Subgroup Centrality Measures

Isabel Chen

based on: “Subgroup Centrality Measures”, Jocelyn R. Bell
(Dept of Mathematical Sciences, United State Military Academy, West Point)

Nov 20 2014
Recall:

A network is a pair \((V, E)\), with associated adjacency matrix \(A\).

Common centrality measures:

- degree
- betweenness, closeness → based on paths
- Katz, subgraph, eigenvector → based on walks
Radial Centrality Measures

\[ f : V \times V \to \mathbb{R} \]

\[ c(a) := \sum_{x \in V} f(a, x) \]

**Degree**

\[ f(a, x) = \begin{cases} 1, & \text{if } (a, x) \in E \\ 0, & \text{else} \end{cases} \]

**Closeness**

\[ f(a, x) = \text{distance between } a \text{ and } x \]

**Eigenvector:** Suppose \( Aq = \lambda q \) where \( \lambda = \rho(A) \),

\[ f(a, x) = \begin{cases} \frac{1}{\lambda} q_x, & \text{if } (a, x) \in E \\ 0, & \text{else} \end{cases} \]
Medial Centrality Measures

\[ f : V \times (V \times V) \rightarrow \mathbb{R} \]

\[ c(a) := \frac{1}{2} \sum_{(x,y) \in V \times V} f(a, \{x, y\}) \]

**Betweenness**

\[ f(a, \{x, y\}) = \frac{\text{\#shortest paths between } x \text{ and } y \text{ that contain node } a}{\text{\#shortest paths between } x \text{ and } y} \]

Note that \( x \neq a \) and \( y \neq a \).
Subgroup Centrality Measures

Let $S \subseteq V$. For any $a \in V$, define

$$c_S(a) = \sum_{x \in S} f(a, x) \quad \text{for } f \text{ radial}$$

or

$$c_S(a) = \frac{1}{2} \sum_{(x, y) \in S \times S} f(a, \{x, y\}) \quad \text{for } f \text{ medial}$$
Local v Global

**Local:**

\[
c_S(a) = \sum_{x \in S} f(a, x) \quad \text{or} \quad \frac{1}{2} \sum_{(x, y) \in S \times S} f(a, \{x, y\})
\]

**Global:**

\[
c_{S'}(a) = \sum_{x \in S'} f(a, x) \quad \text{or} \quad \frac{1}{2} \sum_{(x, y) \in S' \times S'} f(a, \{x, y\})
\]

Observe that

- local w.r.t. \( S \) = global w.r.t. \( S' \)
- global w.r.t. \( S \) = local w.r.t. \( S' \)
For $f$ radial,
\[
   c_S(a) + c_S'(a) = \sum_{x \in V} f(a, x) = c(a)
\]

For $f$ medial,
\[
   c_S(a) + c_S'(a) + Bc_S(a) = \frac{1}{2} \sum_{(x,y) \in V \times V} f(a, \{x, y\}) = c(a)
\]

where
\[
   Bc_S(a) = \frac{1}{2} \sum_{(x,y) \in S \times S'} f(a, \{x, y\})
\]

is the ‘boundary’ measure.
Degree Example

(a) Local degree = 2. Normalized: $\frac{2}{4}$

(b) Global degree = 3. Normalized: $\frac{3}{8}$
Normalization

To account for size of $S$,

- for $f$ radial, normalize by $|S|$ if $a \not\in S$, or $|S| - 1$ if $a \in S$.
- for $f$ medial, normalize by $(\frac{|S|}{2})$ if $a \not\in S$, or $(\frac{|S| - 1}{2})$ if $a \in S$. 
Subgraph Centrality

Note that subgroup centrality measures are defined based on the edge structure of the underlying network \((V, E)\), not the \textit{induced} edge structure of \((S, E')\).

The latter is termed \textit{subgraph centrality} by the author, not to be confused with the walk-based subgraph centrality put forward by Prof Estrada.
Subgroup v Subgraph

Degree

subgroup and subgraph centrality coincide.

Closeness

\[ cl_S(a) = \sum_{x \in S} d(a, x) \]

Subgroup closeness adds the distance from \( a \) only to nodes in \( S \). \textit{It does not matter if a geodesic uses nodes from outside \( S \).}

Note that for ranking purposes, nodes with \textit{high} \( cl_S(a) \) will be ranked \textit{last}. Alternatively, invert when normalizing, then rank as usual (with highest value ranked first).
Total communicability:

\[ \text{TC} = e^{\beta A} \cdot 1 \]

Spectral Decomposition of \( f(A) = e^{\beta A} \):

\[ \text{TC} = f(A) \cdot 1 = Qf(D)Q^T \cdot 1 \]

\[ = \sum_{i=1}^{n} e^{\beta \lambda_i} (q_i^T \cdot 1)q_i \]

\[ = e^{\beta \lambda_1} \left[ (q_1^T \cdot 1)q_1 + \sum_{i=2}^{n} e^{\beta \lambda_i - \lambda_1} (q_i^T \cdot 1)q_i \right] \]

Therefore eigenvector centrality is the limiting case of total communicability, as \( \beta \to \infty \).
Conclusion

▶ Two ways to look at eigenvector centrality:
  ▶ a node’s rank is proportional to the sum of the ranks of its neighbors:

\[
q_a = \sum_{x \in N(a)} \frac{1}{\lambda} q_x
\]

▶ limiting case of walk-based centrality measure

▶ From a walk-based point of view, it is not surprising that eigenvector subgroup measures give different results from eigenvector (induced) subgraph measures.
Betweenness

\[ f(a, \{x, y\}) = \frac{\text{#shortest paths between } x \text{ and } y \text{ that contain node } a}{\text{#shortest paths between } x \text{ and } y} \]

Note that \( x \neq a \) and \( y \neq a \).
Betweenness Example
Betweenness Example

\[0\] \[3.5\] \[1\] \[0.5\]
Subgroup Betweenness

For any $a \in V$ and any $S \subseteq V$,

$$b_S(a) = \frac{1}{2} \sum_{(x,y) \in S \times S} f(a, \{x, y\})$$

Then

$$b_S(a) + b_{S'}(a) + Bb_S(a) = b(a)$$

where $Bb_S(a)$ is the boundary betweenness of $a$ relative to $S$:

$$Bb_S(a) = \sum_{(x,y) \in S \times S'} f(a, \{x, y\})$$
Boundary Betweenness Example

- Local Betweenness $B_S(a) = 0$
- Global Betweenness $B_{S'}(a) = 0$
- Boundary Betweenness (normalized) $B_{bS}(a) = 1$ is largest possible
Local Betweenness and Local Clustering

For $a \in V$ let $S = \{a\} \cup N(a)$.

$$b_S(a) = \frac{1}{2} \sum_{(x,y) \in S \times S} B(a, \{x, y\}) = \sum_{x, y \in N(a) \atop (x, y) \in E} 0 + \sum_{x, y \in N(a) \atop (x, y) \notin E} B(a, \{x, y\})$$
Local Betweenness and Local Clustering

\[ b_S(a) = \sum_{x,y \in N(a)} 0 + \sum_{(x,y) \in E} B(a, \{x, y\}) \leq \sum_{(x,y) \notin E} 1 \]
Local Betweenness and Local Clustering

Therefore

\[ 1 = \frac{1}{\binom{|N(a)|}{2}} \left( \sum_{x,y \in N(a)} 1 + \sum_{x,y \in N(a), (x,y) \in E} 1 \right) \geq b'_S(a) + C_a \]

where \( C_a \) is the local clustering coefficient of \( a \):

\[ C_a = \frac{\# \text{ connected neighbors of } a}{\# \text{ pairs of neighbors of } a} = \frac{\sum_{x,y \in N(a), (x,y) \in E} 1}{\binom{|N(a)|}{2}} \]
The author claims:

\[ b'_S(a) = 1 - C_a \]

“thus local betweenness is a generalization of local clustering”

but in fact,

\[ b'_S(a) \leq 1 - C_a \]

- high clustering \(\rightarrow\) low local betweenness
- low local betweenness \(\not\rightarrow\) high clustering
Analysis on Dolphin Network

- 62 dolphins: 33 males, 25 females, 4 unknown
- Edges represent ‘friendship ties’
- Female dolphin SN100 (highest betweenness) splintered the group into 2 communities:
  - Males split into two groups
  - Females remain cohesive
Dolphin network
<table>
<thead>
<tr>
<th></th>
<th>Degree</th>
<th>Closeness</th>
<th>Betweenness</th>
<th>Eigenvector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beescratch</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Topless</td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>
Male-Dolphin subgraph

\[ n_1 = 8 \]
\[ \delta_1 = 0.464 \]

\[ n_2 = 21 \ (18) \]
\[ \delta_2 = 0.195 \ (0.242) \]
Male-Dolphin subgraph vs subgroup betweenness

Fig. 5. Subgraph and Subgroup Betweenness, Top Four. Subgraph betweenness is represented by solid squares and subgroup by solid disks, sized according to rank. The values indicate the rank in the other measure.
Interpretation

To ‘communicate’ with the male dolphins,

- assuming communication passes through the females as well, the subgroup approach may be more effective (e.g., spreading gossip).
- if females are somehow excluded (e.g., communicable disease affecting and carried solely by male dolphins), then the subgraph approach is relevant.
Subgraph-Eigenvector centrality

Rankings determined by relationships with other males only

many male friends
Subgroup-Eigenvector centrality

Rankings determined by relationships with both **males & females**

not many female friends

both m & f friends
Local vs Global Subgroup-Closeness

Fig. 7. Local and Global Closeness: Top Five. Local closeness is represented by solid disks and global by solid squares, sized according to rank. The values indicate the node’s rank in the other measure.
Interpretation

- High **global** subgroup-closeness $\rightarrow$ message spreads quickly among the **females**
- High **local** subgroup-closeness $\rightarrow$ message spreads quickly among the **males**
Other Interesting Results

Patchback: not in top 10 w.r.t. all other measures.
Number 1: ranked 20th globally and 14th overall.

<table>
<thead>
<tr>
<th>Local</th>
<th>Subgraph</th>
<th>Global</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beescratch</td>
<td>PL</td>
<td>Beak*</td>
<td>Beescratch</td>
</tr>
<tr>
<td>DN63</td>
<td>DN63</td>
<td>Topless*</td>
<td>DN63*</td>
</tr>
<tr>
<td>Oscar</td>
<td>Knit</td>
<td>MN105</td>
<td>Oscar*</td>
</tr>
<tr>
<td>Number 1*</td>
<td>SN96</td>
<td>Jonah*</td>
<td>Beak**</td>
</tr>
<tr>
<td>Upbang*</td>
<td>Oscar</td>
<td>Patchback*</td>
<td>Topless**</td>
</tr>
</tbody>
</table>

Table 1. Closeness. Top five ranked, listed in decreasing order. An asterisk or double asterisk indicates a tie.
Divisions based on structure

Fig. 8. Structural Dolphin Communities. Shape indicates gender: Males are square, females circles, and unknowns are triangles. Community 1 appears in white, Community 2 in black.
Results

Much less difference between subgroup and subgraph measures.

- For both communities, very high or perfect correlation between subgroup and subgraph-closeness, and subgroup and subgraph-betweenness.
- In Community 2, high correlation (0.999) for eigenvector centrality.
- In Community 1, correlation of 0.7342 for eigenvector centrality:
  - Upbang is distance two away from Community 2, therefore ranks highest in the subgroup sense.
  - Gallatin is further away from Community 2, therefore ranks highly only in the subgraph sense.
Conclusion and Extensions

- The different rankings obtained from different measures can (sometimes) provide information about the network.
- Proposed measures reveal properties not immediately apparent from the total graph nor the subgraph.
- Extends easily to weighted networks. It may be relevant to weight edges so that communications ‘prefer’ to travel within the subgroup if possible.
- Other centrality measures: different ’distance’ for closeness; only counting certain paths for betweenness; total communicability?
- Generalize to the temporal setting.
Extension to total communicability

\[ TC := e^A 1 \]

Some approaches:

- Rank is proportional to rank of immediate neighbors:
  \[
  TC_S(a) = \sum_{x \in S \cap N(a)} TC(x) = \sum_{(a,x) \in E} e_x^T e^A 1
  \]
  \[
  SC_S(a) = \sum_{x \in S \cap N(a)} SC(x) = \sum_{(a,x) \in E} \left( e^A \right)_{xx}
  \]

- Walk-based:
  \[
  \sum_{x \in S} \left( e^A \right)_{ax}
  \]

- Distance-based:
  \[
  \sum_{x \in S} \xi_{ax}
  \]
Francesca will tell you more...