**3. Preliminaries**

Let us consider a social network G(V, E), where each user of the network is represented by a node vi ∈ V, and relationships between users are represented by edges (vi, vj) ∈ E. Community structure in social network generally means the natural division of the nodes into subgroups C= {c1, c2,...., ck} such that users within the same subgroup cl are densely connected, whereas connections among subgroups are relatively sparse. Researchers primarily focused into disjoint community structure, where each node belongs to a single community only. However, the social networks often contain community structure with overlaps where a node may hold multiple community memberships.

**Definition 1:** Let C= {c1, c2,...., ck} be a community structure for a network G(V, E). C is said to be *Disjoint Community* if for any two communities ci and cj in C, ci ∩ cj = ∅ , where i ≠ j, otherwise C is said to be the set of *Overlapping Communities*.

**Example 1: <give example of disjoint community>**

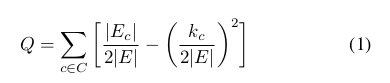
**Definition 2:** Given a disjoint community structure for a network G(V, E), the nodes with inter-community edges are termed as Community Boundary Nodes (CBNs).

**Example 2: <give example of CBN>**

In this paper, we leverage the disjoint community structure of SN for identifying the outlayer in the network.

**3.1. Modularity Measure**

To quantify the quality of disjoint community structure the most widely accepted modularity measure Q [cite] is used which can be mathematically written as:



\*Note: <write the equation>

Where |Ec| and kc represent the total number of edges within a community c and the sum of degree of all nodes within community c, i.e. <write equation>, where du  is the degree of node u.

According to the definition 0 ≤ Q ≤ 1, and the value closer to unity signifies better quality of community structure, i.e., intra-community connections are dense compared to inter-community connections.

**3.2 Problem Definition:**

For a given a social network G(V,E), with known community structures, the goal is to identify the set of outlier nodes Vo ∈ V, that have the tendency to belong in more than one community or whose neighbors either belong to only one community or do not belong to any community.

**3.3. Outlier Score:**

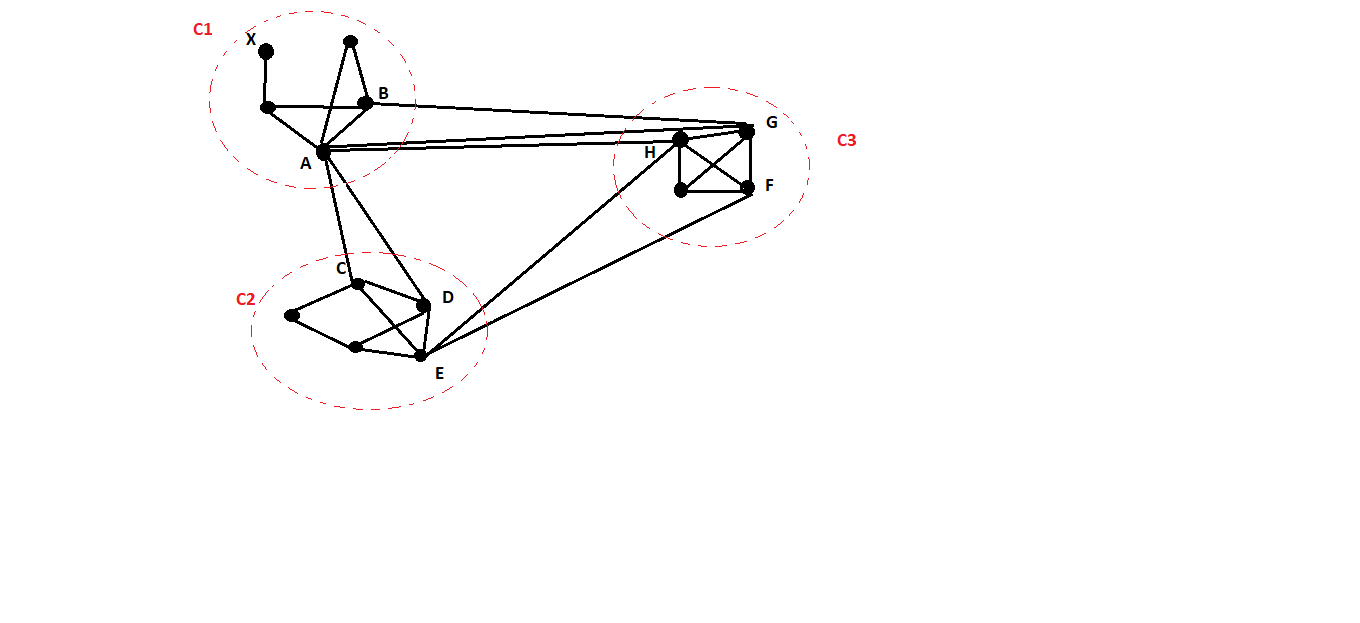
In this work an outlier score S is used based on a node-based metric, called *relative permanence* [cite], that provides a quantitative measure of how much the node belongs to its community and mathematically defined as :

Perm(v)= ( ( 1/e1 + 1/e2 +... + 1/ek ) / D(v) ) \* ( IN(v) / EN(v) )

where v represents a node, each ei , 1 ≤ i ≤ k, represents the number of neighbours of v belonging to the *i*th external group, D(v) represents the degree of node v, EN(v) represents the external neighbor of v and IN(v) represents the internal neighbors of v.

In general, 0 ≤ S ≤ 1, and a lower value of S siginifies that the node has tendency to belong to more than one community, i.e., it is loosely connected within its own community. Where as S=1 signifies that a node is strongly associated its own community. But there may be some nodes with outlier score S=1 whose neighbors belong to only one community, for e.g., the degree-1 nodes, such nodes marginally connected to communities featuring a weak association and may be marked as outliers. Similarly, the nodes having lower S value less than a threshold T may also be marked as outliers.

Let us consider a simple social network graph consisting of 14 nodes and 26 connections, with known disjoint community structure C={c1,c2,c3} as shown in Fig. <> .



In Fig. <> the outlier score of node A is computed as

Perm(A)= ((½+½)/D(A))\*(IN(A)/EN(A))

=((½+½)/7)\*(3/4)=0.107. [ D(A)=7,IN(A)=3,EN(A)=4,ENG(A)={2,2}(i.e.,two external

neighbours in C2 and two external neighbours in C3) ]

Similarly we calculate the values for other boundary nodes as well.

B=C=D=F=0.75

E=G=0.15

H=0.6