Understanding Vulnerability of Communities in Complex Networks

Viraj Parimi, Dr. Tanmoy Chakraborty, Dr. Ponnurangam Kumaraguru

Problem

Given a graph $G(V, E)$, each node i is associated with a cost C_i, and there is a total budget K. For a vertex set $S \subseteq V$, let $G[S]$ be the graph induced by S,

$$\textsc{Cost}(S) = \sum_{i \in S} C_i,$$

$$f(S) = \textsc{Metric}(G, G[\bar{S}])$$

Here, $\bar{S} = V \setminus S$ and Metric can be any measure that quantifies the network structure of the graph. Note that, $f(S)$, a value function is a real-valued *set function*. We need to identify a set of nodes $S \subseteq V$ which,

$$\textbf{maximize} \quad f(S)$$
$$\textbf{subject to:} \quad \textsc{Cost}(S) \leq K$$

Metrics

Modularity - measures the strength of division of a network into communities.

$$\frac{1}{2m} \sum_{i,j} (A_{ij} - \frac{k_i * k_j}{2m}) \delta(c_i, c_j)$$

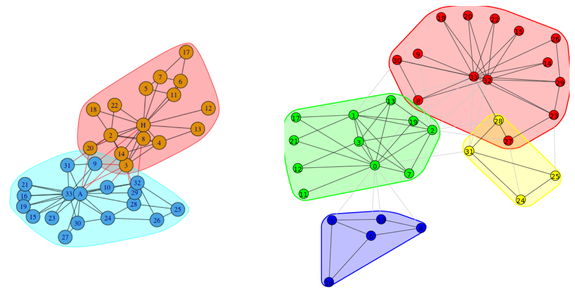
where A_{ij} : Adjacency matrix, k_i : Degree(i), m : $|E|$, c_i : Community label of node i, $\delta(c_i, c_j)$: Function that returns 1 if $c_i = c_j$ and 0 otherwise.

Normalized Mutual Information(NMI) - measures the similarity between two community structures X and Y.

$$\frac{2 \sum_{i=1}^{c_X} \sum_{j=1}^{c_Y} (n_{ij} log(\frac{n_{ij} n}{x_i * y_j}))}{(n-k)H(X) + \bar{y} log(n-k) - \sum_{j=1}^{c_Y} (y_j log(y_j))}$$

where c_X : # of communities in network X, n_{ij} : $|X_i \cap Y_j|$, n : $|V|$, x_i : $|X_i|$, $\bar{y}$: $|Y|$, k : # of nodes removed.

Approach



Original Louvain

- Even the unit cost version of this problem is **NP-Hard**. Can be reduced from **Maximum Vertex Coverage for Bipartite Graphs**.
- Used a small dataset - **UCI Karate Club** network with 2 communities and 34 nodes. Tried heuristics for K = 4.
- Since problem is not restricted by any community detection algorithm, hence applied multiple of them like,
  - Louvain
  - Multilevel
  - Fast Greedy
  - Infomap etc..
- Ground truth data created by generating ${}^nP_r$ combinations of candidate sets and evaluated them individually thereby sorting them based on their effectiveness in changing the community structure.

Method

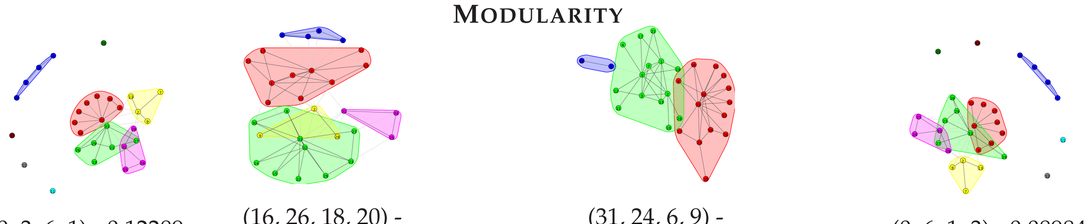
- **Network based Greedy approach**
  Remove nodes that contribute heavily to the network property based on the greedy property such as,
  - Clustering Coefficient
  - Various centrality measures like,
    - Degree Centrality
    - Betweenness Centrality etc..
  - Eccentricity etc..

  Idea is to try an reduce the value function itself by attacking nodes that greatly maximize it.
- **Community based Greedy approach**
  Similar to above method but first identify a community based on community centric greedy measures such as,
  - Link Density
  - Conductance
  - Compactness

  Since we want to change the community structure itself, hence the idea is to target the community first, identifying potential weak structures. This is followed with identifying weak nodes within the identified community similar to previous approach.
- **Genetic Algorithm**
  Generate an initial population of budget size from ${}^nP_r$ combinations and keep improving them over generations based on the fit function similar to the value function.

Results

Modularity



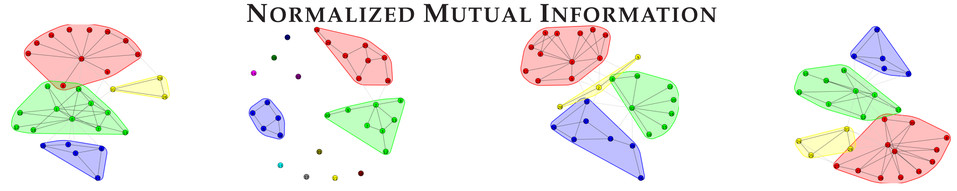
(0, 3, 6, 1) - 0.12289
Exhaustive

(16, 26, 18, 20) - 0.04439
NBG - Eccentricity

(31, 24, 6, 9) - 0.05090
CBG - Link Density, Eigenvector Centrality

(0, 6, 1, 3) - 0.08984
Genetic

Normalized Mutual Information



(32, 33, 5, 6) - 0.38580
Exhaustive

(33, 32, 0, 1) - 0.62733
NBG - Intra Degree

(33, 13, 7, 14) - 0.51916
CBG - Conductance, Coreness

(7, 2, 30, 21) - 0.74314
Genetic

References

[1] My T. Thai Structural Vulnerability Assessment of Community-Based Routing in Opportunistic Networks

[2] Samir Khullar Budgeted Maximum Coverage Problem

[3] M. Girvan Community Structure in Social and Biological Networks

Ack: Sushmita Ruj - ISI Calcutta and Arindam Pal - TCS Research

Findings

Value Function	Modularity				Normalize Mutual Information			
Greedy Metrics	Exhaustive	NBG - Eccentricity	CBG - Link Density, Eigen Centrality	Genetic	Exhaustive	NBG - Intra Degree	CBG - Conductance, Coreness	Genetic
Nodes Selected	0, 3, 6, 1	16, 26, 18, 20	31, 24, 6, 9	0, 6, 1, 3	32, 33, 5, 6	33, 32, 0, 1	33, 13, 7, 14	7, 2, 30, 21
Scores	0.12289	0.04439	0.05090	0.08984	0.38580	0.62733	0.51916	0.74314

In the case of Modularity, the best algorithm that came close to the ground truth values was Genetic Algorithm. But in the case of NMI, the best algorithm was CBG using Conductance and Coreness as its greedy metrics.

A Future Direction

- Apply different formulation involving Leaders, Members and Orbiters.
- Once heuristics is well defined, apply the same on general large networks.
- Check the optimality of the solution.
- Theoretically analyse the problem more.