COMP9727: Recommender Systems

Lecture 6: Social Network Recommendation

Wayne Wobcke

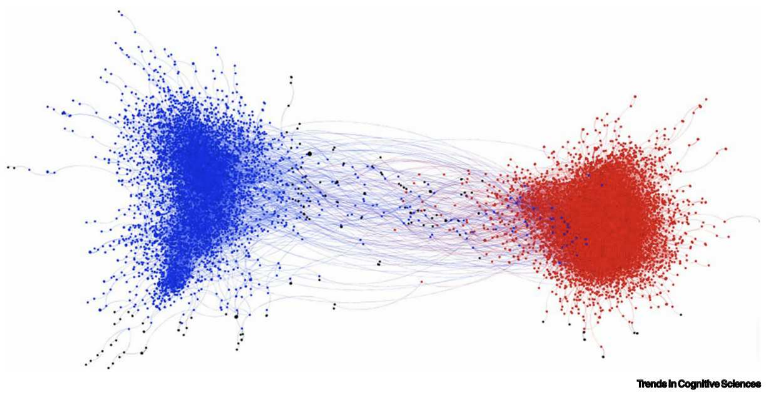
e-mail:w.wobcke@unsw.edu.au

This Lecture

Recommendation within social networks, social network analysis

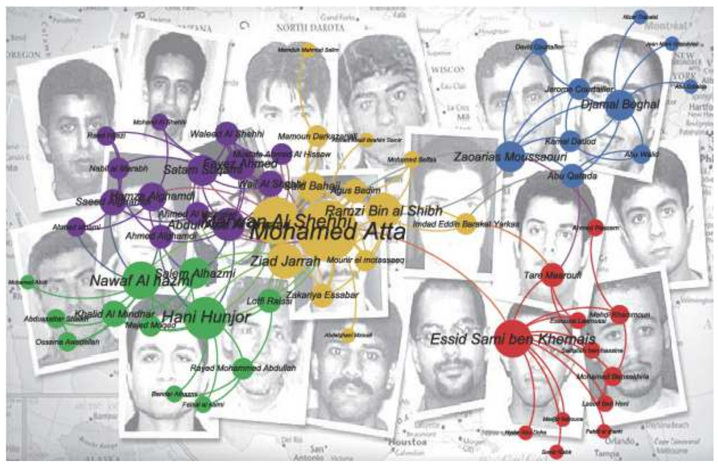
- Social Network Analysis
  - Network properties
  - Strong and weak ties
  - Homophily
- Link Prediction
  - Triadic closure (friend of a friend)
- Reciprocal Recommendation
  - Matching Markets
  - Online dating

Political Polarization on Twitter



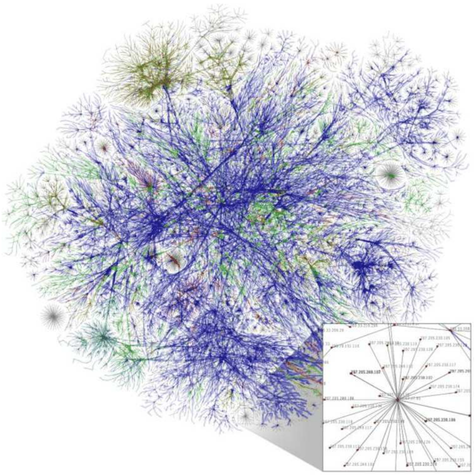
Blue = Democrat; Red = Republican

Terrorist Networks

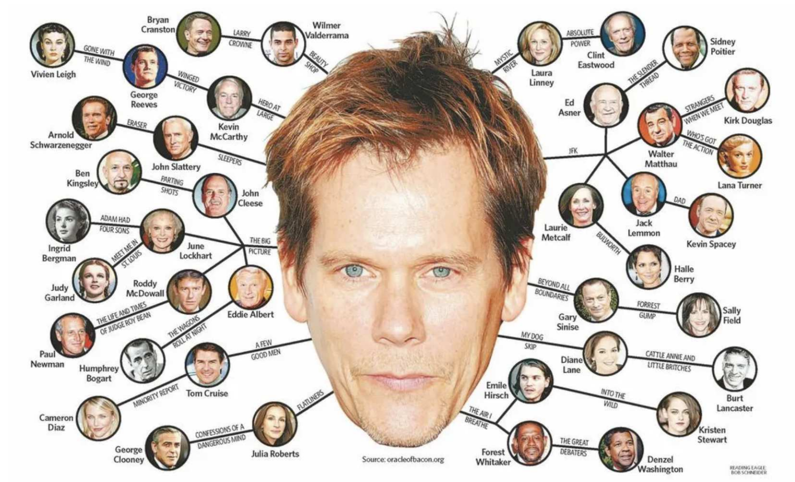


Core-Periphery Structures

Internet (2003)

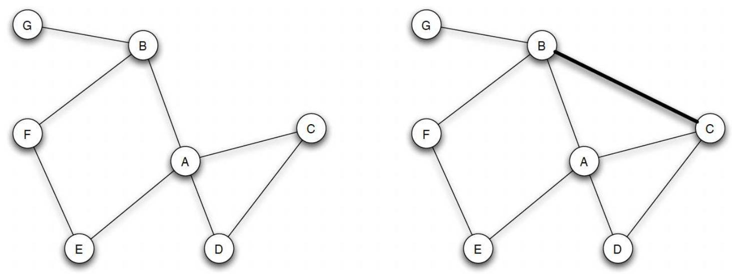


Small World Hypothesis



Triadic Closure

Two people with a friend in common are more likely to become friends



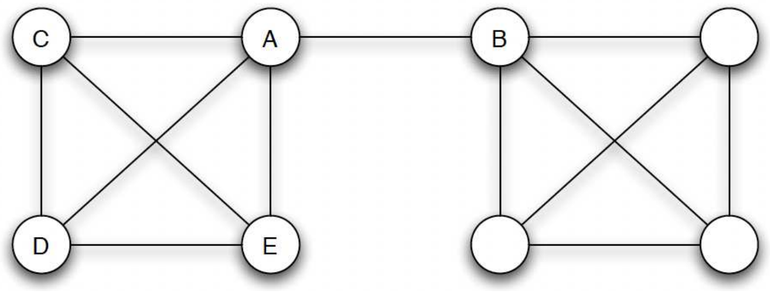
(a) *Before B-C edge forms.* (b) *After B-C edge forms.*

Social Network Concepts

- Gatekeeper node A
  - For some nodes B and C, all paths from B to C go via A
- Clustering coefficient of node A
  - Proportion of pairs of A's friends that are also friends
- Betweenness centrality of A
  - Number of shortest paths in whole graph that go through A
- Bridge A–B
  - Removing A–B breaks up a connected component
- Local bridge A–B
  - A and B have no friends in common
  - Removing A–B means shortest path length from A to $B > 2$

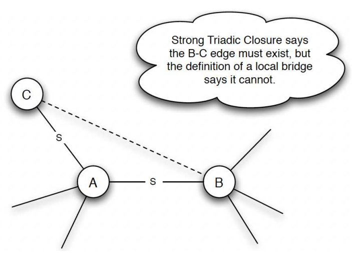
Bridge

Bridge A–B = Removing A–B breaks up a connected component



Strong and Weak Ties

- Strong ties = family, real friends; weak ties = "Facebook friends"
- Weak ties ("connections") necessary for information flow
- Triadic closure more likely with strong ties
- If a node A satisfies Strong Triadic Closure and is involved in at least two strong ties, any local bridge it is involved in must be a weak tie



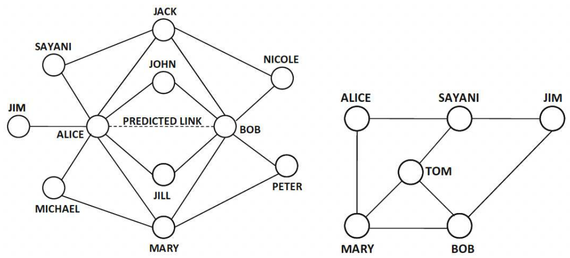
Homophily

"Birds of a feather flock together" ("similar" to oneself)

- On attributes: gender, race, ethnicity, age, class, education
- For: marriage, confidants, friends, co-workers, casual contacts
- Causes: geography, family, organizations, social status, job rank
- Homophily measurement
  - (#observed pairs)/(#expected pairs)
  - e.g. gender: $\frac{\#(observed\ (M,F)\ pairs)}{L.(\#M.\#F)/(N^2/2)}$ with N nodes and L links

Very very very robust, but effects may be weak

Link Prediction



- Neighbours: Let neighbours of A be $S(A)$
  - Jaccard similarity: $\frac{|S(A) \cap S(B)|}{|S(A) \cup S(B)|}$
  - Adamic-Adar measure: $\sum_{k \in S(A) \cap S(B)} \frac{1}{log(|S_k|)}$
- Paths: Let n^t_{AB} be the number of paths from A to B of length t
  - Katz measure: $\sum_{t=1}^{\infty} \beta^t n^t_{AB}$ for suitable β

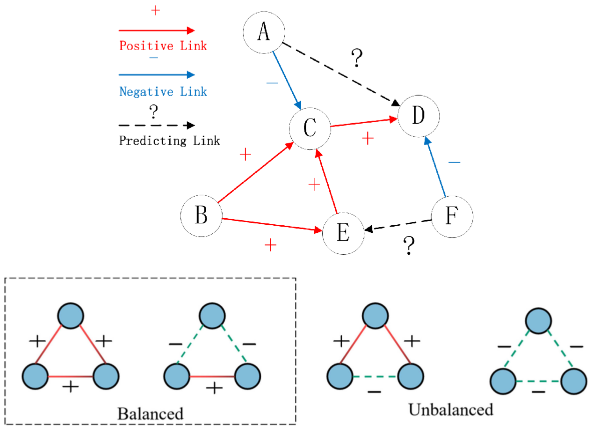
Link Recommendation

- Classification (Content-Based)
  - Training instances are pairs of (features of) nodes
  - Use negative sample set the same size as positive set
- Matrix Factorization (Collaborative Filtering)
  - Adjacency matrix of 0/1 (1 = friend)
  - Or choose negative set (0s) the same size as positive set (1s)
- Network Trust
  - e.g. PageRank where pages are people

Recommend users with highest prediction score

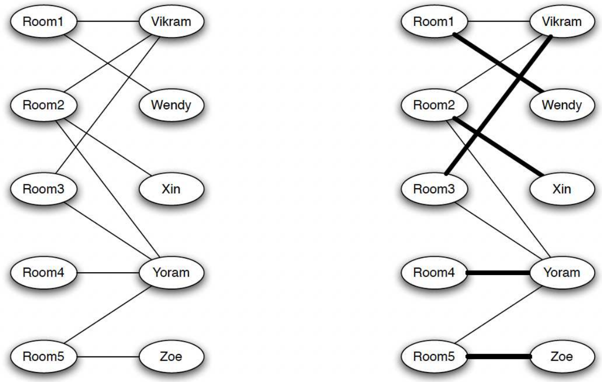
Signed Social Networks

"The enemy of my enemy is my friend"



Matching Markets

Bipartite graph



Stable Matching/Marriage Problem

Suppose there are exactly N users and N "items" and each user has a preference ordering over all items, and each "item" has a preference ordering over all users.

Examples: Students/Universities, People/Jobs, ...

Definition: A matching is a mapping between users and items so that each user is assigned one item, and each item is assigned one user.

Definition: A matching is stable if, for no user A matched with item B, there is a user A' matched with item B' such that A prefers B' to B and B' also prefers A to A'.

Theorem (Gale–Shapley, 1962): A stable matching always exists, given the above assumption.

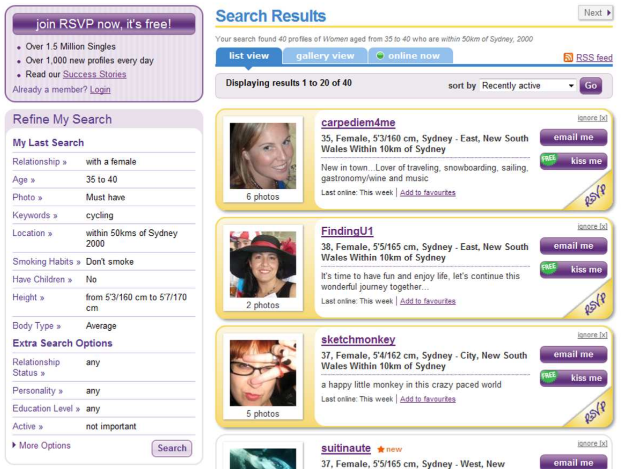
Reciprocal Recommenders

Traditional Recommenders	Reciprocal Recommenders
Success is determined solely by the user seeking the recommendation	Success is determined by both parties to the recommendation
Users do not provide detailed explicit user profiles	Users may provide profiles, however these may be of limited use
Satisfied users are likely to return for more recommendations; better recommendations mean more engagement in the future	Users may leave the system after a successful recommendation; better recommendations might mean less engagement in the future
The same item can be recommended to all users	Popular users should not be recommended to too many users

Bidirectional Collaborative Filtering

- Users are not items
  - Users can reject invitations, cannot be commodified
- Consequences for recommendation
  - Can't recommend the same user (too often)
  - Can't recommend "popular" users (too often)
  - Can't just recommend people a user will find attractive
  - Can't just recommend people who are likely to say "yes"

Online Dating Site



Homophily Metric for Online Dating

For senders $s \subseteq S$ and receivers $r \subseteq R$

$$H(s,r) = \frac{n^+(s,r)}{n(s,r).\frac{|u_s(r)|}{|u_s(R)|}.\frac{n^+(S,r)}{n(S,r)}}$$

where $n^+(s,r)$ $(n(s,r))$ = the number of positive (all) contacts from s to r

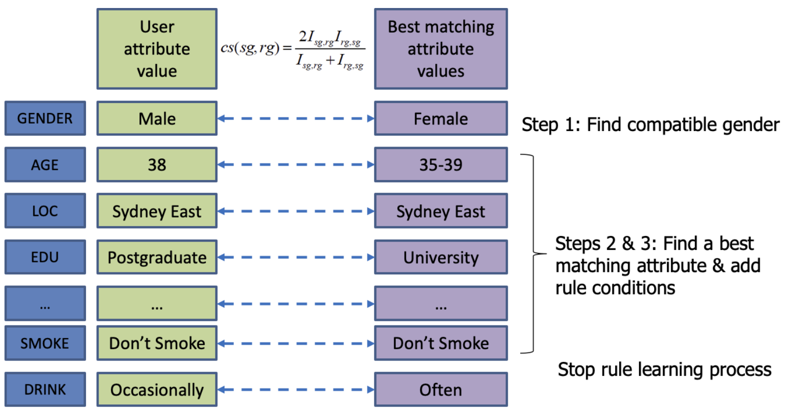
- Expected number of positive contacts from s to r depends on
  1. Activity of the subgroup s as measured by the number of contacts initiated by s
  2. Expected proportion of those contacts that are sent to r, derived from the number of users contacted by s in r compared to R
  3. Expected success rate of those contacts, as estimated by the success rate of all senders S when contacting r

Homophily in Online Dating

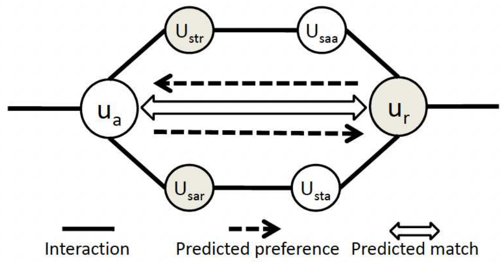
- Age: Male up to 3 years older than female
- Location: Region in Sydney (c.f. socio-economic status)
- Ethnicity: Asian, Middle Eastern, Indian
- Religion: Hindu, Jewish, Born Again Christian
- Education: More education implies higher homophily
- Physical Appearance: Slim/Athletic
- Marital Status: Single, (Separated, Divorced)
- Children and Pets: Children, Want but don't have pets
- Drinking/Diet: Occasional drinkers, Vegetarian/vegan
- Personality: Social
- Politics: Left wing more so than right wing
- Smoking: Non-smoking

Rules Method

For each user attribute value, find best matched values for same attribute



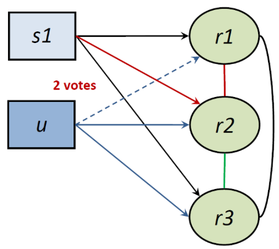
Heterogeneous Link Prediction



u_{sta}: user similar in taste to u_a (both contacted u_{sar})
u_{sta} contacted u_r so u_a would like to contact u_r
u_{saa}: user similar in attractiveness to u_a (u_{str} replied positively to both)
u_r replied positively to u_{saa} so u_r would reply positively to u_a

Interaction-Based Collaborative Filtering

- Based on positive contacts: negative interactions make no difference to results, worse if positive and negative treated the same
- u is similar to $s1$ if they have positive contact in common
- Recommend contacts of users similar to u not already contacted by u
- Ranking is by number of "votes" for candidate $r1$ by similar receivers $r2$ and $r3$

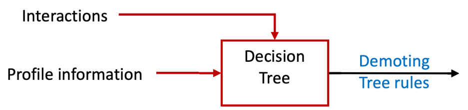


Decision Tree Critic/Popular Users

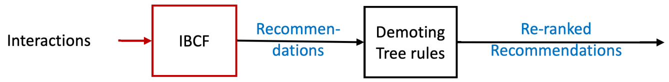
- Example Decision Tree rule
  - If F very popular and M not popular then M→F unsuccessful
- Critic has rules with Success Rate Improvement (SRI) < 1 (demoting)
  - Decision Tree predicts negative interactions accurately
  - Negative rules mostly, but not always, have SRI < 1
- Applying Critic
  - Convert numbers of "votes" into a score between 0 and 1
  - Re-rank all candidates by multiplying this score by rule SRI < 1
  - Effect is to demote candidates popular relative to users

Cascaded Methods

- Building Critic (one off)



- Cascaded Recommender



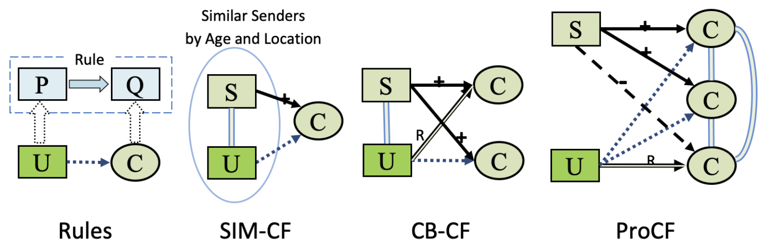
- Decision Tree uses profile features (age, location, body type, etc.) and temporal features (activity, popularity) to generate rules from data
  - Decision Tree model and rule generation done once
  - Rules characterize positive and negative interactions

Hybrid Methods/Cold Start

Exploit rule-based recommender to augment CF

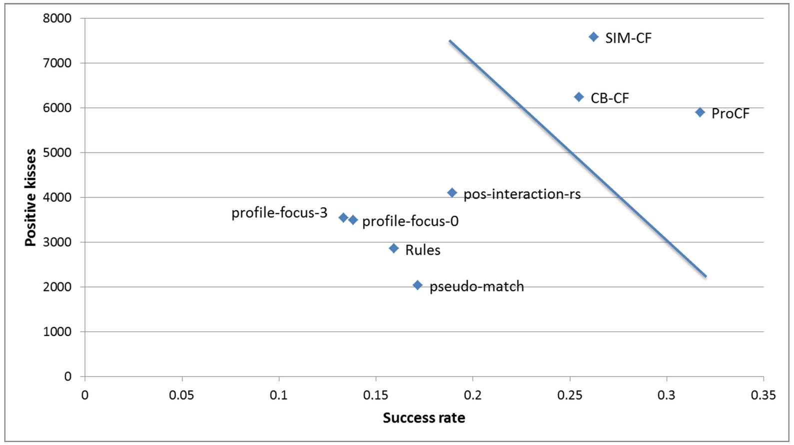
- Hybrid method (SIM-CF)
  - Content-based user similarity + CF + re-rank using critic
  - Senders with similar age and same location as the user
- Content-Boosted method (CB-CF)
  - CF recommendation treating rule-based recommendations as if they were actual contacts + re-rank using critic
- New type of boosted method (ProCF)
  - Probabilistic similarity using rule-based recommendations as if they were actual contacts (no need to re-rank)

Summary of Methods



Rules: C satisfies the conclusion Q of the rule P ⇒ Q generated for U
SIM-CF: C replied positively to a user similar in age and location to U
CB-CF: C replied positively to a user who initiated a successful interaction with a candidate recommended by Rules to U
ProCF: C is a similar candidate to either a Rules recommendation for U or a user who replied positively to U

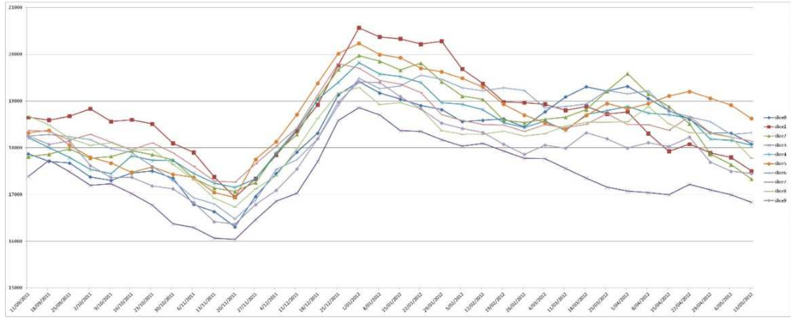
Trial Results



Additional Metrics

Metric	SIM-CF
Lift in communications per active user	5.4%
Contact success rate	26%
Contacts unread or no response	36%
Average/median age difference	3.9/3 years
Average/median distance apart	105/20km
Communications by women/men	780/2069
Communications to non-popular/popular	1125/1724

A/B Testing Pitfall



All slices use same recommendation method!!

Trends consistent going back over a year (e.g. order of slices)

Systematic variation between slices based on user behaviour

Summary

- Much social science interest in social network analysis
  - Confirm hypotheses at scale (homophily, small world)
- Link prediction is not link recommendation
  - But can apply techniques to recommendation
  - And recommendation techniques to link prediction
- Recommender system evaluation requires variety of metrics
  - And a user trial to observe changes in behaviour
- Next lecture sequential/temporal recommendation