Importance of Friend Recommendation

• With our limited time and so many sites to choose from, we often face a dilemma of choosing a few sites over others.
• Users prefer more engaging sites, where they can find familiar faces such as friends, relatives, or colleagues.
• Users leave sites easily when they cannot find friends:
  • 60% of Twitter/Facebook Users quit within the first month

Solution:
Recommend Friends Using Link or Content Information

Finding Friends with Minimum Information

• What is the minimum information available: Username

• Example: What Twitter is doing:
  • The search space for finding Friends is the whole Twitter

• What we propose:
  • Reduce the search space
  • Let $\beta$ for a new user denote the search space reduction ratio.
  • For example, when search space is 1/10 of Twitter, $\beta = 10$
  • Because we have many users, we are interested in the expected Value of $\beta$

• How can we find these search spaces?
  • Observation 1: we have assortative mixing (friends are similar)
  • Observation 2: we can partition Twitter and each new user is assigned to a partition, i.e., no overlapping search spaces

• How do we partition?
  • We can partition based on attribute values
  • For example, language: partitions representing users speaking English/French/etc.

• What attributes should we select?
  • We can select attributes that represent the three general social forces that result in friendships:
    • Homophily (attribute selected: age)
    • Confounding (attributes selected: Language and Location)
    • Influence (Fitting well within the friends crowd)

Predicting the Attributes From Usernames

• We predict the language, age, and the location from the usernames
• The age and location influence the username

1. To train an age classifier, we use a set of 80,000 usernames with their age.
2. For predicting location, we train over a set of 36 million geo-located usernames from Twitter.
3. For language prediction, we train an n-gram language detector over 40 million words.

Why Traditional Methods Fail!

• There is no link or content!
  • No Content: right after a user joins a new site, there is no profile information or site activity.
  • No Link: right after a user joins a new site, the user is a disconnected singleton in the friendship graph.

• This is a universal problem for all sites and for each and every user, right after she joins a new site.

• Isn’t that the cold start problem?
  • In cold-start, one often assumes that either link or content is available.

• Then what do sites do?
  • In Twitter: Recommending Celebrities or political figures in the United States or the world.

Experimental Results

• We collect 135 million friendships from Reddit

State-of-the-art link prediction achieves $\beta$ of [2.4, 5.4] with access to link information

We achieve an expected $\beta$ of [5.49, 31.04] without link or content information

This work was supported, in part, by the ONR Research grants: N000141110527 and N000141410095.