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]
- Effect of ranking monopolies on social networks [Baeza-Yates+2002, Cho+2004, Chakrabarti+2005, Pandey+2005]
- Recommender systems: content-based, collaborative, cold-start [Schein+2002]

Targeting highly networked customers

$$\Pr(X_i \mid \mathbf{X}^k, \mathbf{Y}, \mathbf{M})$$

Probability of a given reaction of *i*-th customer in a social network

Bitvector of marketing campaign over all nodes

Known reaction of pilot customers

Attributes of marketed item and/or relations to attributes of customer *i*

Search for **M** to maximize expected lift in profit

$$ELP(\mathbf{X}^k, \mathbf{Y}, \mathbf{M}) = \sum_i R_i \Pr(X_i = 1 \mid \mathbf{X}^k, \mathbf{Y}, \mathbf{M}) - |\mathbf{M}|\, c$$

Reward/revenue for converting potential customer *i* to purchase

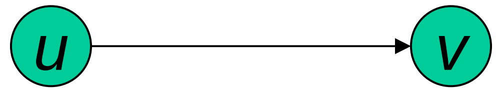
Number of 1s in **M**

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

One-shot, greedy, and hill-climbing search for the best **M**

Epidemics and eigenvalues

If infected, infects each neighbor with probability β every time step



If infected, heals with probability δ every time step

Pr v infected at *t*

Pr *v* gets infection at *t*−1

$$1 - \boxed{p(v,t)} = \left(1 - p(v,t-1)\right)\left(1 - \boxed{r(v,t-1)}\right) + p(v,t-1)\left(1 - r(v,t-1)\right)\delta$$

$$1 - r(v,t) = \boxed{\prod_{(u,v)\in E}\left(1 - \beta\, p(u,t-1)\right)}$$

Pr no neighbor gives infection

$$\mathbf{p}(t) = \left((1-\delta)\mathbf{I} + \beta\mathbf{E}\right)\mathbf{p}(t-1) = \mathbf{S}\mathbf{p}(t-1), \text{ say}$$

Approximate linear propagation model

p(*t*) dies down to zero as $t \to \infty$ if $\lambda_1(\mathbf{S}) < 1$, which happens if

$$\delta / \beta > \lambda_1(\mathbf{E})$$

Hidden links in recommender systems

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

$P \rightarrow Z \rightarrow M$

↓Person \ Movies→				
	0			1
		?	0	
1	0			

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

Person *p* chooses hidden "genre" *z*

Genre induces distrib over movies

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

$$\begin{cases} \Pr(z \mid a,m) \propto \Pr(a \mid z)\Pr(z \mid m) \\ \Pr(z \mid m) \propto \sum_a [a \in m]\Pr(z \mid a,m) \end{cases}$$

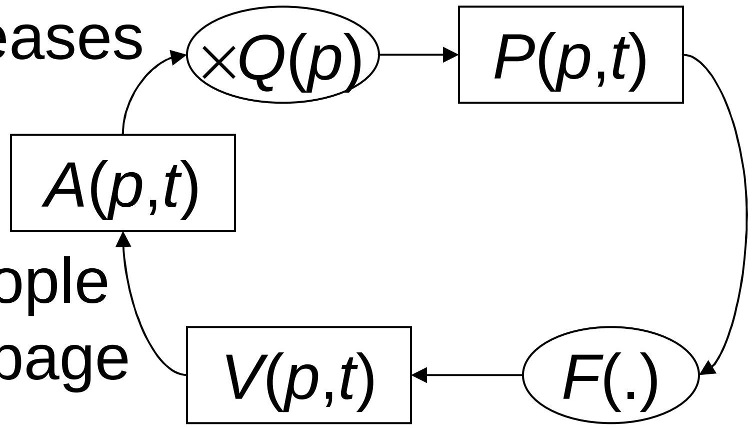
Final recommendation

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

Positive feedback and "Googlearchy"

Awareness increases

Popularity increases

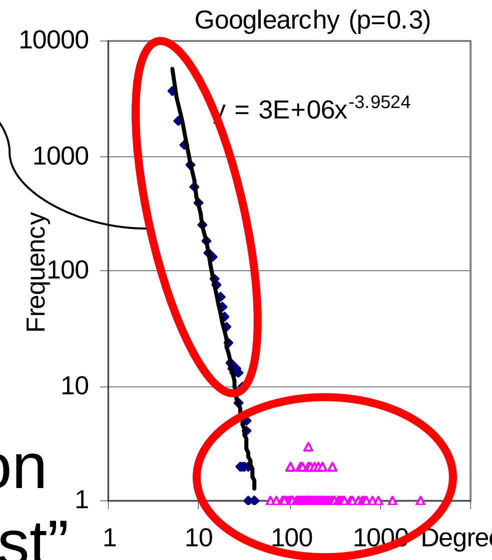


More people visit the page

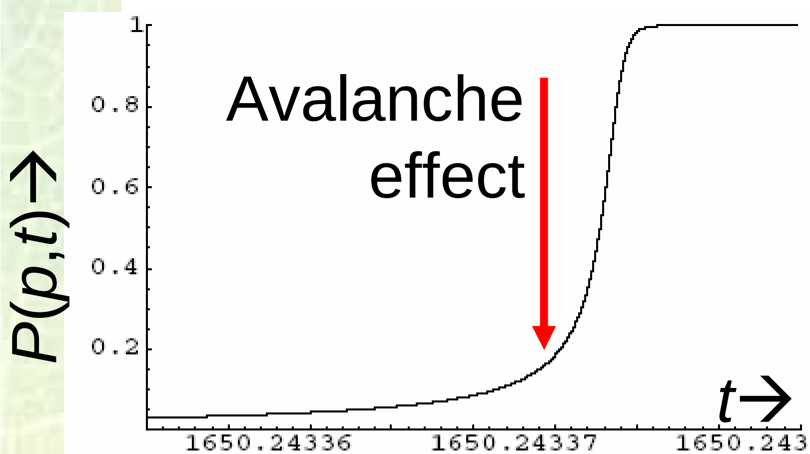
Search engine ranks *p* highly because it is popular

Usual power-law behavior

Degree distribution under Googlearchy



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