Community Detection using a Measure of Global Capacity to Influence

Rumi Ghosh
Kristina Lerman
Influence Based Definition of Community

- Claim: “Community comprises of individuals who have more capacity to influence individuals within the community than outsiders.”

The capacity to have an effect on the character, development, or behavior of someone or something or the effect itself

Influence in a large part is the ability to reach a crucial man through the right channels, the more the channels in reserve, the better

Oxford English Dictionary

Pool & Kochen (1978)
Total Capacity to Influence

Influence Matrix

\[ P = \beta A + \beta \alpha A_1 + \ldots + \beta \alpha^n A_n + \ldots = \beta A (I - \alpha A)^{-1} \]

- **direct attenuation factor**
- **adjacency matrix**
- **indirect attenuation factor**

\[ AxAxA \ldots A \text{ n+1 times} \]

**How Many**

The total influence of \( b \) on \( c \) thus depends on the number of (attenuated) channels from \( b \) to \( c \) or the sum of all weighted paths from \( b \) to \( c \).

**Who**

Do the two (1 hop) paths from node \( a \) to node \( i \) have equal capacity to influence node \( i \)?

**Power of transmission**

Probability of transmission from the immediate neighbors like \( e \) to \( c \) or \( g \) to \( c \) is \( \beta \).
Probability of transmission over 1-hop paths such as \( b \) to \( c \) via \( e \) is \( \beta \alpha \).
Probability of transmission along n-hop path is \( \beta \alpha^{n-1} \).
Influence Based Modularity

- Redefine Modularity
  \[ Q = (\text{connectivity within the community}) - (\text{expected connectivity within the community}) \]

- Greater connectivity implies greater capacity to influence.

- Implies in best division of the network, influence of nodes within their community is more than their influence outside their community.

We have to maximize \( Q \) to obtain the communities the network is divided into.

\[
Q = \sum_{i,j} [P_{ij} - \bar{P}_{ij}] \delta(s_i, s_j)
\]

\(\delta(s_i, s_j) = 1 \text{ if } s_i = s_j, \quad \text{otherwise } \delta(s_i, s_j) = 0\)
Null Model

Given by the capacity to influence in an equivalent random graph which has the same number of nodes $N$ as the original network.

Expressed in terms of a $N \times N$ matrix $\overline{P}$ (where $N$ is the total number of nodes in the network).

When all vertices are placed in a single group axiomatically $Q = 0$.

**Expected influence on a node $j$ ($W_{j}^{in}$)** = Actual influence on node $j$ in the original network.

**Expected capacity of a node $i$ to influence other ($W_{i}^{out}$)** = Actual capacity of node $i$ to influence of the node $i$ in the original network.

$$\sum_{i,j} [P_{ij} - \overline{P}_{ij}] = 0$$

$$W_{j}^{in} = \sum_{i} P_{ij} = \sum_{i} \overline{P}_{ij}$$

$$W_{i}^{out} = \sum_{j} P_{ij} = \sum_{j} \overline{P}_{ij}$$

$$W = \sum_{i,j} P_{ij} = \sum_{i,j} \overline{P}_{ij}$$

Total expected capacity to influence ($W$) = Total actual capacity to influence.
Expected Capacity to Influence

Expected capacity of node $i$ to influence node $j$

$$\bar{P}_{ij} = f_1(W_{i\text{out}})f_2(W_{j\text{in}})$$

where $f_1$ and $f_2$ are some functions.

$$W_{i\text{out}} = \sum_j f_1(W_{i\text{out}})f_2(W_{j\text{in}})$$

$$= f_1(W_{i\text{out}}) \sum_j f_2(W_{j\text{in}})$$

$$W_{j\text{in}} = \sum_i f_1(W_{i\text{out}})f_2(W_{j\text{in}})$$

$$= f_2(W_{j\text{in}}) \sum_i f_1(W_{i\text{out}})$$

$$f_1(W_{i\text{out}}) = C_1 W_{i\text{out}}$$

$$f_2(W_{j\text{in}}) = C_2 W_{j\text{in}}$$

For some constants $C_1$ and $C_2$.
Detecting Community Structure

- Once modularity $Q$ is derived, an algorithm is to be selected that divides the network into communities in a manner that optimizes $Q$.

- Decision version of modularity maximization is NP complete (Brandes et al., 2008).

- We use leading eigenvector method to obtain approximate solution.

\[
W = \sum_{i,j} P_{ij} = \sum_{i,j} C_1 C_2 W_{i\text{out}} W_{j\text{in}} = C_1 C_2 W^2
\]

\[
P_{ij} = \frac{W_{i\text{out}} W_{j\text{in}}}{W}
\]

\[
Q = \sum [P_{ij} - (W_{i\text{out}} W_{j\text{in}}/W)] \delta(s_i, s_j)
\]
Zachary’s Karate Club

Two factions the club got divided into are represented by circles and squares resp. Nodes predicted to belong to the same community are shown by the same color.

Communities found when our algorithm is run till termination

Communities found when Newman’s (2004) algorithm is run till termination

Single Iteration of Newman’s and our algorithm

\[ \alpha = \beta = \frac{1}{N} \]

N=34
College Football

Schedule of Division 1 games for 2000 season (Girvan et al., 2002)

nodes: teams
edges: regular season games between 2 teams they connect

inter-conferences and intra-conferences games not equally distributed

natural communities may be bigger than the conferences

number natural communities predicted changes from 8 at $\alpha=0$ to 4 at $\alpha=0.1$

Purity

Similarity of a node pair = 1 if the two nodes actually belong to the same community (observed), 0 otherwise.

Purity of a detected community = total pairwise similarity of nodes belonging to that community

Maximum total similarity obtained when all teams belonging to same community end up in same predicted communities

Purity of community detection = total pairwise similarity of all detected communities / maximum total similarity

Purity

Similarity of a node pair = 1 if the two nodes actually belong to the same community (observed), 0 otherwise.

Purity of a detected community = total pairwise similarity of nodes belonging to that community

Maximum total similarity obtained when all teams belonging to same community end up in same predicted communities

Purity of community detection = total pairwise similarity of all detected communities / maximum total similarity
Political Books

Political books data compiled by V.Krebbs
nodes: books about politics sold by online bookseller Amazon
edges: co-purchasing of books by the same buyers
49 marked as conservative, 43 as liberal and 13 as neutral

α = 0.08 leads to 2 groups with 6 of the neutral books in one group containing 52 books (46 conservative) and 7 in the other group containing 53 books (43 liberal)
Indicates that 6 of the neutral books maybe conservatively inclined and 7 of them be liberally inclined.

Reduces to community detection using edge based modularity (Newman, 2006) when α = 0