Network Structure and the Long Tail of Electronic Commerce

Arun Sundararajan
(joint work with Gal Oestreicher-Singer)

http://oz.stern.nyu.edu/
Economic ecommerce networks are informative

- Economic networks: generated by actions taken by (human) agents in commercial or economic situations
- Contain information about complex consumer preferences and product characteristics that is explanatory/predictive

Today’s talk

- Uses a component of Amazon’s co-purchase network
- (Weighted) PageRank: measures the “importance” of the network
- Gini coefficient: measures demand inequity
- Compute PageRank and Gini aggregated by category
- Higher PageRank $\Rightarrow$ Lower Gini
- One interpretation: a more influential recommendation network is associated with a flatter demand distribution
Summary of data

Information Rules: A Strategic Guide to the Network Economy (Hardcover)
by Carl Shapiro, Hal R. Varian

List Price: $35.00
Price: $22.05 & eligible for FREE Super Saver Shipping on orders over $25. Details
You Save: $12.95 (37%)
Availability: Usually ships within 24 hours. Ships from and sold by Amazon.com.
Want it delivered Tuesday, May 9? Order it in the next 89 hours and 2 minutes, and choose One-Day Shipping at checkout. See details

142 used & new available from $3.72

Also Available in: List Price: Our Price: Other Offers:
Digital (Download: Microsoft Reader) $29.95 $18.87

Customers who bought this item also bought
The Economics of Information Technology: An Introduction (Raffaele Mattioli Lectures) by Hal R. Varian
Re-Thinking the Network Economy: The True Forces That Drive the Digital Marketplace by Stan Liebowitz
Crossing the Chasm by Geoffrey A. Moore
The Innovator's Dilemma: The Revolutionary Book That Will Change the Way You Do Business (HarperBusiness Essentials) by Clayton M. Christensen
The Economics of Network Industries by Oz Shy
**Summary of data**

**Connected component of co-purchase network**

- Gathered using Amazon’s XML service
- Depth-first crawl starting from popular book
- One graph per day, midnight
- Began August 2005, currently have about 600 graphs collected
Summary of data

**Primary data**

- Invariant: ASIN, Title, Author, List Price, Category, Number of pages, Release Date
- Daily: Sale Price, Sales Rank (3-hour intervals), identity of neighbors, secondary market prices

**Derived data**

- Daily demand (book level)
- Weighted PageRank (book level)
- Gini coefficient (category level)
Network evolution: summary

**A**
- Number of nodes (left scale)
- % of new nodes (right scale)

**B**
Number of new edges, as a % of total number of edges in the network

- **Between two new nodes**
- **With a new “source” node**
- **With a new “sink” node**
- **Between pre-existing nodes**
Network evolution: Degree distributions

In-degree ($k$) vs. probability ($p(k)$)

- 2/1
- 2/9
- 2/17
- 2/26
Properties of co-purchase networks

The network is very highly clustered

Distribution of clustering coefficients for a sample day

Average clustering coefficients over a month
Properties of co-purchase networks

Average shortest paths: 19 degrees of separation...

Distribution of average distances for a sample day

Average average distances over one month

- All nodes
- “Connected” nodes
Relating network structure, demand equity

Outline of approach

• (Weighted) PageRank: measures the “importance” of the network
• Gini coefficient: measures demand inequity
• Compute each aggregated by category
• Associate variations in PageRank with variations in Gini

Data

• 28 daily co-purchase networks
• Analyzed individually, also combined into four 7-day composite networks with edge weights
(Weighted) PageRank

Measure of the influence of the entire network

- Underlying model: Random surfer
- (Approximately) the steady-state probability of arriving at a particular page on the graph when surfing randomly
- Usually used as a network-based measure of page importance...
- ...but can also be interpreted as a measure of the extent to which a network structure influences a page

\[
\text{PageRank}(i) = \frac{(1 - \alpha)}{n} + \alpha \sum_{j \in G(i)} \text{Weight}(j,i) \left( \frac{\text{PageRank}(j)}{\text{OutDegree}(j)} \right)
\]

- For the daily networks, weights are all 1
SalesRank versus PageRank

(displayed for a sample)
The Gini coefficient

- Captures the extent to which demand is concentrated among the highest selling products in a group
- Measured by the area above the Lorenz curve

![Diagram](image-url)
**Demand inequality: the Gini coefficient**

**The Gini coefficient**

- Calculated for 200+ categories, comprising at least 100 books
- Compute average PageRank for each category as well

![Graph A: Science: Chemistry](image)

![Graph B: Computers and Internet: Web Development](image)
Network structure and the “long tail”

(recall: lower GINI => flatter demand, longer tail)

\[
\log[\text{GINI}] = a + b_1 \log[\text{AVGDEMAND}] + b_2 \log[\text{AVGPAGERANK}] \\
+ b_3 \log[\text{PAGERANKVAR}] + b_4 \log[\text{SIZE}] + b_5 \log[\text{AMIXING}]
\]

AVGDEMAND: Average demand for books in the category

AVGPAGERANK: Average PageRank for books in the category

PAGERANKVAR: Variance in PageRank across books in the category

SIZE: Number of books in the category

AMIXING: Fraction of co-purchase links to books within the same category
Network structure and the “long tail”

\[ \log[\text{GINI}] = a + b_1 \log[\text{AVGDEMAND}] + b_2 \log[\text{AVG PAGERANK}] \\
+ b_3 \log[\text{PAGERANKVAR}] + b_4 \log[\text{SIZE}] + b_5 \log[\text{AMIXING}] \]
Network structure and the “long tail”

Products with lower Gini coefficients are those whose demand manifests a more prominent “long-tail”
Network structure and the long tail

Trade-press intuition for our findings

Higher average PageRank =>
Network has more influence on traffic? =>
more equal demand distribution
Economic ecommerce networks are informative

- Economic networks: generated by actions taken by (human) agents in economic situations
- Contain information about complex consumer preferences and product characteristics that is explanatory/predictive

Ongoing research

- Network characteristics (distance, clustering) and outcome variations (random utility versus location models of demand).
- Structural econometric model of network “peer effects”.
- Economic networks and demand prediction
- Identifying the influence of the network on individual products

http://oz.stern.nyu.edu/