Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2009 Proceedings

International Conference on Information Systems (ICIS)

2009

Spreading the Oprah Effect: The Diffusion of Demand Shocks in a Recommendation Network

Eyal Carmi
Tel-Aviv University, carmieya@post.tau.ac.il

Gal Oestreicher-Singer
Tel-Aviv University, galos@post.tau.ac.il

Arun Sundararajan
New York University, asundara@stern.nyu.edu

SPREADING THE OPRAH EFFECT: THE DIFFUSION OF DEMAND SHOCKS IN A RECOMMENDATION NETWORK

Research-in-Progress

Eyal Carmi
Recanati School of Business
Tel-Aviv University
carmieya@post.tau.ac.il

Gal Oestreicher-Singer
Recanati School of Business
Tel-Aviv University
galos@post.tau.ac.il

Arun Sundararajan
Stern School of Business
New York University
asundara@stern.nyu.edu

Abstract

We study the magnitude and persistence of the diffusion of exogenous demand shocks on an ecommerce recommendation network. The demand shocks are generated by book reviews on the Oprah Winfrey Show and in the NYTimes, and the recommendation network is generated by Amazon's copurchase network. We find a strikingly high level of diffusion of exogenous shock through such networks. Neighboring books experience a dramatic increase in their demand levels, even though they are not actually featured on the review. An average of 40% of neighbors, even 4 clicks away see a statistically significant increase in their demand levels; this effect is indicative of the depth of contagion in online recommendation networks following exogenous shocks. We also document how clustered networks "trap" a higher fraction of the contagion closer to the reviewed book, and we provide summaries of the persistence and relative magnitude of the demand inflation of the neighborhood.

Keywords: Social networks, diffusion, economics of IS, electronic commerce, econometrics

Introduction, Research Questions and Related Work

We study the rate and persistence of the diffusion of exogenous demand shocks through the co-purchase recommendation network on Amazon.com. Our goal is to understand whether such online networks form a basis for the contagion of exogenous economic events, and if so, what the magnitude of this influence/contagion is; how far the shocks "spread"; how long they persist; how the magnitude and persistence of the shock is affected by the clustering and local structure of the underlying recommendation network; how these shocks alter the structure of online networks; and under what conditions the shock leads to a fundamental shift in the demand curve for the product rather than simply being a fad-like shift.

Our work is motivated by the observation that online commercial and social interactions have increased dramatically over the past decade, and an important by-product of this process is the emergence of different types of *visible electronic networks*. For example, some online *social* networks make visible relationships between individuals who are friends, colleagues, or trading partners (facebook.com, myspace.com, Friendster.com, just to name a few). More importantly in our opinion, electronic commerce sites are organized as a collection of web pages, each featuring a single product (e.g. book, video, or a content item) which are linked by hyperlinks to other products' pages, creating a *product network* where the products are the nodes. The set of pages that a cognitively bounded consumer pays attention to is determined in part by the structure of such product networks.

One of the oldest examples of an electronically visible product network is the "copurchase" network of Amazon.com. Beyond guiding consumer attention, these networks are also often generated by aggregated consumer actions rather than being predetermined by the firm. Therefore, those networks could potentially change dynamically with shifts in consumers' taste. Both of these observations frame the program of research described in this summary.

Our research questions are summarized as follows:

1. To what extent do exogenous shocks diffuse through an online product network? How far do they propagate, how much do they affect demand of neighboring products, and how long do they persist?
2. The local structure of the network can vary significantly. To what extent can variations in the distance, persistence and magnitude of these diffusions be explained by structural properties of the underlying online product network?
3. To what extent do exogenous shocks alter the structure of online product networks?

Figure 1. An example of the co-purchase links on Amazon.com

The bulk of the findings in this document address the first question; we summarize some preliminary findings on question 2, while question 3 currently remains in progress.

Our research adds to a growing body of work in the IS literature that studies the impact of "social" networks on business outcomes. Prior IS work has established that online social networks can explain the productivity of information workers (Aral et al. 2006), the transfer of organizational knowledge (Huang and DeSanctis 2005; Wu et al. 2008), the adoption of new IT-based products and services (Hill et al. 2006), the impact of product reviews on ecommerce outcomes (Forman et al. 2008), and the success of collaborative open source projects (Singh 2007).

The influence of social networks on diffusion of a product has been studied extensively, mainly in the marketing literature, and the structure of the social network has been linked to the diffusion process (for example, Goldenberg et al. (2009) study the role of hubs in the adoption process, and Watts and Dodds (2007) study the role of "influentials" in global cascades). Most related to our work are studies on multi-product diffusion (e.g., Chintagunta and Haldar 1998; Libai et al. 2008), which measure correlations in sales among products or product categories. However, researchers have traditionally focused on small sets of similar products. Studying cross-selling is one of the main challenges in the field of direct marketing, in which complementarity, co-incidence and heterogeneity are conceptualized as leading to co-purchasing decisions (Edwards and Allenby 2003; Manchanda et al. 1999). To the best of our knowledge, our work is novel in examining how product networks affect multi-product diffusion on a large scale. A more detailed review of this literature is available on request.

Despite the extensive study of social networks, researchers have directed limited attention towards online *product* networks. This is perhaps surprising, given the ubiquity of product networks and their potential influence on business rather than on organizational outcomes. Recent studies of product networks include research by Oh et al. (2008), who study the network of videos on YouTube, linking the social network structure to the diffusion of content; Mayzlin and Yoganarasimhan (2008), who relate network centrality of hyperlinked competing blogs to success and who show that superstar blogs are more likely to have incoming links; and Oestreicher-Singer and Sundararajan (2008), who study how recommendations might affect the demand distribution of products sold online and contribute towards the “long tail” of electronic commerce. Our work contributes a new perspective to this stream of research by focusing on how such networks mediate the diffusion of *new attention* that is brought to an ecommerce site on account of an exogenous event.

A related literature examines the effect of external endorsements on demand in the context of traditional commerce (Boatwright et al. 2007; Reinstein and Snyder 2005) and e-commerce (Deschatres and Sornette 2005; Sorensen and Rasmussen 2004; Sornette et al. 2004). Specifically, Oprah Winfrey’s endorsement was shown to have a powerful economic (and political) impact (Balogh 2008; Illouz 2003; Rooney 2005). Sorensen and Rasmussen (2004) have shown that the publication of a book review in the *New York Times* newspaper is followed by a significant increase in the sales of the reviewed book. Deschatres and Sornette (2005) have also shown that book reviews act as exogenous shocks in the context of sales of the reviewed book.

Data and Constructed Variables

We use data about the daily product, pricing, demand and "network" information for over 250,000 books sold on Amazon.com. Each product on Amazon.com has an associated webpage. Those pages each have a set of "co-purchase links", which are hyperlinks to the set of products that were co-purchased most frequently with this product on Amazon.com. This set is limited to 5 items and is listed under the title "Customers who bought this item also bought …" (See Figure 1 for an illustration).

Figure 2. Left: Screen shot from the "Books Seen on the Show" page on Oprah.com. As of mid-2008, "Books Seen on the Show" data are no longer publicly available on the website. Right: Screen shot of the "Sunday Book Review" section of the online edition of the *New York Times* (NYTimes.com).

Conceptually, the co-purchase network is a directed graph in which nodes correspond to products, and edges to directed co-purchase links. Data about this graph are collected using a Java-based crawler, which starts from a popular book and follows the co-purchase links using a depth-first algorithm. At each page, the crawler gathers and records information for the book whose webpage it is on, as well as the co-purchase links on that page, and terminates when the entire connected component of the graph is collected. This process is repeated daily.

We use data collected between January 2006 and June 2008 for this study. We collected the following data for each book on each day: Amazon Standard Identification Number (ASIN; a unique serial number given to each book by Amazon.com), List Price, Sale Price (the price on the Amazon.com website that day), co-purchases (ASINs of the books that appear on its co-purchases list, 5 ASINs per book), SalesRank (a number associated with each product on Amazon.com, which measures its demand relative to that of other products), authors, and category affiliation.

We also collected information about book reviews that appeared on the *Oprah Winfrey Show* (*"Oprah*") by parsing the corresponding dedicated webpages on the Oprah.com website (see Figure 2). Our dataset ranges from January 2006 to April 2008 and includes more than 400 book reviews. In addition, we collected data about book reviews that appeared between January 2006 and June 2008 in the "Sunday Book Review" section of the *New York Times* by parsing the corresponding dedicated webpages on NYTimes.com; this set comprises more than 2000 book reviews. We collected the title, author and review date using a PHP-based crawler, manually verified the information, and then cross-referenced it with the corresponding network and sales data from Amazon.com.

We have developed several different measures to describe the magnitude of the shock, the network structure, and the link parameters.

SalesRankRatio (SRR): Provides a measure of the magnitude of the event at time *t* and is defined as:

$$SRR_{i,t} = \overline{SR_i} / SR_{i,t}$$

$SR_{i,t}$ is the average daily SalesRank on day *t* of the i-th book.

SalesRankShock (SRS): Measures the change in the SalesRank (SR) of a book following the exogenous shock (the broadcast or printing of the book review). Its formulation is based on the assumption that each book has an "average normal level" of sales prior to the review event. (A similar method is used to assess the magnitude of response in "event" studies.) The SRS is constructed in the following way:

$$SRS_i = \frac{\overline{SalesRank_i}}{SalesRank_{Peak,i}} \equiv \frac{\overline{SR_i}}{SR_{Peak,i}},$$

where $\overline{SR_i}$ is the "average normal level" of the book's SalesRank computed over the two weeks prior to the review date, and $SR_{Peak,i}$ is the peak SalesRank the book reached during a 100-hour window after the review.

Persistency: To measure the persistency of the shock, we define a threshold on the book's SalesRank (or on its SRR) and test how long it takes for the book's SalesRank to go back to within one standard deviation of its average pre-event level. After computing the pre-event mean $\overline{SR_i}$ and standard deviation σ_{SR_i} of each book *i*, we calculate the number of days (post-event) until the SalesRank of the book first exceeds $\overline{SR_i} - \sigma_{SR_i}$. See Figure 3 for an example of SRR and Persistency.

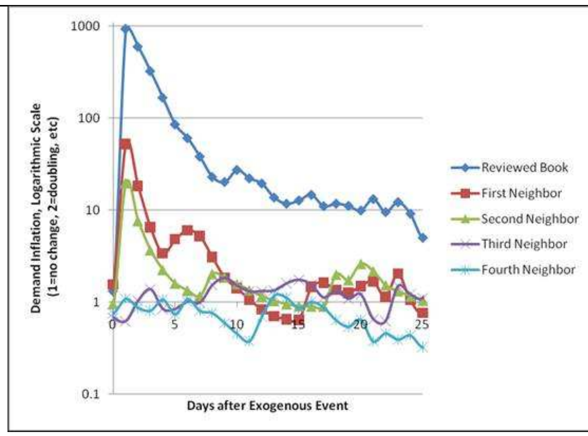

Figure 3. SRR and Persistency for a sample set of neighbors relative to their source.

Aftershock Ratio: Measures which fraction of the shock received by the reviewed book was passed on to neighboring books. This variable is computed by comparing the peak shocks of the reviewed book and of its neighbor as follows:

$$AftershockRatio_i \equiv SRS_i / SRS_{Reviewed},$$

where $SRS_{Reviewed}$ is the reviewed book's *SRS*.

Distance: The number of links on the shortest path from a reviewed book to another book in its network. Therefore, the reviewed book has a distance of *0*, its first neighbors have a distance of *1*, and the second neighbors (books that are two clicks away) have a distance of *2* and so on.

Network Proximity: The simple Distance variable takes into account only the shortest path to a book across the network but does not provide an assessment of how "close" the neighboring book N is to the reviewed book B based on *all possible* paths between N and B. Network proximity addresses this by providing a normalized assessment of how much attention could potentially flow from one book to another based on a weighted summation of all paths.

$$\mathrm{Pr}\,oximity_i = \sum_{k=0} \frac{N_{ik}}{5^k},$$

where N_{ik} is the number of times the i-th book is k clicks away from the reviewed book (*k-neighbor)*. Recall that each book in our network has 5 outgoing links, hence the choice of denominator[1].

Results and Research Plan

Depth of diffusion: A first measure of the depth of diffusion of the exogenous shock in the online product network can be obtained by comparing the average SRR of the reviewed book and of its neighbors in the immediate period (3 days) following the review, to the average SRR of the same products in the immediate period (3 days) prior to the review. We find that on average, first and second neighbors experience a demand shock that is both economically and statistically significant immediately following the event. The demand shock for the third and fourth neighbors seems relatively small in magnitude but is still statistically significant (See Table 1).

Table 1. One-sample Paired T-test results for the SRR before review with those observed after the review date, by distance from the reviewed book.

Distance	3 Days Mean SRR (Before the review)	3 Days Mean SRR (After the review)	P-Value (1% level)
0	1.22	25.34	0.000
1	0.98	2.78	0.001
2	1.01	1.28	0.000
3	1.03	1.18	0.000
4	1.05	1.12	0.000

A closer look at the distribution of this ratio across the different bins of neighbors show a large variance in the magnitude of the shock between books with the same distance, a point we return to later in the summary.

Persistence of diffusion: Our preliminary results suggest that the observed demand shock is *remarkably persistent*. On average, it takes three weeks for the demand of a reviewed book to return to within one standard deviation of its pre-event levels, although there is considerable variation across books. First neighbors of books reviewed on *Oprah* experience a persistency of 7 days on average, or roughly 30% of the persistency observed for reviewed books, as indicated in Table 2. To better understand the influence of the shock, we divided the books into two groups according to whether or not they showed a measurable reaction to the exogenous shock (See Table 2). For books reviewed on *Oprah*, **56%** and **42%** of the first and second neighbors showed positive persistency duration, as did about **40%** of the third and the fourth neighbors. The latter results are striking in their indication of how deep these demand shocks can travel over the recommendation network; nevertheless, there seems to be a considerable variation across books. The distribution of persistency is illustrated in Figure 4A.

Relative magnitude of shock to reviewed book and its neighbors: Figure 4B illustrates the extent to which the magnitude of the demand shock to the reviewed book compares to that of its neighbors. Interestingly, for those neighboring books that are affected by the exogenous shock, there is a considerable fraction that experiences demand inflation comparable to that of the reviewed book. In general, this ratio is higher for first neighbors, but what is striking is that there is very little variation in the distribution of aftershock ratio across second, third and fourth neighbors. This provides further evidence of the extent to which the shock diffuses through the product network in an economically significant way.

[1] The proximity measure assumes all outgoing links are equal.

Table 2. Persistency and Shock statistics for: (a) all books (b) books affected by the exogenous shock; and (c) books that were not affected by the exogenous shock.

Source	Distance	(A) All Books			(B) Affected Books				(C) Books Not Affected		
		#	Persistence (days)	Average SR Shock	#	%	Persistence (days)	Average SR Shock	#	%	Average SR Shock
NYT	0	28	19.25	41.05	28	100%	19.25	41.05	0	0%	0.00
	1	132	2.14	2.43	61	46%	4.62	3.99	71	54%	1.08
	2	415	1.71	1.87	165	40%	4.30	2.95	250	60%	1.15
	3	1046	1.18	1.80	400	38%	3.09	2.88	646	62%	1.13
	4	2487	1.25	1.82	965	39%	3.21	2.90	1522	61%	1.13
Oprah	0	28	21.04	87.23	28	100%	21.04	87.23	0	0%	0.00
	1	119	5.63	6.02	64	54%	10.47	10.21	55	46%	1.15
	2	295	2.00	2.20	121	41%	2.88	3.75	174	59%	1.12
	3	605	1.42	1.88	242	40%	3.55	3.05	363	60%	1.10
	4	1145	1.11	1.77	440	38%	2.90	2.84	705	62%	1.10

A. Persistency **B. Aftershock Ratio**

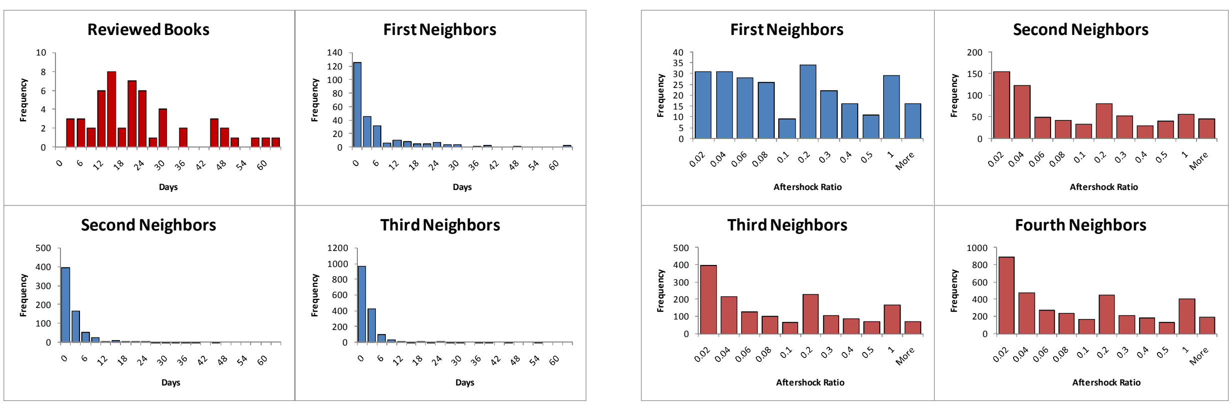

Figure 4. (A) Distribution of persistence, i.e., the number of post-event days for which demand for a book remained one standard deviation above the pre-event average demand, for reviewed books and for first, second and third network neighbors is shown. (B) Distribution for the aftershock ratio for first, second, third and fourth neighbors. The bins for the latter are 0-0.02, 0.02-0.04,…, then 0.1-0.2, 0.2-0.3 and so on.

Some of our other results are summarized below:

1. We find a consistent relationship between the shape of the diffusion curve and the level of clustering in the network. For first and second neighbors, a higher level of clustering in the neighborhood of the reviewed book is associated with both a higher level of SRS (the magnitude of the shock) as well as a higher persistence (the duration for which the effect is observed). However, the opposite correlation is observed for the third and fourth neighbors. This suggests a *fishing net effect* that we are exploring in detail – the triads and cycles created by higher clustering "trap" a larger proportion of the diffused influence closer to the reviewed book.
2. We have computed the total increase in demand for each reviewed book and compared it to the total increase in demand for the book's neighbors. We find that there are two distinct subsets of reviewed books: those whose

demand increase is substantially higher than the total increase for its neighbors, and those for which the total increase in demand for the neighbors is an order of magnitude higher. For both sets, the fraction of demand retained by the reviewed book is positively correlated with clustering. We are investigating this further with the goal of understanding what partitions these sets.

Some salient aspects of our ongoing research and our path forward are summarized below:

1. We are developing an adapted structural duration model (analogous to a hazard rate model) that will allow us to provide more robust estimates of the magnitude and persistence of the duration by interpreting coefficients that assess the probability that the demand has returned to pre-shock levels at different points in time for different categories of neighbors (partitioned as outlined above).
2. We are in the early stages of developing a graph-theoretic analytical model of the process of diffusion that captures consumer attention and choice. We hope to calibrate this model and report on parameters that describe the diffusion process more rigorously, assessing different model specifications on the basis of their fit with the data we have.
3. We are investigating the connections between network properties (such as clustering and triadic closure) in more detail, with the goal of assessing, among other things, whether "fishing net"–like network structures provide higher total revenue to sellers following exogenous shocks, in comparison with tree-like network structures. This analysis will provide important new design guidelines for ecommerce web sites.
4. We are documenting the structural network *changes* in the immediate neighborhood of the reviewed book following the demand shock. This analysis is at a relatively early stage.

Items (1) - (3) are progressing well, and we are confident of reporting on their findings by ICIS 2009. Item (4) is still at a relatively early stage, so we hope to report on some preliminary findings by ICIS 2009.

References

Aral, S., Brynjolfsson, E., and Van Alstyne, M.W. 2006. "Information, Technology and Information Worker Productivity: Task Level Evidence", in *Proceedings of the 2005 International Conference on Information Systems*.

Balogh, S. 2008. "Oprah Winfrey Rallies Voters for Barack Obama." *Heraldsun.com.au* Retrieved February 05, 2008, from http://www.news.com.au/heraldsun/story/0,21985,23159350-663,00.html?from=mostpop

Boatwright, P., Basuroy, S., and Kamakura, W. 2007. "Reviewing the Reviewers: The Impact of Individual Film Critics on Box Office Performance," *Quantitative Marketing and Economics* (5:4), pp 401-425.

Chintagunta, P.K., and Haldar, S. 1998. "Investigating Purchase Timing Behavior in Two Related Product Categories," *Journal of Marketing Research*), pp 43-53.

Deschatres, F., and Sornette, D. 2005. "Dynamics of Book Sales: Endogenous Versus Exogenous Shocks in Complex Networks," *Physical Review E* (72:1), p 016112.

Edwards, Y.D., and Allenby, G.M. 2003. "Multivariate Analysis of Multiple Response Data," *Journal of Marketing Research* (40:3), pp 321–334.

Forman, C., Ghose, A., and Wiesenfeld, B. 2008. "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research* (19:3), pp 291-313.

Goldenberg, J., Han, S., Lehmann, D., and Hong, J. 2009. "The Role of Hubs in the Adoption Process," *Journal of Marketing* (73:2), pp 1-13.

Hill, S., Provost, F., and Volinsky, C. 2006. "Network-Based Marketing: Identifying Likely Adopters Via Consumer Networks," *Statist. Sci.* (21:2), pp 256--276.

Huang, S., and DeSanctis, G. 2005. "Mobilizing Informational Social Capital in Cyber Space: Online Social Network Structural Properties and Knowledge Sharing", in *Proceedings of the 2005 International Conference on Information Systems*.

Illouz, E. 2003. *Oprah Winfrey and the Glamour of Misery: An Essay on Popular Culture*. New York: New York: Columbia University Press.

Libai, B., Muller, E., and Peres, R. 2008. "The Role of within-Brand and Cross-Brand Communications in Competitive Growth," *Journal of Marketing*).

Manchanda, P., Ansari, A., and Gupta, S. 1999. "The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions," *Marketing Science* (18:2), pp 95-114.

Mayzlin, D., and Yoganarasimhan, H. 2008. "Link to Success: How Blogs Build an Audience by Promoting Rivals." Working Paper, Yale School of Management.

Oestreicher-Singer, G., and Sundararajan, A. 2008. *The Visible Hand of Social Networks in Electronic Markets*. SSRN.

Oh, J., Susarla, A., and Tan, Y. 2008. "Examining the Diffusion of User-Generated Content in Online Social Networks." SSRN.

Reinstein, D.A., and Snyder, C.M. 2005. "The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics," *Journal of Industrial Economics* (53:1), March 2005, pp 27-51.

Rooney, K. 2005. *Reading with Oprah: The Book Club That Changed America*. Fayetteville: University of Arkansas Press.

Singh, P.V. 2007. "The Small World Effect: The Influence of Macro Level Properties of Developer Collaboration Networks on Open Source Project Success", in *Proceedings of the 2007 International Conference on Information Systems*.

Sorensen, A., and Rasmussen, S. 2004. "Is Any Publicity Good Publicity? A Note on the Impact of Book Reviews," in: *Working Paper*.

Sornette, D., Deschatres, F., Gilbert, T., and Ageon, Y. 2004. "Endogenous Versus Exogenous Shocks in Complex Networks: An Empirical Test Using Book Sale Rankings," *Physical Review Letters* (93:22), p 228701.

Watts, D.J., and Dodds, P.S. 2007. "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research* (34:4), pp 441-458.

Wu, L., Waber, B.N., Aral, S., Brynjolfsson, E., and Pentland, A. 2008. "Mining Face-to-Face Interaction Networks Using Sociometric Badges: Predicting Productivity in an It Configuration Task", in *Proceedings of the 2008 International Conference on Information Systems*.