Preference Aggregation in Social Networks

Workshop on High Dimensional Network Analytics

Swapnil Dhamal
Joint work with Y. Narahari

Department of Computer Science and Automation
Indian Institute of Science, Bangalore

19 December, 2013
Overview

1. Introduction
2. Empirically
3. Formulation
4. Results
5. Conclusion
Introduction to Homophily in Social Networks

- **What is a social network?**
  Individuals and some friendships

- **What is a primary cause of friendships?**
  Individuals are similar

- **What is a major effect of these friendships?**
  Individuals become similar

- **What is homophily?**
  A bias in friendships towards similar individuals

Social networks and homophily are inseparable.
Preliminaries of Preference Aggregation

- **Voters** have certain preferences over a set of Alternatives

\[ \begin{array}{ccc}
X & Y & Z \\
Y & Z & X \\
Z & X & Y \\
\end{array} \]

\[ \begin{array}{ccc}
Y & Z & X \\
X & Y & Z \\
Y & X & Z \\
\end{array} \]
Preliminaries of Preference Aggregation

- Preference of a voter is a complete ranked list of alternatives

```
Preference of voter i
Y Z X
```

```
Preference of voter j
X Y Z

Preference of voter p
Y X Z
```
Preference Profile $P$ is the multiset of preferences of voters.
Preliminaries of Preference Aggregation

- **Aggregation Rule** \( f \) tells how to aggregate preferences of voters

```
Preference Profile

\[
\begin{array}{ccc}
  i & Y & Z & X \\
  j & X & Y & Z \\
  p & Y & X & Z \\
\end{array}
\]
```

Preference of voter \( i \)

Aggregation Rule

Plurality

\[
\begin{array}{ccc}
  Y & X & Z \\
\end{array}
\]
Preliminaries of Preference Aggregation

- Aggregate Preference $f(P)$ represents preferences of the voters

Preference Profile

Preference of voter $i$

Aggregation Rule

Plurality

Aggregate Preference
\( r \) = number of alternatives

Normalized Kendall-Tau Distance = \( \frac{\text{Number of pair inversions}}{\binom{r}{2}} \)

Distance between \((X, Y, Z)\) and \((Y, Z, X)\) is \(\frac{2}{3}\)
Motivation

- May be infeasible to gather the preferences from all the voters owing to factors like time and interest of the voters.

- Several aggregation rules are computationally intensive.

**Estimate** the aggregate preference of the population by selecting a **subset** of voters and obtaining their preferences.
Social networks influence voting in elections \(^1\) \(^2\)

It is optimal to ignore the network structure \(^3\)

What to believe ???

---

Node selection in voting using attributes of nodes and alternatives without taking social network into account

Node selection in influence maximization, influence limitation, virus inoculation, etc. taking social network into account

---


Overview

1. Introduction
2. Empirically
3. Formulation
4. Results
5. Conclusion
Empirical Observations

- Rankings related to personal and social issues
- Similar rankings by connected nodes
- Highly similar rankings by nodes belonging to big clusters
- First and last alternatives mostly consistent for connected nodes
- Similar rankings by (un)connected nodes for social issues

Social network had a much greater influence on rankings related to personal issues than that related to social ones.
Distribution of Distances

- Histogram for different questions for a given pair and type of issue fit by truncated Gaussian distribution with range $[0, 1]$
- Call discrete version of truncated Gaussian distribution as $\mathcal{D}$

Distance between $\{i, j\}$ followed this distribution with mean $d(i,j)$

$$c(\cdot, \cdot) = 1 - d(\cdot, \cdot)$$
Modeling Homophily for Unconnected Nodes

- $d(i, j)$ known for connected pairs $\{i, j\}$ [0 for $i = j$]

- Initialize $d(i, j) = 1$ for all unconnected pairs

\[d(p, i) + r \cdot d(p, j) < d(i, j)\]  

\[\text{then } d(i, j) = d(p, i) + r \cdot d(p, j)\]

---

All pairs shortest path with update rule \(^7\)

Overview

1. Introduction
2. Empirically
3. Formulation
4. Results
5. Conclusion
Problem Statement

Given a network with a set of nodes $N$ and an aggregation rule $f$, select a subset of nodes $M \subseteq N$ of cardinality $k$, and deduce an aggregate preference that is close enough in expectation to the aggregate preference of $N$ using $f$. 
Representative of a Node

Representative of node $i$ in selected set $M$

$$\Phi(M, i) \sim \arg \min_{j \in M} d(j, i)$$
How to Aggregate Preferences of Selected Nodes?

Swapnil Dhamal (IISc)
How to Aggregate Preferences of Selected Nodes?

Swapnil Dhamal (IISc)
Problem Abstraction

- Objective to minimize $\mathbb{E}[f(P) \Delta f(R)]$
- Find a set $M$ that maximizes $h(M) = 1 - \mathbb{E}[f(P) \Delta f(R)]$
- Given $M$, computing $h(M)$ hard for some aggregation rules
- $h(\cdot)$ not monotone and neither submodular nor supermodular even for simple aggregation rules apart from dictatorship
- Finding $M$ hard even within any approximation factor
- Aggregation rule may be needed to be updated frequently

An approach agnostic to the aggregation rule being used
Weak Insensitivity Property

Deviations for all $i \leq \epsilon$

$\Rightarrow f(P) \Delta f(P') \leq \epsilon$
Weak Insensitivity Property

But the only rule to satisfy this property: \textbf{Dictatorship}
Expected Weak Insensitivity Property

Deviation for all $i$ from distribution with mean $\leq \epsilon$

$$\Rightarrow \mathbb{E}[f(P) \Delta f(P')] \leq \epsilon$$
Empirical Satisfaction of Expected Weak Insensitivity under Distribution $\mathcal{D}$, Kendall-Tau Distance, and the Defined $\Delta$

<table>
<thead>
<tr>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plurality</td>
<td>Veto</td>
</tr>
<tr>
<td>Dictatorship</td>
<td>Borda</td>
</tr>
<tr>
<td>Minmax</td>
<td>Kemeny</td>
</tr>
<tr>
<td>Bucklin</td>
<td>Schulze</td>
</tr>
<tr>
<td>Smith set</td>
<td>Copeland</td>
</tr>
</tbody>
</table>

Survey of Voting Rules

---


Maximize minimum expected similarity

\[ \rho(M) = \min_{i \in N} c(M, i) \]

Maximize average expected similarity

\[ \psi(M) = \text{avg}_{i \in N} c(M, i) \]
Again ...  
Solving the abstracted problem is also NP-hard.

But ...  
Objective functions are non-negative, monotone, and submodular.

That means ...

Greedy hill-climbing gives \( (1 - \frac{1}{e}) \approx 0.63 \) approximate optimal solution.  


Until \( |M| = k \), select \( j \in N \setminus M \) that maximizes \( h(M \cup \{j\}) - h(M) \)
Overview

1. Introduction
2. Empirically
3. Formulation
4. Results
5. Conclusion
### Experimental Results

<table>
<thead>
<tr>
<th>Method name</th>
<th>How to select nodes?</th>
<th>How to aggregate?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greedy-min</td>
<td>Greedy hill-climbing maximize $\rho(\cdot)$</td>
<td>$f(Q')$</td>
</tr>
<tr>
<td>Greedy-avg</td>
<td>Greedy hill-climbing maximize $\psi(\cdot)$</td>
<td>$f(Q')$</td>
</tr>
<tr>
<td>Random-poll</td>
<td>Random</td>
<td>$f(Q)$</td>
</tr>
<tr>
<td>Random-rep</td>
<td>Random</td>
<td>$f(Q')$</td>
</tr>
</tbody>
</table>

\[
\rho(M) = \min_{i \in N} c(M, i) \quad \psi(M) = \text{avg} \left( c(M, i) \right)_{i \in N}
\]

$Q$ : Profile containing only preferences of nodes in $M$
$Q'$ : Profile containing weighted preferences of nodes in $M$
Experimental Results

<table>
<thead>
<tr>
<th></th>
<th>Average case</th>
<th>Worst case</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Personal issues</strong></td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
<tr>
<td><strong>Social issues</strong></td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
</tbody>
</table>
Overview

1. Introduction
2. Empirically
3. Formulation
4. Results
5. Conclusion
Future Work

- Conduct survey on a larger scale
- Models for preferences in social networks
- Study the ‘best’ form of modified preference profile
- Study the problem when voters are strategic
- Consider attributes as well as social network for node selection

For details ...  