Lecture 10: Divide

"What are natural communities, divisions? Who is likely to love/hate my blog?"

COMS 4995-1: Introduction to Social Networks
Thursday, October 11th
Minimizing the conductance of a partition **Bisection**

\[ \phi(S) = \frac{\sum_{i \in S, j \notin S} A_{ij}}{\min\{A(S), A(S')\}} \]

\[ A(S) = \sum_{i \in S} \sum_{j \in V} A_{ij} \]

- Efficient techniques to divide a graph in 2
  - Could be using relaxation method
  - Even proved to approximate NP-hard problems

- Then apply the same algorithm recursively
Outline

* Groups, Clusters, Communities
* Extraction
  o Modularity, Conductance, Hierarchy
* Validation
  o Impact of size, Ground truth
* Evolution & Community model
* Study of network with positive & negative links
Groups, Clusters, Communities

Extraction
  - Modularity, Conductance, Hierarchy

Validation
  - Impact of size, Ground truth

Evolution & Community model

Study of network with positive & negative links
### 2008: Statistical study Comm. Str.

<table>
<thead>
<tr>
<th>DATA SET:</th>
</tr>
</thead>
<tbody>
<tr>
<td>70 graphs (Live journal, web, etc.)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EMPRICAL OBSERVATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decreasing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Studies conductance for size k</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>MODEL:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest fire</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANALYSIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirically reproduced the size dep.</td>
</tr>
</tbody>
</table>

---

Let us define the conductance of $S$:

$$
\phi(S) = \frac{\sum_{i \in S, j \notin S} A_{ij}}{\min\{A(S), A(\overline{S})\}}
$$

Where

$$
A(S) = \sum_{i \in S} \sum_{j \in V} A_{ij}
$$

What about looking at how it depends on size of $S$?

$$
\Phi(k) = \min_{S \subseteq V, |S| = k} \phi(S).
$$
Geographical Networks

(a) Low-dimensional meshes

(b) POWERGRID
Example 1

(a) Zachary’s karate club

(b) ...and it’s NCP plot
Example 2

(c) Network science

(d) ...and it's NCP plot
Example of large network

(a) LIVEJOURNAL
What about explicit communities?

(a) LIVEJOURNAL
Community only up to a fixed size (~100)
  - Usually characterized as “whisker” (isolated components, poorly connected)
No easy way to cut the core
Forest fire is shown to reproduce this behavior
A plausible explanation

Dunbar’s number

Has been observed repeatedly in history

Outline

* Groups, Clusters, Communities
* Extraction
  - Modularity, Conductance, Hierarchy
* Validation
  - Impact of size, Ground truth
* Evolution & Community model
* Study of network with positive & negative links
### 2006: Group formation

<table>
<thead>
<tr>
<th>DATA SET:</th>
<th>DATA SET:</th>
<th>DATA SET:</th>
<th>DATA SET:</th>
<th>DATA SET:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live journal, and DBLP</td>
<td>Contains friendship and explicit communities</td>
<td>Look at triple ((u,C,k))</td>
<td>EMPIRICAL OBSERVATIONS</td>
<td>Law of diminishing returns</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How to join? Effect of # friends

Graphs showing the probability of joining a community or conference when a certain number of friends are already members.
How to join? Effect of conn. friends

- Proportion of adjacent pairs
How to join? Community Growth

All community with pop >100
**2009: Affiliation Networks**

**MODEL:**
Evolution of Bipartite Graph $B(Q,U)$  
Parameters: $c_q$, $c_u$, $\beta$

With proba $\beta$:  
New node $q$, choose $q'$ according to deg  
Copy $c_q$ edges from $q'$  

With proba $1-\beta$:  
New node $u$, choose $u'$ according to deg  
Copy $c_u$ edges from $u'$

Folding of the graph $G(Q,E)$  
Connect all nodes with a common affiliation

**ANALYSIS:**
Proof that it reproduces  
- Heavy-tail degree distribution  
- Densification (if $c_u < c_q \frac{\beta}{(1-\beta)}$)  
- Shrinking Diameter (same cond.)

Presence of popular affiliation that are always nearby

Interesting sparsification methods

Extensions to other folding methods

---

Outline

* Groups, Clusters, Communities
* Extraction
  o Modularity, Conductance, Hierarchy
* Validation
  o Impact of size, Ground truth
* Evolution & Community model
* Study of network with positive & negative links
What if everyone “knows” each other?
- The acquaintance graph is complete
- But some relationship are positive, some negative
- Ex: wikipedia elections, diplomatic relations, slashdot, epinions.

- B trusts C

- B distrusts C
Let’s consider groups of 3 nodes

* Which one(s) seem “paradoxical/unstable”?

![Graphs showing structural balance](image)

5.1. STRUCTURAL BALANCE

- Based on this reasoning, we will refer to triangles with one or three `+`'s as balanced.
- Similarly, there are sources of instability in a configuration where each of A, B, C are mutual enemies (as in Figure 5.1(d)). In this case, there would be forces motivating two of the three people to “team up” against the third (turning one of the three edge labels to -).
- Let’s consider groups of 3 nodes.
- Which one(s) seem “paradoxical/unstable”?

![Graphs showing structural balance](image)
A complete positive/negative graph $G = (V, E^+, E^-)$ is said balanced if the following condition holds:

- For any triangle in the graph, either they are all friends or there is exactly one edge $+$ between them.

Figure 5.2: The labeled four-node complete graph on the left is balanced; the one on the right is not.
Anecdotal evidence

* Diplomatic relation in Europe 1872-1907
  - “The reason for fighting. I never got straight.”

![Graphs showing alliances in Europe from 1872 to 1907](image)

(a) Three Emperors' League 1872–81
(b) Triple Alliance 1882
(c) German-Russian Lapse 1890
(d) French-Russian Alliance 1891–94
(e) Entente Cordiale 1904
(f) British-Russian Alliance 1907
Theorem:

- A complete graph $G=(V,E^+,E^-)$ is balanced iff
- It can be divided in two sets of mutual friends, with complete mutual antagonisms between the set

Proof:

Local moves have global effect

Friends of A

Enemies of A
5.4 A Weaker Form of Structural Balance

What if you may not always league against a common enemy?

Result:

- Mutual friends inside W
- Mutual friends inside X
- Mutual friends inside Y
- Mutual friends inside Z
- Set W
- Set X
- Set Y
- Set Z
- Mutual antagonism between all sets

![Diagram of structural balance](image)

Figure 5.1: Structural balance: Each labeled triangle must have 1 or 3 positive edges.

A, B, and C are mutual friends: balanced.

A is friends with B and C, but they don't get along with each other: not balanced.

A and B are friends with C as a mutual enemy: balanced.

A, B, and C are mutual enemies: not balanced.

Based on this reasoning, we will refer to triangles with one or three +'s as balanced, since they are free of these sources of instability, and we will refer to triangles with zero or two +'s as unbalanced. The argument of structural balance theorists is that because unbalanced triangles are sources of stress or psychological dissonance, people strive to minimize them in their personal relationships, and hence they will be less abundant in real social settings than...
Structural Balance vs. Social Status
So far, we assume that all pairs have an edge
  o What if, in addition, they may simply have no edge

A relaxed definition
  o Missing edge = edge not observed yet
  o A graph is balanced iff all edges currently missing can be fixed so that the resulting graph is balanced
  o Note that this generalizes our previous definition
Examples:
A first result

* This is equivalent to: partition graph in two: A B
  o all edges inside A and B are positive
  o all edges between a node in A and B are negative

* Proof follows previous argument
Again, a more general condition that seems local:

- Property: G does not admit any cycle with an odd number of negative edge
- It has to be true for a graph to be balanced
- It is in fact also a sufficient condition
Proof

* Step 1: reduces to super-nodes graph $G'$

* Cycle with odd negative exists iff true in $G'$
* The final graph only has negative edges!
Step 2: apply a general result
  o Lemma: a graph is bipartite iff it has no odd cycle
What if instead of 100% of triangles are balanced ... 
  o We had $1-\epsilon$ fraction of triangles that are balanced 
  o In other words, conditions is rarely observed.

THM: Assuming $\epsilon < 1/8$ and defining $\delta = \epsilon^{1/3}$, then 
  o for $(1-\delta)$ of nodes, at least $(1-\delta)$ of pairs are friends 
  o or, we can divide $G$ in two groups $X, Y$ such that 
    * In $X$ and $Y$, at least $(1-\delta)$ of pairs are friends  
    * At least $(1-\delta)$ pairs with one end in $X$ and $Y$ are enemies