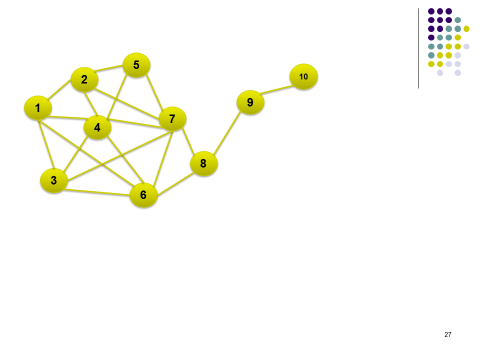
**Sample Questions for Social Media Analytics**

1. In the diagram below, which is the node with highest betweenness centrality? Calculate the betweenness score of this node and show your calculations.



**Answer:** Node 8. For calculations, refer to Zoom recording for session on betweenness centrality. Unscaled betweenness centrality of node 8 is 14.

2. A survey was conducted within cohort of 65 students, which asked each student to indicate which 3 students in the class s/he corresponds with the most on Facebook. The Professor then used this data to draw the social graph and calculated the betweenness centrality of each student. The average betweenness centrality came out to be 200.

If the Professor had looked up each student on Facebook and created a social network between these 65 students using the actual friendship or contact list, would the average betweenness centrality be above 200? Justify your response.

**Answer:** The average betweenness from the real network would be **less** than 200. The real network will be more dense than the one created from the survey. Thus there will be more paths between pairs of nodes, which will lower the average betweenness score.

1. The following tables show non-directional friendships and attributes of people (perceptions about two brands on a 1-7 scale) in a social network across time.

**Snapshot on January 1, 2020**

**Friendships (“Yes” indicates the presence of friendship)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| A | - | Yes |  | Yes |  |
| B |  | - | Yes |  |  |
| C |  |  | - |  |  |
| D |  |  |  | - | Yes |
| E |  |  |  |  | - |

**Attributes**

|  |  |  |
| --- | --- | --- |
|  | Perception of brand X | Perception of brand Y |
| A | 3 | 6 |
| B | 7 | 3 |
| C | 5 | 5 |
| D | 7 | 4 |
| E | 5 | 7 |

**Snapshot on March 1, 2020**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| A | - | Yes |  | Yes | Yes |
| B |  | - | Yes | Yes |  |
| C |  |  | - |  | Yes |
| D |  |  |  | - | Yes |
| E |  |  |  |  | - |

|  |  |  |
| --- | --- | --- |
|  | Perception of brand X | Perception of brand Y |
| A | 6 | 4 |
| B | 7 | 3 |
| C | 5 | 4 |
| D | 7 | 4 |
| E | 5 | 6 |

3a. Design a simple test of whether social influence exists in this network and perform (and show) all calculations (can be shown in Excel). Be specific in your answer and state the test precisely.

Test for social influence (description):

Show all calculations on a spreadsheet:

Your conclusion from the analysis:

**Answer**

**Test of social influence:**

***In March, is the average Euclidean distance between people who were friends in January less than the average distance between them in January? Mathematically:***

|  |  |  |
| --- | --- | --- |
|  | Average distance in January | Average distance in March |
| Friends in January | A | B |
| Strangers in January | C | D |

*Test of social influence: Is* ***(A-B)/A > (C-D)/C***

*Note: Even in March, the average distances are calculated only for those who were friends or strangers in January. Detailed calculations are shown below. The results show that from January to March, the average distance between those who were friends in January reduced 53%, while that between strangers (in January) decreased 16.4%. Thus there is strong evidence of social influence here.*

*If we use median values (somewhat problematic, because it is theoretically possible for the median to remain unchanged while other values may change significantly), the average distance between friends declined by 55%, while that between friends declined only 10.55%. Thus there is a strong evidence of social influence.*

***Note: It is not enough to compare raw numbers since an absolute decrease of .4 in one case may actually be much bigger change than a decrease of .6 in another case. So % decrease (or change) is an absolute must.***

***The distance in March should be calculated for those who were friends or strangers in January. The comparison should NOT be between friends (or strangers) in Jan and friends (or strangers) in March.***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Jan-14 |  |  |  |  |  |
|  | A | B | C | D | E |
| A | 0 | 5 | 2.236068 | 4.472136 | 2.236068 |
| B |  | 0 | 2.828427 | 1 | 4.472136 |
| C |  |  | 0 | 2.236068 | 2 |
| D |  |  |  | 0 | 3.605551 |
| E |  |  |  |  | 0 |
| Average distance between friends in January | | | | | 3.976529 |
| Average distance between strangers in January | | | | | 2.36339 |
| Median distance between friends in Jan | | | |  | 4.038844 |
| Median distance between strangers in Jan | | | | | 2.236 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mar-14 |  |  |  |  |  |  |  |  |
|  | A | B | C | D | E |  |  |  |
| A | 0 | 1.414214 | 1 | 1 | 2.236068 |  |  |  |
| B |  | 0 | 2.236068 | 1 | 3.605551 |  |  |  |
| C |  |  | 0 | 2 | 2 |  |  |  |
| D |  |  |  | 0 | 2.828427 |  |  |  |
| E |  |  |  |  | 0 |  |  |  |
| Avg. distance in March between those who were friends in January | | | | | | |  | 1.869677 |
| Average distance in March between those who were strangers in January | | | | | | | | 1.973603 |
| % decline in distance from January to March between friends in Jan | | | | | | |  | 0.529822 |
| % decline in distance from January to March between strangers in Jan | | | | | | |  | 0.164927 |
|  |  |  |  |  |  |  |  |  |
| Median distance in March between those who were friends in January | | | | | | |  | 1.825141 |
| Median distance in March between those who were strangers in January | | | | | | | | 2 |
| % decline in median distance between those who were friends in January | | | | | | | | 0.548103 |
| % decline in median distance between those who were friends in January | | | | | | | | 0.105546 |

Thus there is strong evidence of social influence in this data.

Note: It is not only unnecessary to convert from numbers to binary by assuming a threshold, but also prone to incorrect assessment.

3b. Design a test and perform the analysis to check the presence of homophily in this specific dataset.

Test of homophily (describe what you will do):

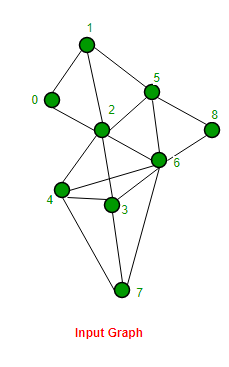
Show all calculations on a spreadsheet:

Your conclusion from the analysis:

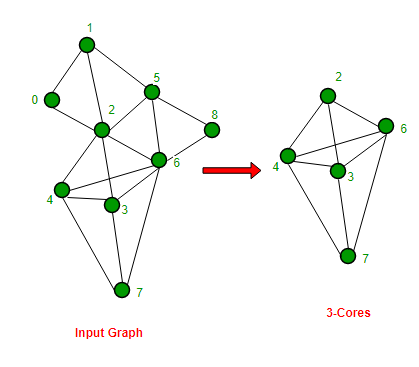
**Partial answer**

**Test of homophily: *Suppose there are X strangers in January. Say, Y of these X (Y less than X) become friends by March. Is the average distance (in January) between the Y folks < the average distance (in January) between the (X-Y) folks?* That is, were folks, who were strangers but “more similar” in January, more likely to become friends than those who were “less similar”?**

1. Find the largest 3-core sub-network from the network below:



**Answer**



1. Consider k-cores and n-cliques. “If we choose k = n-1 (and n > 2), we are likely to find more cores than cliques from a given network.” Do you agree with this statement? Explain your position.

**Answer:** True. An n-clique is a subnetwork of n nodes which are fully connected, while a core of comparable size does not require all nodes to be connected to each other. Naturally it is easier to find the latter.