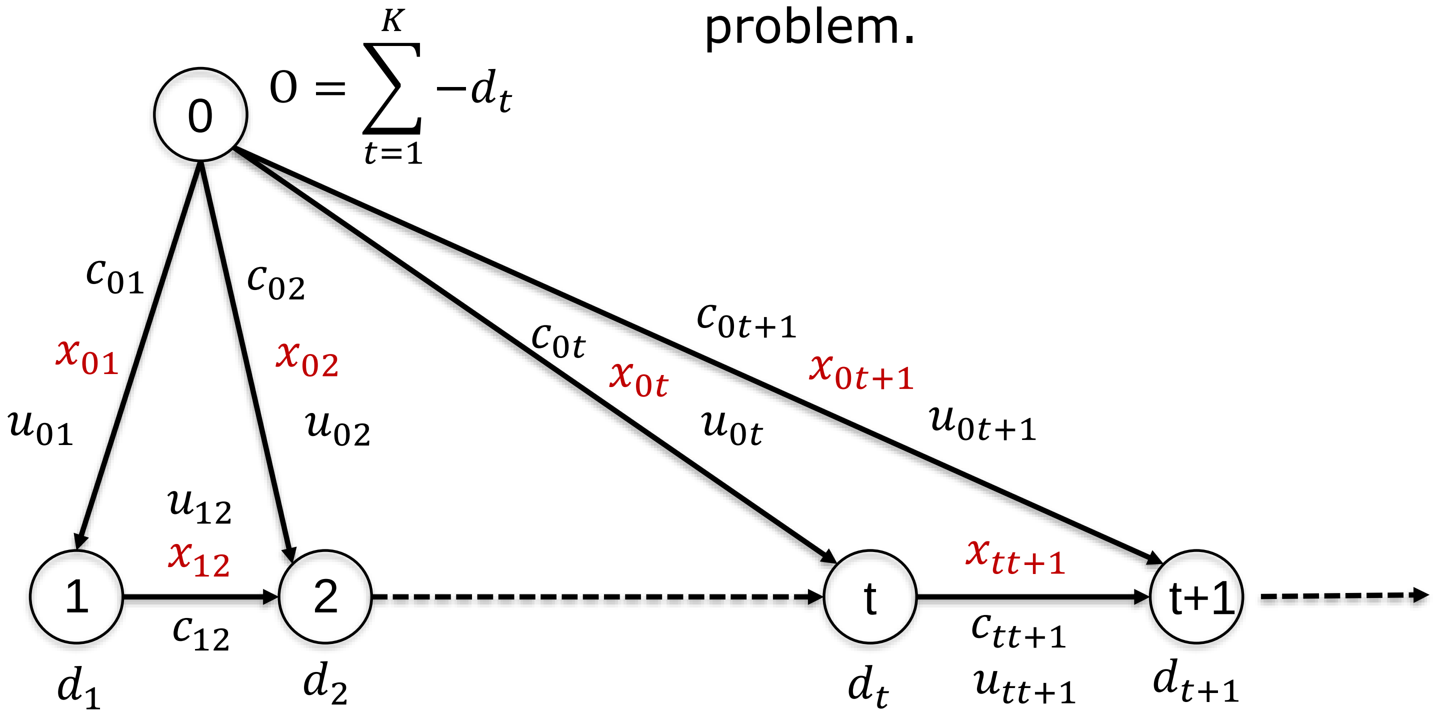
Insert arcs $(t, t+1)$, $t = 1, \dots, K-1$, that represent the storage of goods between two successive periods.

With d_t we denote the (expected) demand at time t, while at node 0 we consider *supply* equal to the sum of all demands.

MIN COST FLOW MODEL: Applications

Suppose further that there are capacity constraints on the arcs that represent the maximum amount of goods that can be either produced or stored in a given period t.

The problem can be formulated as a *minimum cost flow* problem.



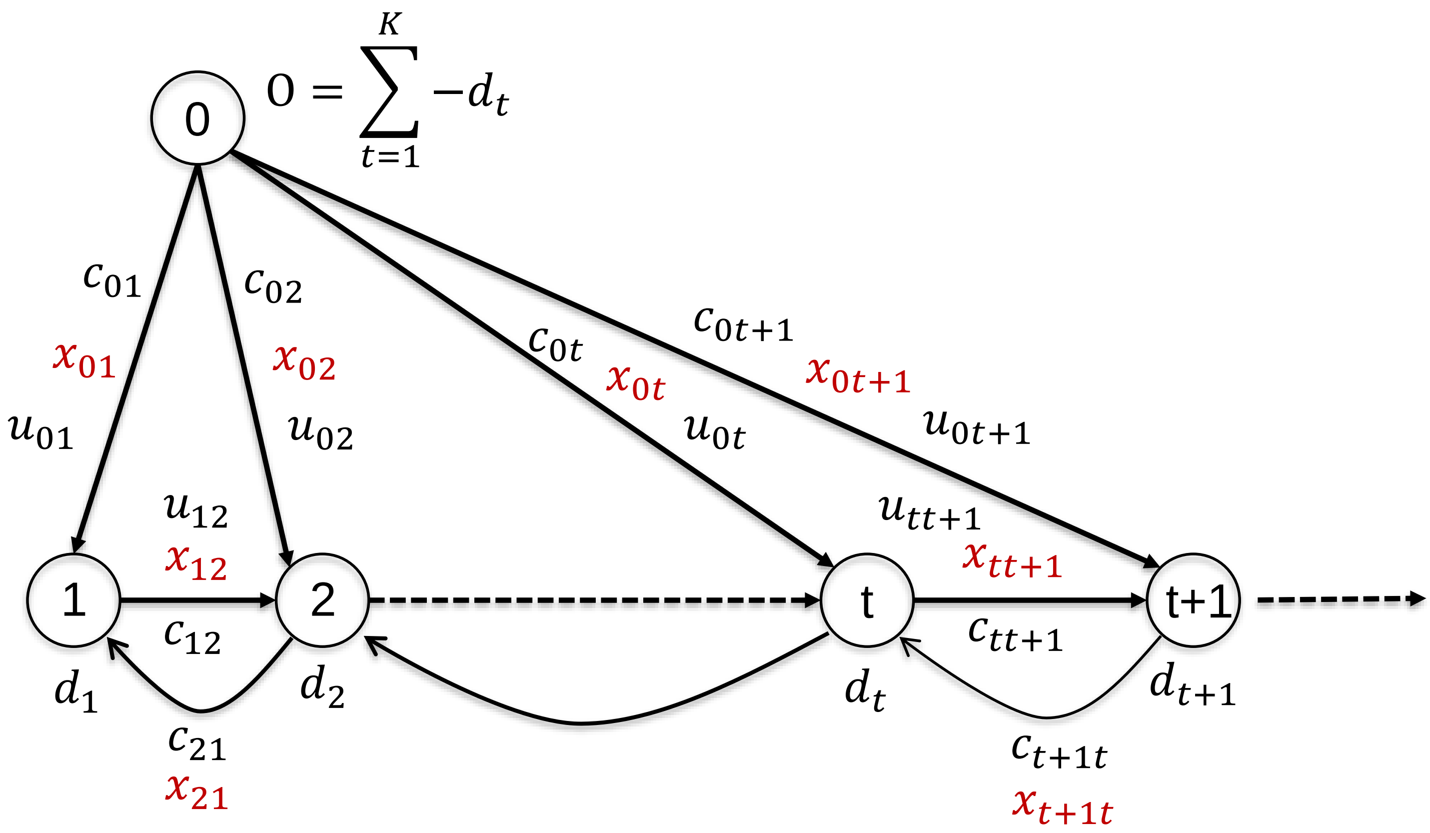
MIN COST FLOW MODEL: Applications

Let us now consider the possibility of not being able to satisfy all demands in all periods. We have the so-called *backorders* scheme, according to which it is planned, for example, to produce in period $t+1$ to satisfy the (un-satisfied) demand of period t.

The consumer at time t will be compensated with a *penalty*. So there will be costs for ‘sending’ the goods from a later period to an earlier period (B_i).

The problem is to find the minimum cost flow of goods that satisfies all demands.

MIN COST FLOW MODEL: Applications



Network with *backorders.*

TRANSPORTATION PROBLEM

The transportation problem is one of the most important linear programming problem and it is a classic application of minimum cost flow models.

It is defined as follows:
Consider a set n of depots, where the generic depot i = 1,...,n has a generic quantity a_i of good that needs to be transported to m markets.

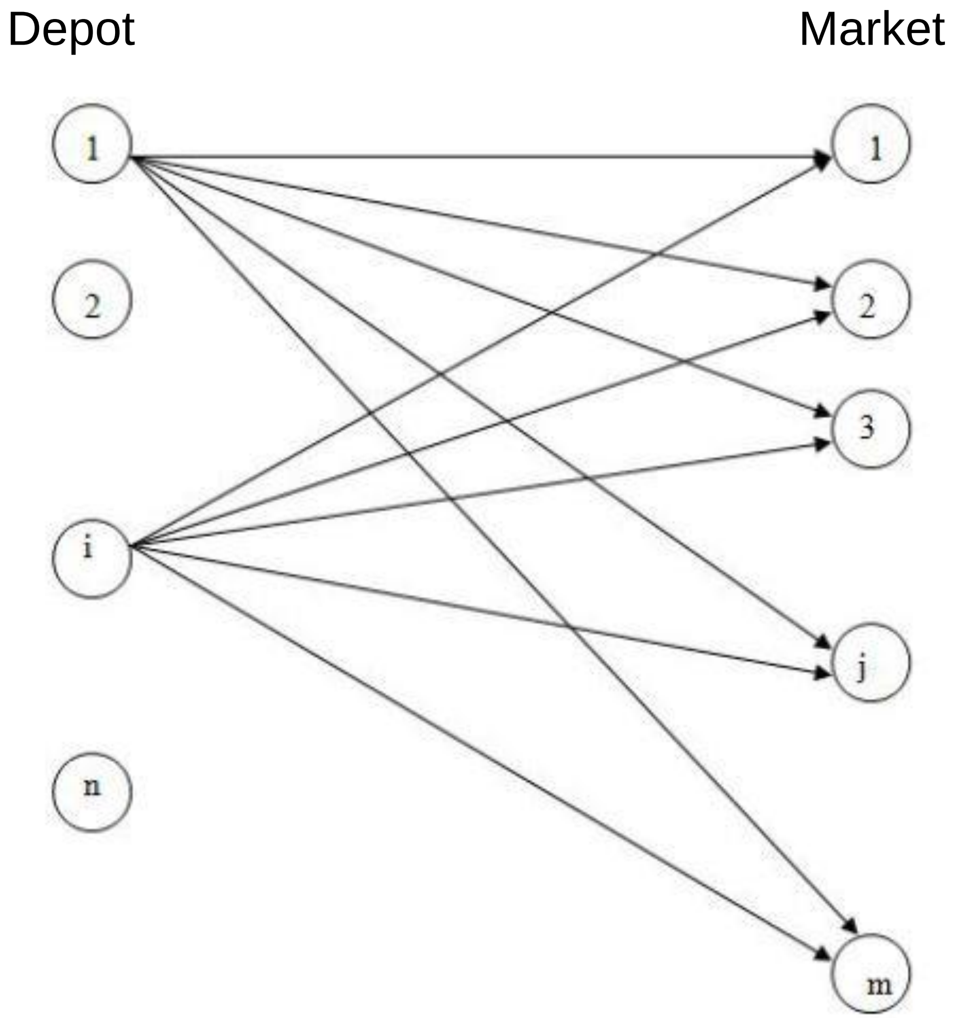
Let the demand for the good in each of the markets j = 1,...,m be denoted by b_j and let c_{ij} be the unit cost of transport between depot i and market j.

Problem:
For each depot i = 1,...,n, determine the quantity of good/flow to be sent to the markets such that:

1. the demand of each market j = 1,...,m is satisfied
2. the constraints on the availability of each depot i =1,...,n are met
3. the total cost of transport is minimized

TRANSPORTATION PROBLEM

Transportation problem



TRANSPORTATION PROBLEM

Transportation problem

Let:

x_{ij} = the (unknown) quantity to be sent from depot i to market j, with i = 1,…,n and j = 1,…,m

The objective function (to minimize):

$$\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij}$$

TRANSPORTATION PROBLEM

Transportation problem

1. the quantity leaving each depot i and shipped to all markets j must not exceed the availability of the goods stored in the warehouse i, i = 1,…,n

$$\sum_{j=1}^{m} x_{ij} \leq a_i \qquad i = 1, \dots, n$$

2. the total amount of shipments arriving at each market j must satisfy market demand

$$\sum_{i=1}^{n} x_{ij} \geq b_j \qquad j = 1, \dots, m$$

3. No negative quantities of goods may be shipped

$x_{ij} \geq 0$ i = 1,…n; j =1,…,m

TRANSPORTATION PROBLEM

Transportation problem

$$Min \qquad \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}x_{ij}$$

$$\sum_{j=1}^{m} x_{ij} \leq a_i \qquad i = 1, \dots, n$$

$$\sum_{i=1}^{n} x_{ij} \geq b_j \qquad j = 1, \dots, m$$

$x_{ij} \geq 0$ i = 1,...n; j =1,...,m

TRANSPORTATION PROBLEM

Necessary and sufficient condition for the existence of at least one solution that satisfies all the constraints of the problem (*feasible solution*)

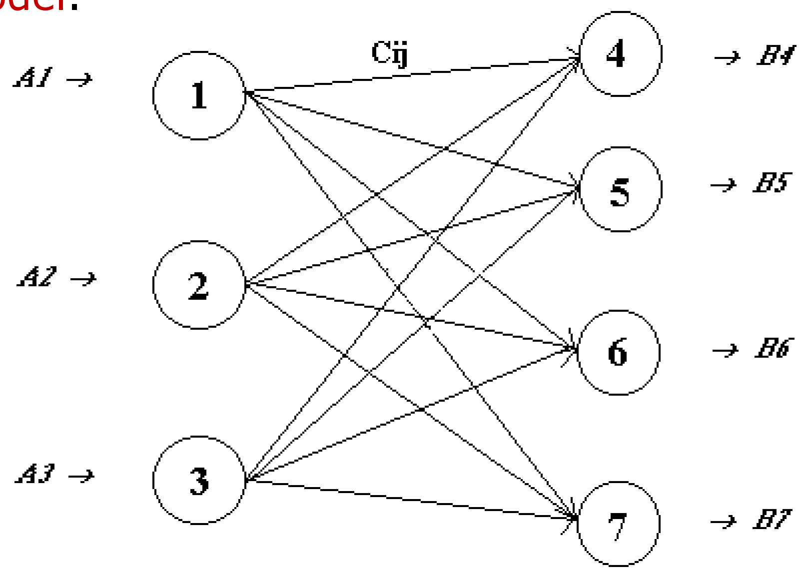
$$\sum_{i=1}^{n} a_i \geq \sum_{j=1}^{m} b_j$$

That is, the total supply is greater than or equal to the total demand.

If, in a transportation problem, the condition of the theorem is satisfied with equality, the model is said balanced, otherwise it is unbalanced.

TRANSPORTATION PROBLEM

The transportation problem can be reformulated as a minimum cost flow model.



At each node j the sum of the incoming flow quantities (x_{ij}) minus the sum of the outgoing flow quantities (x_{jk}) must be equal to the flow demand or supply at that node.

The flow conservation constraints are:

$$\sum_i x_{ij} = b_j \qquad j=4,5,6,7$$

$$-\sum_k x_{jk} = -a_j \qquad j=1,2,3$$

TRANSPORTATION PROBLEM: Example

Energy transport problem (Winston, Albright, 1997). The electricity company Power owns 3 power generation plants and it has to supply the necessary energy to 4 cities.

PLANTS	Energy Supply (milion kw/h)
Plant 1	35
Plant 2	50
Plant 3	40

CITIES	Energy demand (milion kw/h)
city 1	45
city 2	20
city 3	30
city 4	30

TRANSPORTATION PROBLEM: Example

The unitary shipping cost for each pair plant-city is

	city 1	city 2	city 3	city 4
plant 1	8	6	10	9
plant 2	9	12	13	7
plant 3	14	9	16	5

Problem: minimize the total transportation cost satisfying all demands.

TRANSPORTATION PROBLEM: Example

variables

$x_{ij}, \quad i = 1,2,3 \quad j = 1,2,3,4$

It is the amount of energy to be supplied by plant i to city j, expressed in millions of kilowatt per hours.

Constraints at the origins

The amount of energy available in each plant is limited.

$$x_{11} + x_{12} + x_{13} + x_{14} \le 35$$
$$x_{21} + x_{22} + x_{23} + x_{24} \le 50$$
$$x_{31} + x_{32} + x_{33} + x_{34} \le 40$$

Constraints to destinations

The energy demand of each city must be met.

$$x_{11} + x_{21} + x_{31} \ge 45$$
$$x_{12} + x_{22} + x_{32} \ge 20$$
$$x_{13} + x_{23} + x_{33} \ge 30$$
$$x_{14} + x_{24} + x_{34} \ge 30$$

TRANSPORTATION PROBLEM: Example

Objective function

The total cost of supplying energy to the plants must be minimised.

$$\begin{aligned} \min\ & 8x_{11} + 6x_{12} + 10x_{13} + 9x_{14} + \\ & + 9x_{21} + 12x_{22} + 13x_{23} + \\ & + 7x_{24} + 14x_{31} + 9x_{32} + 16x_{33} + 5x_{34} \end{aligned}$$

Non-negativity constraints

The amount of energy sent is positive or zero.

$$x_{11}, x_{12}, x_{13}, x_{14}, x_{21}, x_{22}, x_{23}, x_{24}, x_{31}, x_{32}, x_{33}, x_{34} \geq 0$$

TRANSPORTATION PROBLEM: Example

Objective function

$$\min\ 8x_{11} + 6x_{12} + 10x_{13} + 9x_{14} + $$
$$+ 9x_{21} + 12x_{22} + 13x_{23} + $$
$$+ 7x_{24} + 14x_{31} + 9x_{32} + 16x_{33} + 5x_{34}$$

constraints

$$x_{11} + x_{12} + x_{13} + x_{14} \le 35$$
$$x_{21} + x_{22} + x_{23} + x_{24} \le 50$$
$$x_{31} + x_{32} + x_{33} + x_{34} \le 40$$

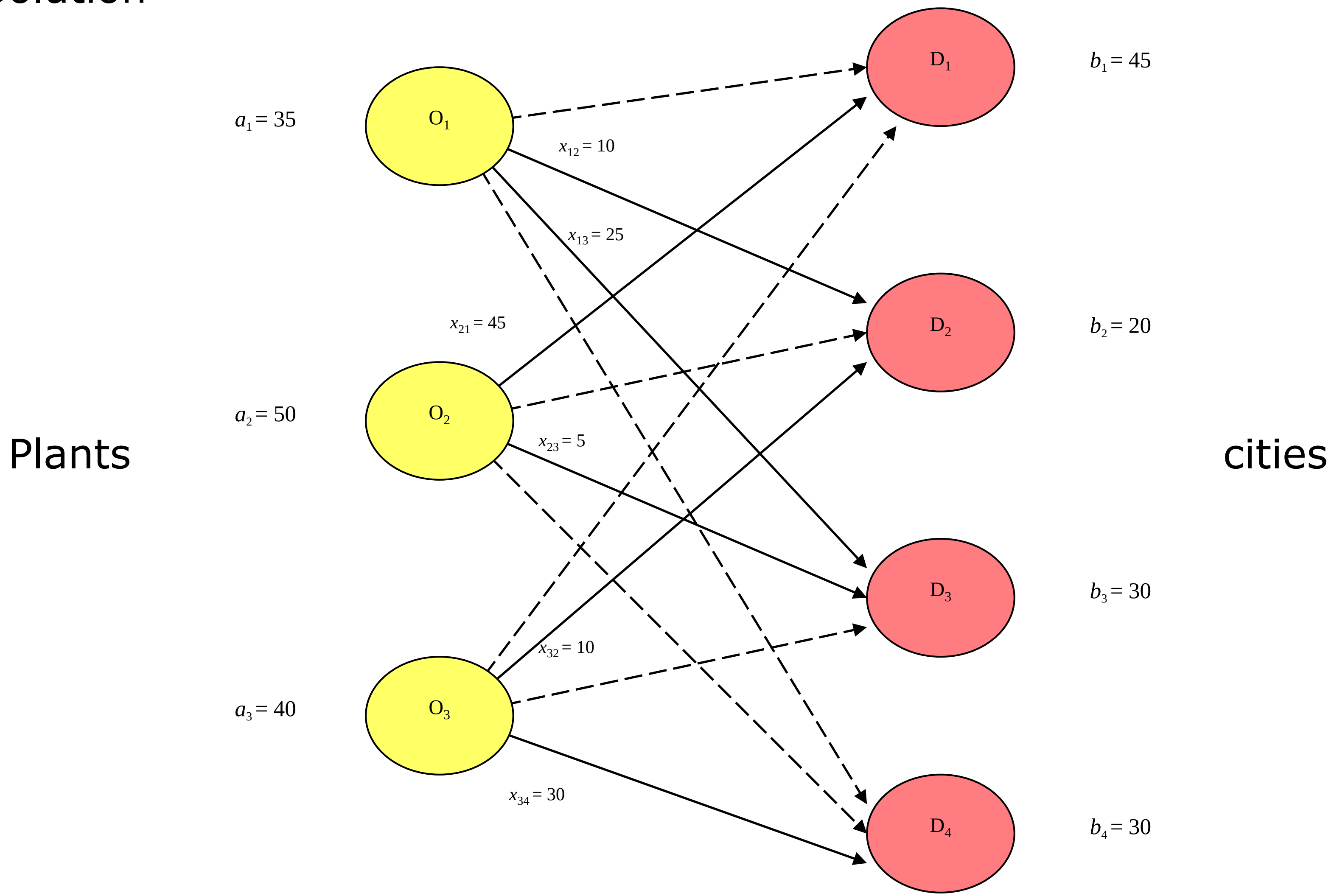
$$x_{11} + x_{21} + x_{31} \ge 45$$
$$x_{12} + x_{22} + x_{32} \ge 20$$
$$x_{13} + x_{23} + x_{33} \ge 30$$
$$x_{14} + x_{24} + x_{34} \ge 30$$

non-negativity

$$x_{11}, x_{12}, x_{13}, x_{14}, x_{21}, x_{22}, x_{23}, x_{24}, x_{31}, x_{32}, x_{33}, x_{34} \ge 0$$

TRANSPORTATION PROBLEM: Example

Solution



SHORTEST PATH PROBLEM

Given a directed graph G=(V,A), representing a transportation network, assume that costs c_{ij} are assigned to each arc (i,j)∈A. They are the cost/length between the pair of adjacent vertices i and j.

Problem: Given two specific nodes u and v, find the path of minimum cost/length between u and v.

SHORTEST PATH PROBLEM

The Shortest Path Problem (SPT) from a node u to a node v, can be formulated as a minimum cost flow problem where there is the supply of a unit of flow in u (b_u = -1), the demand of a unit of flow in v (v = 1) and the balancing of incoming and outgoing flows in all other nodes $i \in V$ ($b_i = 0$).

SHORTEST PATH PROBLEM

We obtain the following program:

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

$$\sum_{k\in P(i)} x_{ki} - \sum_{k\in S(i)} x_{ik} = 0 \qquad i \neq u, v$$

$$-\sum_{k\in S(u)} x_{uk} = -1$$

$$\sum_{k\in P(v)} x_{kv} = 1$$

$$0 \leq x_{ij} \leq 1 \qquad \forall (i,j) \in A$$

SHORTEST PATH PROBLEM

NOTE:

The optimal solution of the problem can only have values 0 or 1 (binary) and will provide a minimum path between u and v by the following interpretation:

$x_{ij} = 1$ The arc (i,j) belongs to the minimum path

$x_{ij} = 0$ The arc (i,j) does not belong to the minimum path

It is not obvious that the solution is binary since the constraint on the variables says that x_{ij} can take values between zero and one. Nevertheless, the class of minimum cost flow problems and, in particular, the Shortest Path problem verifies certain mathematical properties: ALWAYS in any optimal solution of the problem the values of the variables x_{ij} are equal to 1 or 0.