Recommender Systems and Decision Making in the Social Media World

Ido Guy
Yahoo Labs
Israel

DMRS 2015, October 22-23, Bolzano, Italy

High-Level Outline

- Social Recommender Systems – Overview
  - Content Recommendation
  - People Recommendation
- Recommending Social Media Content to Community Owners (Ronen et al., SIGIR 2014)
- Islands in the Stream: Item Recommendation within an Enterprise Social Stream (Guy et al., SIGIR 2015)

Web 2.0 and Social Media

- **Web 2.0 [Oreiley03]**
  - It's all about people
    - Joining online communities
    - Connecting to each other on social networks
    - Creating content as in wikis and blogs a
    - Annotating content with tags, comments, ratings

- **Social Media**
  - Refers to Web 2.0 sites that allow users to share and interact
  - Characterized by
    - User-generated content (e.g., tags)
    - User-centered design
    - Social networks and online communities

Top 10 Sites on the Web (Alexa.com)

1. Google
2. **Facebook**
3. **YouTube**
4. Baidu
5. Yahoo
6. Amazon
7. **Wikipedia**
8. Qq
9. **Twitter**
10. Taobao

Social Overload

- **Facebook** – largest social network site
  - 1,500,000,000 users, half log in every day
    - Over 1B on mobile
  - 190,000,000,000 online “friendships”
  - 74,200,000 pages
  - 4,500,000,000 “likes” per day

- **YouTube** – largest video sharing site
  - 1,000,000,000 users
  - 4,250,000,000 views per day
  - 10,000,000 video hours uploaded per month

- **Twitter** – largest microblogging site
  - 316,000,000 active users
  - 500,000,000 tweets per day
  - 76,000,000 followers of most popular user

Social Overload

- **Information overload –** blogs, microblogs, forums, wikis, news, bookmarked web pages, photos, videos, …
- **Interaction overload –** friends, followers, followees, commenters, co-members, voters, “likers”, taggers, …

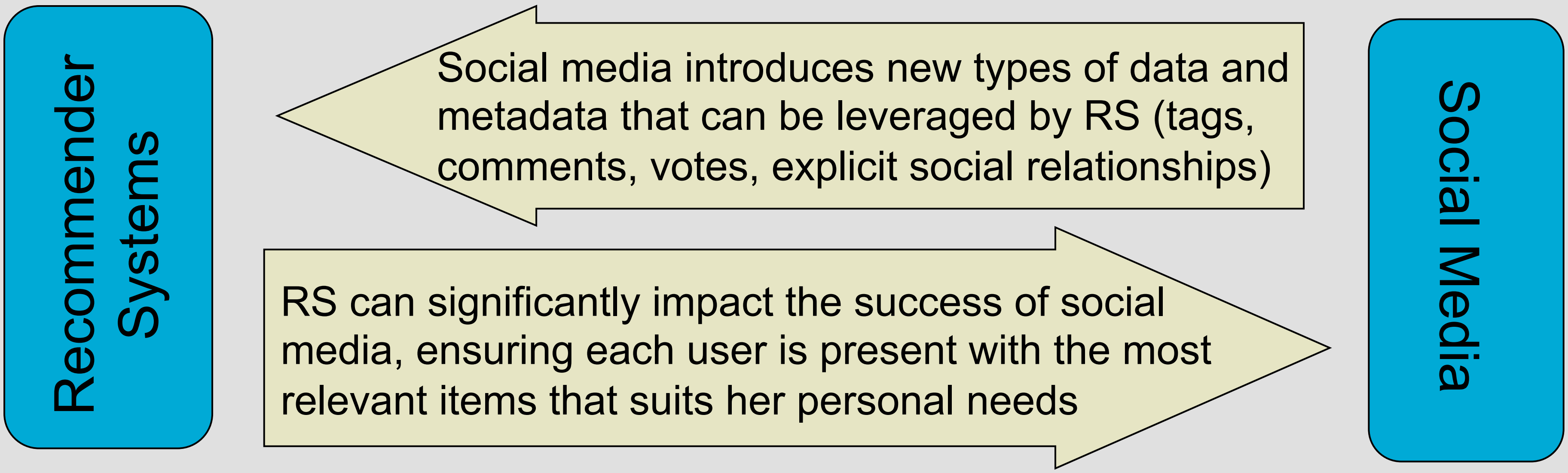
Decision Making by Social Media Users

- What content (blogs, wikis) to read?
- What channels to follow?
- Which social networks to use?
- Who to be friends with?
- Who to follow?
- What groups to join?
- What content to produce?
- How to annotate own and other's content?

Social Recommender Systems

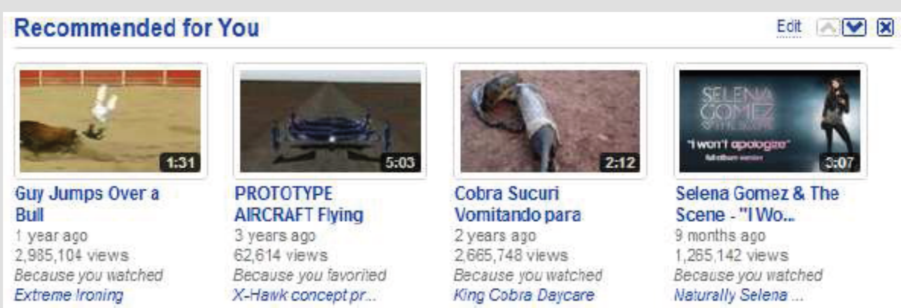
- Recommender Systems that target the social media domain
- Aim at coping with the challenge of social overload by presenting the most attractive and relevant content
- Also aim at increasing adoption and engagement

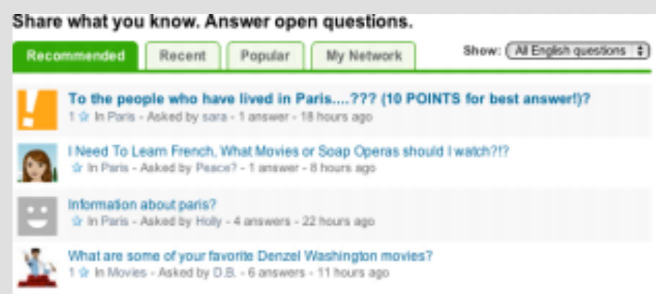
Recommender Systems and Social Media



Content Recommendations

- Video recommendation
  [Davidson et al., RecSys '10]
- News recommendation
  - Digg [Lerman, ICWSM ' 07]
  - Google reader [Liu et al., IUI '10]
- Question recommendation
  [Szpector et al., WWW '13]
- Blog recommendation
  [Arguello et al., ICWSM '08]





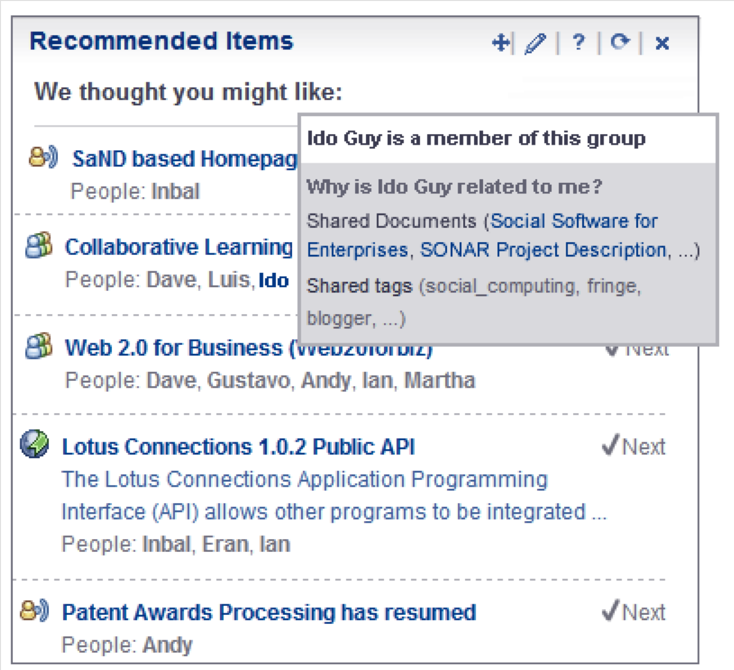
Enhancing CF with Friends

- The user's network of friends and people of interest becomes more accessible in the Web 2.0 era (Facebook, LinkedIn, Twitter,...)
- Such social relationships can be very effective for recommendation compared to traditional CF
  - Recommendation from people the user knows and thus can judge
  - Spare explicit feedback such as ratings
  - Effective for new users
- Various works have shown the effectiveness of friend-based recommendation over CF, e.g.:
  - Movie and book recommendation - Comparing Recommendations Made by Online Systems and Friends [Sinha & Swearingen, 2001]
  - Friends as trusted recommenders for movies [Golbeck, 2006]
  - Club recommendation within a German SNS - Collaborative Filtering vs. Social Filtering [Groh & Ehmig, Group 2007]

Mixed Social Media Item Recommendations

Personalized Recommendation of Social Software Items based on Social Relations [Guy et al., RecSys ’09]

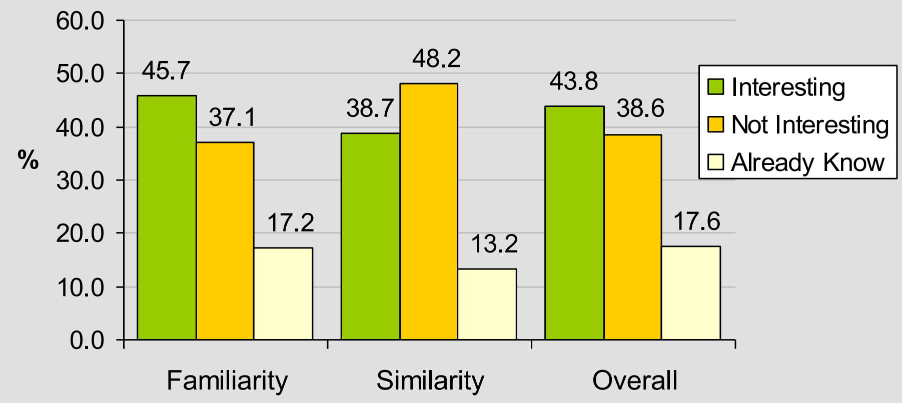
- Social network-based recommendations of blogs, bookmarks, and communities
- Key distinction:
  - Familiarity: co-authorship, org chart, direct connection or tagging, etc.
  - Similarity – co-usage of tags, co-bookmarking, co-membership, co-commenting
- Explanations – showing the “implicit recommender” and her relationship to the user and item

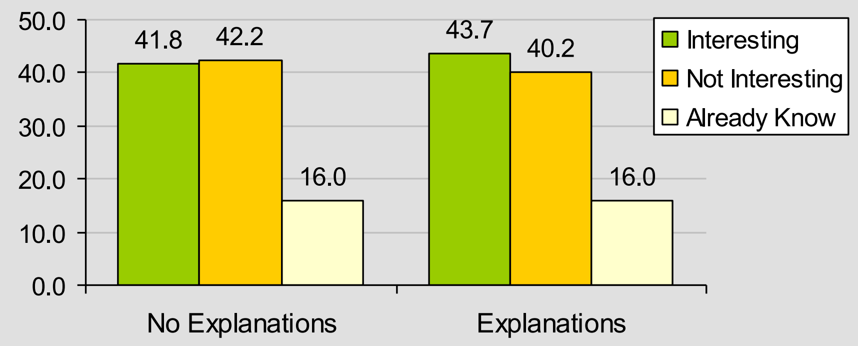


$$RS(u,i) = e^{-\alpha t(i)} \cdot \sum_{v \in N^T(u)} S[u,v] \sum_{r \in R(v,i)} W(r)$$

Mixed Social Media Item Recommendations

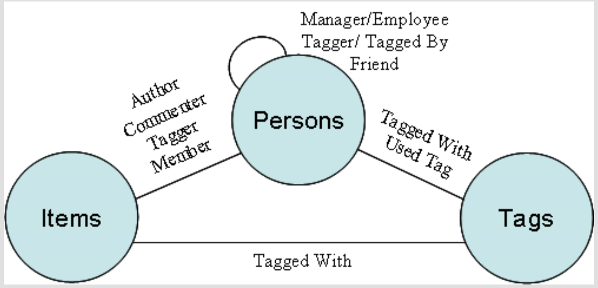
- Evaluation combines a user survey and a live system
- Recommendations from familiar people are significantly more accurate than recommendations from similar people
  - 57% to 43% interest ratio
- Similar people yield more diverse, less expected items
- Explanations have an instant effect increasing interest in recommended items

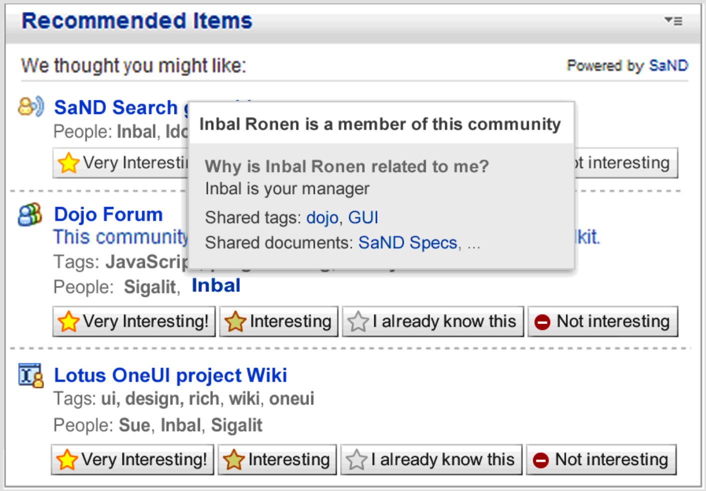




Mixed Social Media Item Recommendations

- Social media recommendation based on people and tags [Guy et al., SIGIR ’10]
- Underlying social aggregation system – SaND
- 5 item types: blogs, bookmarks, communities, wikis, files
- First comprehensive study to compare people-based and tag-based recommenders
- Outgoing and incoming tags





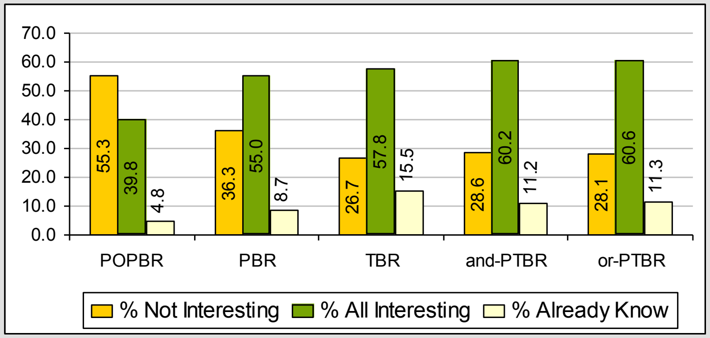
$$RS(u,i) = e^{-\alpha d(i)} \cdot [\beta \sum_{v \in N(u)} w(u,v) \cdot w(v,i) + (1-\beta) \sum_{t \in T(u)} w(u,t) \cdot w(t,i)]$$

Mixed Social Media Item Recommendations

- Direct tag evaluation
  - 65 participants rated 16 tags each
  - Hybrid tags most accurate
  - Incoming slightly outperform outgoing
  - Indirect are least effective

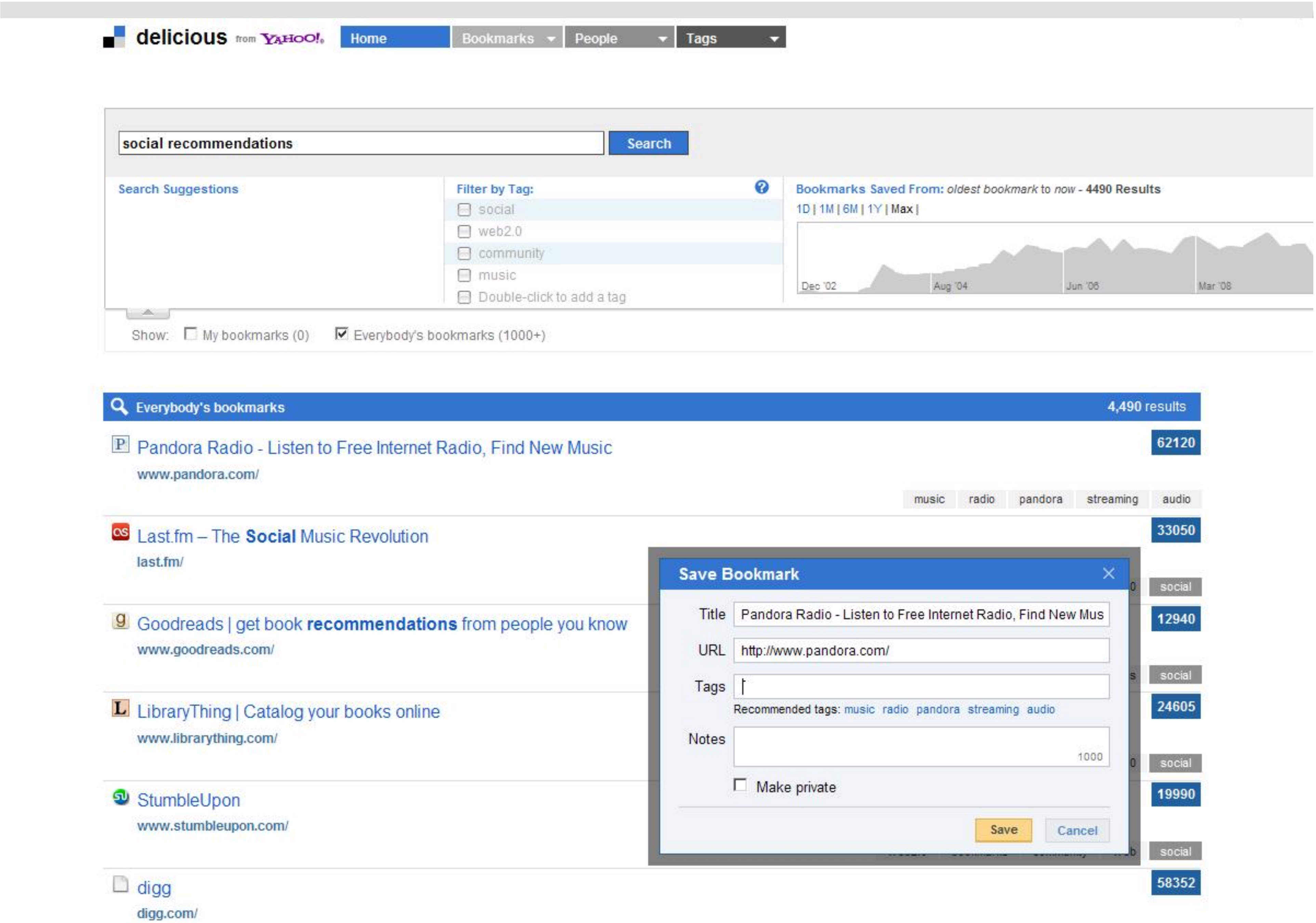
%	Not Interested	Interested	Highly Interested
used	16.84	38.25	44.91
incoming	15.48	31.75	52.78
direct	**7.46**	22.81	**69.74**
indirect	35.38	45.38	19.23

- Large-scale user study
  - 412 participants rated 16 items each
  - All personalization methods outperform popularity
  - Tag-based significantly outperforms people-based in terms of accuracy
  - Yet has less diversity, more expected results, and less effective explanations
  - Hybrid combines the good of both worlds
  - Reaches 70:30 interest ratio for first 16 items



Tag Recommendation

- Adding terms (tags) to objects by the public provides additional contextual and semantic information to various resources:
  - Web pages (e.g. Delicious)
  - Academic publications (e.g. CiteILike)
  - Multimedia objects (e.g. Flickr, Last.Fm, YouTube)
- External tags are useful for many applications
  - search/browse, classification, tag-cloud representation, query expansion
- Tag Recommendation: - recommend appropriate tags to be applied by the user per specific item annotation
  - assist the user in the tagging phase
  - reduce undesired noise in the aggregated folksonomy


delicious from YAHOO!
Home
Bookmarks
People
Tags
social recommendations
Search
Search Suggestions
Filter by Tag:
social
web2.0
community
music
Double-click to add a tag
Bookmarks Saved From: oldest bookmark to now - 4490 Results
1D | 1M | 6M | 1Y | Max |
Dec '02
Aug '04
Jun '06
Mar '08
Show: My bookmarks (0) Everybody's bookmarks (1000+)
Everybody's bookmarks
4,490 results
Pandora Radio - Listen to Free Internet Radio, Find New Music
www.pandora.com/
62120
music radio pandora streaming audio
Last.fm – The Social Music Revolution
last.fm/
33050
social
Goodreads | get book recommendations from people you know
www.goodreads.com/
12940
social
LibraryThing | Catalog your books online
www.librarything.com/
24605
social
StumbleUpon
www.stumbleupon.com/
19990
social
digg
digg.com/
58352
Save Bookmark
Title
Pandora Radio - Listen to Free Internet Radio, Find New Mus
URL
http://www.pandora.com/
Tags
Recommended tags: music radio pandora streaming audio
Notes
1000
Make private
Save
Cancel

Tag Recommendation Approaches

- Popular:
  - Recommend the most popular tags to the user
    - Popular tags already assigned for the target item (Golder 2005)
    - Frequent tags previously used by the user
    - Tags co-occurred with already assigned tags (Sigurbjornsson 2008)
- Collaborative Filtering:
  - Recommend tags associated with “similar” items
  - Recommend tags given by “similar” users
- Hybrid:
  - Recommend tags given by similar users to similar items (Symeonidis08, Rendle10, Carmel 10)

Content-based tag recommendation

- Recommend keywords/phrases from the item's associated text (content, anchor-text, meta data, etc.)
  - e.g .terms with highest *tf-idf* score
- Analyze mutual relationship between content and tags
  - Recommend tags that have the highest co-occurrence with important keywords
  - Language modeling approach (Givon 2010):
    - Estimate the joint tag and keyword probability distribution.
    - This provides an estimation that a given item will be annotated with certain tags, given a background collection of annotated items

Graph-based approaches

- The FolkRank algorithm (Hotho 2006):
  - a resource which is tagged with important tags by important users becomes important
    - The same holds, symmetrically, for tags and users
- We have a graph of connected vertices (resources, users, tags) which are mutually reinforcing each other by spreading their weights
- Graph nodes are scored by random walk techniques:

$$\vec{\omega} = d \cdot A \cdot \vec{\omega} + (1 - d) \cdot \vec{p}$$

w – a weight vector over nodes

A – a row-stochastic matrix of the graph

p - preference vector over the nodes

• For tag recommendation, return the top ranked tags, while setting p to bias the desired pair of user and resource

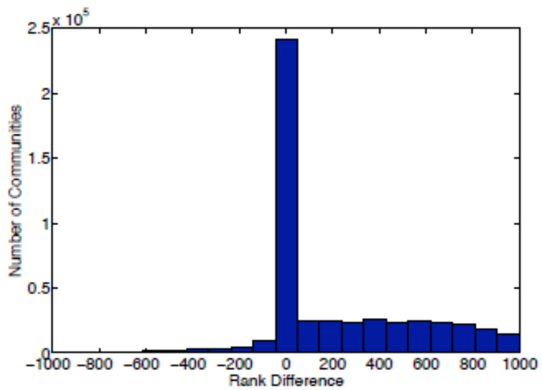
Personalized Community Recommendation

- Collaborative filtering for Orkut communities: discovery of user latent behavior [Chen et al., WWW '09]

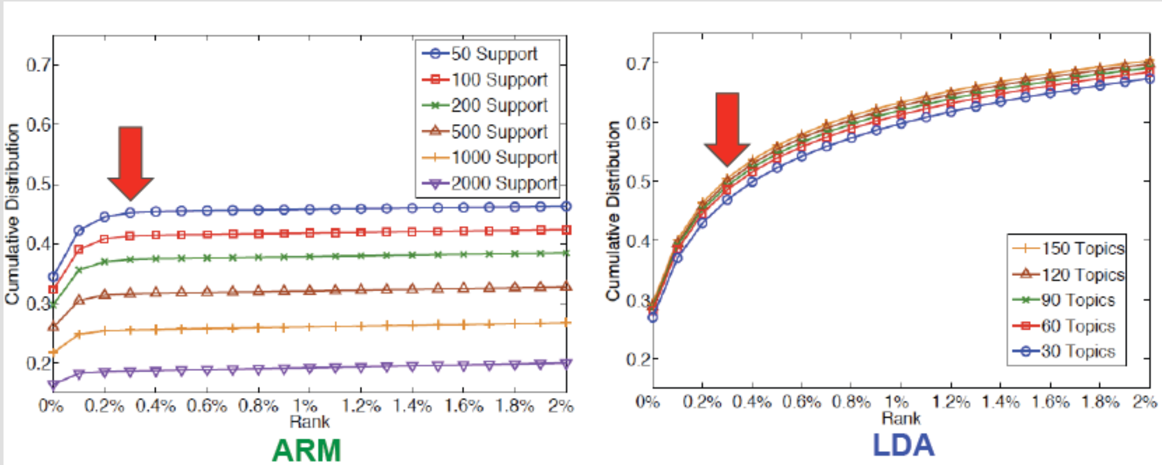
- Personalized community recommendation using CF of two types
  - Association rule mining (ARM) – association between communities shared between many users: users who join X typically join Y
  - Latent Dirichlet Allocation (LDA) – user-community co-occurrences using latent aspects (topics): x is related to y through a semantic feature, e.g., "baseball"
    - Users=docs, communities=topics, membership=co-occurrence
    - Per-topic distribution of users and communities

Personalized Community Recommendation

- Orkut membership data: 492K users, 118K communities
- Top-k recommendation: withhold 1 community the user has joined with k-1 random communities, obtain rank (k=1001)
- ARM is better when recommending lists of up to 3 communities,
- LDA is consistently better when recommending a list of 4 ore more
- In general, LDA ranks communities better than ARM
- LDA is parallelized to improve efficiency

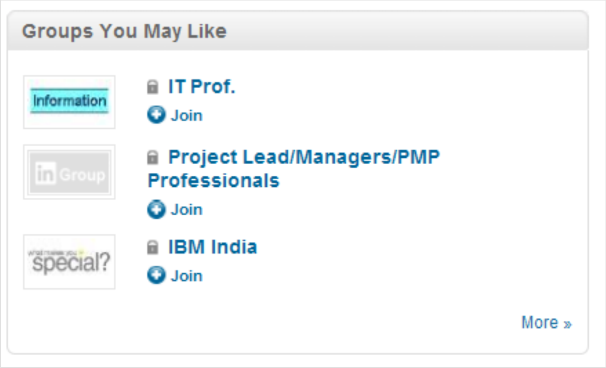


(a) ARM-50 vs. LDA-30



Personalized Community Recommendation

- Flickr group recommendation based on tensor decomposition [Zheng et al., SIGIR '10]
- Group Proximity Measure for Recommending Groups in Online Social Networks (Saha & Getoor, SNA-KDD '08)

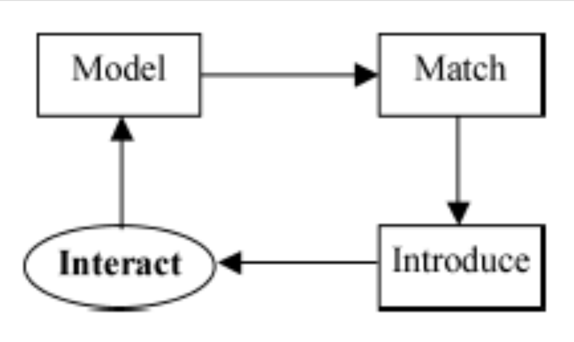


- From LinkedIn Blog ("groups you may like"):
  - Building a virtual profile per group by selecting the most representative features of group members using Information Theory techniques like Mutual Information and KL Divergence.
  - Mapping user's attributes to group's virtual profile
  - Adding more recommendations based on CF

Social Matching

- Social matching: A framework and research agenda [Terveen & Mcdonald, 2005]

- Social matching systems = recommender systems that recommend people to each other
  - Must reveal some amount of personal information
  - Privacy, trust, reputation, interpersonal attraction have greater importance
  - Interaction overload vs. information overload



People Recommendation

- Relationship type
  - Recommending familiar people
  - Recommending interesting people
  - Recommending strangers

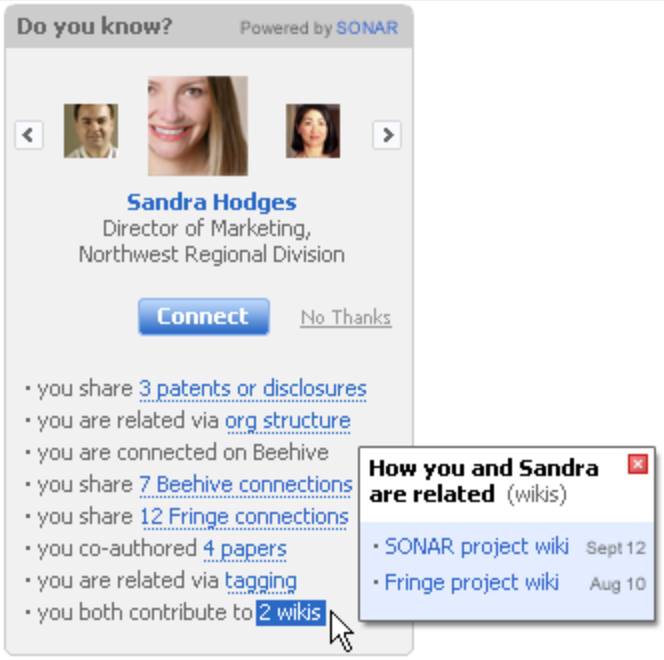
- Relationship lifecycle
  - Regular vs. ad-hoc

- Recommendation technique
  - Content-based
  - Graph-based
  - Interaction-based

Recommending People to Connect with

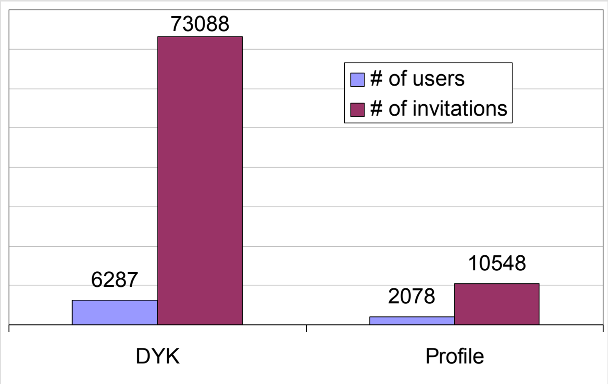
- Do You Know? Recommending People to Invite into Your Social Network [Guy et al., IUI ’09]

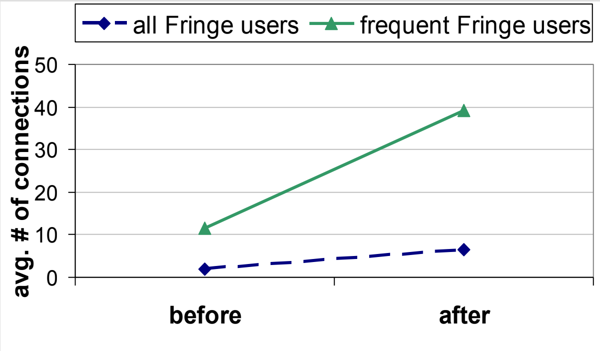
- Recommendation in the enterprise based on the following signals:
  - Org chart relationships
  - Paper and patent co-authorship
  - Project co-membership
  - Blog commenting
  - People tagging
  - Mutual connections
  - Connection in another SNS
  - Wiki co-editing
  - File sharing
- Rich and detailed “evidence”



Recommending People to Connect with

- **Evaluation based on the Fringe enterprise SNS**
- **Dramatic increase in the number of invitations sent and users sending invitations**
  - ***"I must say I am a lazy social networker, but Fringe was the first application motivating me to go ahead and send out some invitations to others to connect"***
- **Evidence increases users' trust in the system and makes them feel more comfortable**
  - ***"If I see more direct connections I'm more likely to add them […] I know they are not recommended by accident"***
- **Substantial increase in friends per user**
- **Sharp decay in usage over time**
  - **Excitement drops, connections exhausted**



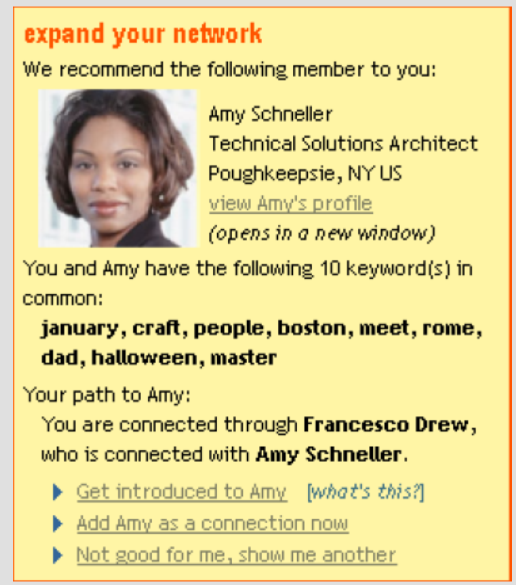


Recommending People to Connect with

- Make new friends, but keep the old: recommending people on social networking sites [Chen et al., CHI ’09]

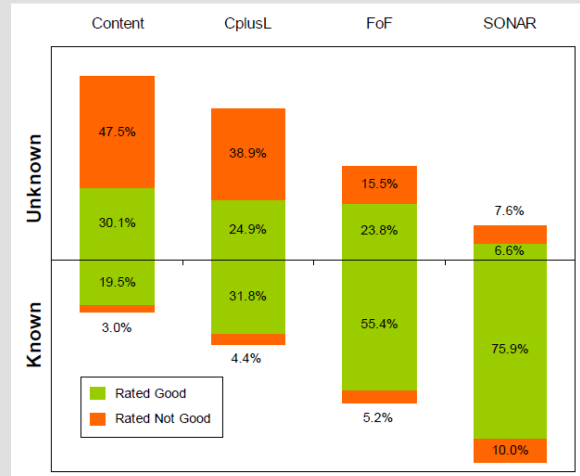
4 Algorithms Compared

- Content Matching (CM)
  - Profile entries, status messages, photo text, shred lists, job title, location, description, tags
  - $vu(wi) = TFu(wi) \cdot IDFu(wi)$
  - Cosine similarity of both users’ word vector
  - Latent semantic analysis did not perform better
    - And does not yield intuitive explanations
- Content-plus-Link (CplusL)
  - Hybrid CM + social link
  - Social link: a sequence of 3 or 4 users
    - a connects to b, a comments on b, b connects to a
- Friend-of-Friend (FoF)
  - Based on number of mutual friends
  - One or more recommendations for 57.2% of the users
- Aggregated Relationships (SONAR)
  - Similar to the “Do you know?” algorithm
  - One ore more recommendations for 87.7% of the users



Recommending People to Connect with

- Evaluation based on the SocialBlue enterprise SNS ("Beehive")
- Survey with 258 participants
  - CM and CplusL yield mostly unknown people, while FoF and SONAR yield mostly known individuals
  - Content similarity vs. relationships alg.
  - The latter are more accurate overall
  - The former are better at discovering new friends
- Controlled field study with 3,000 users
  - SONAR yields most effective results
  - Combine relationships (at first) and content similarity (when the network grows)?



SONAR	FoF	CplusL	Content
59.7%	47.7%	40.0%	30.5%

Table 2. Recommendations resulting in connect actions.

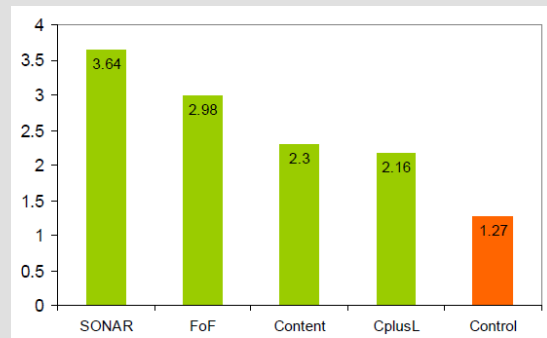
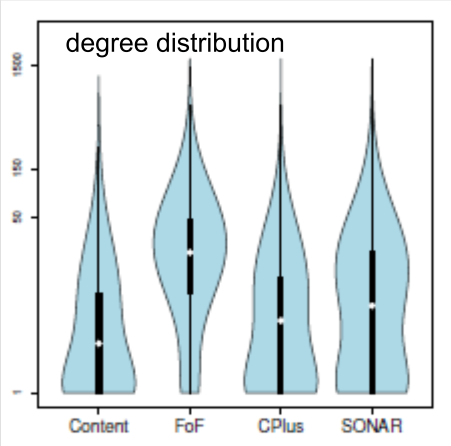
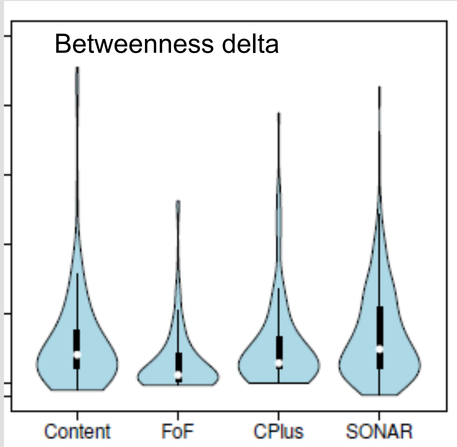


Figure 4. Increase in number of friends.

Recommending People to Connect with

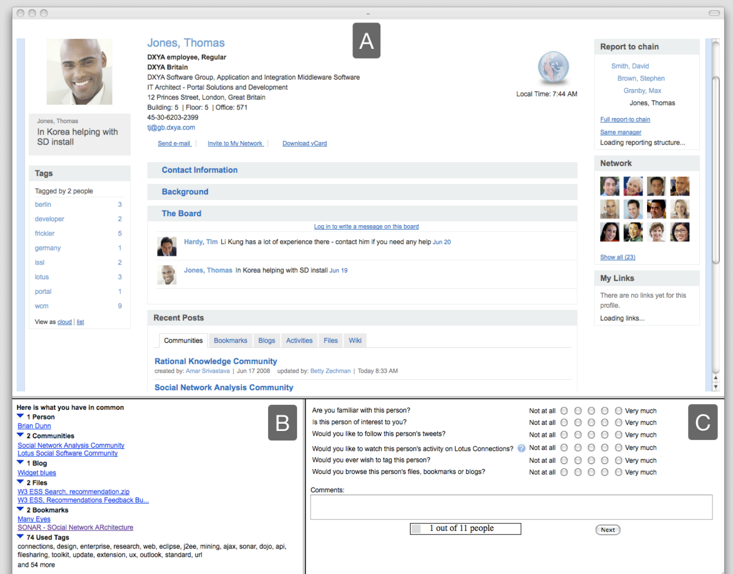
- The network effects of recommending social connections [Daly et al., RecSys '10]
- FoF is highly biased towards well-connected users, leading to high rec. frequency of the same users
- CM is most diverse and often recommends users with few connections only
- CM and SONAR affect betweenness centrality most significantly
- CM is most biased for same country but least biased for same division
- SONAR substantially increases cross-country and intra-division connections
- Highlight network effects when recommending people?





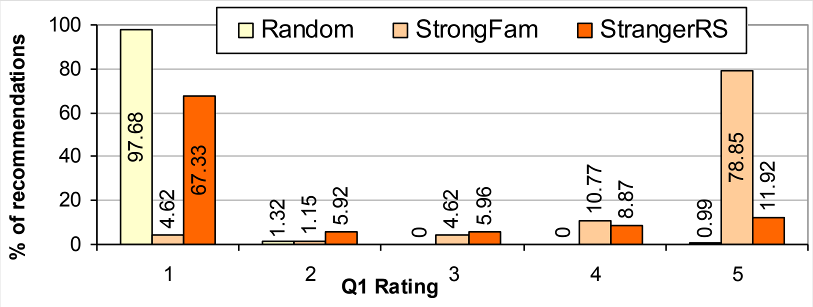
Stranger Recommendation

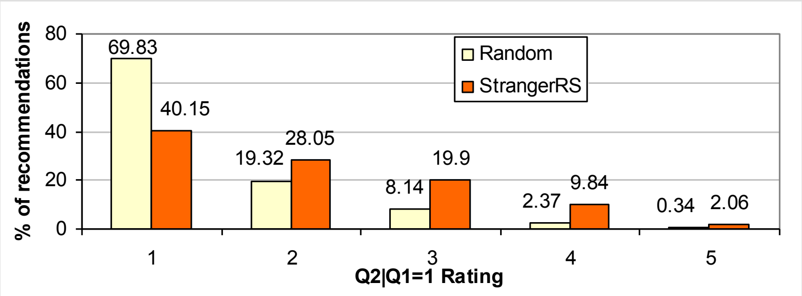
- Do you want to know? Recommending strangers in the enterprise [Guy et al., CSCW ’11]
- Recommendation of people who are unknown yet interesting in the organization
- Maybe useful to
  - Get help or advice
  - Reach new opportunities
  - Discover new routes for career development
  - Learn about new assets that can be leveraged
  - Connect with SMEs and influencers
  - Cultivate organizational social capital
  - Grow own reputation and influence within the organization
- Complements recommendation of people to connect with, as those are quickly exhausted over time



Stranger Recommendation

- Method - subtracting the familiarity network from the similarity network
- Similarity: common things and places: tags, communities, wikis
- Score based on Jaccard's index
- Presentation with evidence
- Two-thirds of the recommendations are strangers
- Significantly more interesting than a random person
- Out of 9 recommendations, 67% got at least one stranger rated 3 or above
- Exploratory recommendation
  - Low accuracy, high value





Recommending People to Follow

- Recommending twitter users to follow using content and collaborative filtering approaches [Hannon et el., RecSys' 10]
- CB, CF, and Hybrid approaches
- User profiles based on
  - Own tweets
  - Followers' tweets
  - Followees' tweets
  - Followers
  - Followees
- Using Lucene to index users by their profile, after applying TF-IDF to boost distinctive terms/users within the profile

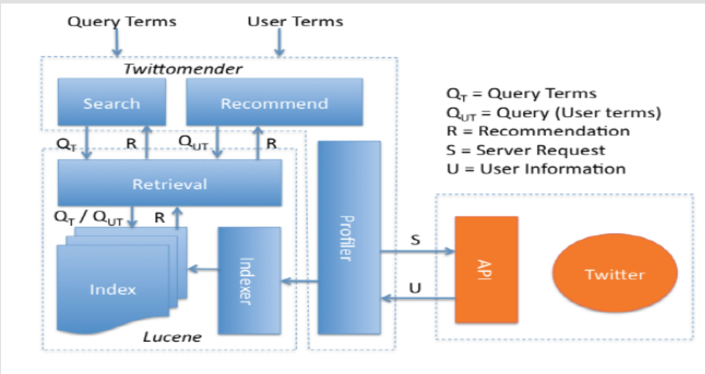


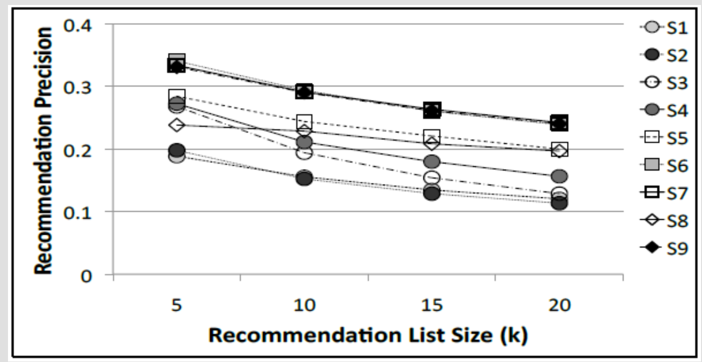
Figure 1: The Twittomender System Architecture.

Recommending People to Follow

- Offline Evaluation, 20K users
  - 19,000 training set Twitter users
  - 1,000 test users
  - Create index per profile and predict followees
  - Measure by precision and position
  - Results show trade-off between the two
  - Slight advantage to followers and tweets of followers
  - Hybrid improves results (precision > 0.3)

Table 1: Evaluation Datasets.

Users	Tweets	Words	Followers	followees
1000	80	15	664	321
19,000	78	14	465	520



- Live User Trial, 34 participants
  - Hybrid approach combining all types
  - 30 recommended Twitter users
  - Indicate whom s/he is likely to follow
    - No actual effect
  - On average, 6.9 out of 30