Unit 12: Cores and Peripheries, Hubs and Authorities, Clustering, Community Detection

ICPSR
University of Michigan, Ann Arbor
Summer 2015
Instructor: Ann McCranie
Cores & Peripheries, Hubs & Authorities

- These are two quite different topics
- But both give nodes to both important actors AND their relation to the structure of the network as a whole
- Not really complete positional analysis as we’ve been discussing, but finding important types of positions nonetheless
Research interested in cores and peripheries

- World systems theory
- Scientific collaboration networks
- Internet research
- Effectiveness of organizational workgroup
- Within personal networks
  - Stability of support networks over time
  - HIV risk among drug users
  - Career Development
What are C&P used for?

- **World Systems Theory**
  - The argument that we should consider all the world’s countries, essentially, as part of one big system (network!) that can be understood as follows:
    - Core (capitalist, industrialized, rich)
    - Semi-periphery (industrializing, mostly capitalist, not as rich with poverty)
    - Periphery (not as industrialized, very poor)

- **Dependency theory**: resources (like trade) flow unequally to the core from the periphery
Cores and peripheries: What are they?

Borgatti and Everett define three intuitive understandings:

1. All members of a network belong to the network, but to greater or lesser extents. More “belonging” or connections are more core. It’s continuous.

2. There are core and periphery members, and nodes are in either one.

3. The core is the “center” of the nodes in Euclidean space, where the nodes in the middle have the most connections to others.

Classic Definition of Core/Periphery: Discrete Mode

- In this, there is a core and a periphery – an actor is either in the core (incident to lots of ties) or not (not incident)

Core and Periphery: Is it there?

- When we already think we know the core and the periphery a priori – we can posit it and test the fit
- In this: do male monkeys form a core and females a periphery?

We can essentially test here whether the average value on the core block is higher than the periphery, relative to the variance within the blocks – an implicit analysis of variance. In this case, the correlation is just .206, which is not significant in a QAP test.

Core and Periphery: How to find

- Posit two groups. Look for one that has many ties among its members and from others. The other may have ties to the core, but not much among itself.

Their proposed solutions: Genetic algorithm that tries to maximize the Pearson product moment correlation between a permuted observed matrix and an ideal. This continues until the algorithm finds the smallest number of errors.

Sociogram of Baker’s (1992) directed journal co-citation database from the field of social work, 1985-1986. Node size indicates in-degree and edge labels indicate the value of the tie. An edge’s value is listed in two located, near each receiving node. Drawn in UCINET.

Abbreviations: AMH (Administration in Mental Health), ASW (Administration in Social Work), BJSW (British J of SW), CAN (Child Abuse and Neglect), CCQ (Care Quarterly), CW (C Welfare), CYSR (Children and Youth Services Rev), CSWJ (Clinical SWJ), FR (Family Relations), IJSW (Indian J S W), JGSW (J Gerontological S W), JSP (J S Policy), JESW (J Education for S W), PW (Public Welfare), SCW (S Casework), SSR (S Service Rev), SW (S W), SWG (S W with Groups), SWHC (S W in Health Care), SWRA (S W Research and Abstracts),
Core and Periphery: Coreness

• Here, each node can be, to a greater or lesser extent, “core” through “coreness”
• A continuous measure.
• Does this through “pairwise” matching in the matrix – pairs have high values are high in coreness, middling for semi-periphery, low for periphery.

Core and Periphery: Coreness

- Here, each node can be, to a greater or lesser extent, “core” through “coreness”

<table>
<thead>
<tr>
<th></th>
<th>SW</th>
<th>SCW</th>
<th>SSR</th>
<th>ASW</th>
<th>CW</th>
<th>JSWE</th>
<th>SWRA</th>
<th>CSWJ</th>
<th>SWHC</th>
<th>CYSR</th>
<th>SWG</th>
<th>JGSW</th>
<th>PW</th>
<th>BJSW</th>
<th>FR</th>
<th>CAN</th>
<th>CCQ</th>
<th>AMH</th>
<th>JSW</th>
<th>JSP</th>
<th>Core</th>
</tr>
</thead>
<tbody>
<tr>
<td>SW</td>
<td>124</td>
<td>106</td>
<td>73</td>
<td>58</td>
<td>58</td>
<td>44</td>
<td>45</td>
<td>43</td>
<td>28</td>
<td>18</td>
<td>19</td>
<td>19</td>
<td>9</td>
<td>8</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>17.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCW</td>
<td>124</td>
<td>36</td>
<td>8</td>
<td>32</td>
<td>21</td>
<td>18</td>
<td>47</td>
<td>20</td>
<td>8</td>
<td>9</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>6.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSR</td>
<td>106</td>
<td>36</td>
<td>21</td>
<td>17</td>
<td>16</td>
<td>39</td>
<td>20</td>
<td>0</td>
<td>14</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CW</td>
<td>73</td>
<td>8</td>
<td>21</td>
<td>0</td>
<td>18</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JSWE</td>
<td>58</td>
<td>32</td>
<td>17</td>
<td>0</td>
<td>11</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>3.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASW</td>
<td>58</td>
<td>21</td>
<td>16</td>
<td>18</td>
<td>11</td>
<td>24</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWRA</td>
<td>44</td>
<td>18</td>
<td>39</td>
<td>20</td>
<td>8</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWHC</td>
<td>45</td>
<td>47</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSWJ</td>
<td>43</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWG</td>
<td>28</td>
<td>5</td>
<td>14</td>
<td>0</td>
<td>70</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>12</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYSR</td>
<td>40</td>
<td>9</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JGSW</td>
<td>18</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PW</td>
<td>19</td>
<td>0</td>
<td>13</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BJSW</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FR</td>
<td>9</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAN</td>
<td>8</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCQ</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LJSW</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMH</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Core and Periphery: Coreness

- Again, this fits an intuitive notion about being closer to the "center of the action"

Some criticisms

• This doesn’t allow us to talk about multiple cores (Everett & Borgatti, 1999)
• How do we measure goodness of fit?
• Directed networks in this arrangement are symmetrize – losing complexity. Also, treating self-loops as missing also truncates data (Boyd, 2010)
• Correlation does not mean that nodes will be identical (concordance) (Garcia Muniz, 2006)
An ideal core/periphery

Core/Periphery in a directed network

Clothing Trade, 2000


Using something similar to the coreness value of Borgatti & Everett, they also use singular value decomposition that allows graph asymmetries and utilizing the diagonal.

Fig. 3. Density matrix for clothing, 2000.
Fig. 4. Core/periphery plot for clothing, 2000.

Table 5
Top 15 Countries in 2000 Clothing Trade on Three Coreness Scores.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>w (Symmetric)</th>
<th>u (Export-Coreness)</th>
<th>v (Import-Coreness)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>USA</td>
<td>0.220</td>
<td>1</td>
<td>0.225</td>
</tr>
<tr>
<td>2</td>
<td>Italy</td>
<td>0.204</td>
<td>2</td>
<td>0.206</td>
</tr>
<tr>
<td>3</td>
<td>Germany</td>
<td>0.204</td>
<td>3</td>
<td>0.197</td>
</tr>
<tr>
<td>4</td>
<td>France</td>
<td>0.204</td>
<td>4</td>
<td>0.189</td>
</tr>
<tr>
<td>5</td>
<td>UK</td>
<td>0.203</td>
<td>5</td>
<td>0.189</td>
</tr>
<tr>
<td>6</td>
<td>Spain</td>
<td>0.181</td>
<td>6</td>
<td>0.184</td>
</tr>
<tr>
<td>7</td>
<td>Hong Kong</td>
<td>0.178</td>
<td>7</td>
<td>0.183</td>
</tr>
<tr>
<td>8</td>
<td>Netherlands</td>
<td>0.172</td>
<td>8</td>
<td>0.182</td>
</tr>
<tr>
<td>9</td>
<td>Japan</td>
<td>0.165</td>
<td>9</td>
<td>0.178</td>
</tr>
<tr>
<td>10</td>
<td>Belgium</td>
<td>0.165</td>
<td>10</td>
<td>0.176</td>
</tr>
<tr>
<td>11</td>
<td>China</td>
<td>0.158</td>
<td>11</td>
<td>0.169</td>
</tr>
<tr>
<td>12</td>
<td>Canada</td>
<td>0.156</td>
<td>12</td>
<td>0.167</td>
</tr>
<tr>
<td>13</td>
<td>Switzerland</td>
<td>0.153</td>
<td>13</td>
<td>0.166</td>
</tr>
<tr>
<td>14</td>
<td>Austria</td>
<td>0.153</td>
<td>14</td>
<td>0.159</td>
</tr>
<tr>
<td>15</td>
<td>South Korea</td>
<td>0.151</td>
<td>15</td>
<td>0.150</td>
</tr>
</tbody>
</table>
Hubs and Authorities

• Developed by Kleinberg (1999)
  • (Right around the time Brin and Page were thinking up what would become Google’s signature PageRank)
  • Called Hyperlink-Induced Topic Search (HITS)

• A good hub chooses many good authorities.
  • Good lists, lots of links, reference site

• Good authorities are chosen by many good hubs.
  • Lots of incoming ties, provide content
Supreme Court Decisions

Using citation date from about 30,000 Supreme Court Cases, they looked at the authority score (as determined by Kleinberg’s solution)

“The authority score of a case depends on the number of times it is cited and the quality of the cases that cite it. Symmetrically, the hub score of a case depends on the number of cases it cites and the quality of the cases cited. Thus, authority scores indicate the degree to which a case is thought to be important for resolving other important issues that come before the Court, while hub scores indicate the degree to which a case is well-grounded in previous important rulings.” (pg 17)

It’s context specific.

### Supreme Court Decisions

#### Table 3
Top five authorities (post-1953) as of 2002 by issue area

<table>
<thead>
<tr>
<th>Case</th>
<th>Authority score</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw</td>
<td>Percentile</td>
</tr>
<tr>
<td><strong>Civil rights</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Brown vs. Board of Education</em>, 347 U.S. 483 (1954)</td>
<td>0.07</td>
<td>99.88</td>
</tr>
<tr>
<td><em>Shapiro vs. Thompson</em>, 394 U.S. 618 (1969)</td>
<td>0.06</td>
<td>99.83</td>
</tr>
<tr>
<td><em>Baker vs. Carr</em>, 369 U.S. 186 (1962)</td>
<td>0.06</td>
<td>99.79</td>
</tr>
<tr>
<td><em>Reynolds vs. Simms</em>, 377 U.S. 533 (1964)</td>
<td>0.05</td>
<td>99.74</td>
</tr>
<tr>
<td><em>United States vs. Raines</em>, 362 U.S. 17 (1960)</td>
<td>0.05</td>
<td>99.70</td>
</tr>
<tr>
<td><strong>Criminal cases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Mapp vs. Ohio</em>, 367 U.S. 643 (1961)</td>
<td>0.08</td>
<td>99.89</td>
</tr>
<tr>
<td><em>Gideon vs. Wainwright</em>, 372 U.S. 335 (1963)</td>
<td>0.06</td>
<td>99.83</td>
</tr>
<tr>
<td><em>Miranda vs. Arizona</em>, 384 U.S. 436 (1966)</td>
<td>0.06</td>
<td>99.81</td>
</tr>
<tr>
<td><em>Katz vs. United States</em>, 389 U.S. 347 (1967)</td>
<td>0.05</td>
<td>99.77</td>
</tr>
<tr>
<td><em>Duncan vs. Louisiana</em>, 391 U.S. 145 (1968)</td>
<td>0.04</td>
<td>99.66</td>
</tr>
<tr>
<td><strong>First Amendment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>N.A.A.C.P. vs. Button</em>, 371 U.S. 415 (1963)</td>
<td>0.15</td>
<td>99.99</td>
</tr>
<tr>
<td><em>New York Times Co. vs. Sullivan</em>, 376 U.S. 254 (1964)</td>
<td>0.13</td>
<td>99.99</td>
</tr>
<tr>
<td><em>N.A.A.C.P. vs. Alabama</em>, 357 U.S. 449 (1958)</td>
<td>0.13</td>
<td>99.98</td>
</tr>
<tr>
<td><em>Speiser vs. Randall</em>, 357 U.S. 513 (1958)</td>
<td>0.13</td>
<td>99.98</td>
</tr>
<tr>
<td><em>Roth vs. United States</em>, 354 U.S. 476 (1957)</td>
<td>0.11</td>
<td>99.97</td>
</tr>
<tr>
<td><strong>Privacy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Griswold vs. Connecticut</em>, 381 U.S. 479 (1965)</td>
<td>0.08</td>
<td>99.90</td>
</tr>
<tr>
<td><em>Roe vs. Wade</em>, 410 U.S. 113 (1973)</td>
<td>0.06</td>
<td>99.80</td>
</tr>
<tr>
<td><em>Eisenstadt vs. Baird</em>, 405 U.S. 438 (1972)</td>
<td>0.05</td>
<td>99.71</td>
</tr>
<tr>
<td><em>Doe vs. Bolton</em>, 410 U.S. 179 (1973)</td>
<td>0.04</td>
<td>99.56</td>
</tr>
<tr>
<td><em>Carey vs. Population Services Int'l</em>, 431 U.S. 678 (1977)</td>
<td>0.03</td>
<td>99.28</td>
</tr>
</tbody>
</table>

*Note: Importance is determined by The Oxford Guide to United States Supreme Court Decisions (Hall, 1999), Congressional Quarterly’s Guide to the United States Supreme Court (Biskupic and Witt, 1997), and the Legal Information Institute (2005).*

Statnet in R demonstration

- Your lab exercises are already posted
- Execute the code in CHUNKS AT A TIME, so you can see how it works.
- I highly recommend you use Rstudio. It will make your experience much more pleasant.
CLUSTERING/COMMUNITY DETECTION AND GRAPH PARTITIONING
Background

- Clustering: a polyvalent term
- New solutions for new questions and new kinds of data
  - Huge, complex networks
  - Major computational issues
  - Introduction of (more) of biological and physical metaphors into the field of social networks
- Physics turned to social network analysis: called it network science
Clustering

• Can refer to a tendency for actors to group together

• Can refer to the “clustering coefficient”, which is the level transitivity in a network

• Newman (and others) distinguish this from “community detection,” but you can see that where there is lots of transitivity, there will be lots of ties, and there will be communities
Graph Partitioning

• An old problem in statistics, computer science
• Grouping vertices in a network into non-overlapping groups.
• The number and size of the partitions is prespecified
  • Sound familiar?
• It can force partitions when, perhaps they should not exist.
Graph Partitioning

- Partitioning is not a trivial matter.
- There are about 648 2-group solution for 10 nodes, but for 100 nodes this rises to about $2.5 \times 10^{29}$ - 250 Octillions
- So, you know, not feasible for direct calculation yet
“Community Detection”

- More appropriate term for what we are talking about here than clustering
- Trying to find the “natural fault lines” in the network with (Newman, 2010) in which community members are highly connected to one another, but not others
- Finding factions, but without predetermining how many or how large they will be
“Community Detection”

- Various solutions to defining “highly connected within and less connected without” have been proposed
- Modularity (Newman 2006)

“This idea, that true community structure in a network corresponds to a statistically surprising arrangement of edges, can be quantified by using the measure known as modularity (17). The modularity is, up to a multiplicative constant, the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random.”
Community detection: modularity
(Newman and Girvan 2004 and Newman 2006)

• Operates not on the matrix of the relation itself, but a characteristic matrix of the matrix, something Newman calls the “modularity matrix”

• You then: calculate the leading eigenvector of the modularity matrix and into groups by the signs of the elements
Finding modularity with two groups

Classic data set in terms of “factions” within a network. Shows a network of karate club (Zachery) members who eventually split into two groups over a club dispute.

Fig. 2. Application of the eigenvector-based method to the karate club network of ref. 23. Shapes of vertices indicate the membership of the corresponding individuals in the two known factions of the network, and the dotted line indicates the split found by the algorithm, which matches the factions exactly. The shades of the vertices indicate the strength of their membership, as measured by the value of the corresponding elements of the eigenvector.

Finding it with more: but how many?


Nice aspect of this algorithm is that it allows you to "maximize" modularity within a network, so that you find an optimal solution – i.e., dividing into more groups would start reducing the modularity of the groups.

Fig. 3. Krebs' network of books on American politics. Vertices represent books and edges join books frequently purchased by the same readers. Dashed lines divide the four communities found by the algorithm, and shapes represent the political alignment of the books (circles are liberal, squares are conservative, and triangles are centrist or unaligned).

Studying modularity over time: Consensus Formation

- Shwed and Bearman, 2010
- Argues that you can see the state of consensus in a given field over time by studying the changing modularity of the citations within that field.
- Three patterns: flat (no contestation), cyclic (key questions have no closure) and spiral (big questions have closure, spins into “higher level” questions.)

Figure 4. Epistemic Rivalry, Size, and Expert Reports in Five Validating Cases
Note: The dashed line refers to the number of papers in the dynamic window and to the logarithmic right-hand-side-axis. The solid line refers to the level of epistemic rivalry, estimated as the modularity score scaled for logged network size, on the left-hand-side-axis. The bars show years in which critical expert committees published a consensus report, or, in panel C, the years Collins identifies as marking the end of controversy and the emergence of consensus.
Other solutions

- Hierarchical clustering (really a class of solutions, we’ve seen them before in blockmodeling)
- Considering betweenness of edges as ways to find groups
- Good reviews:
  - Schaeffer, S. E., (2007) Graph clustering.
Latent Position Clustering Model

- Not discrete group membership
- Vertices can “belong” to different groups to greater or lesser degrees
- The probability of a tie from one actor to another depends on the distance between them in Euclidian “social space” and the “location” of the actor depends on different distributions that correspond to different clusters
- Two estimations for that likelihood: two-stage max. likelihood and Markov Chain Monte Carlo sampling

Krivitsky and Hancock, 2008
Latent Position Clustering Model

- Doesn’t require knowing the number of clusters first
- Can take into account clustering, transitivity and the well-known tendency toward homophily on actor attributes.
- A stochastic blockmodel

Handcock, Raftery, Tantrum 2007
Jenine K. Harris, Kate E. Beatty, Jesse D. Lecy, Julianne M. Cyr, Robert M. Shapiro II, Mapping the Multidisciplinary Field of Public Health Services and Systems Research. American Journal of Preventive Medicine, Volume 41, Issue 1, July 2011, Pages 105-111
Consider blockmodeling

- Allows for many alternative specifications of grouping that aren’t about being in a community – they can be about connecting to similar kinds of actors.
- You can truly get at roles and positions instead of “communities”
- This is an important distinction.
- But, you can use some of the intuitions of blockmodeling – namely, the reduction of the network into recognizable groups, as part of your work here.
Reading

- Chapter 10 in Mark Newman’s “Networks: An Introduction” is excellent and technical

- Shwed and Bearman (2010) “Consensus Formation”