Scalable Preference Aggregation in Social Networks

IFCAM Workshop on Social Networks
Indian Institute of Science, Bangalore

Y. Narahari
Joint work with Swapnil Dhamal

Game Theory Lab
Department of Computer Science and Automation
Indian Institute of Science, Bangalore

January 16, 2014
Overview

1. Introduction and Motivation
2. A Sample Survey
3. Problem Formulation
4. Experimental Results
5. Conclusions
Overview

1. Introduction and Motivation
2. A Sample Survey
3. Problem Formulation
4. Experimental Results
5. Conclusions
What constitutes a social network?
Individuals and friendships
Homophily in Social Networks

- **What constitutes a social network?**
  Individuals and friendships

- **What causes friendships?**
  Similarity of individuals
Homophily in Social Networks

- **What constitutes a social network?**
  Individuals and friendships

- **What causes friendships?**
  Similarity of individuals

- **What do friendships cause?**
  Individuals become more similar

What is homophily?
A bias in friendships towards similar individuals
Homophily plays a key role in social networks.
Homophily in Social Networks

- **What constitutes a social network?**
  Individuals and friendships

- **What causes friendships?**
  Similarity of individuals

- **What do friendships cause?**
  Individuals become more similar

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

- **What constitutes a social network?**
  Individuals and friendships

- **What causes friendships?**
  Similarity of individuals

- **What do friendships cause?**
  Individuals become more similar

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

Homophily plays a key role in social networks.
Preference Aggregation

- **Agents or Voters** have certain preferences over a set of Alternatives.
Preference of a voter is a complete ranked list of alternatives.
Preference Profile $P$ is a vector of preferences of voters.

Preference Profile $P$ is a vector of preferences of voters.

Preference of voter $i$ is $YZX$.

Preference of voter $j$ is $XYZ$.

Preference of voter $p$ is $YZX$.
Aggregation Rule $f$ outputs an aggregate preference for each preference profile.
Aggregate Preference $f(P)$ summarizes the preferences of the voters.

Preference Profile

Preference of voter $i$

Aggregation Rule

Plurality

Aggregate Preference

Y X Z

Y X Z
Normalized Kendall-Tau Distance

- \( r = \) number of alternatives

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

- Distance between \((X, Y, Z)\) and \((X, Z, Y)\) is \(\frac{1}{3}\)
- Distance between \((X, Y, Z)\) and \((Y, Z, X)\) is \(\frac{2}{3}\)
- Distance between \((X, Y, Z)\) and \((Z, Y, X)\) is 1
Motivation for the Work

- Many situations where we need to obtain a satisfactory aggregate preference given the individual preferences: meetings, committees, voting, poll surveys, product ranking, search engine aggregation, collaborative filtering, etc.
Motivation for the Work

- Many situations where we need to obtain a satisfactory aggregate preference given the individual preferences: meetings, committees, voting, poll surveys, product ranking, search engine aggregation, collaborative filtering, etc.

- For large networks, it is infeasible to gather the preferences from all the voters due to a variety of factors: time, lack of interest of the voters, etc.
Motivation for the Work

- Many situations where we need to obtain a satisfactory aggregate preference given the individual preferences: meetings, committees, voting, poll surveys, product ranking, search engine aggregation, collaborative filtering, etc.

- For large networks, it is infeasible to gather the preferences from all the voters due to a variety of factors: time, lack of interest of the voters, etc.

- Most interesting aggregation rules are computationally intensive

**Estimate** the aggregate preference of the population by selecting a *subset* of voters, taking into account the social network.
Social networks do influence voting in elections 1 2

Social networks do influence voting in elections \(^1\) \(^2\)

Network structure can be ignored in many contexts \(^3\)

---


Social networks do influence voting in elections.\(^1\)\(^2\)

Network structure can be ignored in many contexts.\(^3\)

Opinions are divided

---


Node selection in voting using attributes of nodes and alternatives without taking social network into account \textsuperscript{4}


\textsuperscript{5} N.R. Suri and Y. Narahari. \textit{IEEE - TASE}. 2012

- **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

---


Current Art and Research Gaps (2)

- **Node selection in voting** using attributes of nodes and alternatives without taking social network into account \(^4\)

- Node selection in influence maximization, influence limitation, virus inoculation, etc. **taking social network into account** \(^5\) \(^6\)

**Our interest:** **Node selection in voting taking social network into account**

---


Overview

1. Introduction and Motivation
2. A Sample Survey
3. Problem Formulation
4. Experimental Results
5. Conclusions
Personal issues

- Favorite place to meet
- Favorite recent movie
- Favorite food cuisine
Survey Questions

Personal issues

- Favorite place to meet
- Favorite recent movie
- Favorite food cuisine

Social issues

- Most deserving Test batsman for the vacant spot
- Most deserving Prime Minister
- Most likely Prime Minister
- Most deplorable crime
Observations on the Survey Network

- Somewhat similar rankings by connected nodes
- Very similar rankings by connected nodes belonging to big clusters
- First and last alternatives mostly consistent for connected nodes
- Somewhat similar rankings even by (un)connected nodes for social issues

The social network had higher influence on rankings related to personal issues than social issues.
Observations on the Survey Network

- Somewhat similar rankings by connected nodes
- Very similar rankings by connected nodes belonging to big clusters
- First and last alternatives mostly consistent for connected nodes
- Somewhat similar rankings even by (un)connected nodes for social issues

The social network had higher influence on rankings related to personal issues than social issues
Distribution of Distance

- Histogram for different questions for a given pair fit by truncated Gaussian distribution having range \([0, 1]\)
- Considered a discrete version of the truncated Gaussian distribution, \(D\)

Distance between \(i\) and \(j\) followed distribution \(D\) with mean \(d(i, j)\)

\[c(\cdot, \cdot) = 1 - d(\cdot, \cdot)\]
Overview

1. Introduction and Motivation
2. A Sample Survey
3. Problem Formulation
4. Experimental Results
5. Conclusions
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 to the aggregate preference of $N$ using $f$. 

---

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 to the aggregate preference of $N$ using $f$. 
Starting to Solve the Problem

Distance between set $M \subseteq N$ and node $i \in N$

$$d(M, i) = \min_{j \in M} d(j, i)$$

Representative of node $i$ in set $M$

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

\[ f(P) \]  
\[ f(R) \]

Y. Narahari (IISc)  Scalable Preference Aggregation in Social Networks  January 16, 2014  13 / 24
How to Aggregate Preferences of Selected Nodes?

\[ f(P) \]

\[ f(Q) \]
How to Aggregate Preferences of Selected Nodes?

In the context of social networks, we consider nodes with preferences that can be aggregated. Let's denote the set of nodes by $P$ and $Q'$, and the function that aggregates preferences by $f$. The nodes $i$, $j$, $s$, and $t$ represent different users or entities with preferences.

- For node $i$, the aggregated preference is denoted as $f(P)$.
- For node $j$, the aggregated preference is denoted as $f(Q')$.

The diagram illustrates the aggregation process, where $f$ is applied to each node's preferences to yield the aggregated preferences $f(P)$ and $f(Q')$. This process is crucial for understanding how preferences can be combined in a scalable manner within social networks.
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 to the aggregate preference of $N$ using $f$. 
What is ‘Close Enough’?

For any $y \in f(R)$, distance $= \min_{x \in f(P)} \delta(x, y)$

$y \sim U_{f(R)}$, $f(P) \Delta f(R) = E_{y \sim U_{f(R)}} \left[ \min_{x \in f(P)} \delta(x, y) \right]$

Our objective is to minimize $E[f(P) \Delta f(R)]$
For any \( y \in f(R) \), distance = \( \min_{x \in f(P)} \delta(x, y) \)
What is ‘Close Enough’?

For any $y \in f(R)$, distance = $\min_{x \in f(P)} \delta(x, y)$
What is ‘Close Enough’?

For any $y \in f(R)$, distance $= \min_{x \in f(P)} \delta(x, y)$

$y \in f(R)$ u.a.r., $f(P) \Delta f(R) = \mathbb{E}_{y \sim u f(R)} \left[ \min_{x \in f(P)} \delta(x, y) \right]$
What is ‘Close Enough’?

For any \( y \in f(R) \), distance = \( \min_{x \in f(P)} \delta(x, y) \)

\( y \in f(R) \) u.a.r., \( f(P) \Delta f(R) = \mathbb{E}_{y \sim u f(R)} \left[ \min_{x \in f(P)} \delta(x, y) \right] \)

Our objective is to minimize \( \mathbb{E}[f(P) \Delta f(R)] \)
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 to the aggregate preference of $N$ using $f$. 
Issues in Solving this Problem

- Find a set $M$ of size $k$ that maximizes
  $h(M) = 1 - \mathbb{E}[f(P) \Delta f(R)]$

- Given $M$, computing $h(M)$ hard for many aggregation rules

- $h(\cdot)$ not monotone and neither submodular nor supermodular even for simple aggregation rules apart from dictatorship

- Aggregation rule may be needed to be changed frequently (to tackle strategic users)

An approach agnostic to the aggregation rule

$$\rho(M) = \min_{i \in N} c(M, i) \quad \psi(M) = \sum_{i \in N} c(M, i)$$
Weak Insensitivity Property

Deviations for all \( i \leq \epsilon \)

\[ \Rightarrow f(P) \Delta f(P') \leq \epsilon \]
Only Dictatorship seems to satisfy this property
Expected Weak Insensitivity Property

\[ \mathbb{E}[f(P) \Delta f(P')] \leq \varepsilon \]

Deviations for all \( i \) from distribution with mean \( \leq \varepsilon \)
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

---


Objective Functions in the New Problem

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 new problem is also **NP-hard**.
Again ...

Solving the new problem is also **NP-hard**.

However ...

Objective functions are non-negative, monotone, and submodular.
What Next?

Again ...
Solving the new problem is also NP-hard.

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

That means ...
Greedy hill-climbing gives \((1 - \frac{1}{e})\) approximate optimal solution.\(^a\)


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

1. Introduction and Motivation
2. A Sample Survey
3. Problem Formulation
4. Experimental Results
5. Conclusions
## 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>

$$\rho(M) = \min_{i \in N} c(M, i) \quad \psi(M) = \sum_{i \in N} c(M, i)$$

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

<table>
<thead>
<tr>
<th>Personal issues</th>
<th>Average case</th>
<th>Worst case</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="average_personal_issues.png" alt="Graph" /></td>
<td><img src="worst_personal_issues.png" alt="Graph" /></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social issues</th>
<th>Average case</th>
<th>Worst case</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="average_social_issues.png" alt="Graph" /></td>
<td><img src="worst_social_issues.png" alt="Graph" /></td>
<td></td>
</tr>
</tbody>
</table>
Future Work

- Explore other forms of modified preference profile $R$ given $P$
- Conduct a survey on a larger scale
- Study the problem when agents are strategic
<table>
<thead>
<tr>
<th>T</th>
<th>U</th>
<th>H</th>
<th>O</th>
<th>K</th>
<th>N</th>
<th>Y</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>A</td>
<td>T</td>
<td>N</td>
<td>U</td>
<td>O</td>
<td>K</td>
<td>Y</td>
</tr>
<tr>
<td>N</td>
<td>K</td>
<td>T</td>
<td>U</td>
<td>H</td>
<td>Y</td>
<td>A</td>
<td>O</td>
</tr>
<tr>
<td>T</td>
<td>H</td>
<td>A</td>
<td>N</td>
<td>K</td>
<td>Y</td>
<td>U</td>
<td>O</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>T</td>
<td>N</td>
<td>H</td>
<td>Y</td>
<td>U</td>
<td>K</td>
</tr>
<tr>
<td>U</td>
<td>O</td>
<td>Y</td>
<td>K</td>
<td>N</td>
<td>A</td>
<td>H</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>N</td>
<td>A</td>
<td>K</td>
<td>H</td>
<td>Y</td>
<td>U</td>
<td>O</td>
</tr>
<tr>
<td>T</td>
<td>U</td>
<td>H</td>
<td>O</td>
<td>K</td>
<td>N</td>
<td>Y</td>
<td>A</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>T</td>
<td>N</td>
<td>U</td>
<td>O</td>
<td>K</td>
<td>Y</td>
</tr>
<tr>
<td>N</td>
<td>K</td>
<td>T</td>
<td>U</td>
<td>H</td>
<td>Y</td>
<td>A</td>
<td>O</td>
</tr>
<tr>
<td>T</td>
<td>H</td>
<td>A</td>
<td>N</td>
<td>K</td>
<td>Y</td>
<td>U</td>
<td>O</td>
</tr>
<tr>
<td>A</td>
<td>O</td>
<td>T</td>
<td>N</td>
<td>H</td>
<td>Y</td>
<td>U</td>
<td>K</td>
</tr>
<tr>
<td>U</td>
<td>O</td>
<td>Y</td>
<td>K</td>
<td>N</td>
<td>A</td>
<td>H</td>
<td>T</td>
</tr>
</tbody>
</table>

---

Scalable Preference Aggregation in Social Networks

January 16, 2014 23 / 24
Modeling Homophily for Unconnected Nodes

Initializations

- \( d(i, j) \) known for connected pairs \( \{i, j\} \) [0 for \( i = j \)]
- \( d(i, j) = 1 \) for all unconnected pairs

![Diagram](image-url)

\( d(p, j) \)
\( d(p, i) \)

\( \text{All pairs shortest path with update rule} \)

\[
\text{if } d(p, i) + r d(p, j) < d(i, j) \text{ then } d(i, j) = d(p, i) + r d(p, j)
\]
Modeling Homophily for Unconnected Nodes

Initializations

- $d(i, j)$ known for connected pairs $\{i, j\}$ [0 for $i = j$]
- $d(i, j) = 1$ for all unconnected pairs

All pairs shortest path with update rule

\[
\text{if } d(p, i) \oplus_r d(p, j) < d(i, j) \text{ then } d(i, j) = d(p, i) \oplus_r d(p, j)
\]