Your friends are better than you:
Friendship Paradox and its social consequences

Kristina Lerman
USC Information Sciences Institute

http://www.isi.edu/~lerman
Networks shape behavior and perceptions
Networks shape behavior and perceptions

1. Networks shape behavior: form the substrate for social interactions and information flow.

2. Your friends are a small subset of the population.

3. Friends are not a random sample of the population.

4. This distorts your perceptions. (A rare trait can appear very popular)
What color is more popular?
What color is more popular?

Most nodes think blue is popular

Nodes think yellow is more Blue is not especially popular
Local vs global views

100% of this person’s friends think baseball caps are trendy.

K. Schaul, “A quick puzzle to tell whether you know what people are thinking”, Wonkblog, Washington Post
https://www.washingtonpost.com/graphics/business/wonkblog/majority-illusion/
Networks distort local information

- Networks can systematically bias individual perceptions of what is common among peers
  - Example: College students overestimate peers’ alcohol use

How many alcoholic drinks are consumed at a party

Source: Most Students Do PartySafe@Cal
Outline

Friendship paradox
- The many friendship paradoxes in networks
- Origins of paradoxes: a network science view

Perception bias
- Friendship paradox in directed networks
- ... biases perceptions of popularity
- Twitter case study: Some hashtags appear more popular than they are

Polling
- Estimate global popularity with perception bias
- ... with a limited budget
Lots of paradoxes!

Friendship paradox
You friends have more friends than you do, on average [Feld, 1991]
Lots of paradoxes!

**Friendship paradox**
You friends have more friends than you do, on average [Feld, 1991]

**Generalized friendship paradox**
You friends are more $X$ than you are, on average [Hodas et al., 2013, Eom & Jo, 2014]
Lots of paradoxes!

**Friendship paradox**
You friends have more friends than you do, on average [Feld, 1991]

**Strong friendship paradox**
Most of your friends have more friends than you do [Kooti et al., 2014]

**Generalized friendship paradox**
You friends are more $X$ than you are, on average [Hodas et al., 2013, Eom & Jo, 2014]
Lots of paradoxes!

**Friendship paradox**
You friends have more friends than you do, on average [Feld, 1991]

**Strong friendship paradox**
Most of your friends have more friends than you do [Kooti et al., 2014]

**Generalized friendship paradox**
You friends are more $X$ than you are, on average [Hodas et al., 2013, Eom & Jo, 2014]

**Generalized strong friendship**
Most of your friends are more $X$ than you are [Kooti et al, 2014]

**Majority illusion**
Most of your friends have a trait, even when it is rare. [Lerman et al, 2016]
How common is strong friendship paradox?

Very common... almost everyone observes that most of their friends are more popular

<table>
<thead>
<tr>
<th>Network</th>
<th>Type</th>
<th>Nodes</th>
<th>Probability of paradox</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiveJournal</td>
<td>Social</td>
<td>3,997,962</td>
<td>84</td>
</tr>
<tr>
<td>Twitter</td>
<td>Social</td>
<td>780,000</td>
<td>98</td>
</tr>
<tr>
<td>Skitter</td>
<td>Internet</td>
<td>1,696,415</td>
<td>89</td>
</tr>
<tr>
<td>Google</td>
<td>Hyperlink</td>
<td>875,713</td>
<td>77</td>
</tr>
<tr>
<td>ProsperLoan</td>
<td>Social Finance</td>
<td>89,269</td>
<td>88</td>
</tr>
<tr>
<td>ArXiv</td>
<td>Citation</td>
<td>34,546</td>
<td>79</td>
</tr>
<tr>
<td>WordNet</td>
<td>Semantic</td>
<td>146,005</td>
<td>75</td>
</tr>
</tbody>
</table>
But wait, there is more

**Activity**
You post less than most of your friends

**Diversity**
You see less diverse content than most of your friends

**Virality**
You see less viral content than most of your friends

---

**Why?**

**Friendship paradox**
You friends have more friends than you do, on average [Feld, 1991]

**Strong friendship paradox**
Most of your friends have more friends than you do [Kooti et al., 2014]

**Generalized friendship paradox**
You friends are more $X$ than you are, on average [Hodas et al., 2013, Eom & Jo, 2014]

**Generalized strong friendship**
Most of your friends are more $X$ than you are [Kooti et al, 2014]

**Majority illusion**
Most of your friends have a trait, even when it is rare. [Lerman et al, 2016]
Different explanations

**Friendship paradox**
You friends have more friends than you do, on average [Feld, 1991]

**Generalized friendship paradox**
You friends are more X than you are, on average [Hodas et al., 2013, Eom & Jo, 2014]

**Strong friendship paradox**
Most of your friends have more friends than you do [Kooti et al., 2014]

**Generalized strong friendship**
Most of your friends are more X than you are [Kooti et al., 2014]

**Majority illusion**
Most of your friends have a trait, even when it is rare. [Lerman et al, 2016]
Friendship paradox as a byproduct of sampling from a heterogeneous distribution

![Graph showing the distribution of posted tweets and the comparison between your value and your friends' values.](image)
Friendship paradox as a byproduct of sampling from a heterogeneous distribution

\[
\text{mean} = \frac{1}{k} (x_1 + \cdots + x_k) > x \quad \text{Sample mean grows with } k
\]
Strong friendship paradox is a network effect

- To explain strong friendship paradox, need to account for network structure
- Building blocks of network structure
  - $dK$ series framework represents network structure as the joint degree distribution of subgraphs of up to $d$ nodes


**First-order structure (1K)**
- Node degree distribution

**Second-order structure (2K)**
- Pair degree distribution

**Third-order structure (3K)**
- Triplet degree distribution
First-order (1K) structure

- Node degree distribution $p(k)$
  - Probability that a randomly selected node has degree $k$.
- Any heterogeneous degree distribution (variance > 0) will lead to a (weak) friendship paradox
First-order (1K) structure

- Neighbor degree distribution $q(k) \sim kp(k)$
- Probability that a randomly selected neighbor has degree $k$.
- Any heterogeneous degree distribution (variance $> 0$) will lead to a (weak) friendship paradox.
Digg social network
Second-order (2K) structure

- Nodes do not link at random
- Joint degree distribution of connected pairs of nodes $e(k, k')$
  - Probability that a randomly selected edge links nodes with degrees $k$ and $k'$.
- Degree assortativity $r_{2k}$
• ... nodes do not link to random neighbors
• **Neighbor assortativity**: neighbors tend to have similar (or dissimilar) degrees, \( r_{3k} \)
• Networks can have the same 1K and 2K structure but different 3K structure
  • Wu, Percus & Lerman, Neighbor Degree Assortativity in Networks, *in preparation*
Real-world networks have third-order structure

Degree correlations among node’s neighbors in real-world networks are often large
Third-order structure enhances paradoxes

Neighbors’ degrees are not correlated*

Neighbors’ degrees are correlated*

*same 1K and 2K structure
Third-order structure enhances paradoxes

Neighbors’ degrees are not correlated

Neighbors’ degrees are correlated

*same 1K and 2K structure
Fraction of degree-$k$ nodes experiencing the paradox in real-world networks • ; predictions of the $2K$ model · · · and the $3K$ model —.

From friendship paradox to “majority Illusion”

Nodes have a binary trait: active/not, yellow/blue, heavy drinker/teetotaler, ...

Many think that blue is common

Blue does not appear common

Network structure amplifies majority illusion

More nodes will think that blue is very common when:

• Higher degree nodes are more likely to be blue: degree-trait (k-x) correlation
• High degree nodes link to low degree nodes: degree disassortativity \((r_{2k}<0)\)
• Neighbors tend to have similar degree: neighbor assortativity \((r_{3k}>0)\)

Network structure amplifies majority illusion

Fraction of nodes experiencing the majority illusion in a synthetic network with 0.5% active nodes;

\[ \alpha = -2.1, \ n = 100,000, \ k_{max} = 876, \ p_x = 0.005 \]

Friendship paradox in directed networks

friends

in-degree $d_i$

user

followers

out-degree $d_o$
Friendship paradox in directed networks

friends

friends-of-friends

followers-of-friends

user

followers

friends-of-followers

followers-of-followers
Probability a node experiences a paradox

friends have more followers

followers have more friends

friends have more friends

followers have more followers
Local perception bias

- Popularity of a (binary) attribute: probability a random node $v$ has value $f(v)$
  \[ E\{f(v)\} \]

- Local perception of node $v$ about popularity of an attribute $f$ is the fraction of her friends with attribute
  \[ q(v) = \frac{\sum_{u \in \text{friends}(v)} f(u)}{d_i(v)} \]

- Local perception bias: nodes perceive the attribute $f$ to more popular than it actually is
  \[ E\{q(v)\} \geq E\{f(v)\} \]
Global popularity vs local perception on Twitter

Twitter data
- **Time period**
  - Summer 2014
- **Network**
  - 5K users + tweets
  - Their 600K friends + tweets
- **Hashtags**
  - 18M hashtags
  - Focus on 1K most popular hashtags, used by >1K people

Compare perceived popularity of hashtags to their actual popularity
Conditions for local perception bias

Local bias exists if:

- Higher out-degree (high influence) tend to have the attribute.
- Lower in-degree nodes (high attention) tend to follow nodes with attribute.

Theorem: \((E[q(v)] \geq E[f(v)])\) if

\[
\text{Cov}\{f(X), d_o(X)\} \geq 0 \quad \text{and},
\]

\[
\text{Cov}\{f(U), \mathcal{A}(V)|(U, V) \sim \text{Uniform}(E)\} \geq 0.
\]
Polling
What is the right question to ask in a poll?

Estimate the true prevalence of an attribute through polls

- Estimate the fraction of liberals vs conservatives, heavy drinkers vs teetotalers, people who used a hashtag vs not, ...
- ...with limited budget $b$

Polling:

1. Intent Polling (IP): $[b$ random nodes$]$ Will you vote for X?
3. Follower Perception Polling (FPP): $[b$ random followers$]$ What fraction of your friends will vote for X?

- aggregates perceptions of more people
Bias & variance of FPP

- **Bias** of the polling estimate (error) $T$
  \[
  \text{Bias}(T) = \mathbb{E}\{T\} - \mathbb{E}\{f(X)\} = \frac{\text{Cov}(f(X), d_o(X))}{\hat{d}}
  \]

- **Variance** is bounded by $\lambda_2$, second largest eigenvalue of the symmetrized adjacency matrix of the network

- **Mean squared error** of the polling estimate $T$:
  \[
  \text{MSE}\{T\} = \mathbb{E}\left\{[T - \mathbb{E}(f(X))]^2\right\} = \text{Bias}\{T\}^2 + \text{Var}\{T\}
  \]
FPP polling algorithm is more efficient

When used to estimate the true popularity of Twitter hashtags, FPP has lower variance and MSE. For a given budget, i.e., number of nodes sampled, it outperforms other polling methods on many hashtags.
To summarize

• Network structure can systematically bias local perceptions
  • Making a rare attribute appear far more common than it is, under some conditions

• Open questions: What is the impact of network bias on
  • Collective dynamics in networks, e.g., contagious outbreaks
  • Network control and intervention
  • Psychological well-being
    • Your friends are happier than you are (Bollen et al. 2016)
    • Your co-authors are more prestigious than you are (Eom & Jo 2014)
    • Social comparison theory
THANK YOU!

Sponsors

NSF: CIF-1217605
ARO: W911NF-16-1-0306

Questions?
lerman@isi.edu