Beyond Pagerank: Network Effects Between Web Entities (WWW 2005 Panel Discussion)

Soumen Chakrabarti
IIT Bombay

www.cse.iitb.ac.in/~soumen
Web entities, relations, economy

- Information entities: page, ad, href, iframe
- Real-life artifacts: goods, services—these live in complex attribute spaces
- Actors: searcher, author, search engine, vendor, ad author, ad server
- Many interconnected relations
  - wrote(author,href), received(person,email), paid(vendor,adserver,money), bookmarked(searcher,href,datetime), bought(searcher,product),…
- Emails, blogs, Friendster, Orkut, Tribe, LinkedIn, Yahoo360, del.icio.us, …
Research efforts

How to model interactions and use models for well-motivated optimization problems?

- Network value of customers [Domingos+2001, Kempe+2003]
- Viral marketing, network epidemics and spectral analysis [Wang+2003]
- Recommender systems: content-based, collaborative, cold-start [Schein+2002]
Targeting highly networked customers

Probability of a given reaction of \( i \)-th customer in a social network

Known reaction of pilot customers

Search for \( M \) to maximize expected lift in profit

\[
\text{ELP}(X^k, Y, M) = \sum_i R_i \Pr(X_i = 1|X^k, Y, M) - |M|c
\]

Reward/revenue for converting potential customer \( i \) to purchase

One-shot, greedy, and hill-climbing search for the best \( M \)

Bitvector of marketing campaign over all nodes

Attributes of marketed item and/or relations to attributes of customer \( i \)

Number of 1s in \( M \)

Cost of advertising to one customer (may not be constant)
Epidemics and eigenvalues

If infected, infects each neighbor with probability $\beta$ every time step

If infected, heals with probability $\delta$ every time step

\[
\Pr \nu \text{ infected at } t = 1 - p(\nu, t) = (1 - p(\nu, t - 1))(1 - r(\nu, t - 1)) + p(\nu, t - 1)(1 - r(\nu, t - 1))\delta
\]

\[
1 - r(\nu, t) = \prod_{(u, v) \in E} (1 - \beta p(u, t - 1))
\]

\[
p(t) = ((1 - \delta)I + \beta E)p(t - 1) = Sp(t - 1), \text{ say}
\]

\[
p(t) \text{ dies down to zero as } t \to \infty \text{ if } \lambda_1(S) < 1, \text{ which happens if } \delta / \beta > \lambda_1(E)
\]
Hidden links in recommender systems

Is person $p$ likely to enjoy movie $m$?

$\Pr(p, m) = \sum_z \Pr(p) \Pr(z \mid p) \Pr(m \mid z)$

Person $p$ chooses hidden “genre” $z$

Exploit additional linkage: find movies with casting similar to that of movies user has liked before

First, fit a person-actor model from training data

$\Pr(p, a) = \sum_z \Pr(p) \Pr(z \mid p) \Pr(a \mid z)$

Next, “fold in” a movie via EM

$\Pr(z \mid a, m) \propto \Pr(a \mid z) \Pr(z \mid m)$

$\Pr(z \mid m) \propto \sum_a [a \in m] \Pr(z \mid a, m)$

Final recommendation

$\Pr(p \mid m) = \sum_z \Pr(z \mid m) \Pr(p \mid z)$
Positive feedback and “Googlearchy”

Awareness increases \( A(p,t) \) \( \rightarrow \) Popularity increases \( P(p,t) \) \( \times Q(p) \)

More people visit the page \( V(p,t) \) \( \rightarrow \) Search engine ranks \( p \) highly because it is popular

Avalanche effect

Usual power-law behavior under Googlearchy

Degree distribution

\[ y = 3 \times 10^6 x^{-3.9524} \]

Celebrities on “bestseller list”

“Unintended consequences” of promotions owing to dynamic network effects
Challenges

- Tractable yet reliable models
  - Many types of entities, relations, quantities
  - Simplify without losing the essence
- Large-scale clean-room experiments tough
  - Parameter settings can leave permanent effect
  - Long-term observation, control groups
- Highly non-linear dynamical systems
  - Explosive/logistic growth, burstiness, competition
- Data privacy, capture, scaling
  - Distributed ownership, willful distortions