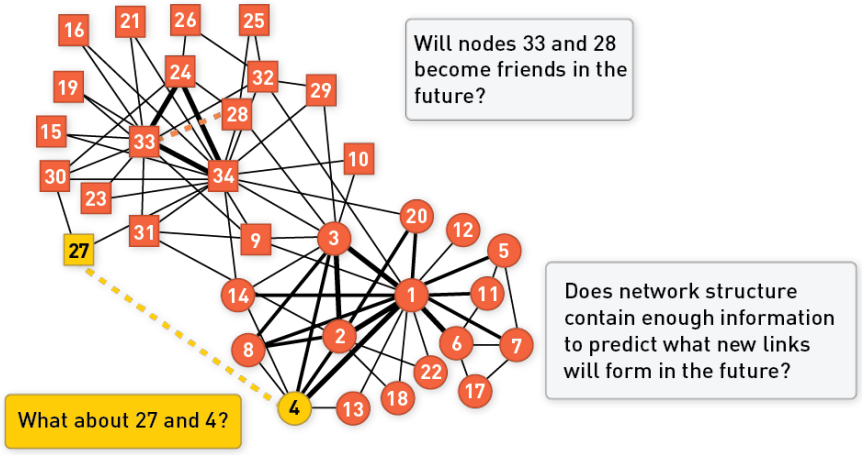
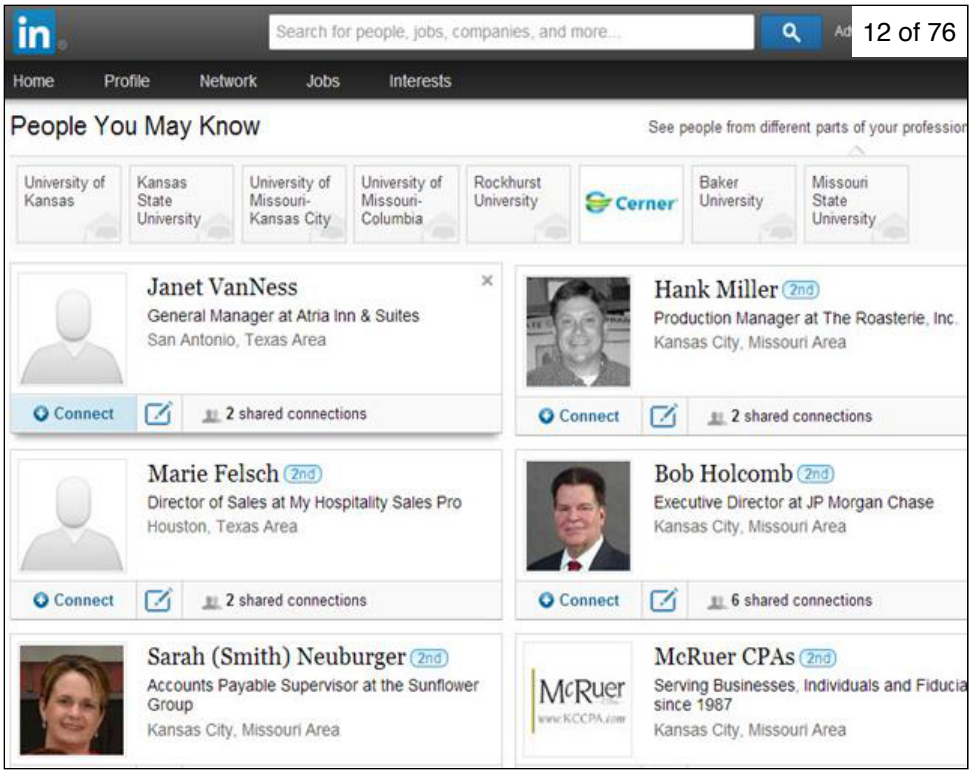
Social Networks and Link Prediction

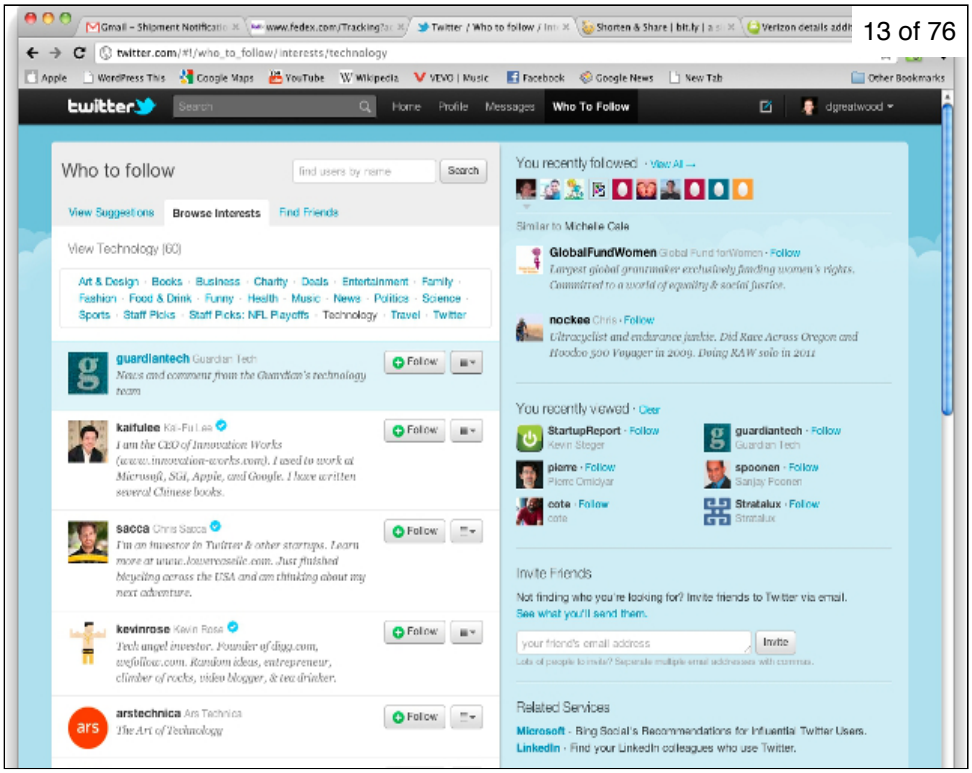
1. Simple heuristic based on friends of friends (FoF)
2. Based on machine learning ranker
3. Hybrid model based on a supervised random walk

- A snapshot of a social network is used to suggest new friends and entities that should link.
- Social network sites such as LinkedIn, Google+, and Facebook use a friends suggestion algorithm to help users broaden their networks.

Link Prediction




Search for people, jobs, companies, and more...
Home
Profile
Network
Jobs
Interests
People You May Know
See people from different parts of your professior
University of Kansas
Kansas State University
University of Missouri-Kansas City
University of Missouri-Columbia
Rockhurst University
Cerner
Baker University
Missouri State University
Janet VanNess
General Manager at Atria Inn & Suites
San Antonio, Texas Area
Connect
2 shared connections
Hank Miller 2nd
Production Manager at The Roasterie, Inc.
Kansas City, Missouri Area
Connect
2 shared connections
Marie Felsch 2nd
Director of Sales at My Hospitality Sales Pro
Houston, Texas Area
Connect
2 shared connections
Bob Holcomb 2nd
Executive Director at JP Morgan Chase
Kansas City, Missouri Area
Connect
6 shared connections
Sarah (Smith) Neuburger 2nd
Accounts Payable Supervisor at the Sunflower Group
Kansas City, Missouri Area
McRuer CPAs 2nd
Serving Businesses, Individuals and Fiducia since 1987
Kansas City, Missouri Area

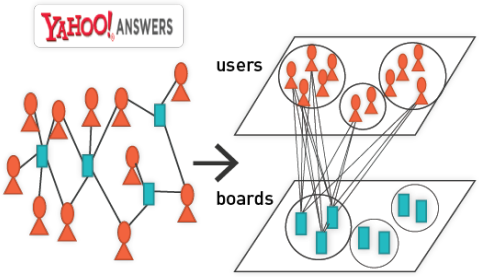

twitter.com/#!/who_to_follow/interests/technology
twitter
Home
Profile
Messages
Who To Follow
Who to follow
Search
View Suggestions
Browse Interests
Find Friends
View Technology (60)
Art & Design · Books · Business · Charity · Deals · Entertainment · Family · Fashion · Food & Drink · Funny · Health · Music · News · Politics · Science · Sports · Staff Picks · Staff Picks: NFL Playoffs · Technology · Travel · Twitter
guardiantech
kaifulee
sacca
kevinrose
arstechnica
You recently followed
Similar to Michele Cala
GlobalFundWomen
nockee
You recently viewed
StartupReport
guardiantech
pierre
spoonen
cote
Stratalux
Invite Friends
Not finding who you're looking for? Invite friends to Twitter via email.
See what you'll send them.
Invite
Related Services
Microsoft · Bing Social's Recommendations for Influential Twitter Users.
LinkedIn · Find your LinkedIn colleagues who use Twitter.

Link Prediction Methods

In order for the proximity measures to make sense while estimating similarity among vertices, we will need to modify these measures.

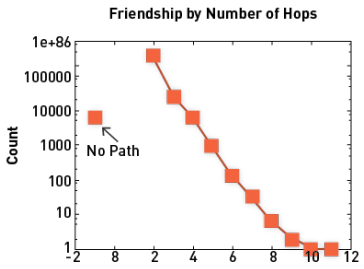
We will consider such proximity measures under three different categories:

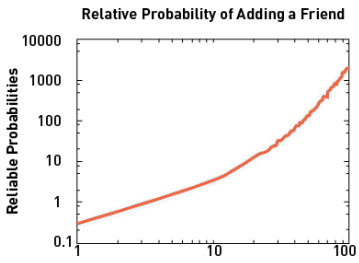
- Node-Neighborhood-Based Methods
  - Common neighbors
  - Jaccard's coefficient
  - Adamic-Adar
- All-Paths-Based Methodologies
  - PageRank
  - SimRank
- Higher-Level Approaches
  - Unseen bigrams
  - Clustering



Friend Recommendation

- Learn to recommend potential friends
- Facebook link creation (L. Backstorm 2011)
  - 92% of new friendships on FB are **friend of a friend**
    - Triadic closure (Granovetter 1973)
  - More **common friends** helps:
    - Social capital (Coleman, 1988)





Friends Suggestions: Facebook

Friends Suggestions

- Approximately 1.5 billion people on Facebook's network
- Each profile has about 150 friends
- That works out to about 20,000 FoFs
- We want to suggest good links, rather than overload the user with suggestions

Heuristics Based on FoF

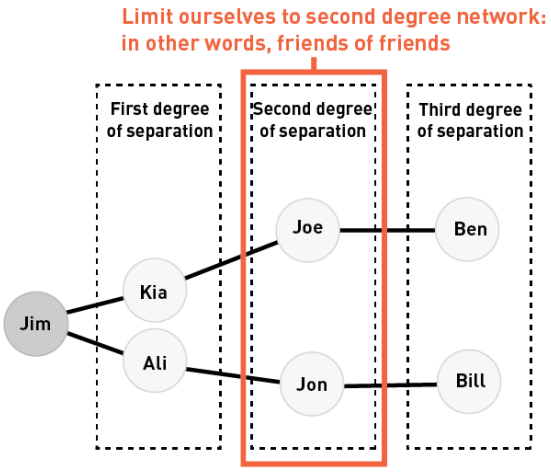


Figure 7.12
An example of FOF where Joe and Jon are considered FoFs to Jim

Algorithm Sketch: Most FoF Will Not Be Good Suggestions

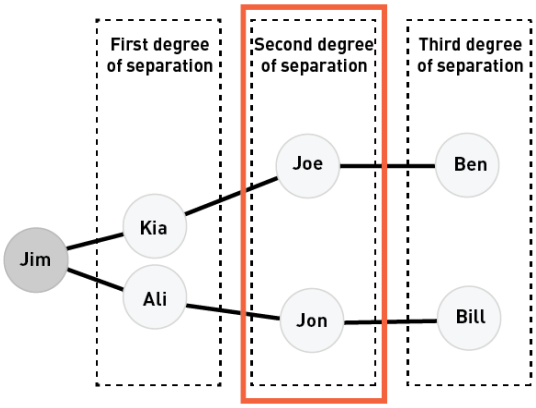


Figure 7.12
An example of FOF where Joe and Jon are considered FoFs to Jim

- FoF is a list of individuals who are indirectly connected to you through friends.
  - Not everybody in this list will be familiar to you.
  - E.g., I lived in Japan for five years, so just because you know me does not mean you know my friends in Japan.

FoF Ranking Algorithm (cont.)

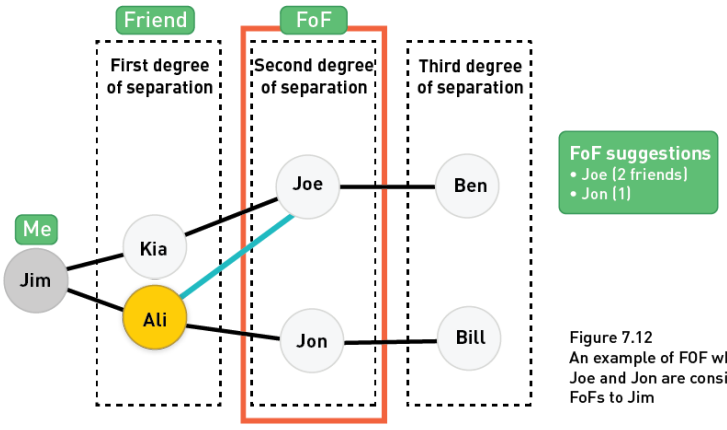


Figure 7.12
An example of FOF where Joe and Jon are considered FoFs to Jim

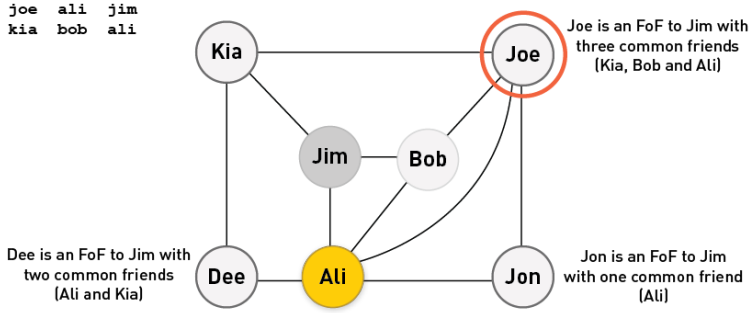
- For each FoF:
  - Determine the number of common friends.
  - Sort suggestions in decreasing order of count.

Implement the FoF Algorithm in MapReduce

- Two MapReduce jobs are required to calculate the FoFs for each user in a social network.
- Job 1: Produce a list of FoFs and number of mutual friends.
  - Job calculates the common friends for each user.
- Job 2: Sort list of FoF suggestions.
  - The second job sorts the common friends by the number of connections to your friends.

Friend Graph as an Adjacency List

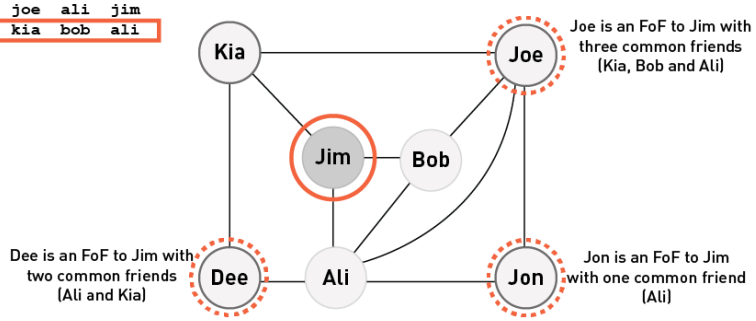
```
$ cat test-data/ch7/friends.txt
joe  jon  kia  bob  ali
kia  joe  jim  dee
dee  kia  ali
ali  dee  jim  bob  joe  jon
jon  joe  ali
bob  joe  ali  jim
jim  kia  bob  ali
```



Generate a List of FoFs for Jim

```
$ cat test-data/ch7/friends.txt
joe  jon  kia  bob  ali
kia  joe  jim  dee
dee  kia  ali
ali  dee  jim  bob  joe  jon
jon  joe  ali
bob  joe  ali  jim
jim  kia  bob  ali
```

Do this in Hadoop
Do this in Spark



Summary on Friend Suggestion Algorithm: FoFs

- First cut at suggesting new connections on social networks
  - Can we do better?
- Limited our exploration new connections to friends of friends
  - Other sources of new connections
  - E.g., both attended the same high school and graduated the same year; worked in the same 50-person company

Algorithm for Ranking 22.5 K FoFs

Link Suggestion Version #2: Use machine learning

22500 fof = 150*150

Two-stage system

- Step 1: Rank based on FoFs based on mutual friends (and possible other criteria such as hometown, high school, company, university).
- Step 2: Score each candidate suggestion and rerank using, say, a logistic regression model.
  - Select top N (say 1000) from step 1 and rescore using a machine learning logistic regression model.
  - Build a friend-connection model.
    - Build models at different levels:
      - Global model, local to a country, local to a type of person

ML Connection Model

- Goal: Expand a network for an individual (or a group)
- Data collection
- Feature engineering
- Modeling using logistic regression or gradient-boosted decision trees (binomial logistic version)
- Evaluation in the lab (using a held-out test set)
- AB test in the wild

Friend-Link Suggestions in Social Graphs

1. Simple heuristic based on friends-of-friends (FoFs)
2. Based on machine learning ranker
3. Next section:
    - Hybrid model based on a supervised random walk