Birds of a Feather Shop Together:
Predicting Adoption with Social Networks

Sharad Goel, Yahoo! Research

Joint work with Dan Goldstein

Thanks to Jake Hofman, Randall Lewis, David Reiley, and Duncan Watts

Social Science Research

Traditionally, social science research has been based extensively on:

- Self-Reported Data: Surveys, Polls
- Laboratory Experiments
- "Data-Free" Methods: Simulations, Theoretical Models, Rhetoric

Social Science Research

While these approaches have at times been quite fruitful, they are subject to some fundamental limitations:

- Surveys are better suited for measuring attitudes rather than actions
- Laboratory experiments have problems of external validity
- Difficult to study interacting individuals, group dynamics

Computational Social Science: An Emerging Discipline

Over the past 20–30 years, large-scale data analysis has been successfully integrated into biology and physics

Social scientists have only recently begun to adopt this approach, but the benefits are already apparent

We can now address problems in that social sciences that were largely thought to be intractable ten years ago

Computational Social Science

An Explosion of Data

- Demographic: Age, Gender
- Behavioral: Retail Purchases, Web Browsing
- Network: Email, IM, Facebook

Computational Social Science

An Explosion of Computing Power

- MapReduce, Hadoop, Pig
- Amazon Elastic Compute Cloud (EC2)
- (Amazon Mechanical Turk, Facebook Application Platform)

Predicting Adoption

Predictions regarding the adoption of products and behaviors—from consumer services to health care—are often based on individual-level attributes, for example, age, sex, and geography.

It has generally not been feasible to consider the actions of an individual's social contacts, despite growing evidence that associates often:

1. Influence one another

2. Tend to be similar on a variety of dimensions, leading to comparable propensities to adopt. [McPherson, Smith-Lovin, Cook]

Predicting Adoption

With the increased availability of communication data, such network-based predictions are now possible.

- Hill, Provost & Volinsky (2006) investigate adoption of a new telecommunications service over the phone-contact network

- Bhatt, Chaoji & Parekh (2010) Investigate adoption of a PC-to-phone service over the IM network

Two Questions

1. Can networks be used to predict adoption of products that are not directly related to communication?

2. How well do network-based predictions perform relative to traditional methods?

Results

We examine the extent to which email and IM network data can help predict matters as diverse as:

- Retail spending
- Joining a recreational league
- Reacting to online advertisements

Results

1. When an individual's contacts make purchases, adopt a service, or click on an ad, that is a strong indicator that the consumer will do the same.

2. The marginal value of network information relative to traditional features ranges from negligible to large, depending heavily on the domain and the strength of the baseline.

Data

The Network:

- We define an edge between two individuals if they have recently communicated via Yahoo! Mail or Messenger
- Restrict to the 100M+ people with at least one network contact. Median degree is 3; mean is 11.
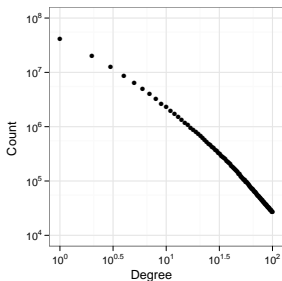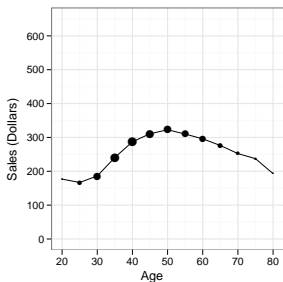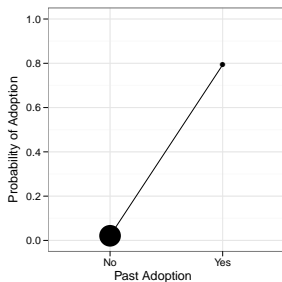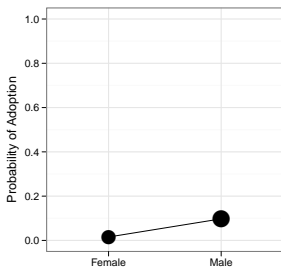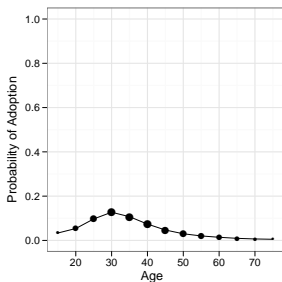
Data

Three Adoption Domains:

- Purchases at an upscale department store chain (600K people)
- Signups for Yahoo! Fantasy Football (10M+ people)
- Clicks on ten front-page display ads (10M+ people)

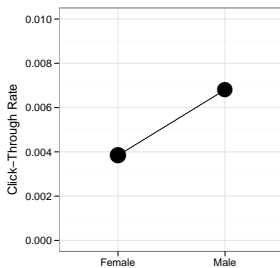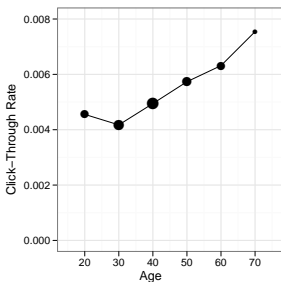For each individual, we know their behavior and the behavior of their contacts.

Sales at an Upscale Department Store
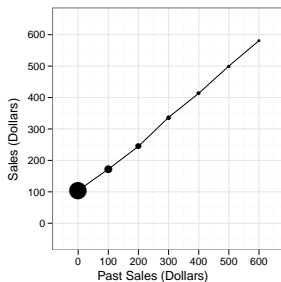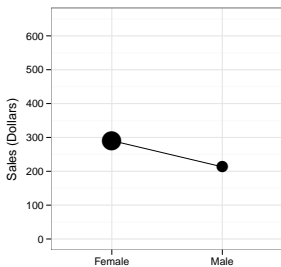
Fantasy Football Signups

Clicks on a Front Page Display Ad
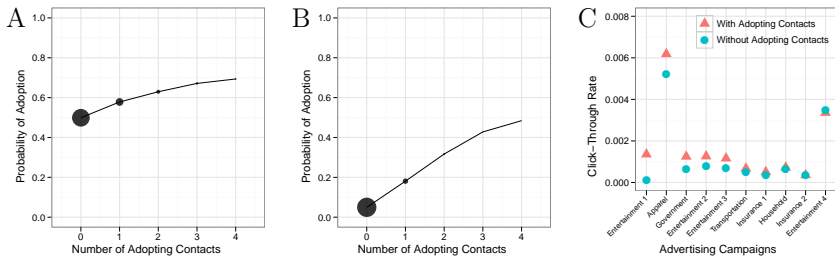
There is a Social Signal

Figure: Department store sales, fantasy football signups, and display ad CTR

The Top-K problem: Identifying the Most-Likely Adopters

It is often useful to identify those individuals most likely to adopt:

- Marketers may find it effective to advertise more heavily—or at least differently—to these high-likelihood candidates
- Health care professionals may attempt targeted interventions

We compare the predictive power of social network data to demographic (age & sex), and demographic-plus-behavioral baselines

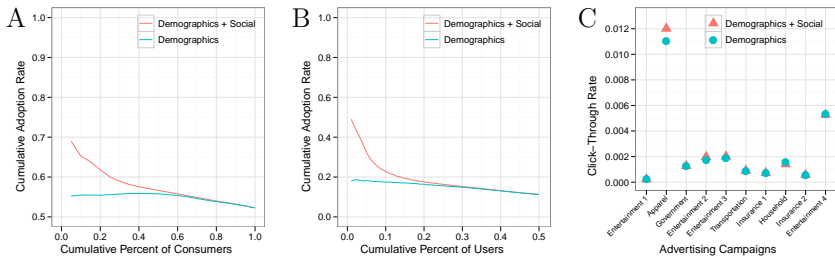## Social Complements Demographic Predictors

Figure: Department store sales, fantasy football signups, and clicks on display ads
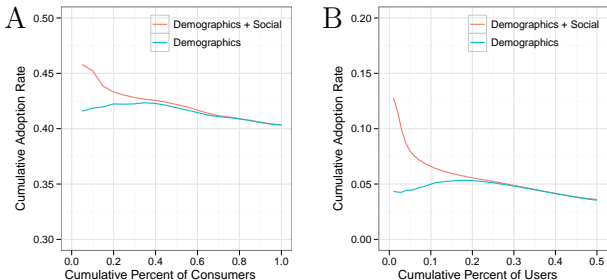
Initial Adoptions

Figure: Department store sales, and fantasy football signups

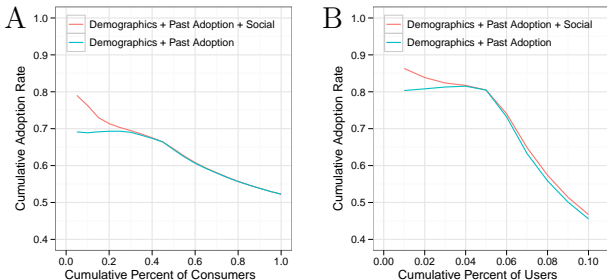## Social Complements Demographics-Plus-Past-Adoption

Figure: Department store sales, and fantasy football signups

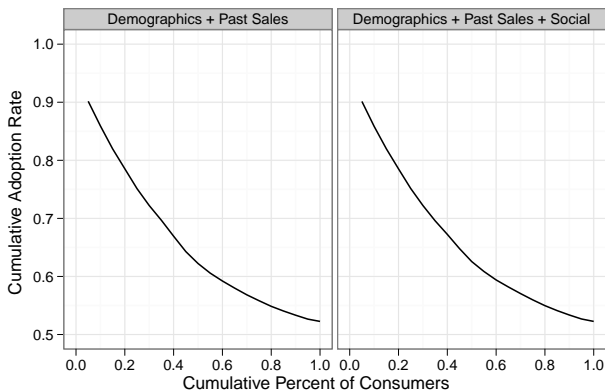## Transactional Data Is Hard to Beat

Figure: Department store sales

A Social Marketing Strategy?

In identifying those most likely to adopt, we've implicitly assumed some individuals have already adopted. What if we're starting from scratch (i.e., no adopters)?

1. Use standard demographic and behavioral targeting to find the first wave of adopters

2. Target the friends of those adopters (really, identify top candidates based on social, demo, behavioral, etc.)

A complication: When adoption rates are low and degrees are small, friends of adopters form a relatively small set. The performance of the strategy can then be modest.

A Social Marketing Strategy?

Consider the following scenario:

1. Average adoption rate in the population is p_0

2. Average adoption rate of those with an adopting friend is p_1 ($p_1 > p_0$)

3. Each consumer has k friends

A simple calculation shows that the average adoption rate of those targeted by the social marketing strategy is:

$$\tilde{p} = p_0 \left(\frac{1 + kp_1}{1 + kp_0} \right)$$

For $p_0 = 0.01$, $p_1 = 0.02$, $k = 10$: $\tilde{p} = 1.09 p_0$

Summary

1. Your contacts' actions are often quite informative of your own (correlation, not necessarily causation).

2. Social features often complement—do not merely substitute for—demographic and behavioral data. There is substantial variation across domains, and the marginal value of social ranges from negligible to large

3. When you can't bootstrap off an existing pool of adopters, a social marketing strategy may yield relatively modest improvements.

Blog Post:
messymatters.com/birdshop