Visual Clustering Analysis of Social Network (40%)

Social networks are ubiquitous. A fundamental problem related to these networks is the discovery of clusters or communities. Intuitively, a cluster is a collection of individuals with dense friendship patterns internally and sparse friendships externally.

The discovery of close-knit clusters in these networks is of fundamental and practical interest and the one of major focused problems in AI and Data Mining. There are many reasons to seek tightly-knit communities in networks, for instance, target marketing schemes can be designed based on clusters, and it has been claimed that terrorist cells can be identified.

This assignment gives students two options with different requirements based on student's disciplinary background, personal preference and existing experience. This is to satisfy the students who are not major in IT and Mathematics (Algorithms).

Option One: (group work):

A group of two students are required to work together to analyze an organization's email network (a type of social networks) through the data clustering and a clustered graph visualization.

Task 1: through data clustering, we can identify abnormal (implicit) network patterns that against the hierarchical structure of the organization,

Task 2: through a clustered graph visualization, we can visually read and quickly understand the data clustering output, including the abnormal network patterns.

Option Two: (individual work):

An individual student is required to visualize an organization's email network (a type of social networks) through graph visualization and clustered graph visualization.

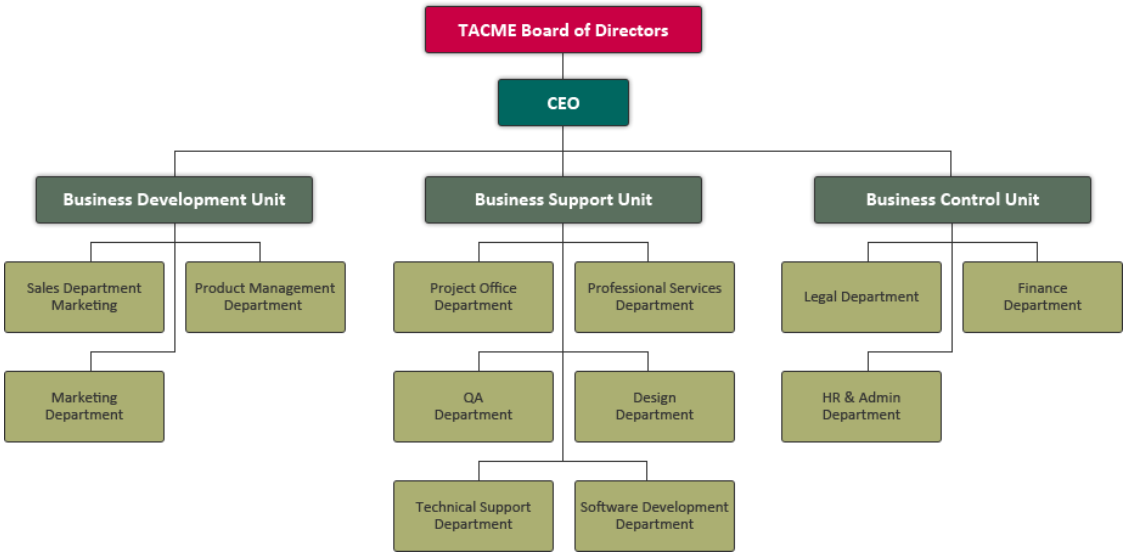
Task 1: using graph visualization to visualize attributed email network,

Task 2: using clustered graph visualization to visualize a given clustered email network that enables readers to quickly understand the data clustering output.

The weight of this assignment is 40%.

Specification:

The following Figure shows the organization structure of TACME, sourced from http://www.tacme.com/corporate_structure.html



A list of staff's ID, Name and Position in TACME

ID	Name	Position
0	James	Director
1	David	Director
2	George	CEO
3	Ronald	Business Development Manager
4	John	Business Support Manager
5	Richard	Business Control Manager
6	Daniel	Sales Department Leader
7	Kenneth	Product Department Leader
8	Anthony	Marketing Department Leader
9	Robert	Project Office Leader
10	Charles	Professional Service Leader
11	Paul	QA Leader
12	Mark	Design Office Leader
13	Kevin	Technical Support Office Leader
14	Edward	Software Development Leader
15	Joseph	Legal Office Leader
16	Michael	Finance Office Leader
17	Jason	HR Office Leader

The **Email communication detail** in a particular month is shown below:

ID	Emails per month	Weight	ID
0	5	1	1
0	6	1	2
1	5	1	2
2	25	2	3
2	36	2	4
2	53	3	5
3	150	4	6
3	213	5	7
3	298	5	8
4	345	6	9
4	123	4	10
4	212	5	11
4	453	7	12
4	156	4	13
4	278	5	14
5	300	5	15
5	78	3	16
5	256	5	17
6	78	3	7
6	145	4	8
7	139	4	8
9	34	2	10
9	134	4	11
9	546	7	12
9	23	2	13
9	145	4	14
10	256	5	11
10	222	5	12
10	190	4	13
10	56	3	14
11	78	3	12
11	112	4	13
12	98	3	14
15	88	3	16
15	128	4	17
16	238	5	17
17	5	1	7
16	15	2	6
16	23	2	7

16	**54**	**3**	8
16	**18**	**2**	9
16	**23**	**2**	11
16	**41**	**2**	13
16	**13**	**2**	14
16	**27**	**2**	10

Weight description:

Quantity	Weight
<10	1
11 – 50	2
51 – 100	3
101 – 200	4
201 – 300	5
301 – 400	6
> 401	7

General Requirement:

Option One: (group work)

Students are required:

1) To draw (visualize) the original email network on the paper (or screen) with the satisfaction of the following Aesthetics Rules: *a) Symmetrical Display, b) Minimization of Edge-Crossings and c) Maximization of Angular Resolution*. In addition, since each edge *e* in the graph is associated with a weight *w(e)*, you need to map the *w(e)* to a graphical attribute, such as color, types of line, size or shapes, to enhance the readbility of the weight

2) To cluster this email network (or graph) into clustered structures by using **Markov Clustering Algorithm**. You need to produce two clustered structures 1) with the weight *w(e)*, 2) without the weight *w(e)* .

3) Discuss the findings. If there is one (or more) abnormal network pattern(s) found, you need to describe them in details.

4) To draw (visualize) these two clustered graphs (one with *w(e)*, another without *w(e)*) on the paper (or screen). Using geometric rectangles (or circles) to bound clusters in the drawing. Make sure that these regions are not overlapped. In addition, these drawings shall also satisfy the general graph drawing aesthetics.

Option Two: (individual work)

Student is required:

1) To draw (visualize) the original email network on the paper (or screen) with the satisfaction of the following Aesthetics Rules: a) Symmetrical Display, b) Minimization of Edge-Crossings and c) Maximization of Angular Resolution. In addition, since each edge e in the graph is associated with a weight w(e), you need to map the w(e) to a graphical attribute, such as color, types of line, size or shapes, to enhance the readability of the weight

2) To draw (visualize) a given clustering of the above email graph, that is: {0, 1, 2}, {3, 4, 5}, {6, 7, 8}, {9, 10, 11}, {12, 13, 14}, {15, 16, 17} on the paper (or screen). Using geometric rectangles (or circles) to bound clusters in the drawing. Make sure that these regions are not overlapped. In addition, these drawings shall also satisfy the general graph drawing aesthetics, including a) Symmetrical Display, b) Minimization of Edge-Crossings and c) Maximization of Angular Resolution.