Modeling the Speed of Diffusion and Learning in Heterogeneous and Segregated Societies

SED July 3 2009   Plenary Talk of Matthew O Jackson

Based on Paper: B. Golub and M.O. Jackson (2008)

Copyright 2009, Do not post, distribute, or reproduce in any way without authors’ permission.
Introduction: Speed

How quickly does information diffuse in a society?

How quickly does a society reach an approximate consensus in opinions?

How quickly does a disease spread?

Financial contagions?

How do the answers depend on network structure?
Main Questions

• How does speed depend on the process?
  • Pure diffusion – spreading out from one node
  • Updating/Learning – repeated processing of information from neighbors

• How does speed depend on the network?
  • Density of links
  • Pattern of links
Networks

• Networks differ in their link density

• Networks differ in how links are spread across nodes: **Homophily**
  – Bias of relationships towards own type

• Technology and globalization are changing networks:
  – More relationships??
  – more/less homophily??
Outline

I Background on networks

II Modeling networks and diffusion/updating processes

III Results on how structure impacts the speed of diffusion and learning
<table>
<thead>
<tr>
<th>Category</th>
<th>Average Degree (# links)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS Friendships (CJP 09)</td>
<td>6.5</td>
</tr>
<tr>
<td>Romances (BMS 03)</td>
<td>0.8</td>
</tr>
<tr>
<td>Borrowing (BDJ...)</td>
<td>3.2</td>
</tr>
<tr>
<td>Co-authors (Newman 01, GLM 06)</td>
<td></td>
</tr>
<tr>
<td>Bio</td>
<td>15.5</td>
</tr>
<tr>
<td>Econ</td>
<td>1.7</td>
</tr>
<tr>
<td>Math</td>
<td>3.9</td>
</tr>
<tr>
<td>Physics</td>
<td>9.3</td>
</tr>
<tr>
<td>Facebook (Marlow 09)</td>
<td>120</td>
</tr>
</tbody>
</table>
Homophily/Segregation Patterns:

- Tendency to associate with others with similar characteristics: age, race, gender, religion, profession....

  - Lazarsfeld and Merton (1954) "Homophily"

Yellow: Whites
Blue: Blacks
Reds: Hispanics
Green: Asian
Pink: Other
White: Missing
### Adolescent Health, High School in US:

<table>
<thead>
<tr>
<th>Percent:</th>
<th>52</th>
<th>38</th>
<th>5</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>86</td>
<td>7</td>
<td>47</td>
<td>74</td>
</tr>
<tr>
<td>Black</td>
<td>4</td>
<td>85</td>
<td>46</td>
<td>13</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Other</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
How much more likely to link to own type?

<table>
<thead>
<tr>
<th>84 schools</th>
<th>Race</th>
<th>Sex</th>
<th>Grade (Age)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg</td>
<td>1.6</td>
<td>1.5</td>
<td>8.5</td>
</tr>
<tr>
<td>stddev</td>
<td>.10</td>
<td>.08</td>
<td>.90</td>
</tr>
</tbody>
</table>
Modeling:

I Random Networks

II Speed of Diffusion/Learning
Multi-Type Random Network Model

- \{1, \ldots, n\} agents/nodes

- Partitioned into groups \( N_1, \ldots, N_m \)

- Node \( i \) in group \( k \) is linked to a node \( j \) in group \( k' \) with probability \( P_{kk'} \) (undirected)

- Homophily: \( P_{kk} > P_{kk'} \) for \( k' \neq k \)
Multi-Type Random network
For Exposition: Islands Model

- Link with a same type agent: probability $p_s$
- Link with a diff type agent: probability $p_d$
- Overall probability of links $p$ (>> $1/n$)
- $m$ equal-sized groups
Example Low Homophily
Example High Homophily
Diffusion/Learning Processes:

(1) Diffusion / gossip / broadcasting / navigation: speed depends on shortest path structure

(2) Updating: repeated discussion and weighing of neighbors’ opinions, iterative best response...
1. Shortest Paths / Pure Diffusion

• Characterize how shortest paths are affected by density and homophily
2. Model of Updating/Learning

French (1956), Harary (1959), DeGroot (1974)...

• At each date talk to neighbors

• update opinion or behavior by taking an average of neighbors’ opinion/behavior

• Iterate on this process
DeGroot Model

- Individuals \{1, \ldots, n\} are in a network

- \(T_{ij} = 1/d_i\) if i is linked to j and = 0 otherwise

- Start with beliefs (behavior, etc.) \(b_i(0)\) in [0,1]

- Updating: \(b_i(t) = \sum_j T_{ij} b_j(t-1)\)

  So: \(b(t) = T b(t-1) = T^t b(0)\)
Example

\[ T = \begin{bmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{2} & \frac{1}{2} & 0 \\
\frac{1}{2} & 0 & \frac{1}{2}
\end{bmatrix} \]
Updating
Updating

\[
\begin{array}{ccc}
0 & 1/3 & 1/3 \\
1/3 & 1/2 & 0 \\
1/3 & 1/2 & 1/3 \\
\end{array}
\]

\[
\begin{array}{ccc}
1/2 & 0 & 1 \\
1 & 1/3 & 1/3 \\
1/3 & 1/2 & 1/3 \\
\end{array}
\]
Updating

1/3 1/3 1/3 1/2 1/2

0 1/3 0 1/3 1/3 1/2 1/2

5/18 1/3 1/6 1/2 1/2 1/2 0 1/3 1/3 1/2 1/2 1/2

5/12 1/3 1/6 1/2 1/2 1/2 0 1/3 1/3 1/2 1/2 1/2
Updating

\[
\begin{align*}
\text{1/3} & \quad \text{1/3} & \quad \text{1/2} & \quad \text{1/2} \\
\text{2} & \quad \text{1} & \quad \text{0} & \quad \text{0} \\
\text{0} & \quad \text{3} & \quad \text{1/3} & \quad \text{1/2} \\
\text{1/3} & \quad \text{1/3} & \quad \text{1/2} & \quad \text{1/2} \\
\end{align*}
\]

\[
\begin{align*}
\text{1/3} & \quad \text{1/3} & \quad \text{1/2} & \quad \text{1/2} \\
\text{2} & \quad \text{1} & \quad \text{0} & \quad \text{0} \\
\text{0} & \quad \text{3} & \quad \text{1/3} & \quad \text{1/2} \\
\text{1/3} & \quad \text{1/3} & \quad \text{1/2} & \quad \text{1/2} \\
\end{align*}
\]

\[
\begin{align*}
\text{5/18} & \quad \text{5/12} & \quad \text{1/6} & \quad \text{1/6} \\
\text{1} & \quad \text{2} & \quad \text{3} & \quad \text{3} \\
\text{1/3} & \quad \text{1/3} & \quad \text{1/2} & \quad \text{1/2} \\
\text{1/3} & \quad \text{1/3} & \quad \text{1/2} & \quad \text{1/2} \\
\end{align*}
\]

\[
\begin{align*}
\text{5/18} & \quad \text{5/12} & \quad \text{1/6} & \quad \text{1/6} \\
\text{1} & \quad \text{2} & \quad \text{3} & \quad \text{3} \\
\text{1/3} & \quad \text{1/3} & \quad \text{1/2} & \quad \text{1/2} \\
\text{1/3} & \quad \text{1/3} & \quad \text{1/2} & \quad \text{1/2} \\
\end{align*}
\]

\[
\begin{align*}
\text{2/7} & \quad \text{2/7} & \quad \text{2/7} \\
\text{1} & \quad \text{2} & \quad \text{3} \\
\text{1/3} & \quad \text{1/3} & \quad \text{1/2} & \quad \text{1/2} \\
\text{1/3} & \quad \text{1/3} & \quad \text{1/2} & \quad \text{1/2} \\
\end{align*}
\]
Convergence

If the network is path-connected and at least one agent places some weight on own opinion, then the DeGroot process has a unique limit (a standard Markov result) and all agents converge to same belief.

The Consensus is $\sum_i b_i(0) \, d_i / \sum_j d_j$
$t = 4$
Other Processes that are equivalent to DeGroot:

• **Myopic best responses**: Each agent wants to match the average behavior of his/her neighbors –

• **Random walk – Markov Chain**: Start at some node/state and randomly transition - how long until distribution of location looks like steady-state distribution?
Consensus Time

\[ CT(T, \varepsilon) = \sup_b \min \{ t : \| T_t b - T^\infty b \| < \varepsilon \} \]

How long until vector of beliefs is within \( \varepsilon \) of its limit? (worst case)
Analysis

- Multi-Type Random Network is Drawn

- Diffusion: What is average path length?

- Learning/Random Walk: How fast does DeGroot process converge?
Theorem on Network Structure (Jackson 08)

Large $n$ results, let $d = np$ (avg degree)

$$\frac{\text{AvgDist}(n)}{\log(n)/\log(d)} \xrightarrow{P} 1$$

link density matters but not homophily
Intuition:

1 step: Reach d nodes,
Ideas:

1 step: Reach $d$ nodes, then $d(d-1)$,
Ideas:

1 step: Reach $d$ nodes,
then $d(d-1)$,
then $d(d-1)^2$, 
Ideas:

1 step: Reach d nodes, then d(d-1), then d(d-1)^2, d(d-1)^3, ...

After k steps, totals roughly d^k
Ideas:

After $k$ steps, reach $d^k$

When do we reach all $n$?

$$d^k = n \quad \text{or} \quad k = \log(n)/\log(d)$$
Ideas:

After $k$ steps, reach $d^k$

When do we reach all $n$?

$d^k = n$ or $k = \frac{\log(n)}{\log(d)}$

*Most* at maximum distance (10, 100, 1000, 10000...)
Small Worlds/Six Degrees of Separation

• $n = 6.7$ billion (world population)

• $d = 50$ (friends, relatives…)

• $\log(n)/\log(d)$ is about 6 !!
Learning Speed?
Learning Speed?
Measuring Homophily

\[ WH = \frac{(p_s - p_d)}{(mp)} \]

extra frequency of linking to own type relative to overall link frequency

[between -1 and 1]
Theorem (Golub and Jackson 08)

\[ CT(T(n), 1/n) \approx \frac{\log(n)}{\log(1/|WH|)} \]

\[ \approx \text{means Prob(within factor of 2)} \rightarrow_n 1 \]
Representative Agents:
Representative Agents:
Consensus Time

Large $n$:

$CT$

$0 \quad 1/2 \quad 1 \quad WH(n)$
So

AvgDist $\approx \log(n) / \log(d)$

CT $\approx \log(n) / \log(1/|WH|)$
## Speed:

![Image of graph with nodes and edges]

<table>
<thead>
<tr>
<th>Process</th>
<th>Independent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Density</td>
</tr>
<tr>
<td>Shortest Path</td>
<td>↑</td>
</tr>
<tr>
<td>Linear Updating</td>
<td>0</td>
</tr>
</tbody>
</table>
Summary

• Pure Diffusion/Distance/Diameter is unaffected by segregation/homophily but is decreased by density
  – If the process on one link is unaffected by neighboring links, then network density matters but not the pattern

• Random Walks/Weighted Averaging/Updating is slowed by segregation/homophily but unaffected by density
  – If the process on one link is relative to the number of neighboring links then network pattern matters but not density