Algorithm for Computing Desirable Cluster

El-Zohny H

El- Morsy H

Mathematics Department
Faculty of science
Alazhar University
Cairo, Egypt.

Abstract

In this paper we will introduce algorithm which compute good cluster in graph by using the density function. We will compute this algorithm for weighted and unweighted graphs. Some examples will be illustrated.

Definitions

Definition of Graph: A graph G (V, E) is a diagram consisting of a finite non – empty set of elements, called "vertices" denoted by V (G) together with a setof unordered pairs of these elements, called "edges" denoted by E(G).The set of vertices of the graph G is called "the vertex-set of G" and the list of the edges is called "the edge-list of G" [1].

Definition of Cluster Graph: Is a group of nodes in a graph that are more connected to each other than they are to the rest of the graph [1].

Definition of Weighted Graph: Is a graph for which each edge has an associated real number Weight [2].

Definition of Algorithm: In mathematics and computer science, an algorithm is an effective method expressed as a finite list of well-defined instructions for calculating a function. In simple words an algorithm is a step-by-step procedure for calculations.

Density of Graph G (V,E) : Is the ratio of the number of edges present to the maximum possible,

$\boldsymbol{d}(\mathbf{G}) = \frac{\boldsymbol{m}}{\binom{\boldsymbol{n}}{\mathbf{2}}}$**, for** $\boldsymbol{n} \in \{0,1\}$, eq. (1) we set d(G) = 0 . A graph of density one is called *complete* [3].

Intra – Cluster Density: We refer to the density of the subgraph induced by the cluster as the *internal* or *intra-cluster density*:

$d_{\text{int}}(\text{c}) = \frac{|\{\{u,v\} \mid v\epsilon c, u\epsilon c\}|}{|c|(|c|-1)}$ eq. (2)

The intercluster density of a given clustering of a graph *G* into *k* clusters*:*C_1, C_2 , C_3 , ………$C_k$is the *average* of the intercluster densities of the included clusters [3]:

$d_{\text{int}}(\text{G} \mid C_1 \ldots\ldots C_k) = \frac{1}{k}\Sigma^{k}_{i=1} d_{\text{int}}\, C_i$ .eq. (3)

Main Results

Definition of Desirable Cluster: Is a cluster which its internal density is greater than the density of graph.

Consider a graph G (N,M) which contains C_1 , C_2 , ……C_k clusters , if the density of graph $d(\text{G}) = \frac{m}{\binom{n}{2}}$, for $n \in \{0,1\}$, where n is the number of vertices, and m is the number of edges, and if the density of each cluster is given by eq. (2) andfor k cluster by eq. (3) , then the density of good cluster is illustrated by the following algorithm.

There are two algorithms computing desirable cluster the first for unweighted graph and the other algorithm for weighted graph.

1. **For un Weighted Graph:**

Input: Un weighted graph G(N,M) which divided into clusters: $C_1, C_2 \ldots\ldots C_i$.

Algorithm:

1. Compute the density of graph G by using eq.(1).
2. Compute the internal density of C_1 by using eq.(2).
3. Find the ratio $\hat{R}= d(G)/d_{int}(C_1)$.
4. If $\hat{R}< 1$, then:
   Output C_1.
5. If $\hat{R} \geq 1$,
   Return to step 2 for C_2
   Output C
   End.

Example 1: For cluster graph shown in Fig.(1) compute desirable cluster

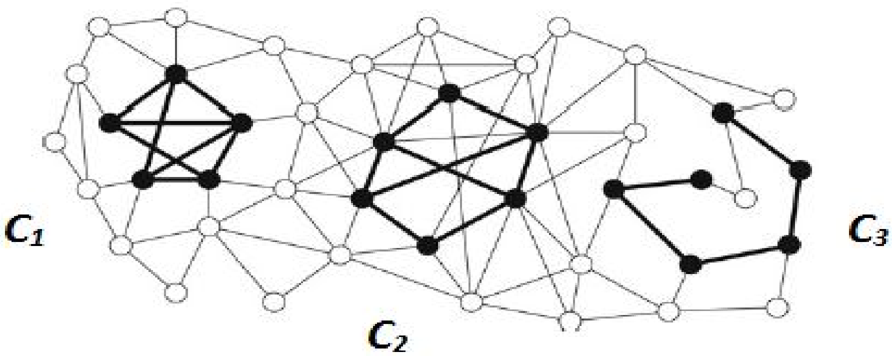

Fig.(1)

First we will compute the density of graph *d(G) by using eq.(1):*

$d(G) = \frac{7+8+5+14}{22 .21} = 0.073$

Then compute the density of C_1 , C_2 , C_3 by using eq.(2):

$d(C_1) = \frac{8}{5 .4} = 0.4$, then the ratio $\hat{R}_1 = \frac{0.075}{0.4} = 0.1825$

$d(C_2) = \frac{8}{6 .5} = 0.3$, ratio $\hat{R}_2 = \frac{0.075}{0.3} = 0.1666$

$d(C_3) = \frac{5}{6 .5} =0.2$, ratio $\hat{R}_3 = \frac{0.075}{0.2} = 0.438$

Since $\hat{R}_1< \hat{R}_2< \hat{R}_3$ then C_1 is the best cluster.

Example 2: For cluster graph shown in Fig.(2) we have:

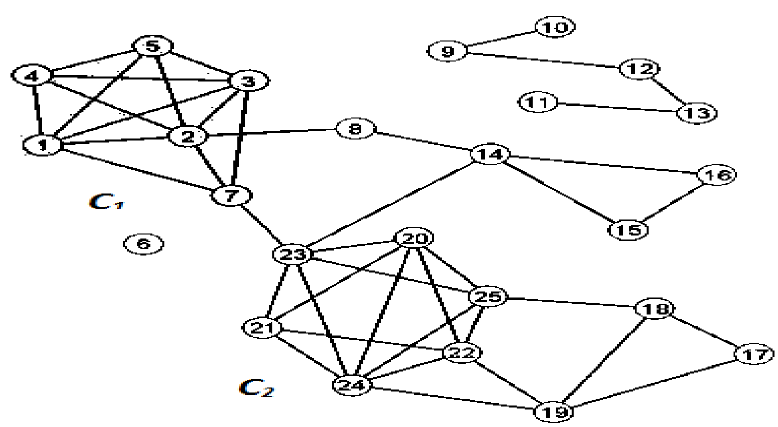

Fig.(2)

The density of graph $d(G) = \frac{42}{24\,.23} = 0.0761$,
$d(C_1) = \frac{12}{6.5} = 0.41$, then the ratio $\hat{R}_1 = \frac{0.076}{0.41} = 0.19$
$d(C_2) = \frac{13}{6\,.5} = 0.433$, ratio $\hat{R}_2 = \frac{0.075}{0.433} = 0.76$
Since $\hat{R}_2 < \hat{R}_1$, Then C_2 is good cluster.

Example 3: A company is divided into three departments, each of which is formed by two or three 5–7 person teams. Each person is represented by a vertex, and an edge is placed between two people if they interact on work-related matters on a daily basis. The teams and the departments are encompassed by dotted lines as shown in Fig.(3):

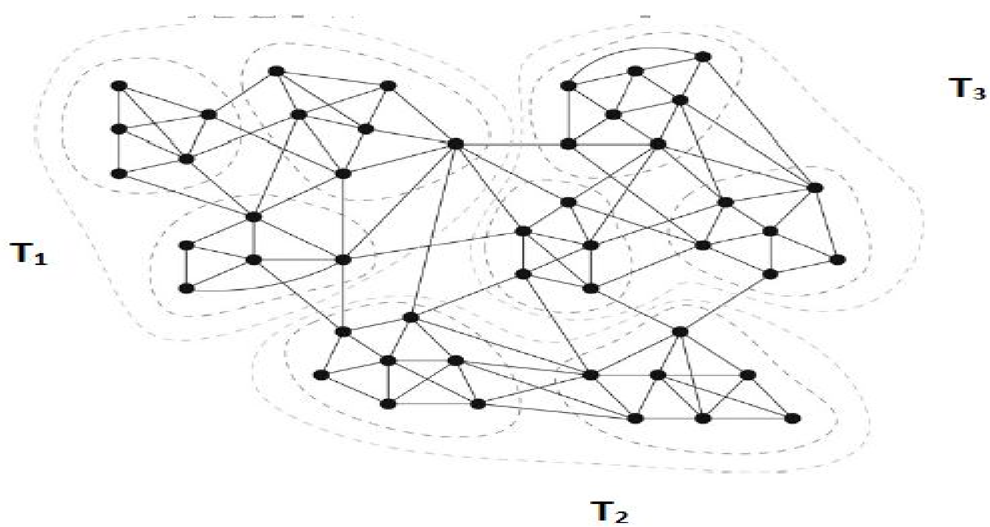

Fig.(3)

The density of graph $d(G) = \frac{115}{48\,.47} = 0.0509$,
$d(T_1) = \frac{38}{16.15} = 0.1583$, then the ratio $\hat{R}_1 = \frac{0.0509}{0.1583} = 0.3219$,
$d(T_2) = \frac{36}{14\,.13} = 0.1978$, ratio $\hat{R}_2 = \frac{0.0509}{0.1978} = 0.2577$.
$d(T_3) = \frac{39}{18\,.17} = 0.1274$, ratio $\hat{R}_3 = \frac{0.0509}{0.1274} = 0.3999$.
Since $\hat{R}_2 < \hat{R}_1 < \hat{R}_3$, then T_2 is the best team of work.

2. For Weighted Graph:

Input: Weighted graph G(N,M) which divided into clusters: $C_1, C_2 \ldots\ldots C_i$ each cluster with m edges and n vertices , and weight w.

Algorithm:

6. Compute the density of graph G by using equation:
   $d(\mathrm{G}) = \frac{1}{\binom{n}{2}} \sum_{\substack{\{u,v\} \epsilon m \\ u,v\, \epsilon\, n}} W\,(u, v)$.
7. Compute the internal density of C_1 by using equation:
   $d_{\mathrm{int}}(\mathrm{c}) = \frac{\sum_{\substack{\{u,v\} \epsilon m \\ u,v\, \epsilon\, n}} W\,(u,v)}{|c|(\,|c|-1\,)}$.
8. Find the ratio $\hat{R} = d(G)/d_{int}(C_1)$.
9. If $\hat{R} < 1$, then:
   Output C_1.
10. If $\hat{R} \geq 1$,
    Return to step 2 for C_2
    Output C
    End.

Example 4: For weighted cluster graph shown in Fig.(2), each edge has certain weight, we can compute the best cluster as follows:

w{1,2}=3 , w{2,3}=5 , w{3,4}=2 , w{4,5}=7 , w{5,3}=1, w{5,1}=2 , w{1,3}=2 , w{1,4}=6 ,w{1,7}= 3 , w{7,2}=2 , w{7,3}=1 , w{2,5}=7 , w{7,23}=7 , w{23,14}=10 , w{14,8}=2 , w{8,2}=4, w{14,15}=2 , w{15,16}=7, w{16,14}=3 ,w{25,18}, w{22,19}=2 ,w{19,24}=3, w{18,19}=5, w{19,17}=3 , w{17,18}=2,w{9,10}=5 ,w{9,12}=3, w{12,13}=9 ,w{11,13}=1,w{23,20}=2,w{21,23}=5 ,w{21,20}=2, w{20,25}=3,w{25,22}=7 ,w{22,20}=6, w{22,24}=1, w{21,24}=1, w{24,25}=3 ,w{20,24}=8 ,w{23,24}=7 ,w{22,12}=5.

$d(G) = \frac{69+50+45}{24\,.23}$= 0.297.

$d(C_1) = \frac{50}{6.5}$ = 1.66 , then the ratio $\hat{R}_1 = \frac{0.297}{1.66}$ = 0.178.

$d(C_2) = \frac{45}{6\,.5}$ = 1.5 , ratio $\hat{R}_2 = \frac{0.297}{1.5}$ = 0.98.

since $\hat{R}_1 < \hat{R}_2$, then C_1 is the desirable cluster.

Theorem 1: For graph G(V,E) if its edges and vertices are classified by its importance the best cluster is the cluster which have the important edges and vertices.

Example 4: This is a simple example for previous theorem.

Consider electric circuit consists of two clusters A,B where each edge describe electric wire with certain current density I ampere , show Fig.(4).

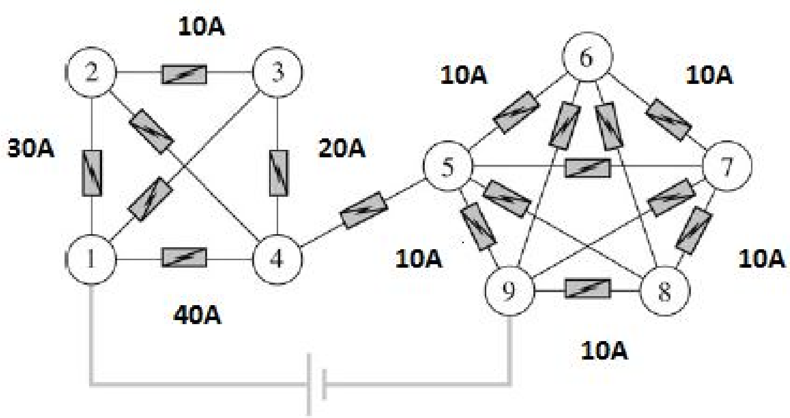

Fig.(4)

The wire {1,3} has current density of 50 ampere , and {2,4} has current 70 A, {6,9} = 20 A , {6,8}=20A , {5,7}=10 A , {7,9}=10 A , {5,8}=15A.

Then the density of the circuit $d(G) = \frac{345}{9\,.8}$= 4.79.

$d(A) = \frac{10+20+40+30+50+70}{4\,.3}$ = 18.33 , then the ratio $\hat{R}_1 = \frac{4.79}{18.33}$ = 0.268.

$d(B) = \frac{125}{5.4}$ = 6.25 , ratio $\hat{R}_2 = \frac{4.79}{6.25}$ = 0.788.

$\hat{R}_1 < \hat{R}_2$ Then cluster A is the best although it has few edges and vertices but with high electric current density.

References

Wilson R.J.:Introduction to graph theory. Oliver and Boyed, Edinburgh 1972.
Susanna S.Epp, Discrete Mathematics With Application, Third Edition,Thomson Learning,Inc,2004.
Satu Elisa ,: Graph clustering , Computer Science Review, (2007) ,27-64.
http://webwhompers.com/network-dynamics/6-network-dynamics/109-limitations-of-traditional-graph-theory.html.