Social Network Dynamics in a Massive Online Game

Aaron Clauset
Assistant Professor, Computer Science and BioFrontiers Institute, University of Colorado Boulder
External Faculty, Santa Fe Institute
joint work

Sears Merritt (Colorado)

Abigail Z Jacobs (Colorado)

Winter Mason (Stevens Inst. Tech.)

funded in part by

James S. McDonnell Foundation

University of Colorado Boulder

DARPA

AFOSR
talk outline

- sources of dynamic social network data
  - online games and social networks
  - massive online game
  - inferring friendships from interactions
  - social networks and performance
  - outlook
data sources for social network dynamics

- Twitter, Facebook, Google+, Pinterest, etc.
- Academic coauthorships & citations
- World Wide Web
- etc.

but...

- links often have low or no cost = unrealistic
- domain functionality can drive social dynamics
- few sources capture “real” social networks (face-to-face time)
talk outline

• sources of dynamic social network data
• online games and social networks
  • massive online game
  • inferring friendships from interactions
• social networks and performance
• outlook
online games

some basic statistics

• 100+ million Americans play online games
• most prefer to play with friends
• broad age distribution (mean = 41)
• 1000s of games, diverse types

rich variety of social interactions
enormous volumes of detailed data
edge weights + dynamics
node attributes + dynamics
largely unexplored in network science
online game social networks

- nodes identified by online pseudonyms
  unique across game / platform & tied to one person (generally)

- edges = online interactions
  interactions = costly
  shared activity, repeated

- nodes attributes
  demographics, online activity, performance, etc.

- edges attributes
  weights, time, character, etc.
talk outline

• sources of dynamic social network data
• online games and social networks
• **massive online game**
  • inferring friendships from interactions
• social networks and performance
• outlook
**Halo: Reach (Bungie, 2010)**

- played online via XBox Live platform
- team combat simulation (FPS)
- 20TB of game data, spanning
  - 18 months of time
  - 17+ million players
  - 1 billion competitions
  - 70% are team competitions
- complex spatial environments
- complex social interactions

*a massive online game*
how it works

• join “party” (of 0-3 friends)

• choose game type and subtype (“competitive / team 4v4”)

• Xbox Live places parties into matches (matchmaking)

• play! (for roughly 10 minutes)

• repeat
what it looks like
what it looks like
a small problem

- we observe interactions not friendships
- interactions = matchmaking + friendships
- no demographic information
a small solution

• anonymous web survey
• 847 participants
• demographic questions
  age, sex, location, education
• psychometric questions
  attitudes, play style, etc.
• friendship survey
  • 14,405 labeled friends
  • 7,159,989 labeled non-friends
talk outline

• sources of dynamic social network data
• online games and social networks
• massive online game
• **inferring friendships from interactions**
• social networks and performance
• outlook
we can observe a sequence of pairwise interactions
\[ \sigma_{ij} = (i, j, t_1), (i, j, t_2), \ldots \]

- can we robustly distinguish friendships from non-friendships?
- this is a general problem for interaction networks

**problems:**
- volume of data varies widely by individual = heavy-tailed distribution in \(|\sigma_{ij}|\)
- friendships are sparse in large networks
- “ground truth” data hard to obtain
what is a friendship?

social interactions:

- friendship = periodic + prosocial interactions
  - diurnal cycle modulates all interactions

recovering latent friendship ties
- supervised learning
- define 9 statistical features
  - which do well?

Merrit, Jacobs, Mason and Clauset, ICWSM 2013
## Statistics to Detect Friendships

### Features of Interaction Time Series:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Autocorrelation</td>
<td>( AC_{x,y} )</td>
</tr>
<tr>
<td>2. Pair Volume</td>
<td>( N_{x,y} )</td>
</tr>
<tr>
<td>3. Fraction of Interactions</td>
<td>( N_{x,y}/N_x )</td>
</tr>
<tr>
<td>4. Schedule Entropy</td>
<td>( H_s(x,y) )</td>
</tr>
<tr>
<td>5. Location Entropy</td>
<td>( H_t(x,y) )</td>
</tr>
<tr>
<td>6. Loc.-Sched. Entropy</td>
<td>( H_{t,s}(x,y) )</td>
</tr>
<tr>
<td>7. Betrayals</td>
<td>( B_{x,y} )</td>
</tr>
<tr>
<td>8. Assistance</td>
<td>( A_{x,y} )</td>
</tr>
<tr>
<td>9. Indirect Assistance</td>
<td>( V_{x,y} )</td>
</tr>
</tbody>
</table>

### Features Grouping:

- **Temporal Features**
  - Autocorrelation
  - Pair Volume
  - Fraction of Interactions

- **Entropy Features**
  - Schedule Entropy
  - Location Entropy
  - Loc.-Sched. Entropy

- **Prosocial Features**
  - Betrayals
  - Assistance
  - Indirect Assistance

---

Merrit, Jacobs, Mason and Clauset, ICWSM 2013
exploring the feature space

classification tree

• 50/50 training/test by survey participant
• cross-validation to control tree size
• highly compact trees, high AUCs (often >0.9)
• key feature is autocorrelation $AC_{x,y}$
  ➤ friendships look like periodic + prosocial interactions

Merrit, Jacobs, Mason and Clauset, ICWSM 2013
lightweight predictors

logistic regression with individual features

- single-feature predictors scale up better on real systems (Facebook, etc.)
- ROC curves
- autocorrelation $AC_{x,y}$ and direct assistance $A_{x,y}$ both highly accurate: AUC > 0.98

Merritt, Jacobs, Mason and Clauset, ICWSM 2013
predictions for low-volume individuals

most people have “shallow” histories

- 90% have less than 200 games
- most users are “casual”
- true for most online social systems
- do predictions fail on these individuals?

Merrit, Jacobs, Mason and Clauset, ICWSM 2013
predictions for low-volume individuals

most people have “shallow” histories

- AUC vs. size of history $N_x$
- periodic + prosocial interactions highly robust and efficient
- total interaction count not good, but not efficient

Merrit, Jacobs, Mason and Clauset, ICWSM 2013
recovering friendships from interactions

some comments:

friendships easy to recover from interactions
results likely to generalize [see Jones et al. PLoS ONE (2013)]
clarifies “friendship” = periodic + prosocial interactions
players structure their behavior to enable friend-friend interactions
raises significant privacy concerns
talk outline

- sources of dynamic social network data
- online games and social networks
- massive online game
- inferring friendships from interactions
- social networks and performance
- outlook
extracting friendship network from interaction network

• ideally, use generative model
• for now, a threshold: friendship if $AC_{x,y} \geq t_c$
• choose threshold $t_c$ by matching sampled with recovered degree distribution
• but, survey is a biased sample and, sampling bias is unknown
• do we match head or tail?
• try both

Social network of a massive online game

Survey data only

Fraction of vertices with degree at least $k$

Degree, $k$

Matches head

Matches tail
social network of a massive online game

- choose threshold $t_c$ by matching sampled with recovered degree distribution
- but, survey is a biased sample and, sampling bias is unknown
- do we match head or tail?
- try both
  - inferred degree distributions
    - no power laws (shocking!)
  - mean degree = 2.4-3.8
component sizes

- 17M people in interaction graph
- 4.7-8.4M in friendship network
- largest component is 11-31% of people
local structure

• vertex-level correlation coefficient

• many near-cliques
  well-defined groups of friends

• many star graphs
  socialites?

• roughly similar to other online
  social networks
a functional role for friendship?

- does friendship impact individual or team performance?
impact of friendship on performance

• among survey respondents
• individual behavior vs. number of friends on team
• you perform better & nicer when you collaborate with friends

Mason and Clauset, CSCW 2013
impact of friendship on performance

- among survey respondents
- team performance vs. number of friends on team
- team performs better the more friendships it contains
talk outline

• sources of dynamic social network data
• online games and social networks
• massive online game
• inferring friendships from interactions
• social networks and performance
• outlook
• social networks: interactions or friendships?
  • interactions highly system dependent
    driven by interface, user goals, online context, etc.
  • friendships are more general
    periodic + prosocial interactions

• user labeled data is crucial
  but always a biased population sample. no panacea

• general procedure:
  1. collect interaction data (big, low fidelity)
  2. get user labeled data (small, biased, high fidelity)
  3. model friendships from interactions (supervised)
  4. extract underlying social network, dynamics
outlook

• online games novel window on human social dynamics

• Halo network is big, detailed, dynamic
  • what large-scale structure? communities?
  • what large-scale temporal patterns?
  • generative models?
  • differences / similarities with other social networks?
  • coupling of performance and friendships?

• interactions ≠ friendships

• friendships shape individual and team performance

fin
web survey
web survey

![Graph showing age distribution with density on the y-axis and age on the x-axis.](image)

![Bar chart showing antisocial behavior with age brackets min-18, 19-23, 24-max.](image)
web survey

population

survey participants

population