Cold Start Link Prediction

Vincent Leroy[1]
B. Barla Cambazoglu[2]
Francesco Bonchi[2]

[1]INSA/UEB
Rennes, France

[2]Yahoo! Research
Barcelona, Spain

Link Prediction

Liben-Nowell & Kleinberg

- Given a snapshot of a social network, predict links that will appear in the next time window
- Purely based on graph features

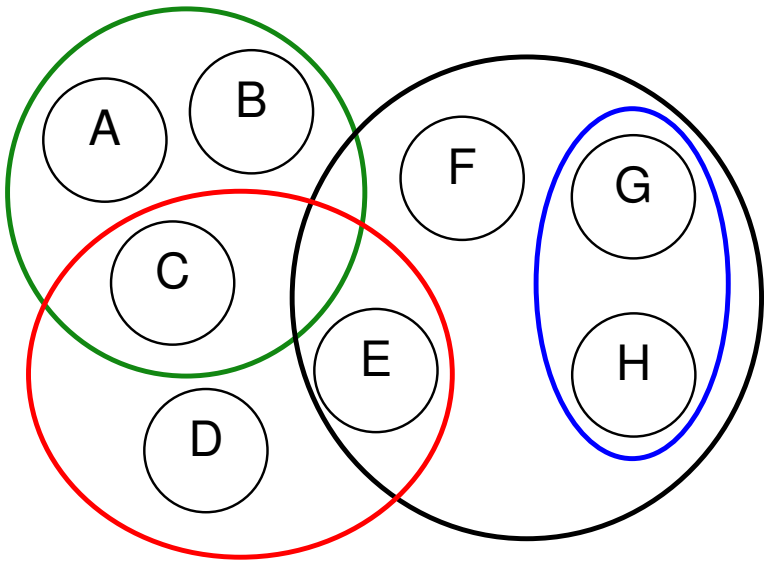
Cold Start Link Prediction

- In some communities, the social network is
  - Hidden and/or private (i.e., not explicit)
  - Very sparse
- Such social networks have many applications
  - Recommendation
  - Viral marketing
- Link prediction starting with an initially empty network

Approach

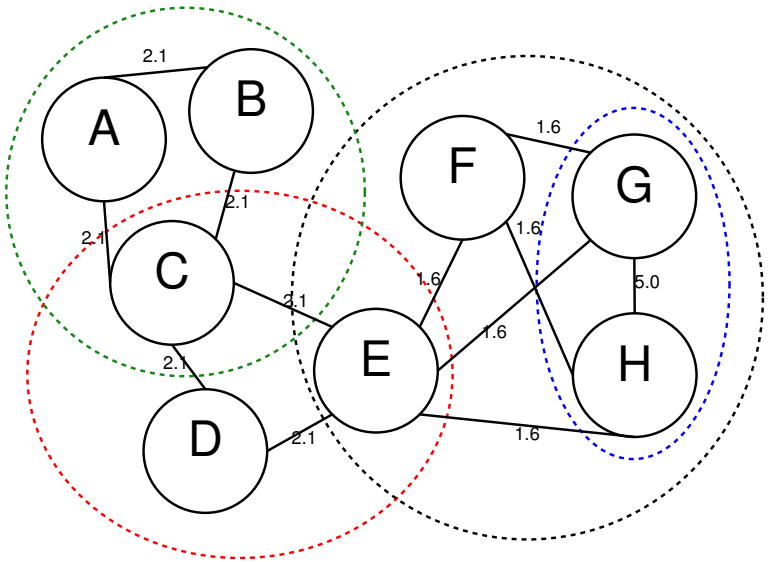
- Two-phase approach
  1. Using available features, bootstrap a social graph
  2. Link prediction through graph-based features
     - Refine link prediction
     - Increase recall
- Experiments on Flickr data
- Features based on group memberships of users

Input Data



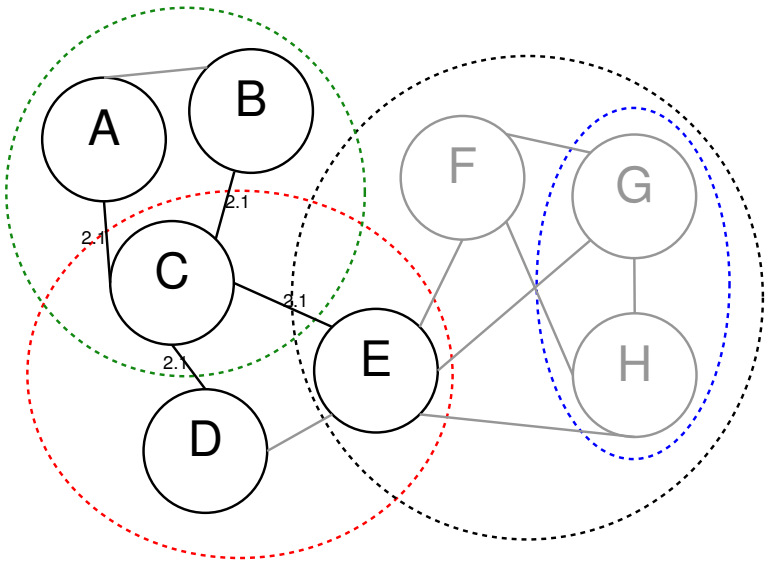
- Large collection
  - 198 thousand users
  - 70 thousand groups
  - 28 million links
  - 39 billion potential links

Phase 1



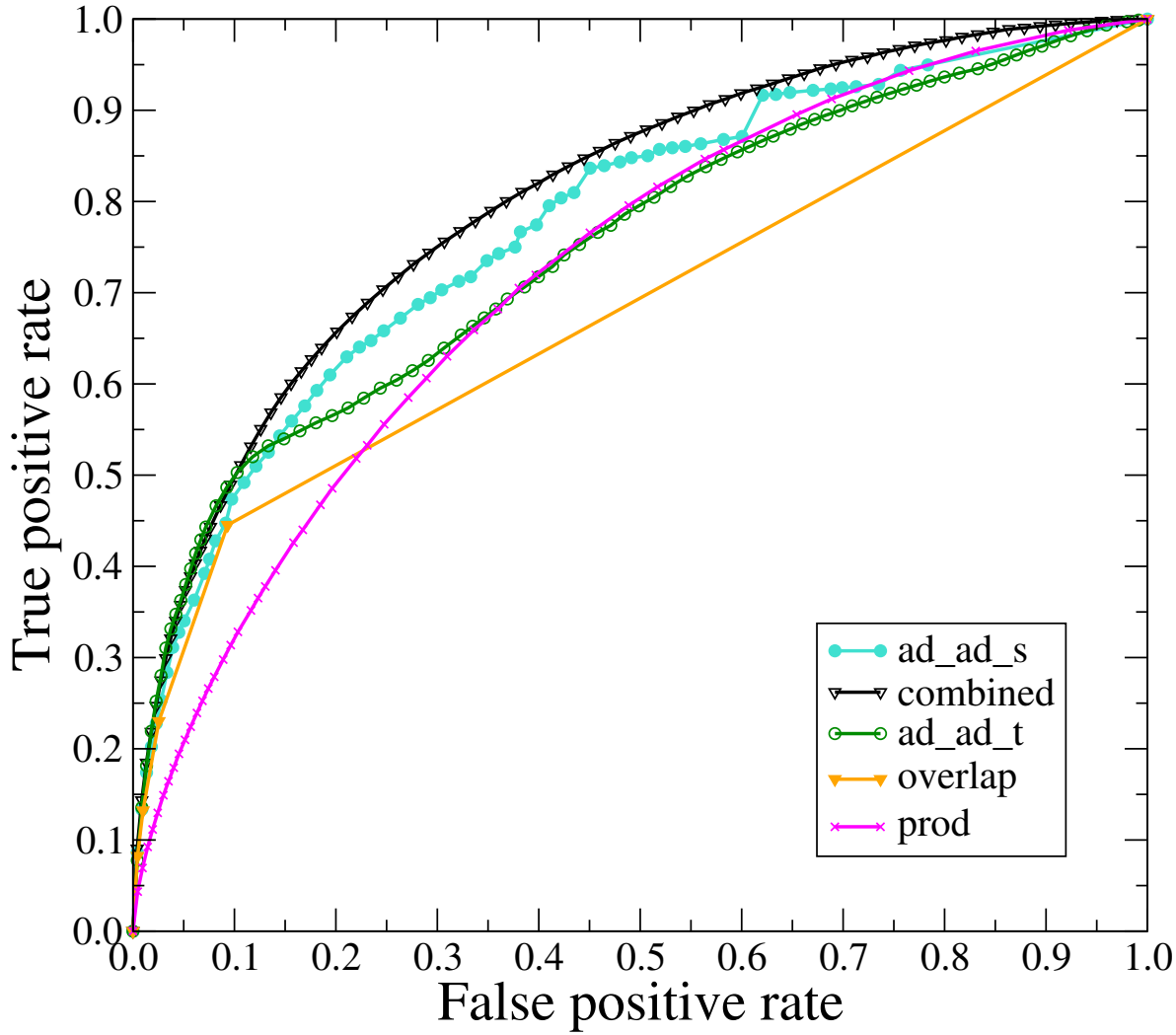
- Simple, generic features
  - Number of groups
  - Common groups
  - Group sizes
  - Inter-arrival time

Phase 1: Prediction for C

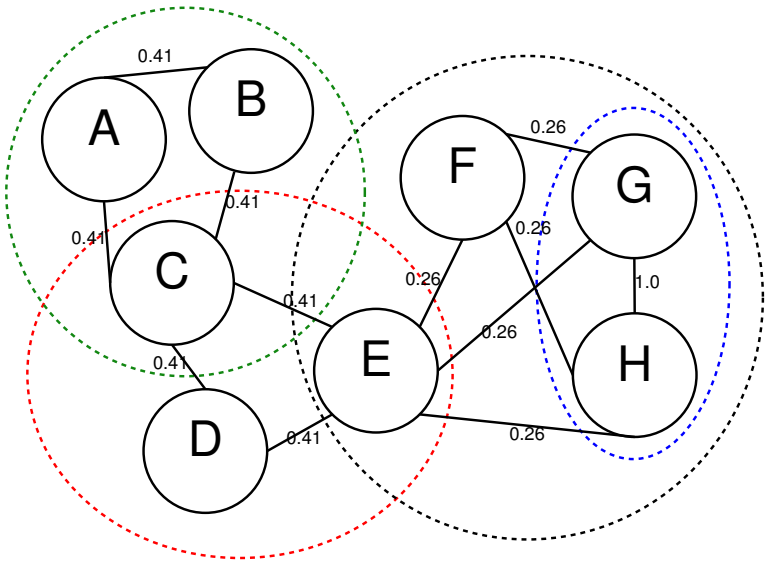


- Links between users sharing at least one group
  - Limited recall
  - Needs to be refined (take other links into account)

Phase 1: Evaluation

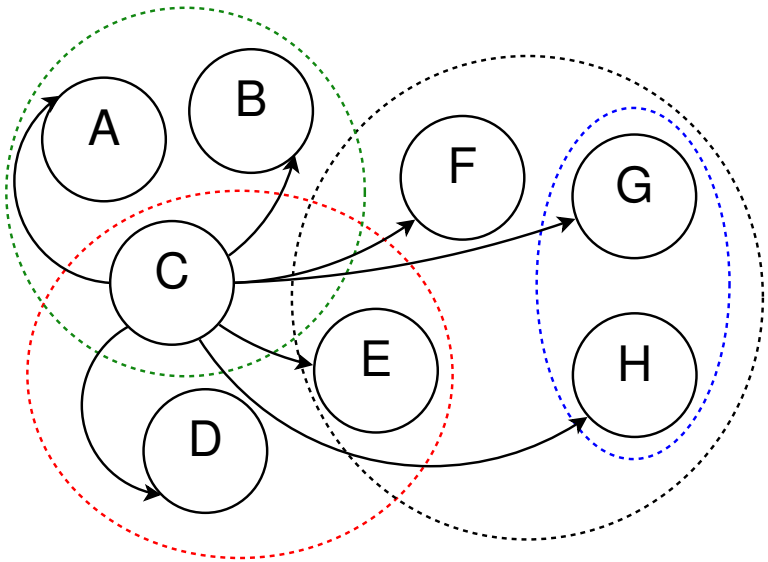


Phase 2



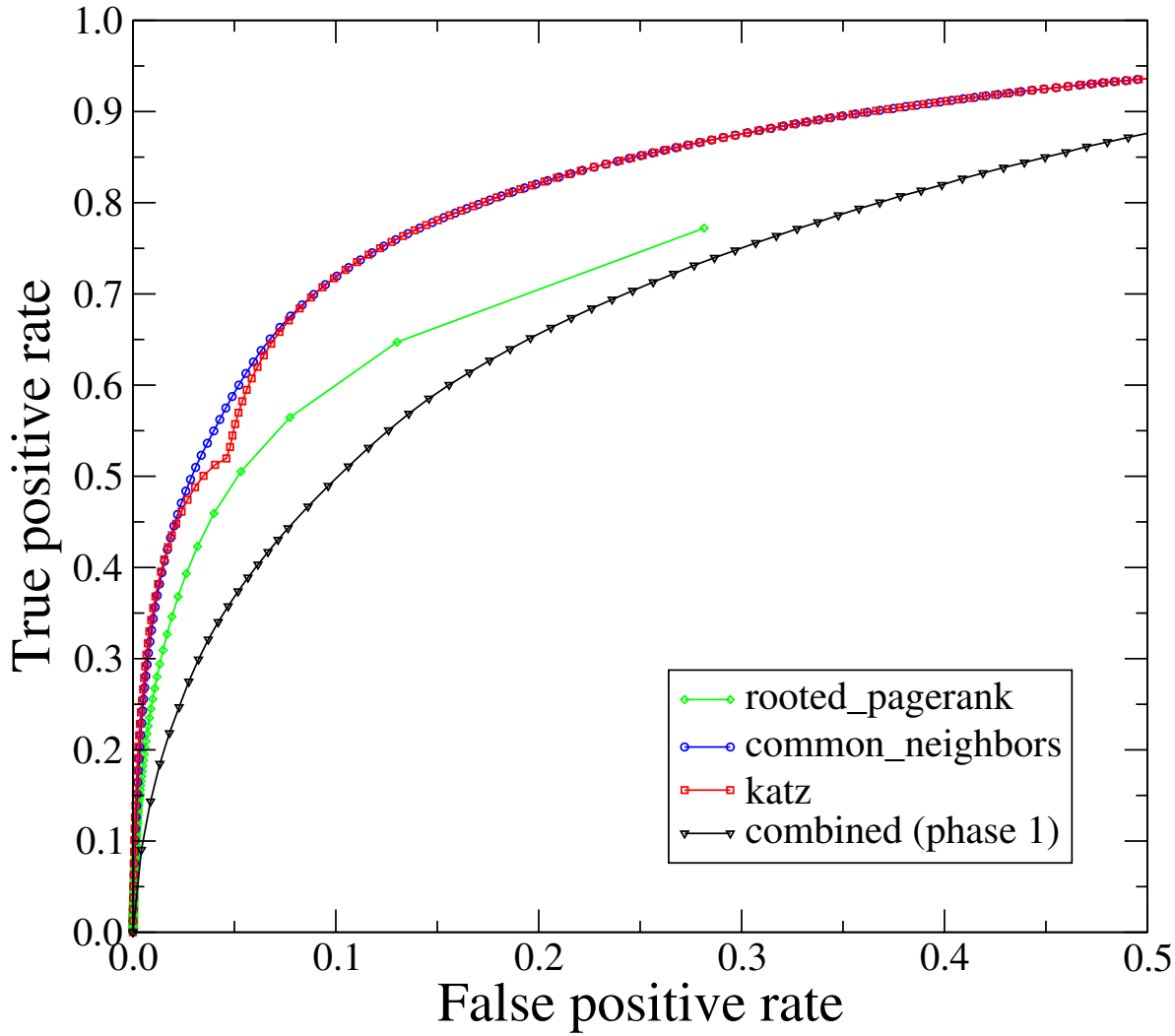
- Conversion process
  - Combine scores on path
  - Convert scores to probabilities
  - Probabilistic graph

Phase 2: Prediction for C



- Graph-theoretic features
  - Common neighbors
  - Katz
  - Rooted PageRank

Phase 2: Evaluation



Conclusion

- Generic approach for cold start link prediction
- Applied on real data
- Groups provide sparse information but are widely used
- More specific features could provide better results
- Applications in the area of privacy
- Improve conversion between scores and probabilities