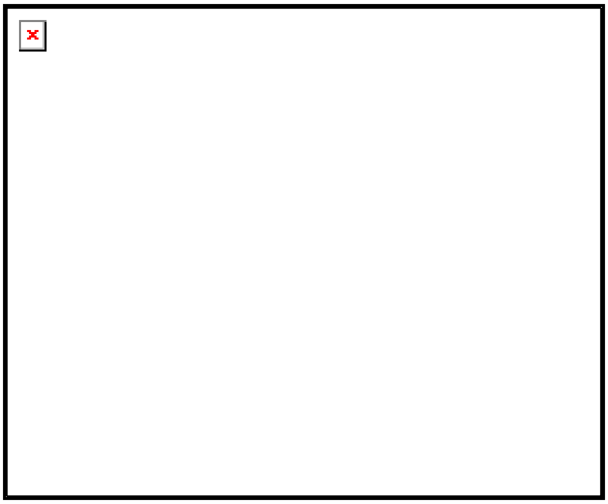
Section 5.2 Directed Graphs

Purpose of Section: We introduce the **directed graph** and its adjacency matrix, and then show how directed graphs arise in connection with tournaments, social networking, the Google Page Rank system, and dynamic programming.

Introduction

Thus far in our discussion of graphs we have not associated a "direction" with the edges. In many models, however, such as traffic flow problems with one-way streets, we restrict the direction of movement along the edges to a single direction. A **directed graph** (or **digraph**) is defined similarly as a (undirected) graph, except the edges are "directed" from one vertex to another. We call these types of edges **directed edges** (or **arcs**). If a directed edge goes from vertex u to vertex v, then u is called the **head** (or **source**) and v is called the **tail** (or **sink**) of the edge, and v is said to be the **direct successor** of u and u is the **direct predecessor** of v. See Figure 1.



A directed graph with directed edges joining the vertices
Figure 1

The applications of directed graphs range from transportation problems in which traffic flow is restricted to one direction and one-way communication problems, to asymmetric social interactions, athletic tournaments, and even the World-Wide Web. Before talking about applications, however, we introduce the concept of the adjacency matrix of a directed graph.

Definition The **adjacency matrix** of a directed graph with n vertices is an $n \times n$ matrix $M = (m_{ij})$, where

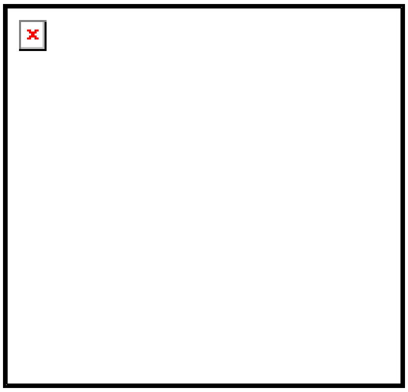
$$m_{ij} = \begin{cases} 1 & \text{if there is a directed edge from vertex } i \text{ to vertex } j \\ 0 & \text{if there is not a directed edge from vertex } i \text{ to vertex } j \end{cases}$$

For example, the adjacency matrix for the digraph in Figure 2 is

$$M = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Tournament Graphs (Dominance Graphs)

A **tournament graph** (or **dominant graph**) is a directed graph where every pair of vertices is joined by *exactly one* directed edge. In other words, for each pair of vertices u and v there is a directed edge from u to v or a directed edge from v to u. Figure 2 shows a tournament graph with five vertices.



Tournament graph with five vertices
Figure 2

Tournament graphs are aptly named since they model round-robin tournaments in tennis, baseball, and so on, where every team plays every other team exactly once (and assume no ties).

The number of directed edges "leaving" a vertex u is called the **out-degree** of the vertex and denoted by $\text{od}(u)$. In the directed graph in Figure 2, we have $\text{od}(5) = 3, \text{od}(1) = 2$. If the vertices of the graph are athletic teams, and a directed edge from vertex u to vertex v means team u beats team v, and the out-degree of u represents the number of wins for team u. On the other hand, the number of directed edges arriving at a vertex v is called **in-degree** of v and denoted by $\mathrm{id}(v)$. In the graph in Figure 2, we have $\mathrm{id}(5)=1$, $\mathrm{id}(1)=2$. In connection with athletic tournaments, the in-degree of vertex v represents the number of losses for team v.

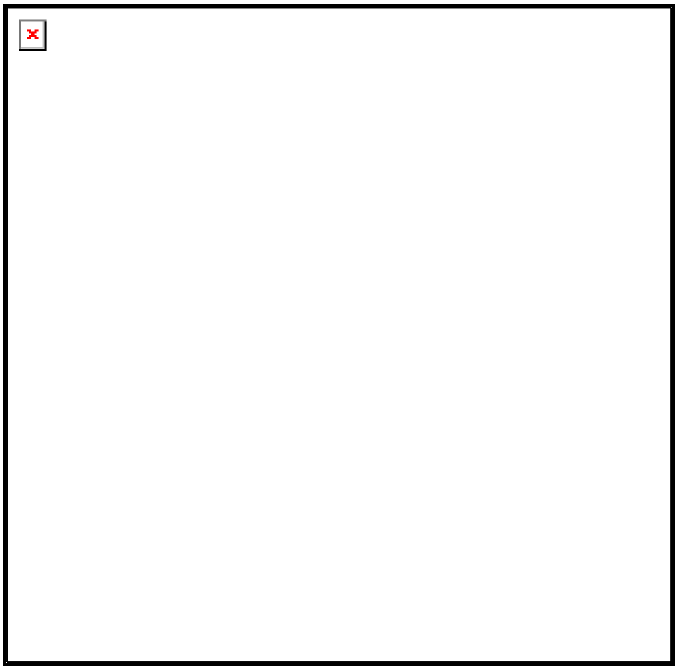
Tournament graphs are even used by sociologists, who call them dominance graphs, and are used to model social interactions.

Dominance Graphs in Social Networking

Many social situations involve people or groups of people (countries, cultures, universities, and so on) in which some individuals or group "dominates" others. The word "dominate" can refer to many kinds of dominance: cultural, physical, political, economic, and so on. Nowadays, online social networking services such as Facebook, Twitter, LinkedIn, and so on, bring people together and offer interesting dynamics between individuals.

Suppose a sociologist wishes to study dominance patterns in a close-knit group of college women. The group consists of five woman: Amy (A), Betty (B), Carol (C), Denise (D), and Elaine (E). The sociologist begins by conducting interviews with each pair to determine their pair-wise dominance.[1] If person A dominates person B, we denote this by writing $A \to B$. (We assume in this simple model that for each pair of students, one person dominates the other.) After conducting the interviews, the sociologist draws the dominance graph which represents pair-wise dominations of the entire group, which is shown in Figure 3. Note that Amy dominates Betty and that Denise dominates Carol.

[1] The determination of one-on-one dominance can be carried out by a series of questions, although it can be subjective in some instances.



Dominance graph
Figure 3

The adjacency matrix for the dominance graph is

		Amy	Betty	Carol	Denise	Elaine
	Amy	0	1	1	0	1
	Betty	0	0	0	0	0
$M =$	Carol	0	1	0	0	1
	Denise	1	1	1	0	0
	Elaine	0	1	0	1	0

The number of 1's in each row is the out-degree of the row vertex and represents the number of **first-stage dominances** of that individual. Note that vertices Amy and Denise each have a "score" of 3, vertices Carol and Elaine each have a score of 2, and Betty has a score of 0. In other words, Amy and Denise each dominate three people in first stage dominances, whereas Carol dominates two people in the first stage, and Betty dominates no one.

The goal is to find the **group leader**[2]. If one person dominates every person in the group, then we would call that person the group leader. If no one person dominates every other person, then the person with the most first-stage

[2] If these were athletic teams playing in a round-robin tournament, we would want to know who should be declared the winner.

dominances is declared group leader. If two or more people *tie* with the most first-stage dominances, then we must resort to **second-stage dominances**. In our example, Amy and Denise are tied, each with 3 first-stage dominances.

To understand second-stage dominances, note that Elaine dominates Denise and Denise dominates Amy. Hence, we say that Elaine has a second-stage dominance over Amy and we denote this by Elaine $\rightarrow$ Denise $\rightarrow$ Amy. To find the number of second-stage dominances, consider the square[3] of the adjacency matrix M^2, which is

$M^2 =$

	Amy	Betty	Carol	Denise	Elaine
Amy	0	2	0	1	1
Betty	0	0	0	0	0
Carol	0	1	0	1	0
Denise	0	2	1	0	2
Elaine	1	1	1	0	0

To interpret M^2, observe in the dominance graph in Figure 3 that Amy $\rightarrow$ Elaine $\rightarrow$ Denise, which is the only second-stage dominance of Amy over Denise. This fact is indicated by the 1 in row Amy, column Denise of M^2. Also note that Amy has two second-stage dominances over Betty (Amy $\rightarrow$ Carol $\rightarrow$ Betty) and (Amy $\rightarrow$ Elaine $\rightarrow$ Betty), indicated with a 2 in row Amy, column Betty of M^2. In other words, the elements of M^2 give the second-order dominances of row members over column members. From M^2 we see that Denise has five second-state dominances (sum of the numbers in row Denise) while Amy has four (sum of the numbers in row Amy). Thus, we declare Denise the group leader.[4]

We could continue our analysis of group dominance by computing the sum $M + M^2$, whose elements give the total number of first-stage and second-stage dominances of one person over another.

---

[3] It is not necessary that the reader know how to find the square of a matrix. One can simply look at the graph to count the second-order dominances.

[4] Since Amy and Denise both have at least three second-stage dominances, we could compute $M^3(G)$, which would give the number of **third-stage dominances**. One suspects, however, that although we might define third-order dominances in theory, it is difficult to observe them in reality.

$M + M^2 =$

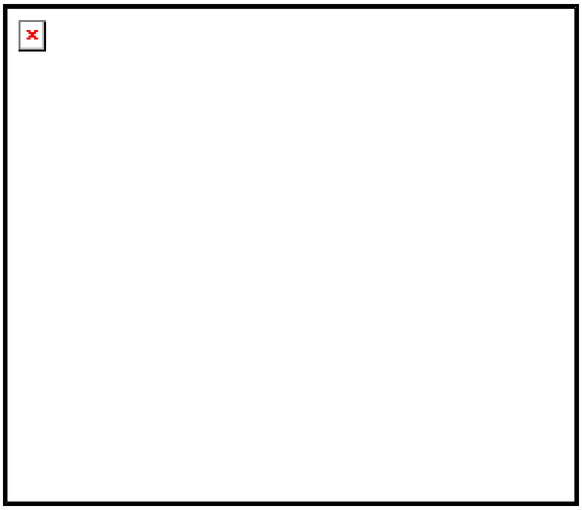
	Amy	Betty	Carol	Denise	Elaine
Amy	0	3	1	1	2
Betty	0	0	0	0	0
Carol	0	2	0	1	1
Denise	1	3	2	0	2
Elaine	1	2	1	1	0

Here, Amy has 7 first- and second-stage dominances over the members in the group (Can you find them?) while Denise has 8, so we call Denise the leader of the group.

The group leader in the social network depended on the out-degrees of the vertices. For the problem of ranking webpages, we focus on the in-degrees or the number of links pointing at a webpage

Google's PageRank System

Google's search engine models the internet as a directed graph, where vertices are webpages and the edges are links between webpages. The strategy behind Google's PageRank system is based on simulating the actions of several online individuals and determining where these individuals will most likely be after a period of time. Consider the tiny internet of four webpages as drawn in Figure 4 with several individuals online.

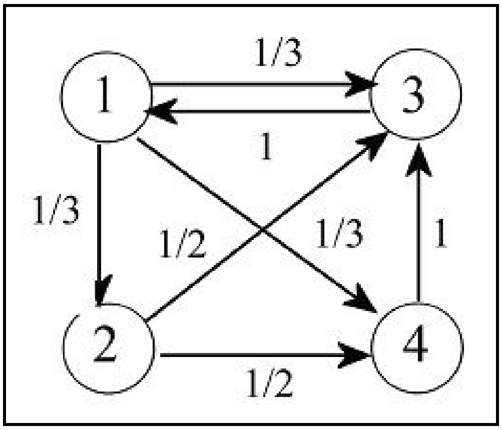


Web graph of a small internet
Figure 4

Assume the movement of the individuals from webpage to webpage obeys the following **transition matrix**

		webpage 1	webpage 2	webpage 3	webpage 4
	webpage 1	0	0	1	0
$T =$	webpage 2	1/3	0	0	0
	webpage 3	1/3	1/2	0	1
	webpage 4	1/3	1/2	0	0

which gives the probabilities that someone currently online in the column webpage will move to the row webpage. Note that there are equal 1/3 probabilities that someone on webpage 1 will move to one of the webpages 2,3, or 4. A person at webpage 2 will move to either webpage 3 or 4 with 1/2 probability, and individuals at webpage 4 will move to webpage 3 with probability 1. We draw these probabilities alongside the edges in the directed graph in Figure 5.



Probabilities of moving from page to page
Figure 5

Initially, assume the fraction of individuals at each website is 0.25, which we represent by the **PageRanks**

$$R_0 = (0.25,\ 0.25, 0.25, 0.25)$$

which gives the PageRank of each page. We now simulate the movement of the online individuals to determine where they most likely will be after one "click of the mouse." Their most likely locations will be the product of the transition matrix times the initial PageRank distribution, or

$$R_1 = TR_0 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.08 \\ 0.45 \\ 0.21 \end{bmatrix}$$

The new PageRank numbers tell us, probabilistically speaking, that 25% of online people will be at page 1, 8% at page 2, 45% at page 3, and 21% at page 4. The individual fractions in the PageRank come from the computations

$$\begin{aligned}
&\text{fraction on page } 1 = (1)(0.25) = 0.25 \\
&\text{fraction on page } 2 = (1/3)(0.25) = 0.08 \\
&\text{fraction on page } 3 = (1/3)(0.25) + (1/2)(0.25) + (1)(0.25) = 0.45 \\
&\text{fraction on page } 4 = (1/2)(0.25) + (1/3)(0.25) = 0.21
\end{aligned}$$

The new fraction of visitors on a given page is found by multiplying the fraction of individuals at the different pages times the probability people will move to the given page.

We now continue this process again and again, finding

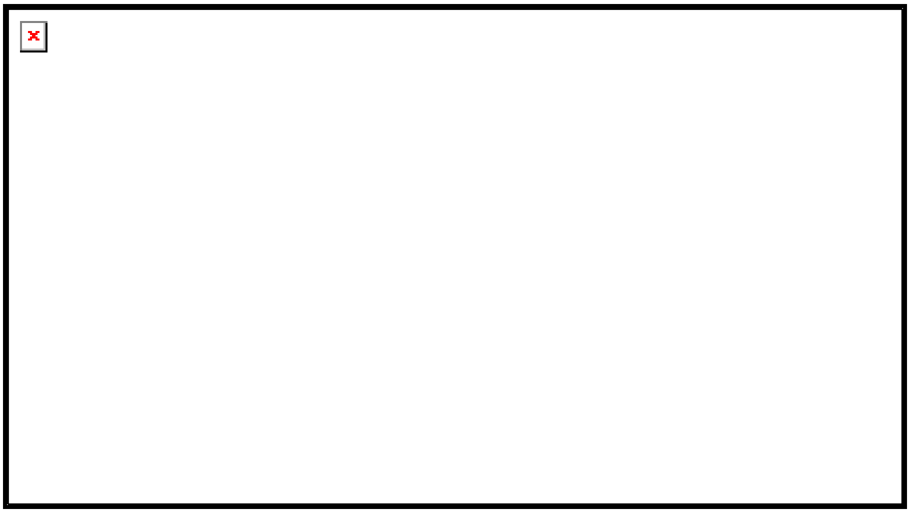
$$R_2 = \begin{bmatrix} 0.45 \\ 0.08 \\ 0.33 \\ 0.12 \end{bmatrix}, R_3 = \begin{bmatrix} 0.33 \\ 0.16 \\ 0.32 \\ 0.19 \end{bmatrix}, R_4 = \begin{bmatrix} 0.32 \\ 0.12 \\ 0.37 \\ 0.19 \end{bmatrix}, \ldots, R_{100} = \begin{bmatrix} 0.35 \\ 0.13 \\ 0.35 \\ 0.17 \end{bmatrix}$$

which means in the long run 35% of individuals will be on webpage 1, 13% on webpage 2, 35% on webpage 3, and 17% on webpage 4. For that reason Google would rank these pages accordingly.

Important Note: As the reader might suspect, Google's actual PageRank system includes several bell and whistles in addition to this basic description. The mathematics behind Google's search engine is a Markov Chain. The probabilistic model describing the movement of individuals through the internet is called a **Markov Chain**. The states of the Markov Chain are the webpages and Google wants to know the steady state of the Markov Chain, which determines which states are most popular.

Dynamic Programming

Many applications of graph theory relate to the movement of objects -- be it cars, trucks, airplanes, or even internet packets from one location to another -- and whatever objects are being moved involve costs. Figure 6 shows a directed graph which represents a collection of possible paths from the "start" to the "end" with numbers on the edges representing the cost of traversing the edge. The problem is to find the path that minimizes the total cost of going from the start to the end.



Minimum path problem
Figure 6

Although a quick examination of this graph will convince you the minimum path is $1-3-4-7-9$, resulting in a minimum cost of 13, for larger graphs with maybe 1000 nodes a visual inspection of the graph would probably provide no useful information about the shortest path.

To solve this problem, we use a powerful technique called **dynamic programming**, introduced by Richard Bellman in the 1950s. The general philosophy of the technique is to subdivide complex problems into smaller parts, solve the component parts and use these results to solve the larger problem.

In the current problem, the strategy is to work backwards, solving sub-problems of finding the shortest distance from a given vertex to the end vertex. We begin by introducing the quantity

$$s_i = \text{ shortest distance from vertex } i \text{ to the end vertex}$$

In the graph in Figure 6, we have $s_9 = 0,\ s_7 = 4,\ s_6 = 9,\ s_8 = 10$. Our goal is to find s_1 We now let

$$d_{ij} = \text{distance from vertex } i \text{ to vertex } j$$

which means $d_{12} = 3,\ d_{46} = 5$ and so on. Hence,

$d_{ij} + s_j =$ shortest distance from vertex i to the end vertex traveling first to vertex j followed by the shortest path to the end vertex.

Hence, to find the shortest distance s_i from vertex i to the end vertex, we compute the minimum

$$s_i = \min_j \left\{ d_{ij} + s_j \right\}$$

taken over all vertices j connected by an edge to vertex i. For example, to find the shortest distance from vertex $i = 4$ to the end vertex, we find the minimum of the distances

4 to 5 distance + min distance home from 5 $= d_{45} + s_5 = 3 + 8 = 11$
4 to 6 distance + min distance home from 6 $= d_{46} + s_6 = 5 + 9 = 14$
4 to 7 distance + min distance home from 7 $= d_{47} + s_7 = 5 + 4 = 9$
4 to 8 distance + min distance home from 8 $= d_{48} + s_8 = 8 + 10 = 18$

Taking the minimum of these sub-problems, the shortest distance from vertex 4 to the end vertex is $s_4 = 9$. We continue this process, moving backwards in the graph and finding the minimum distances $s_8, s_7, \ldots, s_1$ from each vertex, although not necessarily descending in perfect order. We now advise the reader to get out pencil and paper compute the shortest distance home from each vertex, eventually getting s_1, the shortest distance from the starting vertex to the final vertex. The computations also yield the path that gives this shortest distance.

$$s_9 = 0$$

$$s_8 = d_{89} + s_9 = 10 + 0 = 10 \qquad (8 \to 9)$$

$$s_7 = d_{79} + s_0 = 4 + 0 = 4 \qquad (7 \to 9)$$

$$s_6 = \min \begin{cases} d_{68} + s_8 \\ d_{69} + s_9 \end{cases} = \min \begin{cases} 5 + 10 \\ 9 + 0 \end{cases} = 9 \qquad (6 \to 9)$$

$$s_5 = d_{57} + s_7 = 4 + 4 = 8 \qquad (5 \to 7)$$

$$s_4 = \min\begin{cases} d_{45}+s_5 = 3+8 = 11 \\ d_{46}+s_6 = 5+9 = 14 \\ d_{47}+s_7 = 5+4 = 9 \\ d_{48}+s_8 = 8+10 = 18 \end{cases} = 9 \qquad (4 \to 7)$$

$$s_3 = \min\begin{cases} d_{34}+s_4 \\ d_{36}+s_6 \end{cases} = \min\begin{cases} 2+9 = 11 \\ 12+9 = 21 \end{cases} = 11 \qquad (3 \to 4)$$

$$s_2 = \min\begin{cases} d_{25}+s_5 \\ d_{24}+s_4 \end{cases} = \min\begin{cases} 10+8 \\ 3+11 \end{cases} = 14 \qquad (2 \to 4)$$

$$s_1 = \min\begin{cases} d_{12}+s_2 \\ d_{13}+s_3 \end{cases} = \min\begin{cases} 3+14 \\ 2+11 \end{cases} = 13 \qquad (1 \to 3)$$

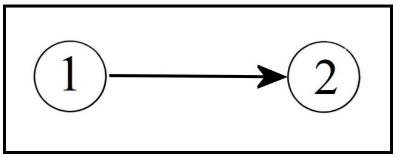
Hence, the shortest distance from vertex 1 to vertex 9 is 13 and retracing the path from start to end we find path that gives the shortest distance is

$$(1 \to 3),\ (3 \to 4),\ (4 \to 7),\ (7 \to 9) \ \text{ or } \ 1 \to 3 \to 4 \to 7 \to 9\,.$$ ▌

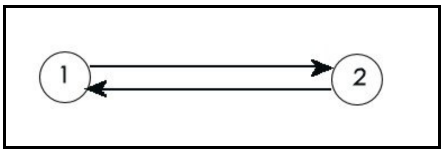
Problems

For the digraphs in Problems 1-6, find the adjacency matrix M. Then compute M^2 and $M+M^2$ and verify that the elements of these matrices agree with the number of dominations in the graphs.

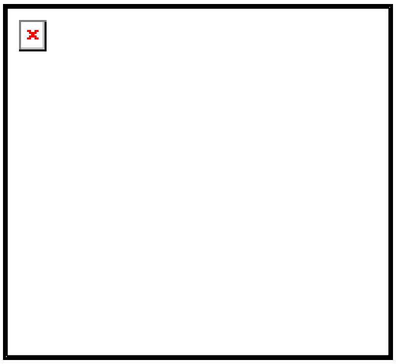
1.



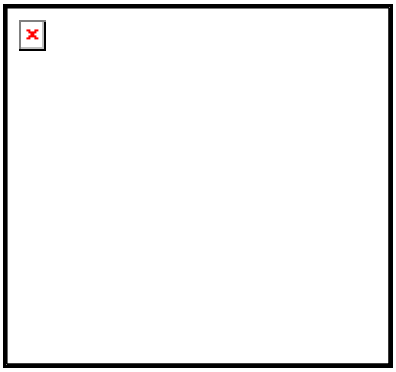
2.



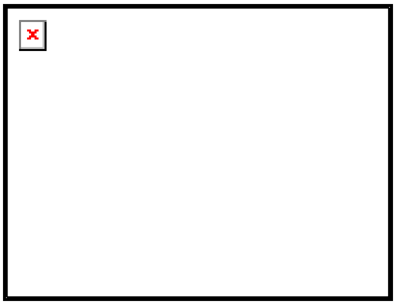
3.



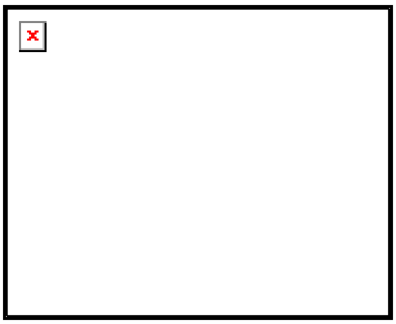
4.



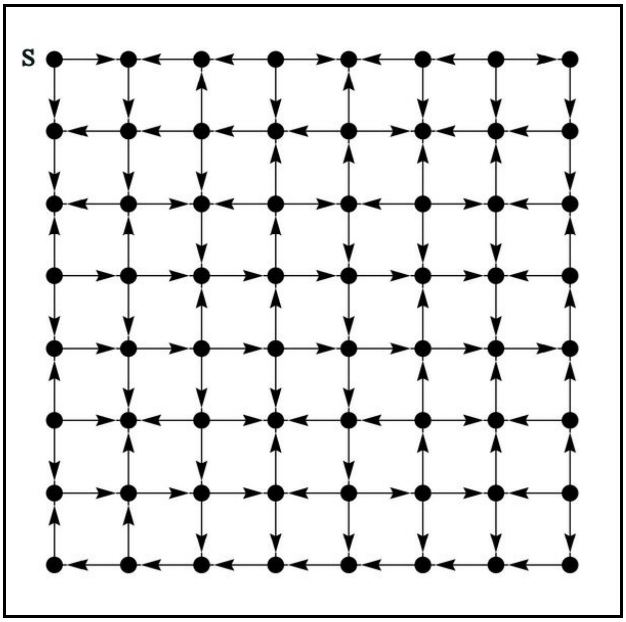
5.



6.



5. **Fun Problem** Find all vertices that are reachable from the starting vertex S along a directed path.



6. **Group Dominance**

The graph shown in Figure 7 shows the dominance of a group of four classmates: Andy. Betty, Charlie, and Denise.

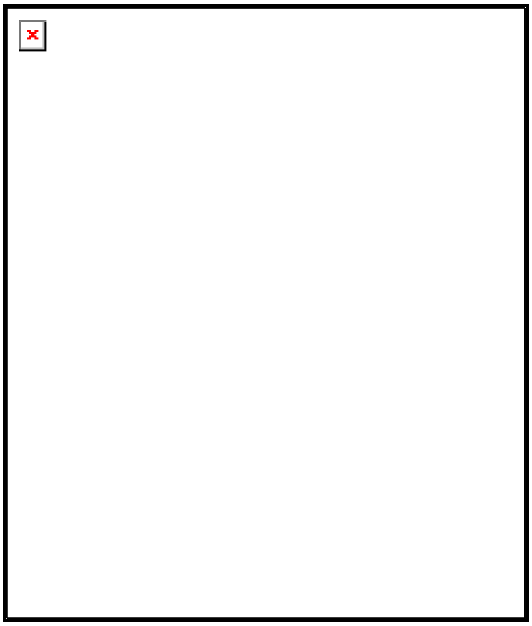
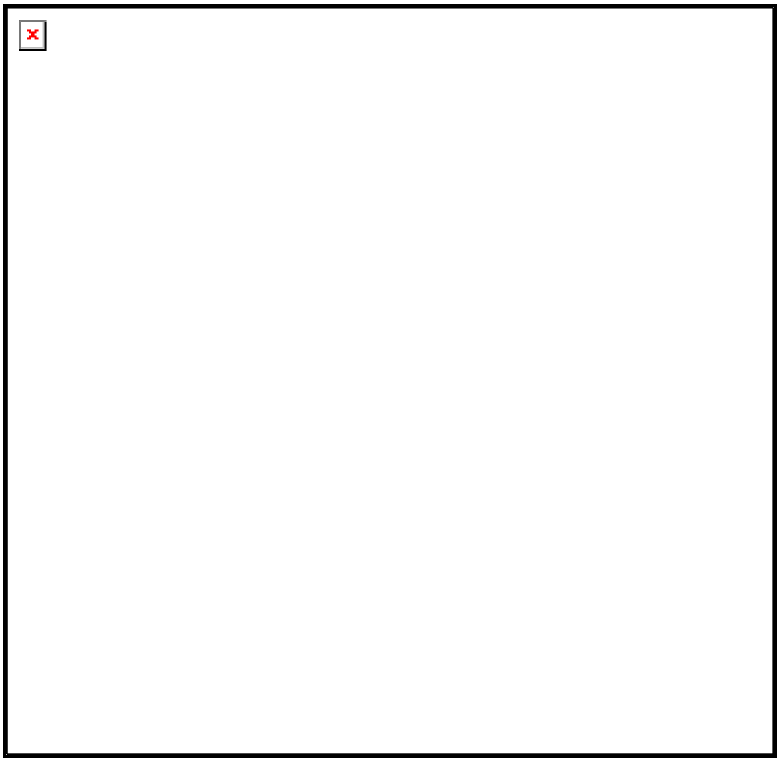


Figure 7

a) Construct the adjacency matrix M for this graph.
b) Is there a first-stage dominance leader?
c) Compute M^2 and interpret its elements.
d) Who is the group leader?
e) Which person is dominated by the most other people?

7. **Round-Robin Tournaments**[5]

The tournament graph in Figure 8 shows the results of a round-robin tournament between five baseball teams: the University of Texas, Texas A&M, Texas Tech, Baylor, and SMU.



Round-robin tournament graph
Figure 8

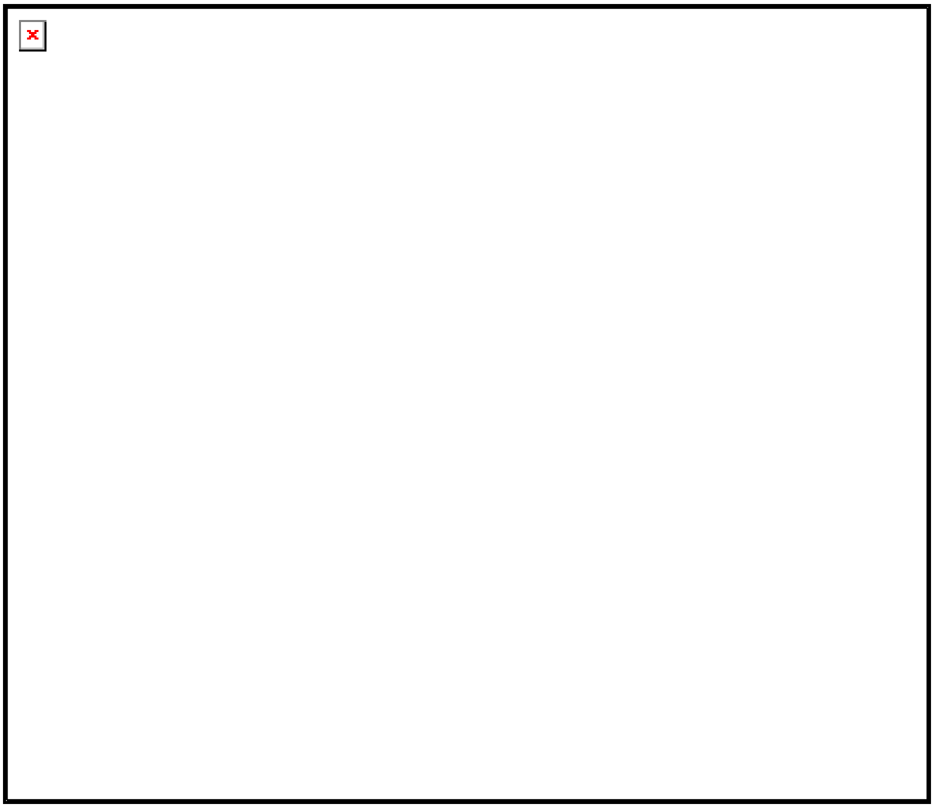
a) Construct the adjacency matrix M for this graph.
b) Is there a consensus leader for the group?
c) Compute M^2 and interpret its elements.
d) Which team is the conference winner?

[5] The author apologizes to any Longhorn fan. The graph was given to the author by an Aggie alumni.

10. **Landau's Theorem** A theorem by Landau states that if some vertex u in a dominance graph has a larger out-degree than all other vertices, then either u dominates all other vertices v, or if it does *not* dominate a given vertex v, then u dominates a third vertex w which in turn dominates v. What does the theorem say in the language of round-robin tournaments? Verify this theorem for the dominance graphs in Problems 1-6.

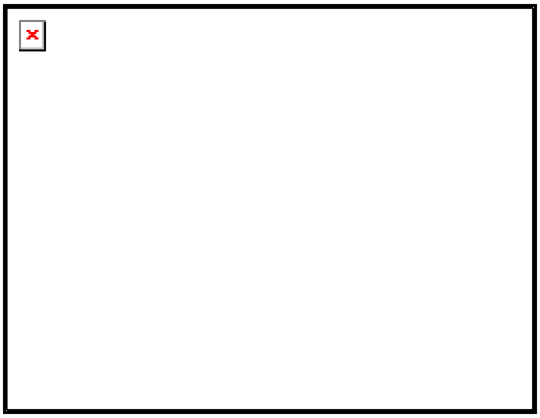
11. **Landau's Theorem in the Yankee Conference** Suppose the football teams in the Yankee Conference play every other team exactly once during the season. At the end of the season Maine has won more games than any other team. However, Maine lost to Vermont. What does Landau's theorem say in the language of the Yankee Conference?

12. **Dynamic Programming** Use dynamic programming to find the shortest way to accomplish the project in Figure 9.



Shortest time problem
Figure 9

13. **Webpages** The following dominance graph illustrates a tiny internet of four webpages where the vertices represent the webpages and the directed edges represent links from one webpage to another. Rank the webpages from first to last.



ΓΣΘΨΞΠΩ