AudioScrobbler

- Audioscrobbler and last.fm
  - Collaborative filtering
  - What is a neighbor?
  - What is the network?

Duke Scrobbler

- Integration with facebook

What can we do with real data?

- How do we find a graph’s diameter?
  - This is the maximal shortest path between any pair of vertices
  - Can we do this in big graphs?

- What is the center of a graph?
  - From rumor mills to terrorists
  - How is this related to diameter?

- Demo GUESS (as augmented at Duke)
  - IM data, Audioscrobbler data

My recommendations at Amazon
And again...

Collaborative Filtering

- **Goal:** predict the utility of an item to a particular user based on a database of user profiles
  - User profiles contain user preference information
  - Preference may be **explicit** or **implicit**
    - Explicit means that a user votes explicitly on some scale
    - Implicit means that the system interprets user behavior or selections to impute a vote
- **Problems**
  - Missing data: voting is neither complete nor uniform
  - Preferences may change over time
  - Interface issues

Google’s PageRank

Imagine a “pagehopper” that always either
- follows a random link, or
- jumps to random page
PageRank ranks pages by the amount of time the pagehopper spends on a page:
- or, if there were many pagehoppers, PageRank is the expected “crowd size”

Everyday Examples of Collaborative Filtering...

- Bestseller lists
- Top 40 music lists
- The “recent returns” shelf at the library
- Unmarked but well-used paths thru the woods
- The printer room at work
- Many weblogs
- “Read any good books lately?”
- ....
- Common insight: **personal tastes are correlated**:
  - If Alice and Bob both like X and Alice likes Y then Bob is more likely to like Y
  - especially (perhaps) if Bob knows Alice