Social Search and Discovery
Using a Unified Approach

**Traditional search**
- Content: Documents
- Task: Find relevant documents
- Ranking: Traditional IR techniques, link analysis

**Social search**
- Content: "Web 2.0" user-generated content (e.g., blogs), metadata (tags, comments, ratings), person-document relationships
- Task: Find relevant people, tags
- Ranking: Leverage the "wisdom of the crowd"

**Technical Approach**
- Search space is expanded to include relationships between objects
- Objects and relationships are indexed and used to compute most relevant search results
- Results come from expanded object space:
  - People who are highly related to the topic
  - Relevant documents, people, blogs, etc.
- A tag cloud defining the topic of your query

**Social Ranking**
- Ranking of documents:
  - Traditional text similarity relevance
  - User-contributed metadata (e.g., tags) adds text to documents
  - Document static-score based on its popularity (bookmarks, comments, etc.)
- Ranking of people (and tags):
  - A person is related to a query if related to documents matching the query
  - The score of the person is a function of the relevance of the matching document
  - Different types of person-document relationships get different weights, e.g., author has higher weight than tagger
  - IEF lowers rank of a person that matches every query, not just this one
  - Personalization: boost people you know

**Data Sources**
- Profiles - IBM’s internal BluePages application contains 475,000 profiles. BluePages serves 3.5 million searches per week and 1.5 million profile views per day.
- Communities - IBM Communities hosts 900 communities. IBM Forums contain 147,000 threads and 410,000 messages.
- Blogs - IBM’s BlogCentral hosts 27,300 weblogs (420 group blogs) with 62,000 entries, 60,000 comments, and 10,800 distinct tags.
- Bookmarks - IBM’s social-bookmarking system Dogear has 327,000 bookmarks from 8,511 users. One-third are intranet links and only 2.5% are private.

**Enterprise Social Search**
- Related people are people related to the result set: authors, commenters, and/or taggers of one or more documents in the result set. This is a ranked list.
- Related tags are a "tag cloud" of tags associated with documents in the result set.
- Additional facets. Search results can be narrowed by additional dimensions.

**Evaluation**
- **Related people**
  - Large-scale user study: 600 respondents from 116 IBM locations in 38 countries.
  - For each query, found the “related people” and asked each to rank their relevance to topics (some believed relevant, some not).
  - Calculated agreement of each person’s and algorithm’s judgment (using NDCG).
  - High agreement shown, improved by more refined algorithms:

**Relevant documents**
- Standard IR evaluation methodology:
  - Picked 50 real users’ queries
  - Executed them on several variants of the search engine
  - Relevance level judged by humans

Using social data and static scores based on social data **significantly** improves search accuracy: