Learning Human Behavior From Data
(Part 4)

SNA
Lecture #11
Where are networks?

Nature organizes itself as networks

“network science”

Aspartame sweetener

Protein reactions

Neurons in the brain
Where are networks?

... and so do humans

An effective way of organizing

Organization chart

Trust ties

Regular permutation group arises from Arunta marriage system.
Levels of analysis

• Dyad level – O(n²)
  – Units are pairs of persons
  – Variables are things like presence of absence of a certain kind of tie between each pair of persons in network

• Node level – O(n)
  – Units are persons
  – Variables are things like the number of friends each person has

• Group/network level – O(1)
  – Units are whole networks (e.g., teams, firms or countries)
  – Variables are things like the density of trust ties, or the average number of degrees of separation between members of the group
## Types of studies

<table>
<thead>
<tr>
<th></th>
<th>Dyad Level</th>
<th>Node Level</th>
<th>Group Level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theory of Networks</strong></td>
<td>Understanding who becomes friends with whom</td>
<td>Explaining why some people are more liked than others</td>
<td>Explaining why some groups have more centralized network structures</td>
</tr>
<tr>
<td>(Antecedents)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Network Theory</strong></td>
<td>Predicting similarity of opinion as a function of friendship</td>
<td>Explaining why some employees rise through the ranks faster than others as a function of social ties</td>
<td>Predicting team performance as a function of structure of trust network within team</td>
</tr>
<tr>
<td>(Consequences)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Perspectives on networks

• Contextual
• Relational
• Structural
• Positional

• Let’s see what each of these mean
Contextual Perspective

• Importance of an individual’s environment
  – To explain individual outcomes, must take into account the node’s social environment in addition to internal characteristics
  – In SNA, the environment is conceptualized as network
Relational Perspective

• SNA puts the focus on relationships between individuals
• Contagion
• Social capital
Relational Perspective - Contagion

- Individuals influence each other
  - Infect each other with diseases, ideas, behaviors
- As a result, we observe network autocorrelation — the tendency for adjacent nodes to have similar characteristics such as opinions, ways of dressing, food preferences
- Flows of information, money
Relational Perspective - Contagion

Discovery of HIV: Sexual contacts among gay men w/ unusual cancers, traced by Bill Darrow of the CDC

The case of AIDS
Relational Perspective – Social Capital

• Why are some individuals more successful than others?
  – Attributes such as intelligence, motivation
  – Human capital

  – Who they know, who they owe
  – Social capital

• Social ties provide access to resources the individual doesn’t own/control directly
Structural Perspective

• It’s not just relational (ties) but structural (pattern of ties)
  – To understand function, need to know more than list of elements.
  How are they connected?
Positional Perspective

• A node’s position in a network determines in part the opportunities and constraints that it will face
  – Risk of news, risk of infection
  – Sense of identity
  – Individual social capital

• Social ties provide conduits along which traffic can flow
• A node’s position in the network has significant implications for
  – How early it encounters something flowing
  – How frequently it receives what is flowing
  – With what certainty it is reached
Summarizing the network perspectives

- **Contextual**
  - It’s the environment!

- **Relational**
  - By environment, we mean ties to others

- **Structural**
  - It’s a network
  - Concepts and metrics for characterizing the network

- **Positional**
  - Location, location, location
What is a network?

• A set of nodes
  – Aka actors, agents, points, vertices

• A set of ties (social relations) of a given type among actors
  – Aka links, lines, edges, arcs
  – Ties can be kinship, friendship, communication, etc

• Each type of tie defines a different network
  – We talk about “the friendship network”, “the advice network”
Nodes

• Individuals or collectives
  – Individual actors can be persons, chimps, cows, etc
  – Collective actors can be teams, organizations, countries, etc.

• Node we are talking about is ego, any node ego is connected to is an alter
  – The set of nodes an ego is connected to, along with ties among them, is an ego network

• In North American social sciences, nodes generally restricting to entities with agency/volition and behavior
• European “actor-network theory” (ANT) mixes persons and objects in network
Ties (edges)

• Typically a relationship between pairs of nodes
  – When we distinguish a single conversation of three people from three conversations between three pairs, we typically think of this as participation in a conversation, not a 3-way tie

• Often measured as binary, present/absent data, but can also be values such as Likert response scales, frequencies, magnitudes of flows (e.g., $ value of trade between countries)

• Ties can be directed or undirected
  – Directed ties go from the first node to the second as in A gave advice to B (and B may or may not have given advice to A)
  – Undirected ties have no direction, as in A and B are members of the same club.
How to represent networks?

- Graph theory
- Matrix algebra
- Relational set theory
Graph theoretic representation

• A network is represented by a graph
• A graph G(V,E) consists of a set of nodes V and a set of edges (ties) E
  – If the ties are directed, we call the edges “arcs”

• If u and v are nodes in V, we can indicate the presence of a tie one of two ways:
  – $u \rightarrow v$ (if directed) or $u - v$ (if undirected)
  – $(u,v) \in E$
    • If the network is undirected then writing $(u,v) \in E$ is the same as $(v,u) \in E$, and we can also write $u - v$ to indicate the tie
Matrix representation

- X is (often) a square matrix in which both rows and columns refer to nodes, and $x_{ij} = 1$ (or $x(i,j) = 1$) indicates the presence of a tie from I to j and $x_{ij} = 0$ indicates no tie was reported from I to j.
  - The row person does it to the column person. E.g., if A represents Liking, then $x(b,a) = 1$ means that b likes a.

- If we are being careful, we distinguish $x_{ij} = 0$ (which indicates the absence of a tie) from $x_{ij} =$ missing value (which means we don’t know if there is a tie or not).

- If the network is directed, then the matrix need not be symmetric, meaning that $x_{ij}$ may not equal $x_{ji}$ for all I and j.
Matrix transpose

• The transpose of matrix $A$ is denoted $A'$ or $A^T$ and is the matrix obtained by interchanging rows and columns of $A$
  – This means that $a(i,j) = a'(j,i)$

• Has the effect of reversing all the arcs in the network, so if $A$ represents “gives money to”, then $A'$ represents “is given money by”
Matrix transpose

<table>
<thead>
<tr>
<th>Gives money to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary</td>
</tr>
<tr>
<td>Mary</td>
</tr>
<tr>
<td>Bill</td>
</tr>
<tr>
<td>John</td>
</tr>
<tr>
<td>Larry</td>
</tr>
</tbody>
</table>

Transpose of “gives money to”

<table>
<thead>
<tr>
<th>Gives money to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary</td>
</tr>
<tr>
<td>Mary</td>
</tr>
<tr>
<td>Bill</td>
</tr>
<tr>
<td>John</td>
</tr>
<tr>
<td>Larry</td>
</tr>
</tbody>
</table>

Aka “is given money by”

Is given money by

<table>
<thead>
<tr>
<th>Is given money by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary</td>
</tr>
<tr>
<td>Mary</td>
</tr>
<tr>
<td>Bill</td>
</tr>
<tr>
<td>John</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Is given money by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary</td>
</tr>
<tr>
<td>Mary</td>
</tr>
<tr>
<td>Bill</td>
</tr>
<tr>
<td>John</td>
</tr>
</tbody>
</table>
Matrix transpose

• The transpose of matrix A is denoted A’ or $A^T$ and is the matrix obtained by interchanging rows and columns of A
  – This means that $a(i,j) = a'(j,i)$

• Has the effect of reversing all the arcs in the network, so if A represents “gives money to”, then A’ represents “is given money by”
Frequently measured ties

• Who knows whom (acquaintance) Especially for large populations
• Friendship, socializing outside of work
• Advice, communication, seeking information from
• Boss of, work dependency (my input is your output)
• Co-membership in boards, co-participation in social events
• Emails, tweets, electronic interactions
Centrality

• Structural importance
  – An aspect of a node’s position in a network
Centrality

In each of the following networks, X has higher centrality than Y according to a particular measure.

- Indegree
- Outdegree
- Betweenness
- Closeness
Indegree vs outdegree

• Suppose relation is seeks advice from
  Outdegree = how many people you seek help from
  Indegree = how many people seek you out
how many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?
Closeness

• What if it’s not so important to have many direct friends?
• Or be “between” others
• But one still wants to be in the “middle” of things, not too far from the center
Announcements

• Homework 3 progress presentations
  – April 13th in class
  – Need to show ~50% progress towards completion
    • Data chosen
    • Ground truth established
    • Features chosen and how to extract them
    • Experimental methodology
Assigned Reading – Paper 7

Measuring User Influence in Twitter: The Million Follower Fallacy

Meeyoung Cha et al. 2010