CSE225 Data Structures, 2018(FALL)

PROJECT #3

Deadline December 27, 2018, 16:00

Social Networks

This project is a programming assignment in C which aims to find influencer peaople in a social graph. Your program will read a data file containing a list of people names and the friendness they have. You will build a graph from the given data file. The vertices of the graph will be the people and there will be an edge between each person who have a friendness relationship.

You will process the graph and make the necessary calculations. The output of your program will be

- **a)** a representation of the graph you generated (can be viewed like adjacency matrix) and
- **b)** the centrality degrees .

The input will be in the following format:

Cem; Ayşe, Ferit, Dundar
Ayşe; Cem, Ferit, Dundar, Belma
Belma; Ayşe, Dundar, Edip
Edip; Belma, Dundar, Gamze
Dundar; Ayse, Belma, Cem, Ferit, Gamze, Edip
Gamze; Dundar, Edip, Ferit, Halit
Ferit; Ayşe, Cem, Dundar, Gamze, Halit
Halit; Ferit, Gamze, Ilke
Ilke; Halit, Jale
Jale; Ilke

a) (25 points) The output of your program must be in the following form:

As the output, the resulting graph can be displayed using either of the following formats:

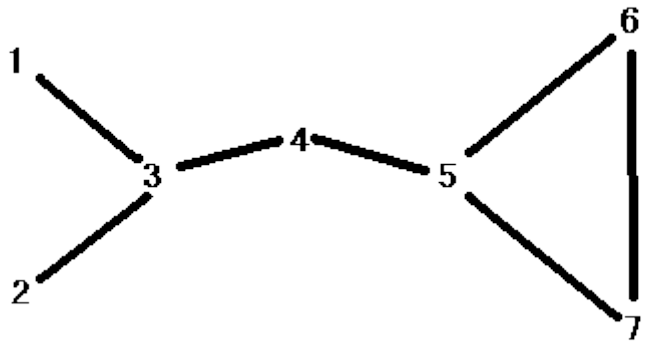
As an adjacency matrix:

The first 2 rows have been done for you:

	Cem	Ayşe	Belma	Edip	Dundar	Gamze	Ferit	Halit	Ilke	Jale
Cem	0	1	0	0	1	0	1	0	0	0
Ayşe	1	0	1	0	1	0	1	0	0	0
Belma										
Edip										
Dundar										
Gamze										
Ferit										
Halit										
Ilke										
Jale										

b)Build your graph and calculate the following values, Degree centrality(20 points), Closeness centrality(20 points), Betweenness centrality(20 points).

Example:



Degree centrality: Degree centrality of a node refers to the number of edges attached to the node. In order to know the standardized score, you need to divide each score by n-1 (n = the number of nodes). Since the graph has 7 nodes, 6 (7-1) is the denominator for this question.

Node	Score	Standardized Score
1	1	1/6
2	1	1/6
3	3	3/6 = 1/2
4	2	2/6 = 1/3
5	3	3/6 = 1/2
6	2	2/6 = 1/3
7	2	2/6 = 1/3

Closeness centrality: You need to calculate the inverted score after you count the total number of steps to a node. In order to know the standardized score, you need to divide a score by (n-1), then take inverse. Note that the most central node is node 4 while the most central node for degree centrality is node 3 and 5.

Node	Score	Standardized Score
1	1/16	6/16 = 3/8
2	1/16	6/16 = 3/8
3	1/11	6/11
4	1/10	6/10 = 3/5
5	1/11	6/11
6	1/15	6/15 = 2/5
7	1/15	6/15 = 2/5

Betweenness centrality: To calculate betweenness centrality, you take every pair of the network and count how many times a node can interrupt the shortest paths (geodesic distance) between the two nodes of the pair. For standardization, I note that the denominator is (n-1)(n-2)/2. For this network, (7-1)(7-2)/2 = 15. Note that node 5 has a little smaller centrality score that node 3 and 4 because the connection between node 6 and 7 reduces the controllability of node 5.

Betwennes Centrality:

$$C_{Betwennes}\ (v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}\ (v)}{\sigma_{st}}$$

Where, $C_{Betwennes}\ (v)$ is the betwenness centarlity of node V, σ_{st} is the number of shortest paths between all source and target pairs, $\sigma_{st}\ (v)$ is the number of shortests paths between all source and target pairs those pass through from node V.

Source	Target	Intermedia Nodes	Path
1	2	3	1-3-2
1	3	-	1-3
1	4	3	1-3-4
1	5	3,4	1-3-4-5
1	6	3,4,5	1-3-4-5-6
1	7	3,4,5	1-3-4-5-7
2	3	-	2-3
2	4	3	2-3-4
2	5	3,4	2-3-4-5
2	6	3,4,5	2-3-4-5-6
2	7	3,4,5	2-3-4-5-7
3	4	-	3-4
3	5	4	3-4-5
3	6	4,5	3-4-5-6
3	7	4,5	3-4-5-7
4	5	-	4-5
4	6	5	4-5-6
4	7	5	5-6-7
5	6	-	5-6
5	7	-	5-7
6	7	-	5-7

$C_{Betwennes}\ (1) = 0$

$C_{Betwennes}\ (2) = 0$

$C_{Betwennes}\ (3) = 9/15$

$C_{Betwennes}\ (4) = 9/15$

$C_{Betwennes}\ (5) = 8/15$

$C_{Betwennes}\ (6) = 0$

$C_{Betwennes}\ (7) = 0$

After making standardization (n-1)(n-2)/2=15

$C_{Betwennes}\ (1) = 0$

$C_{Betwennes}\ (2) = 0$

$C_{Betwennes}\ (3) = \frac{9}{225} = 0.04$

$C_{Betwennes}\ (4) = \frac{9}{225} = 0.04$

$C_{Betwennes}\ (5) = 8/225$=0.035

$C_{Betwennes}\ (6) = 0$

$C_{Betwennes}\ (7) = 0$

c)(15 points) What do you think about the information flow on this graph? What do you think the most powerful/critical node of this graph?,nk that this is a centralized graph? Why, why not?Do you think that

For PROJECT SUBMISSON:

1 page report + Code (by email also with hard copy to department secretary)

(name_surname.docx) both by email (**cse225.marmara.2018 at gmail dot com**)with the code by the deadline and 1 page report submission to department secretary.

with the following contents

REPORT:

a)Adjacency Matrix

b)

Source	Degree Centrality	Closeness Centrality	Betwenness Centrality

1			
2			
3			
4			
5			
6			
7			

Betwennes Centrality

Source	Target	Intermedia Nodes	Path
1	2		
1	3		
1	4		
1	5		
1	6		
...	...		

c)Comments

CODE:

Code (name_surname.c)

When I run the code, the output format will be the same as in the report.

I will review and run your code and compare the results of the output with the results in the report.

If I get different results, you will fail the course and have dicisplinary penalty.

Ref: http://www.sscnet.ucla.edu/soc/faculty/mcfarland/soc112/cent-ans.htm

The main goal of this project is to be familiar with Graphs. So, you need to use Graph data structure.

CODE SUBMISSION:

You should use the following email address in order to submit your code:
cse225.marmara.2018 at gmail dot com.