

# The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets

Gal Oestreicher-Singer

Recanati Business School, Tel Aviv University, Tel Aviv 69978, Israel,  
galos@tau.ac.il

Arun Sundararajan

Stern School of Business, New York University, New York, New York 10012, asundara@stern.nyu.edu

Online commercial interactions have increased dramatically over the last decade, leading to the emergence of networks that link the electronic commerce landing pages of related products to one another. Our paper conjectures that the explicit visibility of such “product networks” can alter demand spillovers across their constituent items. We test this conjecture empirically using data about the copurchase networks and demand levels associated with more than 250,000 interconnected books offered on Amazon.com over the period of one year while controlling for alternative explanations of demand correlation using a variety of approaches. Our findings suggest that on average the explicit visibility of a copurchase relationship can lead to up to an average threefold amplification of the influence that complementary products have on each others’ demand levels. We also find that newer and more popular products “use” the attention they garner from their network position more efficiently and that diversity in the sources of spillover further amplifies the demand effects of the recommendation network. Our paper presents new evidence quantifying the role of network position in electronic markets and highlights the power of basing (virtual) shelf position on consumer preferences that are explicitly revealed through shared purchasing patterns.

*Key words:* social network; social media; peer effects; homophily; slotting; selection; electronic markets; electronic commerce; product networks

*History:* Received November 19, 2008; accepted January 13, 2012, by Pradeep Chintagunta, marketing.  
Published online in *Articles in Advance* June 15, 2012.

## 1. Introduction

Online commercial interactions have increased dramatically over the last decade, leading to the emergence of visible *product networks* that explicitly link related products to one another. Most electronic commerce sites are organized as a collection of webpages, each featuring one focal product (for example, a book, a DVD, or a computer). These product pages are hyperlinked to other product pages, creating a network whose nodes are individual products. Perhaps the oldest example of a visible electronic “product network” is the copurchase network on Amazon.com (illustrated in Figure 1).<sup>1</sup>

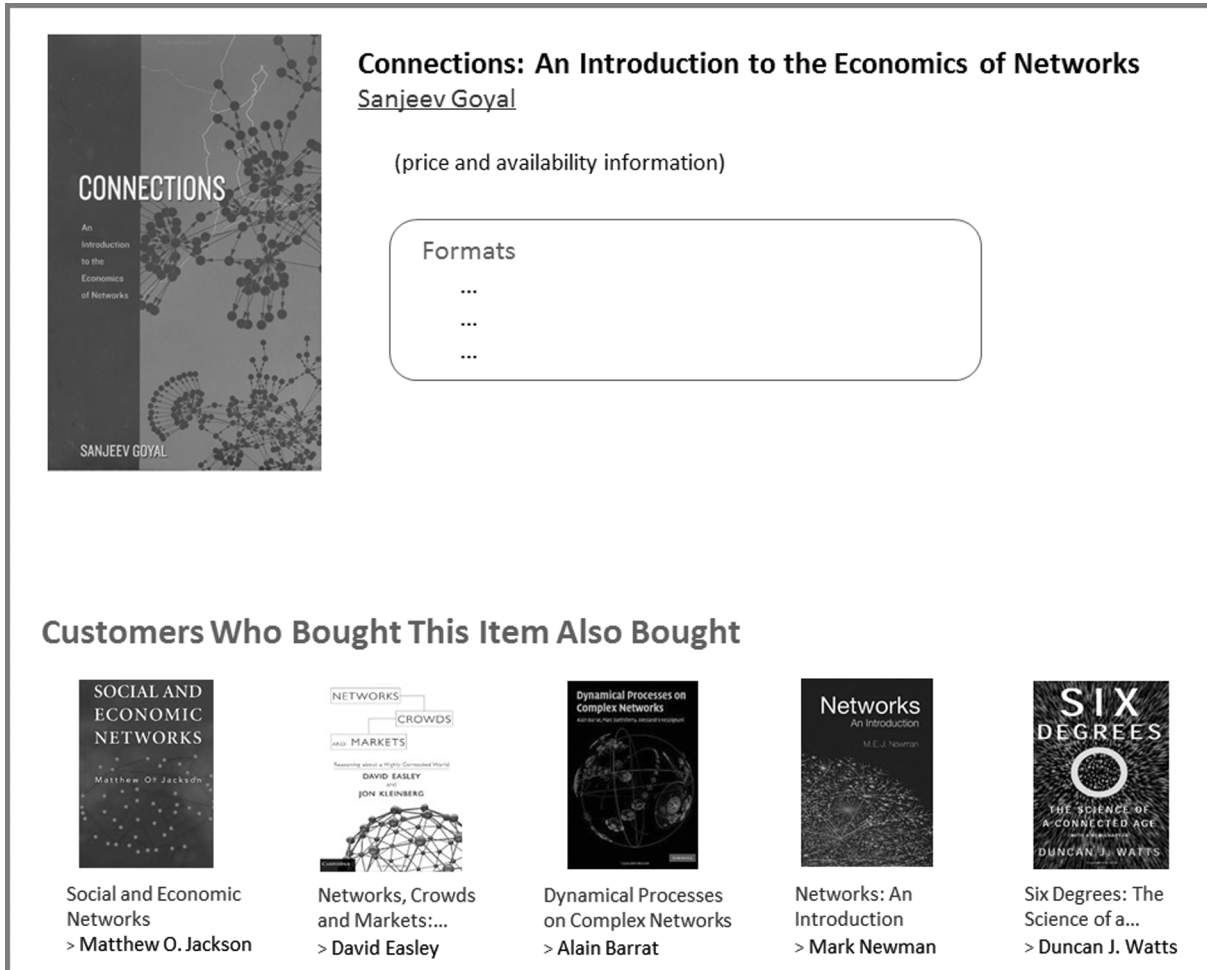
Connections, economic or otherwise, between products are not new. Groups of complementary products are frequently purchased together and influence one

another’s demand levels in different (hidden) ways. What is new, however, and unique to electronic interaction, is that the associations among products are visible, embodied in hyperlinks that can be observed by consumers as they make their purchase decisions. This visibility is thus likely to affect both the magnitude and the nature of influence that products have on each others’ demand levels.

More precisely, every product on an electronic commerce site has a network position, one that is determined by the products and other pages it links to and by those that link to it. If one imagines the process of browsing an electronic commerce site as being analogous to walking the aisles of a physical store, then the *aisle structure* of the electronic commerce site is defined by this graph of interconnected products, and the network position of a product in this graph is its virtual “shelf position.” It is thus natural to expect a product’s virtual shelf position in this electronic network of aisles to affect a product’s demand (more on this later). However, the ensuing direction and extent of the influence of a copurchase network is not immediately clear. For example, the level of attention paid to popular products may increase

<sup>1</sup> Amazon.com provides hyperlinks that connect products, under the heading “Consumers who bought this item also bought...” (also see §3.1). Although Amazon was one of the first to introduce a recommendation network, today almost every major electronic commerce website (Barnes & Noble, YouTube, Yelp, iTunes, etc.) implements a recommendation system that can be modeled as a product network.

**Figure 1** Schematic of an Amazon.com Product Landing Page, Illustrating the “Copurchased” Products That Form the Basis for Our Recommendation Networks



because such products are bought more frequently and thus are more likely to show up more often in a copurchase network. In contrast, such networks might redirect demand toward niche products by making consumers aware of items that were previously not so frequently visible to them. Network visibility might influence demand more intensively for newer products that consumers are less likely to have seen in the past. Alternatively, it might have a greater impact on familiar products, ones that a consumer is more comfortable purchasing if offered unexpectedly. Less expensive products might be influenced more, especially if the influence originates at a more expensive product. This influence might diminish or grow over time.

Clearly, these are empirical questions. The objective of our paper is to answer questions of this kind by measuring how the visibility of electronic product networks alters the influence that products have on one another’s demand and to provide an approach for similar inference in other visible electronic networks. To accomplish this, we use data about the copurchase

networks for more than 250,000 products sold on Amazon.com. To better identify the actual effect of the visible presence of the hyperlinks, we control for various observed and unobserved sources of demand complementarity, including author and category affiliation and year of publication. We also attempt to control for unobserved sources of complementarity by constructing, for each product, three alternative sets of complementary products. First we construct a complementary set based on observed future hyperlinks on Amazon.com. That is, for each product, we construct a complementary set based on “links from the future”—copurchase hyperlinks that are not necessarily visible today but that will be visible in the near future. Such products that are not linked today but that will be linked in the days to follow are assumed to be “as complementary” to the focal product as the items currently present on its webpage (as evidenced by the link that is eventually formed). However, because they are not yet displayed to the consumer, they can serve as a proxy for a set of products that

is complementary but not visible. A second complementary set is constructed using data about the product networks on the Barnes & Noble (B&N) website. The B&N website features a copurchase network similar to the one presented on Amazon. However, products linked on B&N do not necessarily appear on Amazon.com and hence are *invisible* to Amazon.com consumers. Finally, we construct a third complementary product set based on a weighted sum of the demand levels of all products in our data set, calculated as follows: For each pair of products in our sample, we estimate the probability of a link between the two products. We then weight each product's demand according to this propensity of being linked to the focal product and sum those weights to be the complementary "set" of that focal product.

Each of these complementary sets serves as a proxy for demand correlation that might exist regardless of a visible hyperlink being present. Our identification strategy is based on the idea that the set of visible links is a subset of each of those complementary sets. This enables us to identify the influence of a visible hyperlink on demand correlation, after accounting for unobserved complementarity.

Our empirical findings suggest that the visibility of the product network can result in up to a threefold average increase in the influence that complementary products have on one another's demand. We also analyze how the magnitude of influence varies across products of different vintage, products of different categories, and products of varying popularity. A number of interesting implications emerge from these sensitivity analyses, including, among other things, that newer and more popular products "use" the attention they garner from their network position more efficiently and that links from more diverse sets of sources increase the effect of the network on sales. Each of these findings is discussed in detail in the paper. Finally, we relate the benefits of superior network position to the informativeness of the recommendation hyperlinks that define the network.

The rest of this paper is organized as follows. Section 2 provides a summary of related literature. Section 3 describes our data and presents our empirical strategy and the construction of the different complementarity sets. Section 4 describes our empirical results that quantify visible influence and how this influence varies with vintage, category, and popularity, discussing in the process a variety of explanations for our findings. Section 5 concludes, discusses the study's limitations, and outlines directions for future research.

## 2. Related Literature

Our work initiates a deeper understanding of *product* networks. Recently, "social" networks have become

the focus of widespread attention from researchers across fields as diverse as business, economics, political science, and physics;<sup>2</sup> in contrast, the limited attention given to product networks is perhaps surprising, given their economic importance. There are a handful of similar studies of networks of "things," which include an analysis of spillovers in a network of video clips on YouTube by Susarla et al. (2012), a network of blogs by Mayzlin and Yoganarasimhan (2012),<sup>3</sup> and a network of news reports by Dellarocas et al. (2009). Stephen and Toubia (2010) show how the social network position of sellers in an electronic marketplace affects the sales of their products, but they do not study a network of products and therefore do not account for the effect of sales of network neighbors on the sales of a focal product. Similarly, Goldenberg et al. (2012) study the interaction between product networks and social networks in the context of YouTube. Our own related work (Oestreicher-Singer and Sundararajan 2012b) shows that the network of books on Amazon.com has a flattening effect on the demand distribution and contributes to the "long tail" of electronic commerce, but does not address individual demand spillovers. Our current paper thus contributes to this stream of research by analyzing and quantifying the incremental amplification in individual demand that is attributable to the visibility of product networks.

Our results associate *online* product network position with variation in observed sales; the idea that the demand levels for different products are interrelated is fairly well established in the context of traditional "real-world" retail commerce. It is widely recognized that purchases across categories are correlated among consumer goods that are complements or substitutes (Seetharaman et al. 2005, Shocker et al. 2004). This interconnection is a widespread phenomenon that can occur in a variety of ways. For example, a "loss leader" drives purchases for other products (Hess and Gerstner 1987), the existence of software may affect the demand for hardware and vice versa (Binken and Stremersch 2009), cross-brand word of mouth affects the growth of competing brands (Libai et al. 2009), and the demand for a subbrand can affect the consumption of other members of the brand portfolio (Aaker 2004). Such inter-product correlations are of much interest to marketers because they can affect issues such as optimal pricing decisions (Niraj et al. 2008), predicting the sales of new products (Sriram et al. 2010), assessing cross-selling opportunities (Li et al. 2005), or understanding competitive

<sup>2</sup> A complete review of this literature is beyond the scope of this paper; for an extensive review of the study of social networks in economics, the reader is referred to Jackson (2009), Kempe (2011), and Newman et al. (2006).

<sup>3</sup> The network of blogs can also be thought of as a social network.

dynamics (Wedel and Zhang 2004). Cross-selling is clearly a continuing challenge in marketing, wherein complementarity, coincidence, and heterogeneity are conceptualized as factors leading to copurchasing decisions (Edwards and Allenby 2003, Manchanda et al. 1999). To the best of our knowledge, our work is novel in examining how online product networks affect multiproduct demand correlations on a large scale.

Visible hyperlinks alter demand because they redirect consumer attention. A more nuanced explanation for the effect of the visible hyperlinks may be provided by the literature on observational learning, which studies how individuals might draw inferences (about product quality, for example) from mere observation of others' choices or actions (Foster and Rosenzweig 1995). Several recent papers (Cai et al. 2009, Conley and Udry 2010, Moretti 2011, Zhang 2010) study the marketing implications of observational learning. In the context of electronic commerce, Sun (2012) finds that greater variance in rating information increases demand for products with low average ratings. Salganik et al. (2006) show that previous download information increases both inequality and unpredictability of success in the context of online music downloads. Tucker and Zhang (2011) study how consumer observations of site ratings (as measured by a popularity score) alter the traffic to rated sites, showing that vendors with narrower appeal benefit more from such information. This is in contrast with our findings in this paper, which suggest that popular products are more efficient at converting the attention they receive because of a larger number of visible in-links. This difference could be because we are studying a different and larger product category (consumers may simply be more inclined to explore unusual or niche alternatives when browsing wedding sites, the category Tucker and Zhang focus on) or could reflect a difference in the driver of demand redirection (attention versus learning). Rather than directly assess the impact of visible information on consumers' choices, our work studies it indirectly by measuring the impact of visible links on the correlation in demand between different products. A challenge in identifying observational learning is the need to separate it from other sources of quality information, such as word of mouth or editorial reviews; the choice dynamics are often complex in such models (see Zhang 2010). Finally, in a recent paper, Chen et al. (2011) use changes in the information presented to consumers on Amazon as a natural experiment to separate observational learning and word of mouth. Interestingly, they found that although negative word of mouth is more influential than positive word of mouth, positive observational learning information significantly increases sales, but negative observational learning information has no measured effect.

We are able to control for a small number of alternative explanations by including both visible hyperlink effects and complementarity effects, but our analysis of aggregate purchase decisions does not claim to model or identify the consumer actions that lead to the measured demand shifts, nor do we have the level of granularity in our data for such a measurement. Our approach, however, contributes to a growing literature on the identification of peer effects (Manski 1993) in networks using observational data (see, for example, Bramouille et al. 2008, Aral et al. 2009, Ghose and Han 2011).

### 3. Data and Identification Strategy

#### 3.1. Data

We use data on recommendation networks for more than 250,000 books sold on Amazon.com. Each product on Amazon.com has an associated webpage. As discussed in §1, each page has a set of "copurchase links," which are hyperlinks to the set of products that were copurchased most frequently with that product on Amazon.com. This set is listed under the title "Customers who bought this product also bought" and is illustrated in Figure 1. Conceptually, the copurchase network is a directed graph in which nodes correspond to products and edges correspond to directed copurchase links. We collect data about this graph using a Java-based crawler, which starts from a popular book and follows the copurchase links using a depth-first search algorithm. At each page, the crawler gathers and records information for the book whose webpage it is on, as well as the copurchase links on that page, and terminates when the entire connected component of the graph is collected. This is repeated daily. An illustrative part of the network is depicted in Figure 2.

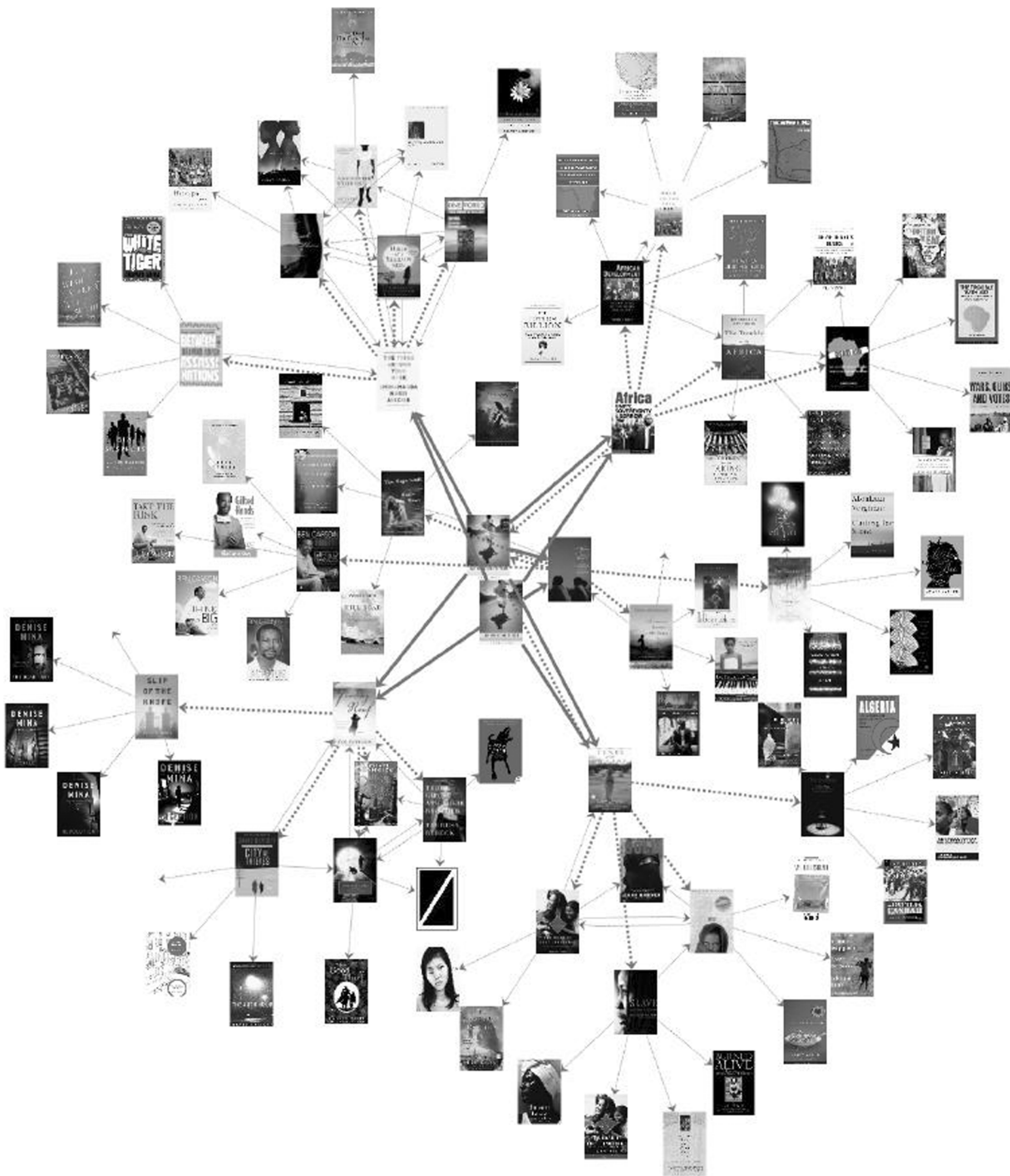
We focus on books because they have a large number of individual titles, the product set is relatively stable (compared to electronics, for instance), and books seem to be a class of products for which the recommendations defining the network we study would actually matter. Our data collection began in August 2005 and continued for about five years. The graph is traversed and recorded every day. A second crawler collects demand information for all books on the graph every 3 hours for the 24-hour period following the collection of the graph. Our algorithms for data collection are outlined in the Web appendix (available at <https://www.box.com/s/85dd1a0329f26a2b9c83>).

The following data are available for each book in the copurchase network, for each day.

**ASIN:** A unique serial number given to each book by Amazon.com. Different editions and different versions have different ASIN numbers.

**List price:** The publisher's suggested price.

Figure 2 Subset of the Recommendation Network, Highlighting the Network Structure



**Sale price:** The price on the Amazon.com website that day.

**Copurchases:** ASINs of the books that appear as its copurchases.<sup>4</sup>

<sup>4</sup> Our work is based on data from 2007, when Amazon.com provided just five copurchase links per product. Currently Amazon.com provides a list of up to 100 such links for each book. Users are initially exposed to the top six due to screen size limitations, and they can then click on a link to view the next six products.

**Sales rank:** The sales rank is a number associated with each product on Amazon.com that measures its demand relative to that of other products. The lower the number is, the higher the sales of that particular product. Sales rank is not an exact measure of sales, but previous research has used it as a proxy, suggesting methods to convert it into a sales measure. Thus, the demand computed is based on the sales rank data generated by Amazon.com and following

a log-linear conversion model suggested by Goolsbee and Chevalier (2003) and by Brynjolfsson et al. (2003). Conversion details are available in the Web appendix.

**Category affiliation:** Amazon.com uses a hierarchy of categories to classify its books. Thus, each book is associated with one or more hierarchical lists of categories, starting with the most general category affiliation and ending with the most specific one. For example, the book *The Search* by John Batelle is associated with the following category hierarchy:

*Subjects > Business and Investing > Biographies and Primers > Company Profiles*

**Author:** The name of the author or authors of the book.

**Publisher:** The name of the publisher of the book.

**Publication date:** The date of publication of the book (by that publisher).

**Secondary market offers:** This includes the following details: (i) total new offers, or the total number of sellers offering a new copy of the book on the secondary market; (ii) total used offers, or the total number of sellers offering a used copy of the book; (iii) the lowest price for a new copy offered by any of the above-mentioned sellers; and (iv) the lowest price for a used copy.

The component of the copurchase network we study changes substantially over time. It contains new nodes every day (more than 6,500 per day, on average), and there are frequent daily changes to the edges between existing nodes. There are also occasional large shifts in the component's size, caused by one or more clusters of nodes detaching from the large connected component; this was often accompanied by a different set of clusters of nodes attaching to this component. Despite the variation in the graph's composition, its in-degree distribution remained quite stable throughout the month. Between 18% and 20% of the books have one incoming link, a little more than 30% have two or three incoming links, roughly the same fraction have between four and seven incoming links, and the in-degree distribution of the remaining 15% or so follows a power-law distribution.

### 3.2. Empirical Strategy

This section outlines our empirical strategy and what we focused on as key alternative explanations for why the demand for interconnected products sold on the Web might be correlated. The "outcome" of interest in our estimation (and hence the dependent variable) is the demand for an individual product, which we refer to as the *focal* product.

We highlight three types of products whose demand may be correlated with the focal product's demand. The first is the set of *visible network neighbors*,

the set of products that link to the focal product through visible hyperlinks.<sup>5</sup> The existence of visible hyperlinks between pages may redirect consumer attention in a way that alters the demand for the products that these hyperlinks "point" to.<sup>6</sup> The second set of products—which we label the *complementary product set*—comprises those whose demand levels are likely to be directly related to the demand for the focal product, even in the absence of visible hyperlinks. By definition, the set of visible network neighbors is included in the complementary product set. We do not make any specific assertion about what underlying economic factors drive the correlations in demand between the focal product and the complementary set; we merely assume that a change in the demand for products in the complementary set may cause a change in demand for the focal product.

Of course, it is impossible to identify the entire universe of complementary products for any given product. In what follows, we will use three different methods of identifying complementary products: by "looking into the future;" by using data from a competing retailer; and by treating the entire set of products as potentially complementary, weighting the complementarity of each by an estimated likelihood that each product is connected to the focal product.

Finally, we consider products that encounter *similar environmental conditions* to those of the focal product. The demand correlations between these products and the focal product are the result of having similar (hidden or observed) individual characteristics or facing similar environmental conditions. For example, people get interested in gardening in spring, and one might therefore expect a correlated increase in the sales of books in the subject category "gardening." Alternatively, an author may engage in marketing activities that promote several of her books, increasing (in a correlated way) the demand for those books. Similarly, a set of books might be assigned as mandatory reading for a graduate course. In all these cases, there is no direct effect that the demand for one product has on the other, and the correlated changes in demand are caused by some other external factor, analogous to what is termed the self-selection effect in social contexts.

<sup>5</sup>Note that although each product only has five outgoing links, it may have many more incoming links.

<sup>6</sup>As discussed in our introductory section, this shift in demand may also be interpreted as evidence of observational learning. We do discuss this connection further when interpreting our results. However, because our data are insufficiently granular to separate the different potential drivers of this effect (if measured), we do not attempt to model different ways in which visibility might alter demand influence.

The focal product’s demand is affected by a number of factors besides the demand for related products. Naturally, demand is influenced by a *product’s individual characteristics*. These product characteristics are divided into two sets: those that are intrinsic or individual to the product, and those that are network-specific. The intrinsic characteristics controlled for in our model are the following:

**Price:** We include both the list price and the sale price of the product. List price incorporates information about the value of a book, whereas sale price (and its magnitude relative to the list price) speaks to affordability.

**Secondary market offers:** This includes the following: (i) total new offers, or the total number of sellers offering a new copy of the book on the secondary market; (ii) total used offers, or the total number of sellers offering a used copy of the book; (iii) the lowest price for a new copy offered by any of the sellers above; and (iv) the lowest price for a used copy.

**Vintage:** This is the number of years since the book was published. We separate books into two groups based on their age—books that were published after January 1st, 2006 (tagged as “recent”) and books that were published before that date. We also partition and bin the books more finely in our dedicated analysis of vintage (see §4.4).

The network-based characteristics include the following:

**In-degree of the book:** The number of other books that point to the focal book.

**Assortative mixing:** The percentage of incoming links from products in the same category.

Finally, in the context of a product network, one might expect the demand for a product to depend on the *characteristics of its neighbors* and the *characteristics of its complementary set*, as well as the characteristics of its associated product dyads (for example, the level of similarity between a focal product and each product it is linked to). After all, although the visibility of a network of hyperlinks between two products potentially redirects consumer attention, the eventual impact of a hyperlink is mediated by the fraction of consumers who actually click on the link and purchase the product. For example, if the product on the source page is expensive relative to the focal product, there may be a higher propensity to explore alternatives by following a hyperlink. Similarly, if two products are of the same category, consumers might be more likely to click on a link between them. We therefore control for the average characteristics of the neighboring products and of the complementary products as well as the *dyad characteristics*.

Combining all of the above leads to the following equation:

$$y_i = \alpha_0 + \alpha_1 \sum_{j \in S_n(i)} y_j + \alpha_2 \sum_{j \in S_c(i)} y_j + \sum_{u=1}^K \alpha_{3,u} \left( \frac{1}{N(i)} \sum_{j \in S_n(i)} x_j \right) + \sum_{u=1}^K \alpha_{4,u} \left( \frac{1}{M(i)} \sum_{j \in S_c(i)} x_j \right) + \sum_{u=1}^K \alpha_{5,u} x_j + \varepsilon_i, \quad (1)$$

where  $y_i$  is the demand for product  $i$ ;  $S_n(i)$  is the set of (visible) network neighbors of product  $i$  (the set of products with incoming links to product  $i$ );  $S_c(i)$  is the set of complementary products of product  $i$ ;  $N(i)$  and  $M(i)$  are the sizes of sets  $S_n(i)$  and  $S_c(i)$ , respectively; and  $x_j$  is the level of characteristic  $j$  of product  $i$  (for example,  $x_{1i}$  is product  $i$ ’s price). A summary of variables is provided in the Web appendix.

A more compact exposition of the model is

$$\mathbf{y} = \alpha_0 \mathbf{1} + \alpha_1 G_n \mathbf{y} + \alpha_2 G_c \mathbf{y} + G_n^* X \alpha_3 + G_c^* X \alpha_4 + X \alpha_5 + \boldsymbol{\varepsilon}, \quad (2)$$

where  $\mathbf{1}$  is an  $n$ -vector of ones;  $\mathbf{y}$  is the vector  $(y_i)$ ;  $G_n$  is the  $n \times n$  interaction matrix for the set  $S_n$ , defined as  $G_n(i, j) = 1$  if  $j \in S_n(i)$  and  $G_n(i, j) = 0$  otherwise;  $G_c$  is the  $n \times n$  interaction matrix for the set  $S_c$ , defined as  $G_c(i, j) = 1$  if  $j \in S_c(i)$  and  $G_c(i, j) = 0$  otherwise—note that by definition  $S_n(i) \subset S_c(i)$ ;  $G_n^*$  is the  $n \times n$  interaction matrix for the set  $S_n$ , defined as  $G_n^*(i, j) = 1/N(i)$  if  $j \in S_n(i)$  and  $G_n^*(i, j) = 0$  otherwise;  $G_c^*$  is the  $n \times n$  interaction matrix for the set  $S_c$ , defined as  $G_c^*(i, j) = 1/M(i)$  if  $j \in S_c(i)$  and  $G_c^*(i, j) = 0$  otherwise; and  $X$  is the matrix of product characteristics—this is an  $n \times K$  matrix, with  $K$  being the number of characteristics, and  $x_{ij}$  is the value of characteristic  $j$  for product  $i$ .<sup>7</sup>

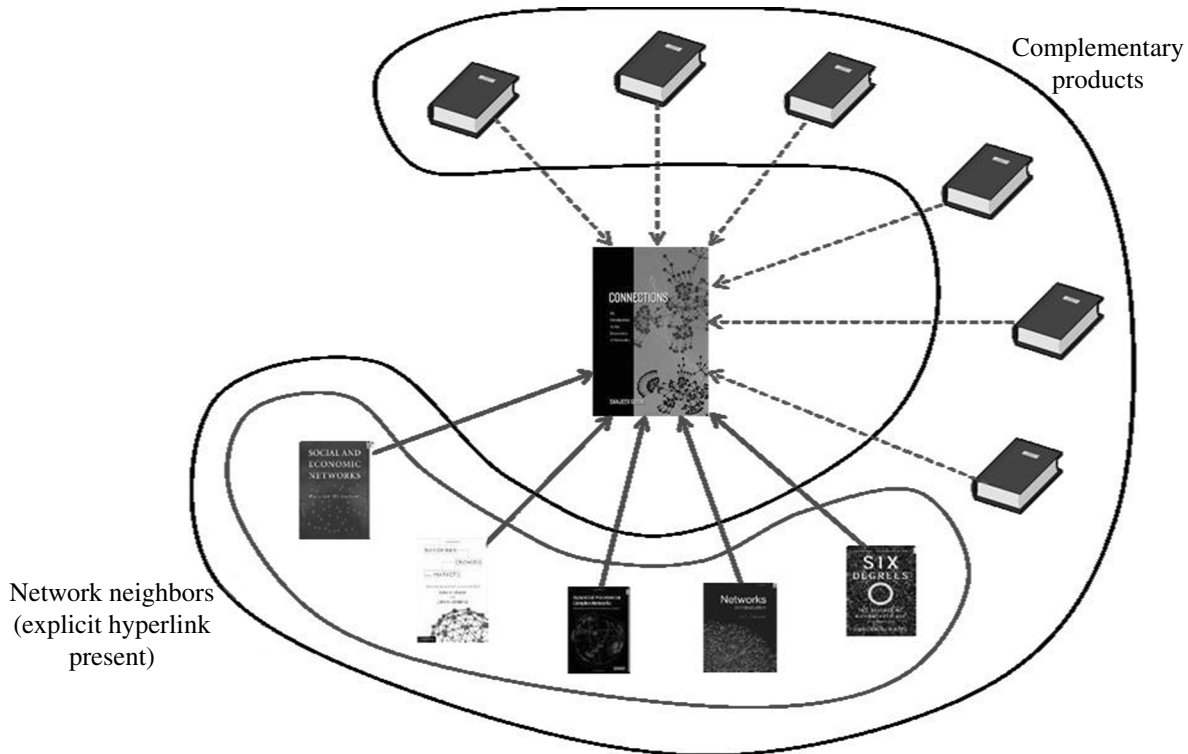
Figure 3 illustrates our empirical strategy; the computation of  $G_n$  and  $G_c$  is explained further when we discuss the different complementarity sets we construct.

In this equation  $\alpha_1$  measures the visible hyperlink effects;  $\alpha_2$  measures the complementary product effects;  $\alpha_{3,u}$  and  $\alpha_{4,u}$  measure the effect of characteristic  $u$  of the network neighbors and of the complementarity products set, respectively; and  $\alpha_{5,u}$  measures the effect of the book’s own characteristic  $u$  on its demand.<sup>8</sup> We control for correlated effects by using

<sup>7</sup>Note that multiplying  $G_n$  with a vector produces the sum over the elements of  $S_n(i)$  (for example,  $G_n \mathbf{y}$  sums the demand for the product’s network neighbors), whereas multiplying  $G_n^*$  with a vector produces an average of the values over the elements of  $S_n(i)$  (for example,  $G_n^* \mathbf{x}$  averages the values of the characteristics of the product’s network neighbors). Similarly, multiplying  $G_c$  with a vector produces the sum over the elements of  $S_c(i)$  and multiplying  $G_c^*$  with a vector produces averages over the elements of  $S_c(i)$ .

<sup>8</sup>Because  $i$  is not included in its own network or its own complementary sets, the term  $\sum \alpha_{5,u} x_i$  must be added to account for

Figure 3 Schematic of Our Empirical Strategy



Note. Each product has a set of visible network neighbors, defined by  $S_n(i)$ , and a set of complementary products, defined by  $S_c(i)$ .

fixed effects. We describe the groupings that drive unobserved heterogeneity more completely in our next section. Because  $S_n(i)$  is a subset of  $S_c(i)$ , comparing the relative influence of  $S_n(i)$  and  $S_c(i)$  gives us an assessment of the additional effect that the demand for a network neighbor has on the demand for a focal product. For products that appear both in  $S_n(i)$  and in  $S_c(i)$ ,  $\alpha_1$  will represent the correlation arising from the presence of the hyperlink and  $\alpha_2$  will represent the correlation from complementarity. In a companion paper (Oestreicher-Singer and Sundararajan 2012a) we provide a theoretical analysis of the conditions under which the model above is identified.

We now summarize some sources of demand endogeneity and how we deal with them. First, as men-

tioned, the demand for two products may be correlated because linked products are similar in observed and unobserved ways, and this similarity among linked products may be responsible for observed comovement in demand rather than any influence mediated by the visible hyperlink. We partially account for this possibility by controlling for observed similarities in characteristics across linked products (author, category, and vintage) and by constructing complementary sets for each product. Second, the demand for linked products may be influenced by shared external factors (as in the gardening and reading list examples above). We attempt to account for these issues by controlling for observed similarities in characteristics across linked products and by constructing complementary sets that might capture unobserved similarities among products. Third, the formation of the network of interest is endogenous: correlation in demand, the outcome variable of interest, results in the formation of the “copurchase” link. Toward addressing this issue, we use data from a different retailer (B&N) and also estimate the probability of link formation among products in our data set, a process that we believe partially opens the “black box” of network formation.

As discussed briefly above, we create three different “complementary sets” for each product, which we use as control groups when estimating the influence of the visible set of network neighbors.

the effect of the book’s own characteristics on demand. Observe that the systematic part of this model is similar to that of a spatial autoregressive model extended to allow for exogenous effects. Similar representations for measuring influence appear in Lee (2007) (for group interaction) and Bramoulle et al. (2008) (for network-based interaction), although their models are somewhat more limited than ours. Spatial autoregressive models have also been used by Yang and Allenby (2003) in the context of consumers coordinating on buying domestic versus imported automobiles, and by Trusov et al. (2009) when separating the effects of target marketing from word-of-mouth effects in the context of adopting a social networking site. A recent survey of spatial models in marketing is available in Bradlow et al. (2005).



**3.2.1. Complementary Set Based on Future Amazon.com Data.** This set includes all books that are current network neighbors of the focal book, along with the set of products that are not visible on the given day but will become visible network neighbors of the focal book during a short future time interval. Intuitively, at any given point in time, products that will become members of the visible network of a focal product in the future are already complementary to the focal product today. They have just not been sufficiently copurchased to show up in the “top five” list. Because they are complementary but not yet visible, they serve as a good benchmark for the added effects of visibility. The major advantage of this method is that it is based on the same recommender algorithm used to create the network of visible neighbors. A shortcoming of this control group stems from this very advantage—it is based on the same endogenous link-formation process that generates the visible network neighbors. Further, the use of this set assumes that levels of complementarity do not vary over the interval of interest.<sup>9</sup> The construction of this set is explained further with examples in the Web appendix.

**3.2.2. Complementary Set Based on B&N Data.** The B&N retail website features a copurchase network similar to the one presented to Amazon’s customers. For each focal product, we collect data about all visible copurchase links on the B&N website. We combine these data with our Amazon.com data to create a superset of all products with a copurchase link pointing to the focal product on Amazon.com, on the B&N website, or both, and use this superset as the complementary set. Although our demand data capture sales for the focal product exclusively through the Amazon website, it seems logical that the set of products copurchased by B&N customers are also complementary to this focal product. However, because they are presented on a different website, they are unlikely to be noticed by a consumer browsing Amazon.com.

This set better controls for time-varying complementarity (as both sets are collected on the same day). Furthermore, because the data are collected from different websites, the complementary set is

not generated by the same endogenous network-formation process as the visible network neighbor set. One limitation of the use of this as a representative invisible set of complementary products is that B&N may have consumers with differing demographics or preferences. We note, however, that we have been unable to find evidence that supports this claim, and prior research (for example, Chevalier and Mayzlin 2006) has used data from both sites interchangeably based on the similar underlying assumption of consumer homogeneity across the sites. A second concern is that Amazon.com may simply have better copurchase link generation because it has richer historical data and a larger customer base. We partially alleviate this concern by using demand data from B&N as alternative input for Equation (1) (correspondingly, we use the B&N neighbors as the visible network neighbors and the Amazon.com neighbors as the invisible complements). The direction and significance of our results persist.

**3.2.3. Complementary “Set” Based on Likelihood of Link Formation.** This approach exploits the information we have about the products that are visibly complementary in order to assess the likelihood of the others being complementary. We follow Manchanda et al. (2004) to assess, for each focal product, the probability that each of the other products will have a copurchase link terminating at the focal product (this link formation is the nonrandom process on our context). Specifically, for each focal product, and using the visible copurchase links and a random sample of nonlinked products, we estimate a logit model whose dependent variable is whether a copurchase link from another product will terminate at the focal product. All observed covariates of the focal product and the other product (including the different product characteristics and the dyad characteristics) are incorporated as independent variables. We then use this model to estimate, for each product in our data set, the probability of its being linked to the focal product. We weight each product’s demand according to its probability of being linked, and we sum these weighted demand levels to be the “influence” of the complementary set. Thus,  $G_c$  is the  $n \times n$  interaction matrix generated, defined as  $G_c(i, j) = 1/\text{Prob}(i, j)$  for all  $j$ .

Naturally, this method can only be applied to books for which we have enough positive training examples, that is, books with a sufficiently large number of incoming links based on which a model of linking can be estimated. We choose the 300 books in our sample that have more than 30 incoming links. Also note that this method requires some changes to Equations (1) and (2).  $G_c$  is now a weighted sum of the demand of all products in our set.  $G_n$  is similarly altered and is

<sup>9</sup> We use a two-week sliding window as the “future interval.” This choice was made after careful consideration and in trying to obtain a set that is the best representation of the real complementarity set. In making this choice, we balanced two different trade-offs. The first has to do with the length of the time interval. Clearly, choosing a longer time interval increases the number of products in the set. However, this does not necessarily improve the accuracy of this set, because it also increases noise due to seasonality. The second trade-off is between having a constant set of complementary products rather than using a sliding window. We chose to use a sliding window because it seemed more important to omit links that have already appeared in the past and whose presence might have influenced purchasing patterns in the present.

the weighted sum of all network neighbors' demand. This results in the following equation:

$$y_i = \alpha_0 + \alpha_1 \sum_{j \in S_n(i)} \text{Prob}(i, j) \cdot y_j + \alpha_2 \sum_j \text{Prob}(i, j) \cdot y_j \\ + \sum_{u=1}^K \alpha_{3,u} \left( \frac{1}{N(i)} \sum_{j \in S_n(i)} x_j \right) + \sum_{u=1}^K \alpha_{4,u} \left( \frac{1}{M(i)} \sum_{j \in S_n(i)} x_j \right) \\ + \sum_{u=1}^K \alpha_{5,u} x_j + \varepsilon_i,$$

where  $\text{Prob}(i, j)$  is the estimated likelihood of a link being formed between product  $i$  and product  $j$ .

A benefit of this method is that it attempts to open the "black box" of the network formation process, perhaps partially accounting for this source of endogeneity. Additionally, if a link between a product and the focal product is the result of an exogenous factor and cannot be explained by observable characteristics (as in the motivating example of the reading list earlier in the section), then the estimated probability of this link forming will be low. Hence, that book will have a small weight in the  $G_n$  computation and thus a lower impact on the estimated coefficients. This does not eliminate the bias, but it reduces its impact on the estimation results.

We use a two-stage model (2SLS) with instrumental variables. As is customary, in the first stage, the endogenous independent variable is estimated using all exogenous instrumental variables, and in the second stage, these estimated values are used as independent variables. Because our model includes two endogenous variables (the visible network neighbors and the complementary sets), in the first stage we generate estimated values for both these variables. Our chosen instrumental variables are based on the secondary market information we collected for each day, which, as we mention above, includes the variety of offers (the number of new offers and the number of used offers) and prices (the lowest price for the new and used offers) in that marketplace. We use the secondary market information of the network neighbors and of the complementary set as instruments for the demand of those sets. This supply-side information serves as a valid instrument, as it is natural to assume that the secondary market price of the neighboring product will be correlated with the demand for the neighbor (this was also shown by Ghose et al. 2006), but not with the demand of the focal product. Note that although the secondary market products are offered on the Amazon.com webpage, their presence and prices are determined exogenously by numerous individual sellers and are not controlled by Amazon.com. Furthermore, all our demand information is exclusively for Amazon.com's own sales and does not include demand for any of these secondary

market sellers. The variety and price information about the visible network neighbors' and complementary products' secondary market are therefore valid instruments for computing the demand for visible network neighbors ( $S_n$ ) and the demand for the complementary set ( $S_c$ ) in the first stage. We have also tested the instruments' validity using the Hansen-Sargan test (Hansen 1982) and the Stock-Yogo critical values (Stock and Yogo 2005).

Finally, to partially account for simultaneity, on each day we exclude the books for which all incoming links are "bidirectional." That is, each focal product (book) that remains in our sample has at least one outgoing hyperlink to a product that does not in turn have a hyperlink terminating at the focal product's page. This does not eliminate all cycles from our data, but it does exclude "short" cycles. About 15% of our observations are eliminated by this refinement. We also exclude the books for which the network neighbors set ( $G_n$ ) and the complementary products set ( $G_c$ ) are identical. This refinement eliminates an additional 1% or so of the books. (Neither of these refinements, aimed at addressing specific sources of endogeneity, alters our regression coefficients substantially.)

## 4. Estimation and Empirical Results

We present several sets of results in this section. The first set quantifies the relative magnitude of the effect of visible network neighbors. These results are based on estimates of Equation (2), using each of the three complementary sets with appropriate controls for observed and unobserved correlated effects. The second set of results is based on repeating this analysis while treating sale price as endogenous. The subsequent analysis uses quantile regression to examine how these results vary with product popularity. A summary of variables is provided in the Web appendix.

### 4.1. Full Model: The Average Effect of the Visible Neighbors

Based on the variables we have constructed as described above, our model translates to the following equation:

$$\log(y) = \alpha_1 \log(z_n) + \alpha_2 \log(z_c) + \alpha_{4,1} \log(x_1) \\ + \alpha_{4,2} \log(x_2) + \alpha_{4,3} \log(x_3) + \alpha_{4,8} \log(x_8) \\ + \alpha_{3,1} \log(q_1) + \alpha_{3,2} \log(q_2) + \alpha_{3,3} \log(q_3) \\ + \alpha_{3,3} \log(q_8) + \alpha_{4,1} \log(x_1) + \alpha_{5,1} \log(w_1) \\ + \alpha_{5,2} \log(w_2) + \alpha_{5,3} \log(w_3) \\ + \alpha_{5,8} \log(w_8) + \alpha_6 R, \quad (3)$$

**Table 1** Summary of Estimates for a Sample Day of Data with the Complementary Set Based on Future Amazon.com Data

	Basic 2SLS model	Control for recency: No fixed effects	Fixed effects: Category I	Fixed effects: Category II	Fixed effects: Author	Fixed effects: Recency
$\ln(z_n)$ Total demand for network neighbors	0.61*** (0.10)	0.52*** (0.100)	0.64*** (0.09)	0.65*** (0.10)	0.41* (0.21)	0.51*** (0.10)
$\ln(z_c)$ Total demand for complementary set	0.39*** (0.10)	0.46*** (0.1)	0.33*** (0.09)	0.32*** (0.09)	0.52** (0.21)	0.47*** (0.10)
$\ln(x_1)$ List price	0.74*** (0.02)	0.71*** (0.01)	0.78*** (0.01)	0.78*** (0.01)	0.72*** (0.03)	0.70*** (0.01)
$\ln(x_2)$ Sale price	-0.78*** (0.02)	-0.76*** (0.01)	-0.83*** (0.01)	-0.83*** (0.01)	-0.79*** (0.04)	-0.75*** (0.01)
$\ln(x_3)$ In-degree of the product	-0.54*** (0.01)	-0.52*** (0.01)	-0.51*** (0.01)	-0.51*** (0.01)	-0.51*** (0.02)	-0.53*** (0.01)
$\ln(x_6)$ Assortative mixing of the product	-0.01** (0.01)	-0.01** (0.01)	-0.01** (0.01)	-0.01*** (0.01)	-0.01 (0.01)	-0.01** (0.01)
$\ln(q_1)$ Avg list price of network neighbors	0.16 (0.13)	0.23 (0.14)	0.13 (0.14)	0.13 (0.14)	0.35 (0.25)	0.22 (0.14)
$\ln(q_2)$ Avg sale price of network neighbors	-0.23* (0.13)	-0.28** (0.13)	-0.21 (0.13)	-0.21 (0.13)	-0.39 (0.24)	-0.27** (0.13)
$\ln(q_3)$ Avg in-degree of network neighbors	-0.21*** (0.08)	-0.15* (0.07)	-0.24*** (0.07)	-0.24*** (0.07)	-0.11 (0.16)	-0.14* (0.07)
$\ln(q_6)$ Avg assortative mixing of network neighbors	-0.02 (0.03)	-0.02 (0.02)	-0.01 (0.02)	-0.01 (0.02)	0.06 (0.06)	-0.02 (0.02)
$\ln(w_1)$ Avg list price of complementary set	-0.60*** (0.14)	-0.68*** (0.14)	-0.50*** (0.14)	-0.51*** (0.14)	-0.58** (0.26)	-0.68*** (0.14)
$\ln(w_2)$ Avg sale price of complementary set	0.71*** (0.13)	0.77*** (0.13)	0.62*** (0.13)	0.63*** (0.13)	0.65*** (0.24)	0.78*** (0.13)
$\ln(w_3)$ Avg in-degree of complementary set	-0.24*** (0.08)	-0.29*** (0.08)	-0.19** (0.08)	-0.18** (0.08)	-0.27* (0.16)	-0.29*** (0.08)
$\ln(w_6)$ Avg assortative mixing of complementary set	0.03 (0.03)	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)	-0.06 (0.03)	0.02 (0.03)
<i>R</i> Recency		0.13*** (0.01)				
Constant	-0.61** (0.285)	-0.57** (0.28)				
Observations	111,811	111,765	111,799	111,802	28,383	111,811
<i>R</i> -squared	0.60	0.60	0.57	0.57	0.49	0.59
Number of groups (fixed effects)			358	390	2,338	91

Note. Standard errors in parentheses.  
 \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

where  $z_n$  is the demand for the set  $s_n$ , and  $z_c$  is the demand for the set  $s_c$ . We estimate this equation using each of our three complementary sets as  $z_c$ . Table 1 details our results for one sample day of the data, using the complementary set that was defined on the basis of future links on Amazon.com.<sup>10</sup> Table 2 presents the results with the B&N-based complemen-

tary set; and Table 3 presents the results with the weighted sum as a complementary set.

For each complementary set, we control for the observed correlated effects using the within transformation. We report on the associated fixed-effects grouping along a few different dimensions: first-level category, second-level category, year of publication, and author.

In each table, the first column shows coefficients estimated using the base 2SLS model. The second column of coefficients (labeled “Control for recency”) includes an additional variable, a dummy that is 1 if the book was published recently (after 2006). The third column (“Fixed effects: Category I”) reports on estimates that control for unobserved heterogeneity by category, or alternatively viewed, correlated

<sup>10</sup> It is important to clarify that the equation corresponds to a snapshot of the network. Therefore, information from the future is only used to identify the books that are part of the complementary set ( $S_c(i)$ ). However, when estimating the equation, we use the present demand—both for the visible network neighbors ( $z_n$ ) and for the complementary set ( $z_c$ ). That is, there is no time lag between the demand for the different sets and the demand of the product. As a consequence, the model can also be written as  $y_i = \alpha_0 \mathbf{1} + \alpha_1 G_n y_i + \alpha_2 G_c \hat{y}_i + G_n^* X_i \alpha_3 + G_c^* X_i \alpha_4 + X_i \alpha_5 + \epsilon$ .

**Table 2** Summary of Estimates for a Sample Day of Data with the Complementary Set Based on Barnes & Noble.com Data

	Basic 2SLS model	Control for recency: No fixed effects	Fixed effects: Category I	Fixed effects: Category II	Fixed effects: Author	Fixed effects: Recency
$\ln(z_n)$ <i>Total demand for network neighbors</i>	0.88*** (0.01)	0.86*** (0.01)	0.83*** (0.01)	0.84*** (0.01)	0.78*** (0.03)	0.84*** (0.01)
$\ln(z_c)$ <i>Total demand for complementary set</i>	0.16*** (0.01)	0.17*** (0.01)	0.18*** (0.01)	0.18*** (0.01)	0.18*** (0.02)	0.18*** (0.01)
$\ln(x_1)$ <i>List price</i>	0.74*** (0.02)	0.73*** (0.02)	0.77*** (0.02)	0.79*** (0.02)	0.71*** (0.04)	0.72*** (0.01)
$\ln(x_2)$ <i>Sale price</i>	-0.75*** (0.02)	-0.74*** (0.02)	-0.78*** (0.02)	-0.79*** (0.02)	-0.69*** (0.05)	-0.73*** (0.02)
$\ln(x_3)$ <i>In-degree of the product</i>	-0.52*** (0.01)	-0.49*** (0.01)	-0.47*** (0.01)	-0.47*** (0.01)	-0.44*** (0.03)	-0.49*** (0.01)
$\ln(x_6)$ <i>Assortative mixing of the product</i>	-0.01* (0.01)	-0.01** (0.01)	-0.01* (0.01)	-0.01** (0.01)	-0.01 (0.012)	-0.013** (0.01)
$\ln(q_1)$ <i>Avg list price of network neighbors</i>	-0.23*** (0.03)	-0.22*** (0.03)	-0.17*** (0.03)	-0.17*** (0.03)	-0.04 (0.06)	-0.22*** (0.02)
$\ln(q_2)$ <i>Avg sale price of network neighbors</i>	0.28*** (0.03)	0.26*** (0.03)	0.21*** (0.03)	0.22*** (0.03)	0.03 (0.06)	0.27*** (0.02)
$\ln(q_3)$ <i>Avg in-degree of network neighbors</i>	-0.44*** (0.01)	-0.41*** (0.02)	-0.41*** (0.01)	-0.41*** (0.01)	-0.35*** (0.02)	-0.40*** (0.01)
$\ln(q_6)$ <i>Avg assortative mixing of network neighbors</i>	-0.01 (0.01)	-0.01* (0.01)	-0.01 (0.01)	-0.01* (0.01)	-0.01 (0.02)	-0.01* (0.01)
$\ln(w_1)$ <i>Avg list price of complementary set</i>	-0.04*** (0.01)	-0.05*** (0.01)	-0.05*** (0.01)	-0.05*** (0.01)	-0.04** (0.01)	-0.04*** (0.01)
$\ln(w_2)$ <i>Avg sale price of complementary set</i>	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
$\ln(w_3)$ <i>Avg in-degree of complementary set</i>	-0.02** (0.01)	-0.04*** (0.01)	-0.04*** (0.01)	-0.04*** (0.01)	-0.04** (0.02)	-0.04*** (0.01)
$\ln(w_6)$ <i>Avg assortative mixing of complementary set</i>	-0.01 (0.01)	-0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	-0.02 (0.02)	-0.01 (0.01)
<i>R</i>		0.16*** (0.01)				
<i>Constant</i>	-0.14*** (0.03)	-0.17*** (0.03)				
Observations	68,613	68,592	68,595	68,599	18,084	68,613
<i>R</i> -squared	0.65	0.65	0.62	0.62	0.53	0.65
Number of groups (fixed effects)			336	378	2,028	87

Note. Standard errors in parentheses.

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

effects when books are grouped according to first-level category. The coefficients in the fourth column (“Fixed effects: Category II”) control for unobserved heterogeneity where the grouping is by second-level category. The set of results in the “Fixed effects: Recency” column control for unobserved heterogeneity by year of publication. The column titled “Fixed effects: Author” reports on estimations that control for unobserved heterogeneity by author; that is, books written by the same author are grouped together. These estimations include only authors who wrote 10 books or more and hence use smaller data sets. Note that in the case of the complementary set based on estimated likelihood of linking, there were not enough data points to run the regression with fixed effects by author.

We carry out these estimations for 30 randomly chosen daily data sets corresponding to 30 days in 2007. In what follows, we will refer to the results of the day presented in Tables 1–3. Comparable results are obtained for the other 29 days.

The estimation results highlight a striking theme: both the set of visible network neighbors and the set of complementary products (in its different constructions) defined earlier are significant, both statistically and economically. Most importantly, the estimated coefficients of  $z_n$  establish that the visibility of a recommendation hyperlink has a significant influence on the demand for products in the network, even after controlling for other possible sources of comovement. Moreover, in most estimated models the magnitude of the coefficient for the visible network neighbors

**Table 3** Summary of Estimates for a Sample Day of Data with the Complementary “Set” Based on Likelihood of Link Formation

	Basic 2SLS model	Control for recency: No fixed effects	Fixed effects: Category I	Fixed effects: Category II	Fixed effects: Recency
$\ln(z_n)$ Total demand for network neighbors	0.67** (0.29)	0.64** (0.31)	0.77*** (0.2)	0.35 (0.35)	0.50* (0.27)
$\ln(z_c)$ Total demand for complementary set	0.63** (0.31)	0.61** (0.30)	-0.07 (0.2)	0.42 (0.48)	0.65** (0.28)
$\ln(x_1)$ List price	1.33 (0.82)	1.22 (0.81)	1.04 (1.24)	1.96* (1.18)	0.49 (0.88)
$\ln(x_2)$ Sale price	-1.31* (0.74)	-1.24* (0.73)	-1.46 (1.14)	-1.79* (1.07)	-0.63 (0.77)
$\ln(x_3)$ In-degree of the product	1.16* (0.64)	1.07* (0.62)	0.60 (0.87)	1.13 (0.98)	0.57 (0.62)
$\ln(x_8)$ Assortative mixing of the product	-0.01 (0.13)	0.01 (0.13)	-0.30 (0.19)	-0.01 (0.20)	-0.03 (0.16)
$\ln(q_1)$ Avg list price of network neighbors	0.69 (2.31)	0.61 (2.32)	9.13*** (3.21)	-3.93 (5.37)	4.17* (2.28)
$\ln(q_2)$ Avg sale price of network neighbors	-1.04 (2.09)	-0.96 (2.09)	-8.71*** (3.01)	3.10 (5.01)	-4.40** (2.14)
$\ln(q_3)$ Avg in-degree of network neighbors	-0.62 (0.48)	-0.57 (0.50)	-0.76 (0.55)	0.14 (0.75)	-0.30 (0.47)
$\ln(q_8)$ Avg assortative mixing of network neighbors	0.018 (0.14)	0.02 (0.13)	0.30 (0.19)	0.09 (0.30)	0.11 (0.14)
$\ln(w_1)$ Avg list price of complementary set	-3.24 (3.74)	-3.16 (3.72)	-13.80*** (3.88)	1.10 (6.12)	-6.44* (3.85)
$\ln(w_2)$ Avg sale price of complementary set	3.88 (3.37)	3.82 (3.34)	13.50*** (3.61)	-0.01 (5.48)	7.01** (3.55)
$\ln(w_3)$ Avg in-degree of complementary set	-1.03 (0.89)	-1.01 (0.88)	0.74 (0.69)	-1.34 (1.67)	-0.81 (0.77)
$\ln(w_8)$ Avg assortative mixing of complementary set	0.03 (0.09)	0.03 (0.09)	0.028 (0.17)	-0.09 (0.15)	-0.05 (0.11)
<i>R</i> Recency		0.15 (0.25)			
Constant	-2.57 (2.99)	-2.59 (2.94)			
Observations	100	100	70	72	91
<i>R</i> -squared	0.69	0.70	0.79	0.61	0.70
Number of groups (fixed effects)			22	25	11

Note. Standard errors in parentheses.

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

$(z_n)$  is more than double the coefficient for the complementary set  $(z_c)$ . That is, adding a link between two complementary books (say between book A and book B) amplifies the observed covariance between the demand levels for the books by up to three times.

Another intriguing result is that the coefficient of the in-degree  $(x_3)$  variable is negative in both Tables 1 and 2. This is interesting and to some extent counter-intuitive. Naturally, in-degree is positively correlated with demand (a correlation coefficient of 0.4).<sup>11</sup> However, when controlling for the incoming traffic from

network neighbors  $(z_n)$ , we find that higher in-degree is associated with lower demand. That is, the same level of total incoming traffic from fewer, more popular sources is associated with higher demand. For example, a product with 2 incoming links, each with a demand level of 100 units, gains more benefit from the recommendation network than a product that has 20 incoming links, each with a demand level of 10 units. In other words, the variety of recommendation sources matters.

This finding seems especially interesting when considered in the context of previous research in marketing on the dispersion of word-of-mouth communication and seeding efforts. The marketing literature suggests that there is a positive effect that arises from the dispersion of word-of-mouth sources (Godes and

<sup>11</sup> This is true for at least two reasons. First, it is likely that more popular products (that is, products that sell more) will have higher in-degree values, and second, it is likely that the high in-degree signals high incoming traffic from network neighbors.

Mayzlin 2004) and the dispersion of product seeding (Libai et al. 2005, Sundararajan 2008). However, this literature tends to model or measure what is essentially a social process operating on a single product and explaining its spread. In contrast, our data examine the effects of a corresponding dispersion of aggregate consumer traffic across different products; therefore, the departure from prior work may not be as surprising as it first seems.

A possible explanation in our context is that more popular products have a more “mainstream audience,” whose interests are easier to predict, and hence this audience is, on average, more likely to find a recommended product interesting. Put differently, the conversion rate of recommendations that originate from more popular products is higher. Another explanation might be that recommendations that originate from more popular products are more informative because the sample of consumers on which the recommendation is based is larger. It is reasonable to assume that hyperlinks on popular pages are based on more instances of copurchasing and are therefore more representative of shared propensity to buy, and thus are more likely to lead to better matches and a higher attention-to-sales conversion rate. Additionally, in line with the theory of observational learning, consumers may find recommendations on the pages of popular products more informative, assuming those recommendations are based on greater numbers of previous purchases. Because our data measure how the impact of increased traffic from one source compares to a corresponding increase in traffic from several less popular sources, it is also useful to note that the literature on search and navigation in online environments does not assume a dispersion effect one way or the other (for example, Brin and Page 1998).

**4.1.1. Assortative Mixing.** All our measures of assortative mixing are with respect to the visible network of hyperlinks. Simply put, the neighborhood of a book is more assortatively mixed if more of the book’s neighbors belong to the same category as the book itself; see Newman (2003) for further details. The level of assortative mixing for the neighborhood of the book itself ( $x_g$ ) has a large and negative coefficient associated with it. This implies that if product A has a recommendation link to product B, then holding everything else constant (including product B’s demand), the demand for product B will be higher if products A and B are in different categories. Consider the following intuitive explanation, which relates to why a consumer might traverse recommendation hyperlinks. At any point in time, the consumer has the option to keep traversing, or to randomly jump to some other product by executing

a search. If products A and B are in the same category, the consumer is more likely to keep searching because the information yielded by the recommendations doesn’t seem to be too “new.” However, if product B is sufficiently different from product A, the consumer may be more motivated to click on the recommendation hyperlink and purchase the product.

#### 4.2. Full Model: Controlling for Price Endogeneity

The coefficients of sale price ( $x_2$ ) and list price ( $x_1$ ) are both significant and have opposite signs. As expected, the coefficient of the sale price (the price that consumers actually pay) is negative. However, the coefficient for the list price is positive. Naturally, there are many unobserved characteristics of a book that would influence its list and sale price and that might also influence its demand. One possible explanation is that after accounting for actual price (sale price), a higher list price simply implies a higher discount level, and a higher discount level is associated with higher demand.

Price is determined strategically (and thus endogenously) by Amazon.com; as a consequence, we repeat our analysis while instrumenting sale price simultaneously with the network influence variables. In the first stage, we generate estimated values for the sale price of the product based on the secondary market prices (those of the new and those of the used offers) and estimated values for the demand of the network neighbors and the complementary set (using only the number of new and used offers for those products). We then estimate our complete model in a second stage using the three estimated variables as independent variables. Because the secondary market is composed of individuals and small retailers, and because a consumer is very unlikely to buy the same product from more than one seller, it is reasonable to assume that the secondary market prices are exogenous to our model. The results of this estimation are presented in Table 4 (one column for each complementary set definition) and are directionally similar to the results presented above.

#### 4.3. Variation by Category

We expected to find divergences in the effect of the influence of the network across different categories of products. Therefore, we estimated a variety of interaction models, interacting category dummies at different levels of aggregation with our influence variables. However, we did not uncover any evidence of significant differences in the influence of the network on the demand for products across categories.

#### 4.4. Vintage: Newer Products Are Influenced More by the Network

Our next set of results measures how the influence of recommendations varies across products of different

**Table 4** Summary of Estimates for a Sample Day of Data (Sale Price Treated as Endogenous)

	Amazon CS	B&N CS	Weighted CS
$\ln(z_n)$	0.64***	0.89***	5.21**
Total demand for network neighbors	(0.11)	(0.02)	(2.33)
$\ln(z_c)$	0.38***	0.17***	-4.83*
Total demand for complementary set	(0.11)	(0.02)	(2.57)
$\ln(x_1)$	0.955***	0.78***	0.36
List price	(0.03)	(0.04)	(1.23)
$\ln(x_2)$	-1.02***	-0.79***	-0.72
Sale price	(0.04)	(0.05)	(1.11)
$\ln(x_3)$	-0.58***	-0.53***	0.95
In-degree of the product	(0.01)	(0.01)	(0.66)
$\ln(x_6)$	-0.01	-0.01*	-0.13
Assortative mixing of the product	(0.01)	(0.01)	(0.163)
$\ln(q_1)$	0.105	-0.41***	6.86
Avg list price of network neighbors	(0.795)	(0.12)	(11.83)
$\ln(q_2)$	-0.17	0.47***	-7.10
Avg sale price of network neighbors	(0.83)	(0.12)	(10.80)
$\ln(q_3)$	-0.22***	-0.44***	-5.07
Avg in-degree of network neighbors	(0.08)	(0.01)	(3.36)
$\ln(q_8)$	-0.02	-0.01	0.54
Avg assortative mixing of network neighbors	(0.02)	(0.01)	(0.68)
$\ln(w_1)$	-1.15	0.03	-5.22
Avg list price of complementary set	(0.79)	(0.06)	(13.18)
$\ln(w_2)$	1.28	-0.07	(10.80)
Avg sale price of complementary set	(0.83)	(0.06)	6.06
$\ln(w_3)$	-0.23***	-0.03**	4.77
Avg in-degree of complementary set	(0.08)	(0.01)	(3.81)
$\ln(w_6)$	0.04	-0.01	-0.56
Avg assortative mixing of complementary set	(0.03)	(0.01)	(0.71)
Constant	-0.61*	-0.12***	-4.83*
	(0.35)	(0.04)	(2.71)
Observations	111,811	68,613	100
R-squared	0.596	0.65	0.61

Notes. Standard errors in parentheses. CS, complementary set.  
 \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

vintage. We obtain these results by estimating the following linear interaction model:

$$\begin{aligned} \log(y) = & \alpha_1 \log(z_n) + \alpha_2 \log(z_c) + \alpha_{4,1} \log(x_1) \\ & + \alpha_{4,2} \log(x_2) + \alpha_{4,3} \log(x_3) + \alpha_{4,8} \log(x_8) \\ & + \alpha_{3,1} \log(q_1) + \alpha_{3,2} \log(q_2) + \alpha_{3,3} \log(q_3) \\ & + \alpha_{3,3} \log(q_8) + \alpha_{4,1} \log(x_1) + \alpha_{5,1} \log(w_1) \\ & + \alpha_{5,2} \log(w_2) + \alpha_{5,3} \log(w_3) + \alpha_{5,8} \log(w_8) \\ & + \alpha_6 R + \alpha_7 R \cdot \log(z_n) + \alpha_8 R \cdot \log(z_c). \end{aligned} \quad (4)$$

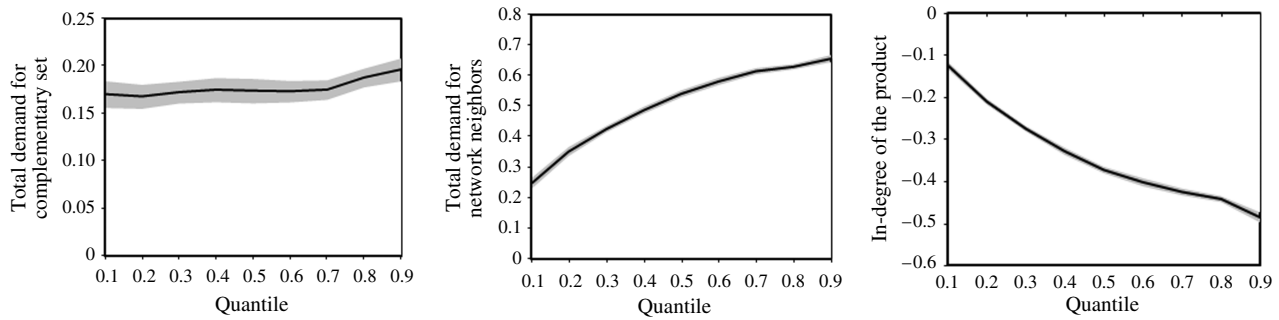
The results of this estimation are presented in Table 5, one column for each complementary set. The positive and significant coefficient of the corresponding interaction variable indicates that recently published books are more influenced by neighboring products. These

**Table 5** How Vintage Affects Network Influence

	Amazon CS	B&N CS	Weighted CS
$\ln(z_n)$	0.39***	0.84***	0.74**
Total demand for network neighbors	(0.09)	(0.01)	(0.31)
$\ln(z_c)$	0.59***	0.18***	0.47
Total demand for complementary set	(0.09)	(0.01)	(0.29)
$R \cdot \ln(z_n)$	0.61***	0.10***	-0.23
Interaction between recency and demand for network neighbors	(0.17)	(0.02)	(0.28)
$R \cdot \ln(z_c)$	-0.58***	-0.003	0.47
Interaction between recency and demand for complementary set	(0.17)	(0.02)	(0.34)
$\ln(x_1)$	0.71***	0.74***	1.36*
List price	(0.02)	(0.02)	(0.82)
$\ln(x_2)$	-0.77***	-0.74***	-1.39*
Sale price	(0.02)	(0.02)	(0.74)
$\ln(x_3)$	-0.53***	-0.50***	0.92
In-degree of the product	(0.01)	(0.01)	(0.60)
$\ln(x_6)$	-0.01**	-0.01*	0.01
Assortative mixing of the product	(0.005)	(0.006)	(0.13)
$\ln(q_1)$	0.19	-0.23***	0.99
Avg list price of network neighbors	(0.14)	(0.03)	(2.27)
$\ln(q_2)$	-0.28**	0.28***	-1.24
Avg sale price of network neighbors	(0.13)	(0.03)	(2.07)
$\ln(q_3)$	-0.14*	-0.41***	-0.65
Avg in-degree of network neighbors	(0.07)	(0.01)	(0.45)
$\ln(q_8)$	-0.02	-0.01*	-0.02
Avg assortative mixing of network neighbors	(0.03)	(0.007)	(0.15)
$\ln(w_1)$	-0.64***	-0.05***	-3.81
Avg list price of complementary set	(0.14)	(0.008)	(3.45)
$\ln(w_2)$	0.77***	0.005	4.35
Avg sale price of complementary set	(0.13)	(0.01)	(3.12)
$\ln(w_3)$	-0.30***	-0.042***	-0.90
Avg in-degree of complementary set	(0.08)	(0.01)	(0.76)
$\ln(w_6)$	0.03	-0.004	0.04
Avg assortative mixing of complementary set	(0.03)	-0.05***	(0.09)
R	0.13***	-0.05	-3.81
Recency	(0.03)	(0.03)	(2.85)
Constant	-0.71**	-0.14***	-0.82
	(0.29)	(0.032)	(3.20)
Observations	111,765	68,592	100
R-squared	0.605	0.653	0.705

Notes. This regression includes interaction terms for recency with network neighbors' demand and for recency with the complementary set demand ( $Recency = 1$  for books published after 2006). Standard errors in parentheses. CS, complementary set.  
 \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

findings are aligned with recent literature on information acquisition and observational learning (specifically, Moretti 2011, who shows that in the case of products for which consumers have a strong prior belief of quality, the effect of observational learning on sales will be smaller) and with theories of bounded rationality. Simply put, a consumer is likely to receive more valuable information from a link pointing to a newer book. It is more likely that the consumer

**Figure 4** How Influence Varies with Popularity

*Note.* This figure presents the coefficients of different quantiles for the demand for the complementarity set ( $z_c$ ) (left-most figure), the demand for the network neighbors ( $z_n$ ) (middle figure), and the focal product's in-degree (right-most figure).

was not aware of a newer book, or that this consumer is aware of an older book and either has chosen not to purchase it or already owns it, both of which situations yield no new demand. Similar results are obtained when a continuous age variable is used; these results are available on request.

#### 4.5. Popularity: Bestsellers Utilize the Network's Influence More Efficiently

It is reasonable to conjecture that the effect of network neighbors will differ across products with different levels of popularity. An econometric challenge in testing this conjecture quantitatively is that the popularity of a book is measured by its demand, which is the dependent variable in our regression equation. In our final set of results we therefore use quantile regression to estimate how the magnitudes of influence of our two different sets—the network neighbors and the complementary products—vary across products with different levels of popularity. This model estimates the relationship between our independent variables and the conditional quantiles of our dependent variable, providing a more complete picture of the conditional distribution of the dependent variable in question and enabling us to assess how influence varies across demand quantiles.<sup>12</sup>

The results of a sample quantile regression for one day of data are summarized in Table 6 and Figure 4. Notice that variation in the demand for the complementarity set ( $z_c$ ) has a stable effect across quantiles on demand for destination products.<sup>13</sup> However, the demand effect of the visible network neighbors ( $z_n$ ) increases substantially as the popularity of the focal product increases,<sup>14</sup> almost threefold as one moves

from the least popular to the most popular deciles. Thus, more popular books are influenced more positively by the demand for their network neighbors, even though the influence of the complementarity set does not vary with popularity.

## 5. Concluding Remarks

Electronic commerce sites have been organized as product networks for over a decade, and the density of these networks grows over time. It seems natural to expect these networks—the most visible and significant “store design” variable available to online retailers—to influence how consumer attention is directed and the ensuing choices consumers make. Our analysis of one such product network—the visible copurchase recommendation network of Amazon.com, perhaps the oldest such network—has provided new and extensive empirical evidence of the magnitude and variation of this influence. More specifically, our results show that the visibility of networks amplifies the shared purchasing of complementary products. We document how this influence varies along a number of different dimensions, which include product popularity and vintage. To the best of our knowledge, these are the first empirical results that identify the impact of visibility of a product network and associate it with variations in products' demand.

As the importance of electronic commerce continues to grow, the ability to control cross-product effects in electronic markets has become a key strategic marketing lever for firms, one that perhaps receives less attention than viral marketing or social media strategy. Our results demonstrate that a clear understanding of the extent to which networks impact demand and the ways in which the influence of these networks varies across different types of products can be critical. The differences we have highlighted across product popularity, vintage, and pricing may be of specific interest to managers navigating the new terrain of online network-driven marketing.

<sup>12</sup> This is in contrast with ordinary least-squares regression, which models the relationship between one or more covariates  $X$  and the conditional mean of a response variable  $Y$  given  $X$ . For more information about quantile regressions, see Koenker and Hallock (2001).

<sup>13</sup> The differences between the quantiles are not statistically significant.

<sup>14</sup> The differences are all statistically significant.



**Table 6** Quantile Regression Results

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\ln(z_n)$	0.295	0.413	0.472	0.528	0.58	0.617	0.641	0.678	0.74
<i>Demand for network neighbors</i>	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.00)***	(0.01)***
$\ln(z_c)$	0.269	0.264	0.277	0.278	0.266	0.263	0.267	0.26	0.245
<i>Demand for complementary set</i>	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.00)***	(0.01)***
$\ln(x_3)$	-0.229	-0.317	-0.387	-0.414	-0.443	-0.473	-0.496	-0.529	-0.588
<i>In-degree</i>	(0.00)***	(0.01)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.19)***	(0.00)***	(0.01)***
$\ln(x_2)$	-0.387	-0.414	-0.437	-0.464	-0.478	-0.499	-0.496	-0.511	-0.577
<i>Sale price</i>	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.02)***	(0.01)***	(0.01)***	(0.02)***	(0.26)***
$\ln(x_1)$	0.324	0.349	0.371	0.391	0.411	0.44	0.437	0.453	0.496
<i>List price</i>	(0.12)***	(0.01)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***
$\ln(x_4)$	0.094	0.116	0.125	0.13	0.143	0.149	0.154	0.162	0.168
<i>Total number of new offers</i>	(0.00)***	(0.00)***	(0.00)***	(0.01)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.01)***
$\ln(x_5)$	0.074	0.072	0.074	0.077	0.079	0.086	0.088	0.096	0.108
<i>Total number of used offers</i>	(0.01)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.01)***	(0.00)***	(0.00)***
$\ln(x_6)$	0.032	0.033	0.036	0.038	0.04	0.037	0.043	0.462	0.05
<i>Lowest price for new offers</i>	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
$\ln(x_7)$	0.042	0.045	0.047	0.049	0.052	0.055	0.055	0.058	0.066
<i>Lowest price for used offers</i>	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
$\ln(x_8)$	0.013	0.014	0.012	0.013	0.009	0.005	-0.011	-0.015	-0.023
<i>Assortative mixing</i>	(0.01)*	(0.01)*	(0.01)*	(0.01)*	0.01	0.01	0.01	0.01	(0.01)**
$\ln(q_2)$	0.037	0.076	0.075	0.095	0.115	0.143	0.15	0.136	0.153
<i>Avg sale price of network neighbors</i>	(0.02)*	(0.02)***	(0.01)***	(0.01)***	(0.02)***	(0.02)***	(0.02)***	(0.03)***	(0.03)***
$\ln(q_1)$	-0.086	-0.138	-0.136	-0.148	-0.162	-0.198	-0.195	-0.192	-0.209
<i>Avg list price of neighbors</i>	(0.02)***	(0.02)***	(0.01)***	(0.02)***	(0.02)***	(0.02)***	(0.02)***	(0.26)***	(0.03)***
$\ln(q_3)$	-0.169	-0.226	-0.228	-0.308	-0.345	-0.377	-0.414	-0.453	-0.522
<i>Avg in-degree of neighbors</i>	(0.01)***	-0.01	(0.00)***	(0.00)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
$\ln(q_4)$	0.004	-0.025	-0.045	-0.055	-0.068	-0.082	-0.086	-0.105	-0.123
<i>Avg number of new offers for neighbors</i>	0.01	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
$\ln(q_5)$	0.041	0.043	0.044	0.036	0.037	0.037	0.037	0.435	0.05
<i>Avg number of used offers for neighbors</i>	(0.01)***	(0.00)***	(0.01)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.01)***	(0.01)***
$\ln(q_6)$	0.024	0.027	0.024	0.019	0.016	0.015	0.016	0.017	0.012
<i>Avg lowest price of new offers for neighbors</i>	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)*	(0.01)*	0.01	0.01	0.01
$\ln(q_7)$	0.002	0.001	-0.002	-0.007	-0.009	-0.005	-0.008	-0.002	-0.002
<i>Avg lowest price of used offers for neighbors</i>	0.01	(0.01)***	0.01	0.01	(0.01)*	0.01	0.01	0.01	0.01
$\ln(q_8)$	-0.001	-0.001	-0.013	-0.015	-0.016	-0.017	-0.019	-0.021	-0.025
<i>Avg assortative mixing of neighbors</i>	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.01)***
<i>Constant</i>	2.485	2.8668	3.097	3.32	3.48	3.68	3.84	4.14	4.61
	(0.07)***	(0.05)***	(0.01)***	(0.05)***	(0.05)***	(0.06)***	(0.07)***	(0.05)***	(0.08)***
<i>Observations</i>	242,626								
<i>R-squared</i>	0.29								

Note. Standard errors in parentheses.

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

We acknowledge that our study has some limitations. First, although we control for observed product characteristics such as author, category, and vintage, there may be unobserved sources of demand correlation that are separate from the visible network effect. To this end, we constructed three different complementary sets for each product: a set based on Amazon.com’s future links, a set based on data from the B&N website, and a set based on a weighting system in which we evaluated the likelihood of each product in the database being linked to the focal product. Although those sets may control for much of the unobserved demand correlation, they cannot fully account for all sources of bias. Specifically, we are

unable to fully control for the endogenous link formation process.

Furthermore, as in many studies of network influence, our estimation equations are endogenous, as demand for products is included as both a dependent variable and an independent variable. We therefore used exogenous secondary market information to instrument both network neighbors and complementarity sets. We report on this analysis and discuss how we have experimented with a variety of other models in the Web appendix.

A more complete panel data analysis, as well as an analysis of the “diffusion” of influence through the copurchase network, remains part of our future

research plan. Data limitations preclude evaluating other omitted variables of consequence, such as the author's level of fame, or marketing expenditure levels. Despite these limitations, our contribution may be widely relevant to managers while also seeding a number of new directions for further research. Our hope is that these limitations are viewed not as a liability but as a path toward future research that extends our question while strengthening the theory and evidence that relate influential electronic networks to marketing inquiry.

Finally, we have provided the first evidence that influence across products is substantial when products are “slotted” according to shared consumer purchasing patterns. We do not investigate whether this strategy dominates manufacturer slotting fees or strategic retailer placements, but economic intuition suggests that delegating this decision indirectly to consumers mitigates the information asymmetry associated with other slotting approaches, eliminates contracting and inequity issues involved with manufacturer slotting fees, and simultaneously reduces the overhead that retailers might face when engaging in elaborate and programmatic slotting strategies that rely on limited and noisy data. It has been conjectured that basing merchandising on “what the consumer would want and benefit from most” is a profitable marketing approach in the long run. As the opportunities for exploiting and integrating collective consumer wisdom into marketing strategy are expanded by the increased popularity of social networks, user-generated content, prediction markets, and other Web 2.0 technologies, establishing that this intuition is right (or wrong)—either theoretically or empirically—represents an excellent new direction for investigation.

### Acknowledgments

The authors thank Pradeep Chintagunta; an anonymous associate editor; three anonymous referees; Albert Laszlo Barabasi; Eyal Carmi; Vasant Dhar; Nicholas Economides; Tulin Erdem; Sanjeev Goyal; William Greene; Panos Ipeirotis; Vardit Schwartz-Landsman; Charles Manski; Roy Radner; and seminar participants at Cambridge University, Columbia University, New York University, Northwestern University, Tel-Aviv University, and the Wharton School. The authors also thank Liron Sivan for excellent research assistance. A precursor to this paper won the award for Best Paper at the INFORMS 2008 Conference on Information Systems and Technology.

### References

- Aaker DA (2004) *Brand Portfolio Strategy* (Free Press, New York).
- Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci. USA* 106(51):21544–21549.
- Binken JLG, Stremersch S (2009) The effect of superstar software on hardware sales in system markets. *J. Marketing* 73(2):88–104.
- Bradlow E, Bronnenberg B, Russell G, Arora N, Bell D, Duvvuri S, Hofstede F, Sismeiro C, Thomadsen R, Yang S (2005) Spatial models in marketing. *Marketing Lett.* 16(3):267–278.
- Bramoulle Y, Djebbari H, Fortin B (2008) Identification of peer effects through social networks. *J. Econometrics* 150(1):41–55.
- Brin S, Page L (1998) The anatomy of a large-scale hypertextual Web search engine. *Comput. Networks ISDN Systems* 30(1–7):107–117.
- Brynjolfsson E, Hu YJ, Smith MD (2003) Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Sci.* 49(11):1580–1596.
- Cai H, Chen Y, Fang H (2009) Observational learning: Evidence from a randomized natural field experiment. *Amer. Econom. Rev.* 99(3):864–882.
- Chen Y, Wang Q, Xie J (2011) Online social interactions: A natural experiment on word of mouth versus observational learning. *J. Market. Res.* 48(2):238–254.
- Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* 43(3):345–354.
- Conley TG, Udry C (2010) Learning about a new technology: Pineapples in Ghana. *Amer. Econom. Rev.* 100(1):35–69.
- Dellarocas C, Katona Z, Rand WM (2009) Media, aggregators, and the link economy: Strategic hyperlink formation in content networks. Working Paper 10-13, NET Institute, New York.
- Edwards YD, Allenby GM (2003) Multivariate analysis of multiple response data. *J. Marketing Res.* 40(3):321–334.
- Foster AD, Rosenzweig MR (1995) Learning by doing and learning from others: Human capital and technical change in agriculture. *J. Political Econom.* 103(6):1176–1209.
- Ghose A, Han S (2011) An empirical analysis of user content generation and usage behavior on the mobile Internet. *Management Sci.* 57(9):1671–1691.
- Ghose A, Smith MD, Telang R (2006) Internet exchanges for used books: An empirical analysis of product cannibalization and welfare impact. *Inform. Systems Res.* 17(1):3–19.
- Godes D, Mayzlin D (2004) Using online conversations to study word-of-mouth communication. *Marketing Sci.* 23(4):545–560.
- Goldenberg J, Oestreicher-Singer G, Reichman S (2012) The quest for content: The integration of product networks and social networks in online content exploration. *J. Marketing Res.* Forthcoming.
- Goolsbee A, Chevalier JA (2003) Measuring prices and price competition online: Amazon and Barnes and Noble. *Quant. Marketing Econom.* 1(2):203–222.
- Hansen LP (1982) Large sample properties of generalised method of moments estimators. *Econometrica* 50(4):1029–1054.
- Hess JD, Gerstner E (1987) Loss leader pricing and rain check policy. *Marketing Sci.* 6(4):358–374.
- Jackson MO (2009) Networks and economic behavior. *Annual Rev. Econom.* 1:489–511.
- Kempe D (2011) Structure and dynamics of information in networks. Lecture notes, University of Southern California, Los Angeles. Accessed June 6, 2012, <http://www-bcf.usc.edu/~dkempe/teaching/structure-dynamics.pdf>.
- Koenker R, Hallock KF (2001) Quantile regressions. *J. Econom. Perspectives* 15(4):143–156.
- Lee LF (2007) Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *J. Econometrics* 140(2):333–374.
- Li S, Sun B, Wilcox RT (2005) Cross-selling sequentially ordered products: An application to consumer banking services. *J. Marketing Res.* 42(2):233–239.

- Libai B, Muller E, Peres R (2005) The role of seeding in multi-market entry. *Internat. J. Res. Marketing* 22(4):375–393.
- Libai B, Muller E, Peres R (2009) The role of within-brand and cross-brand communications in competitive growth. *J. Marketing* 73(3):19–34.
- Manchanda P, Ansari A, Gupta S (1999) The “shopping basket”: A model for multicategory purchase incidence decision. *Marketing Sci.* 18(2):95–114.
- Manchanda P, Rossi P, Chintagunta P (2004) Response modeling with nonrandom marketing-mix variables. *J. Marketing Res.* 41(4):467–478.
- Manski C (1993) Identification of endogenous social effects: The reflection problem. *Rev. Econom. Stud.* 60(3):531–542.
- Mayzlin D, Yoganasimhan H (2012) Link to success: How blogs build an audience by promoting rivals. *Management Sci.* 58(9):1651–1668.
- Moretti E (2011) Social learning and peer effects in consumption: Evidence from movie sales. *Rev. Econom. Stud.* 78(1):356–393.
- Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev.* 45(2):167–256.
- Newman MEJ, Barabasi AL, Watts DJ (2006) *The Structure and Dynamics of Networks* (Princeton University Press, Princeton, NJ).
- Niraj R, Padmanabhan V, Seetharaman PB (2008) A cross-category model of households’ incidence and quantity decisions. *Marketing Sci.* 27(2):225–235.
- Oestreicher-Singer G, Sundararajan A (2012a) Influence in social networks: An identification result. Working Paper EC12-01, Stern School of Business, New York University, New York.
- Oestreicher-Singer G, Sundararajan A (2012b) Recommendation networks and the long tail of electronic commerce. *MIS Quart.* 36(1):65–83.
- Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.
- Seetharaman PB, Chib S, Ainslie A, Boatwright P, Chan T, Gupta S, Mehta N, Rao V, Strijnev A (2005) Models of multi-category choice behavior. *Marketing Lett.* 16(3):239–254.
- Shocker AD, Bayus BL, Kim N (2004) Product complements and substitutes in the real world: The relevance of “other products.” *J. Marketing* 68(1):28–40.
- Sriram S, Chintagunta PK, Agarwal MK (2010) Investigating consumer purchase behavior in related technology product categories. *Marketing Sci.* 29(2):291–314.
- Stephen A, Toubia O (2010) Deriving value from social commerce networks. *J. Marketing Res.* 7(2):215–228.
- Stock JH, Yogo M (2005) Testing for weak instruments in linear IV regression. Andrews DW, Stock JH, eds. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (Cambridge University Press, New York), 80–108.
- Sun M (2012) How does variance of product ratings matter? *Management Sci.* 58(4):696–707.
- Sundararajan A (2008) Local network effects and complex network structure. *B.E. J. Theoret. Econom.* 7(1):Article 46.
- Susarla A, Oh JH, Tan Y (2012) Social networks and the diffusion of user-generated content: Evidence from YouTube. *Inform. Systems Res.* 23(1):23–41.
- Trusov M, Bucklin RE, Pauwels K (2009) Effects of word-of-mouth versus traditional marketing: Findings from an Internet social networking site. *J. Marketing* 73(5):90–102.
- Tucker C, Zhang J (2011) How does popularity information affect choices? A field experiment. *Management Sci.* 57(5):828–842.
- Wedel M, Zhang J (2004) Analyzing brand competition across subcategories. *J. Marketing Res.* 41(4):448–456.
- Yang S, Allenby GM (2003) Modeling interdependent consumer preferences. *J. Marketing Res.* 40(3):282–294.
- Zhang J (2010) The sound of silence: Observational learning in the U.S. kidney market. *Marketing Sci.* 29(2):315–335.