

# The Network Value of Products

Traditionally, the value of a product has been assessed according to the direct revenues the product creates. However, products do not exist in isolation but rather influence one another's sales. Such influence is especially evident in e-commerce environments, in which products are often presented as a collection of web pages linked by recommendation hyperlinks, creating a large-scale product network. The authors present a systematic approach to estimate products' true value to a firm in such a product network. Their approach, which is in the spirit of the PageRank algorithm, uses available data from large-scale e-commerce sites and separates a product's value into its own intrinsic value, the value it receives from the network, and the value it contributes to the network. The authors demonstrate their approach using data collected from the product network of books on Amazon.com. Specifically, they show that the value of low sellers may be underestimated, whereas the value of best sellers may be overestimated. The authors explore the sources of this discrepancy and discuss the implications for managing products in the growing environment of product networks.

**Keywords:** product value, cross-selling, electronic commerce, recommendation systems, social networks

An interesting development in retail in recent years is the emergence of online product networks, in which a large number of items—represented by a collection of web pages—are linked to one another. In most cases, the links between product pages are generated by online recommendation systems that use collaborative filtering algorithms. The products' web pages are the nodes of the product network, and the recommendation hyperlinks are the edges. Imagine browsing an e-commerce site as being analogous to walking the aisles of a physical store; the placement of a product in the network of interconnected web pages constitutes its virtual “shelf placement.” Amazon.com has created what is probably the best-known online product network: a co-purchase network, in which each product page shows prospective customers the other products that were purchased by buyers of the same product. This mechanism, which can substantially affect consumer search (Kim, Albuquerque, and Bronnenberg 2011), has been increasingly used by diverse sellers such as Zappos.com, Hotels.com, and Walmart.com to facilitate consumer navigation and to cross-sell products to customers, thus more fully exploiting the inherent relationships among products.

Here, we highlight the issue of assessing the value of a product in such environments. Understanding the full value

of a product (a category or a brand) can help managers recognize which products to offer (or to stop offering) to customers, which products to promote, and how to better price different products. Furthermore, it can change advertising strategies—for example, by influencing advertisers' bidding behavior in pay-per-click environments. Traditionally, marketers have assessed the value of a product or brand according to the direct revenues the product creates, for example, based on the expected discounted cash flow in measurement methods such as that used by Interbrand (Clifton, Simmons, and Ahmad 2009). Yet the true value generated by a product that is part of a network, which we label “network value,” should take cross-product effects into account. Specifically, it should consider both the revenues an item generates by directing traffic to other items and the revenues an item is not “entitled” to due to traffic directed to it by other items.

In this article, we propose a method for assessing the network value of items in a given large-scale product network, using an approach that is in the spirit of the PageRank algorithm (Brin and Page 1998) popularized by Google for assessing the popularity of web pages. For each focal item, we differentiate the *intrinsic value* portion of its revenue, which is self-generated by the item (and incorporates the marketing activity of the retailer), and the *incoming value* portion of its revenue, which is driven by the incoming recommendation links pointing from other items to the focal item. We assume that the incoming value of a given product “belongs” to the items that point to that product rather than to the product itself. Thus, we define a product's *network value* (i.e., the value that takes into account its network relationship) as the sum of its intrinsic value and the value it generates for its neighbors through its outgoing links, which we label *outgoing value*. The approach we present here is applicable to large-scale databases and can be implemented relatively straightforwardly, relying on observable data. This is of notable importance given that a product's value is of interest not only to the retailer but also to external parties

---

Gal Oestreicher-Singer is Assistant Professor (e-mail: galos@tau.ac.il), and Liron Sivan is a doctoral candidate (e-mail: sivanlir@post.tau.ac.il), Faculty of Management, Recanati Business School, Tel Aviv University. Barak Libai is an Associate Professor, Interdisciplinary Center (IDC) (e-mail: libai@idc.ac.il). Eyal Carmi is senior software engineer, Google Inc. (e-mail: ecarmi@google.com). Ohad Yassin is Director of Engineering and Operations, SAP (e-mail: ohad.yassin@gmail.com). The authors thank Ron Shachar, Jacob Goldenberg, and Eitan Muller for their helpful comments. This study benefited from support from the Israeli Institute for Business Research, Israel National Science Foundation, and the Net Institute. Financial support from the Google Inter-University Center for Electronic Markets and Auctions is gratefully acknowledged. Leigh McAlister served as area editor for this article.

---

such as product manufacturers, whose access to internal sales data may be limited.

We illustrate our approach by applying it to a large product network of books collected from Amazon.com. We focus on the differences in network value across books and specifically examine how the network value of high-selling items differs from that of low-selling items. We show that for the network we analyze, the value of low-selling items is underestimated compared with that of best sellers, and we use the product network value approach to demonstrate how the distribution of the number of incoming and outgoing links, and the extent to which these links are influential, contribute to this phenomenon.

The contribution of this research to the literature is thus twofold. First, this work broadens the analysis of interproduct purchase effects, focusing on the need to consider products as part of a large network. Such a network view can extend the scope of marketing applications such as market-basket analysis and cross-selling analysis, which have traditionally focused on dyads of products because of the complexity associated with larger-scale investigation (Blattberg, Kim, and Neslin 2008).

Second, this research highlights the need to consider the different means by which a product generates value for a firm. There is an analogy to the case of social networks, in which there is an increasing understanding that a customer's value to a firm stems not only from his or her purchases but also from his or her social influence—that is, through word of mouth and imitation (Hogan, Lemon, and Libai 2003; Kumar, Petersen, and Leone 2010; Libai, Muller, and Peres 2012). In the same spirit, we aim to understand how the value that items obtain from and provide to the network affects their overall value. This enables a new perspective on a firm's product portfolio based on the types of value each product contributes and receives.

We organize the remainder of this article as follows: After providing some background to our research, we introduce our theoretical model for computation of the network value of products. We then apply our model to data from Amazon.com; specifically, we examine its implications for the estimated value of the different revenue tiers. We then discuss the implications of our results and directions for further research.

## Background

Several research streams relate to the work presented here. First, marketing researchers have expressed much interest in consumers' social effects and how the flow of influence in a social network is driven by the network structure and actors' characteristics (Van den Bulte and Wuyts 2007; Zubcsek and Sarvary 2011). For example, recent work in this area has shown how an actor's connectivity drives the actor's influence (Goldenberg et al. 2009; Hinz et al. 2011; Iyengar, Van den Bulte, and Valente 2011; Katona, Zubcsek, and Sarvary 2011). Our study takes a parallel look at influence in product networks and shows how the integration of incoming and outgoing links can explain an item's overall

"influence" in a product network. We also show how the assessment of influential nodes may differ between social and product networks.

A second relevant stream of research deals with the effects associated with interrelated products. It is widely recognized that purchases across categories are correlated among consumer goods that are complements or substitutes for one another (Raeder and Chawla 2011; Seetharaman et al. 2005) and that such interrelated effects can be found in multiple cases, such as the effect of a "loss leader" (Hess and Gerstner 1987), software/hardware effects (Binken and Stremersch 2009), and cross-brand word of mouth (Libai, Muller, and Peres 2009). To date, the investigation of interproduct associations has largely focused on dyads or on a small number of entities; for example, researchers have shown that pricing of bacon affects the demand for eggs, and vice versa (Niraj, Padmanabhan, and Seetharaman 2008). However, the demand for these two categories may be correlated with demand for other categories, which are in turn related to additional categories. This is the case for large-scale online product networks, and therefore, new approaches to explore such relationships are needed.

The third research stream consists of a group of studies that have begun to explore product networks. One of the basic challenges in dealing with product networks is to show that the links can indeed create an effect beyond the underlying correlation between items. Two recent studies have demonstrated the effect of links. First, Stephen and Toubia (2010) use data from a unique field experiment in the context of an online social commerce network in which they had data before and after links had been formed among sellers. This enabled them to show how such links affect sales and how, for example, a higher number of incoming links increases the profitability of the connected seller.

Second, Oestreicher-Singer and Sundararajan (2012b) investigate a product network of books on Amazon.com, which is similar to the data analyzed here. They control for alternative explanations of demand correlation using a variety of approaches and show that the explicit visibility of a co-purchase relationship leads to a notable amplification of the influence that complementary products have on one another's demand levels. Another significant finding of their study, which we discuss in greater detail subsequently, is that best sellers are better able to benefit from such links. That is, the visible incoming links of a best seller create more sales than those of lower-selling items.

Other research in this stream demonstrates the effect of product network recommendations on search (Kim, Albuquerque, and Bronnenberg 2010, 2011) and on sales (De, Hu, and Rahman 2010) and how link design can affect the effectiveness of those recommendations (Bodapati 2008). The current research complements this stream by aiming to better understand how value is actually created at the item level: given that recommendation links indeed affect consumption, how can we assess the different levels of value that an item contributes to and takes from the network, and how does this distribution of value differ for different items?

# Modeling Network Value of a Product

## The Setting

We consider a large-scale network of interlinked products. The “outdegree” of product  $u$  represents the number of links that originate from product  $u$  and point to other products, and the “indegree” is the number of links that point to  $u$  from other products. To demonstrate our approach, we use the example of the recommendation product network of books on Amazon.com. In that network, outdegree and indegree are determined by the links Amazon.com creates on the basis of co-purchases of books.

The problem we analyze is when a firm needs to understand the actual value contribution and the types of value generated by each product in its database. Note that our aim is not to analyze the optimal policy of the firm in shaping the network, which is an intriguing issue but beyond the scope of this research. Rather, we accept the structure of the network and the overall sum of revenue of all items in a product network as given. We examine how to redistribute this sum. Therefore, our approach is applicable not only to retailers who can manipulate the product network but also to external parties, such as product manufacturers (e.g., publishers of books sold on Amazon.com), that are not able to affect the links in the product network and must accept the network as given.

Table 1 provides a summary of the variables we use in this section. We divide the revenue of a product into two parts: (1) The intrinsic value portion of the revenue is self-generated by the item. It can be conceptualized as the revenue the product would be expected to yield on that website if it were not connected to others. (2) A product’s incoming value is driven by the recommendation links that point to that product from other products. Thus, for product  $u$ ,

$$\text{Revenue}(u) = \text{Intrinsic Value}(u) + \text{Incoming Value}(u).$$

We assume that an item’s total value consists of the product’s intrinsic value together with the product’s contribution to the incoming values of the products it recommends. We label the latter contribution as the “outgoing value” of the focal product. We refer to the product’s total contribution to the firm (i.e., its intrinsic value together with its outgoing value) as its “network value”:

$$\text{Network Value}(u) = \text{Intrinsic Value}(u) + \text{Outgoing Value}(u).$$

This view is consistent with previous work assessing the value of customers in a network by distinguishing customers’ intrinsic value and the value they provide to the network (Domingos and Richardson 2001).

## PageRank as a Benchmark

Our aim is to develop an approach that will reallocate the value a product generates according to the full recommendation system that the product is a part of. Probably the best-known computational tool that allows for a full network approach is PageRank (Brin and Page 1998), which is essentially an eigenvector centrality measure. This measure has been used for various applications involving ranking web pages. The best known application is Google’s ranking system, but PageRank has also been used for various academic research purposes—for example, for understanding optimal advertising on the web (Katona and Sarvary 2008).

The original PageRank algorithm provides a ranking of the importance of a web page in the hyperlinked structure of the web based on the following model:

$$(1) \quad \text{PageRank}(u) = \sum_{v \in \text{In}(u)} \frac{\text{PageRank}(v)}{\text{Out}(v)},$$

where  $\text{In}(u)$  is the set of web pages (nodes) linking to node  $u$ , and  $\text{Out}(v)$  is the number of outgoing links from node  $v$ .

**TABLE 1**  
**Definitions**

Term	Definition
$\text{In}(u)$	The set of web pages (nodes) linking to node $u$
$\text{Out}(v)$	The set of web pages (nodes) to which node $v$ links
$P(u)$	The price of product $u$
$\text{Demand}(u)$	The number of units of product $u$ sold
$\text{Revenue}(u)$	The total revenue a book generates (price $\times$ units)
$\text{Impressions}(v)$	The number of people visiting product $v$ ’s page (also frequently referred to as page views)
$\alpha_{v \rightarrow u}$	The RCR associated with the dyad $(v, u)$ , which represents the probability that a customer exposed to a link to product $u$ on $v$ ’s page will purchase product $u$
$\text{Intrinsic Value}(u)$	The portion of product $u$ ’s revenue that is not generated by incoming links from other items in the network
$\text{Incoming Value}(u)$	The sales of product $u$ that are attributed back to the network
$\text{Outgoing Value}(u)$	The contribution of product $u$ to the incoming values of products it recommends
$\text{Network Value}(u)$	The augmented value generated by a product that is part of a network. The sum of the product’s intrinsic value and the value it generates for its neighbors through its outgoing links
$k_u$	The percentage of eventual buyers who viewed book $u$ and purchased book $u$

Notes: RCR = recommendation conversion rate.



Intuitively, PageRank is based on a simple model of behavior: a consumer who “surfs” the network randomly follows any one of the links on a page with equal probability.<sup>1</sup> The algorithm is computed iteratively and thus takes into account the effect of the entire network on each page.

Mathematically, in each iteration, the algorithm divides a page's PageRank evenly among its successors (i.e., the pages it links to) in the network. Thus, the ranking of a page is ultimately the stationary probability that a random surfer who begins at a random page will visit the specific page. Therefore, a page can gain a high ranking by having either many pages or a few highly ranked pages that point to it. Although it is widely used as a measure of a node's importance to a network, fundamentally, PageRank provides a proxy for the extent to which the network directs traffic to the node in question. PageRank can therefore be used as a benchmark value for the effect of the network on the traffic to a product's page (and thus its demand).

### A Product Network Value Model

The approach we use to determine product value is similar to PageRank, with a fundamental difference: we focus on the traffic (value) a product *creates* for other products, not only on the traffic it receives. Furthermore, similar to PageRank, we want to take into account that different links (recommendations) generate different levels of traffic; thus, it is not enough to simply evaluate numbers of links. For example, in the context of Amazon.com, a link from Dan Brown's best seller *The Da Vinci Code* is likely to be a more fruitful recommendation compared with one from a lower-selling book.

We define impressions(*v*) as the number of people visiting product *v*'s page (also frequently referred to as “page views”) and observe its outgoing links. It is evident that not every link exposure leads to a purchase. We therefore define  $\alpha_{v \rightarrow u}$  as the recommendation conversion rate (RCR) associated with the product dyad (*v*, *u*); this RCR represents the probability that exposure to a link on *v*'s page will result in a purchase of *u*. This probability is a combination of the probability that a link will be clicked on (frequently referred to as “click-through rate”) and the probability that the user's visit to the next page will result in a purchase. Essentially, the RCR can be thought of as a “cross-selling conversion rate.”

We can now define the incoming value of a product, that is, the sales that are attributed back to other products in the network, as

$$(2) \text{Incoming Value}(u) = \sum_{v \in \text{In}(u)} \alpha_{v \rightarrow u} \times \text{Impressions}(v) \times P(u),$$

where  $P(u)$  is the price of product *u*. Note that the greater the volume of traffic directed to the product from neighbor-

ing products (i.e., the greater the number of impressions of its neighbors or the link's RCR), the larger the fraction of the product's revenue that stems from its incoming value rather than its intrinsic value. For example, the incoming value of a book on Amazon.com that is recommended by many best sellers should be greater than that of a book that earns similar revenue despite not getting many recommendations or receiving recommendations from books that are not purchased often.

The remaining revenue generated by an item is by definition its intrinsic value (i.e., the revenue portion that is not generated by incoming links from other items in the network):

$$(3) \text{Intrinsic Value}(u) = \text{Revenue}(u) - \text{Incoming Value}(u)$$

$$= P(u) \left[ Q(u) - \sum_{v \in \text{In}(u)} \alpha_{v \rightarrow u} \times \text{Impressions}(v) \right].$$

The outgoing value of item *u* is then the sum of all revenues that item *u* generates by recommending other products:

$$(4) \text{Outgoing Value}(u) = \sum_{w \in \text{Out}(u)} \alpha_{u \rightarrow w} \times \text{Impressions}(u) \times P(w),$$

where  $\text{Out}(u)$  is the set of web pages (nodes) to which node *u* links. Adding the intrinsic value of item *u* to the product's outgoing value, we obtain an expression for *u*'s network value:

$$(5) \text{Network Value}(u) = \text{Intrinsic Value}(u) - \text{Outcoming Value}(u)$$

$$= \left[ Q(u) - \sum_{v \in \text{In}(u)} \alpha_{v \rightarrow u} \times \text{Impressions}(v) \right] \times P(u) + \sum_{w \in \text{Out}(u)} \alpha_{u \rightarrow w} \times \text{Impressions}(u) \times P(w),$$

where each product *v* points to product *u* and contributes to its incoming value, and product *u*'s outgoing value stems from each product *w* to which *u* points. Note that a product cannot recommend itself, and thus, by definition, *u* is unequal to *v*.

Two things should be noted about the preceding model. First, our aim here is to redistribute the value in an existing product network in which the overall revenue is given. Removal of a product or a link from the network would alter the network structure. As a result, the total revenue for the new network—along with the intrinsic value, outgoing value, and incoming value (and thus the network value) of the remaining products—would need to be recomputed. Similarly, if a link were removed, the incoming value of the product that the link pointed to would likely decrease (while its intrinsic value would remain the same), which would probably result in lower demand for the product.

Second, the values assessed with this approach are platform dependent. In particular, the intrinsic value of a product may depend on the specific website. Although such value is clearly related to the inherent attractiveness of the item, intrinsic value may still differ between websites, as it is affected by promotion, price, ratings, and reviews and the ease of reaching the item from within the website (e.g., using search tools available). In addition, the incoming and

<sup>1</sup>This model is often extended to include the possibility that the surfer might not follow one of the links on the page but rather jump to a random page with probability  $(1 - d)$  (this probability is also referred to as the “damping factor”); in this case,

$$\text{PageRank}(u) = \frac{1 - d}{N} + d \sum_{v \in \text{In}(u)} \frac{\text{PageRank}(v)}{\text{Out}(v)}.$$

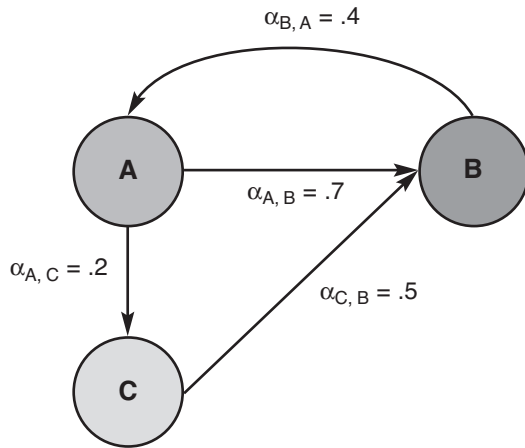
outgoing links can vary among platforms, and so the network value of the same item may differ greatly between networks. However, a comparison of two networks we had data on (Amazon.com and Barnesandnoble.com; see the “Robustness Checks” section) suggests that the difference may not be that large.

### An Illustrative Example

We next use a simple example to illustrate the application of the model. Consider Figure 1, in which three linked products in an online e-commerce site are presented, and the RCRs,  $\alpha_{v \rightarrow u}$ , which represent the probability for each dyad  $(v, u)$  that exposure to a link on  $v$  will result in a purchase of  $u$ , are given.

Table 2 provides additional information on this small product network: for each item, the table presents the quantity sold (column 1), the price per unit (column 4), and consequently the total revenue (column 5). The quantity of purchases of a product indicates the number of impressions it receives (and thus recommendation exposure); however, impressions can also stem from people who look at the product but do not purchase it. To move from quantity to impressions, we can use a number that is often reported for different e-commerce web sites: the conversion rate, which is the percentage of page visits that actually result in a purchase (to differentiate this number from the cross-product RCR, we label it the “classical conversion rate”). Thus, to get to the number of impressions (column 2), we can divide the quantity (column 1) by the classical conversion rate (column 3).

**FIGURE 1**  
Sample Graph Containing Three Products



To understand the full value of each product, we first compute the intrinsic and incoming values of each product. These refer, respectively, to the portion of the revenue that is self-generated by the item (intrinsic) and the portion of the revenue driven by the recommendation links pointing from other items to the focal item (incoming). These values are presented in columns 6 and 7 of Table 2. For example, using Equation 2, we compute the incoming value of product B as follows:

$$\begin{aligned} \text{Incoming Value}(B) &= \sum_{v \in \text{In}(B)} \alpha_{v \rightarrow B} \times \text{Impressions}(v) \times P(B) \\ &= .07 \times 13.33 \times \$150 + .05 \times 28.57 \times \$150 \\ &= \$139.96 + \$214.28 = \$354.29. \end{aligned}$$

We compute B’s intrinsic value using Equation 3 as follows:

$$\begin{aligned} \text{Intrinsic Value}(B) &= \text{Revenue}(B) - \text{Incoming Value}(B) \\ &= \$750 - \$354.29 = \$395.71. \end{aligned}$$

Note that product C is responsible for most of its own revenue (i.e., it has a high intrinsic value), whereas product B “owes” almost half its revenue to network traffic (i.e., it has a high incoming value).

The second step of our approach assigns the incoming value of each product back to its incoming links. For example, product B’s incoming value is \$354.29, which is assigned back to products A and C, in proportion to the strength of their recommendations (which results in \$139.97 being assigned to product A and \$214.28 being assigned to product C). Similarly, product A’s incoming value (\$25) is assigned back to B (its only incoming link), and product C’s incoming value (\$5.33) is assigned back to A (its only incoming link).

After assigning the incoming values back to the incoming links, we can compute the outgoing value of each product (see Equation 4). A product’s outgoing value is the sum of the incoming values that the focal product generates for its neighbors. Table 2, column 8, presents the outgoing value of each product. In our example, B was assigned product A’s incoming value, such that

$$\begin{aligned} \text{Outgoing Value}(B) &= \sum_{w \in \text{Out}(B)} \alpha_{B \rightarrow w} \times \text{Impressions}(B) \times P(w) \\ &= .04 \times 6.25 \times \$100 = \$25. \end{aligned}$$

The last step is computing the network value of each product as the sum of its intrinsic value and its outgoing

**TABLE 2**  
Network Value Algorithm Example, Iteration 1

	1	2	3	4	5	6	7	8	9
			Classical Conversion						
Item	Quantity	Impressions	Rate	Price (\$)	Revenue (\$)	Intrinsic Value (\$)	Incoming Value (\$)	Outgoing Value (\$)	Network Value (\$)
A	10	13.33	.75	100	1,000.00	975.00	25.00	145.33	1,120.33
B	5	6.25	.80	150	750.00	395.71	354.29	25.00	420.71
C	20	28.57	.70	20	400.00	394.67	5.33	214.29	608.95
Total					2,150.00	1,765.38	384.62	384.62	2,150.00

value (Equation 5). Table 2, column 9, presents the network value of each product. Note that the total network value over the entire network is equal to the total revenue. That is, our model simply redistributes the revenues among the products. For example, for product B the network value is

$$\begin{aligned}\text{Network Value(B)} &= \text{Intrinsic Value(B)} + \text{Outgoing Value(B)} \\ &= \$395.71 + \$25 = \$420.71.\end{aligned}$$

Note that product C generates almost all its own revenues, and only 1.3% of its revenues are generated by the recommendation made by product A. Product B, in contrast, is very dependent on external recommendations, which generate 47.2% of its revenue. After attributing the revenues back to the items that generated them, we observe that the revenue that product C generates by recommending other items is equal to more than 41% of the revenue it generates through its own sales. Finally, note that the total network value of product C is much higher than its revenue.

### **Iterations and Convergence**

A fundamental question when exploring influence in networks is that of the “ripple” effect: To what extent can we assume that the network value created by an item spreads in a contagion-like way into the network, beyond the first degree of separation? Take, for example, a product network in which item A recommends item B, which recommends item C, which recommends item D. Consider item B in that network. When applied once, the model attributes to item B a proportion of the revenue from sales of item C, and item A is attributed a proportion of the revenue from sales of item B. This assumes that the recommendation effect of an item stops at the books it recommends and no ripple effect process occurs. However, the picture may be more complicated if there is some effect beyond the first degree of separation. Some of the revenue from sales of item C that is attributed to item B should actually be attributed backward to item A, which generated part of item B’s traffic to begin with. Indeed, B’s actual contribution to C’s revenue should be decreased by the proportion of A’s contribution to B. In the same manner, B has some part in C’s contribution to D’s revenue, in that some of C’s value comes from the incoming value driven by B. In other words, an item is entitled to a share of another item’s network value, not just its revenue.

We dealt with this issue by building an iterative process that enables the outgoing value of a given item to be “pushed” back to other items at higher degrees of separation (for details of this approach, see the section “The Convergence Process” in the Web Appendix at [www.marketingpower.com/jm\\_webappendix](http://www.marketingpower.com/jm_webappendix)). It is important to note here that although theoretically we could envision effects that stretch deep inside the product network, we assume a strong decay across degrees of separation. This is consistent with findings from the social network literature that show that influence is locally bound, with some researchers suggesting three degrees of separation as the typical limit (Christakis and Fowler 2009). It is also in line with findings that suggest that the average shopping basket on sites such as Amazon.com and Barnesandnoble.com contains fewer than three items (De los Santos 2008). Consistent with previous

research, influence decays across the network exponentially (Carmi, Oestreicher-Singer, and Sundararajan 2009; Deschates and Sornette 2005). With our approach, almost all the effect is confined to the close network, and only a relatively small part travels through to higher degrees of separation.

## **Applying the Network Value Assessment: The Issue of Revenue Tiers**

Next, we demonstrate how the application of the product network value approach can be used to study the full value of items in an online product network of books. We ask the following questions: To what extent does the revenue of an item—as reflected in the item’s sales rank—indicate its network value? Can we use the product network value approach to understand the full value of best sellers as compared with low sellers?

### **The Amazon.com Co-Purchase Network**

*Product database.* We created a database of product data including pricing, sales rank, rating, and co-purchase network information for more than 900,000 books sold on Amazon.com on a particular day in 2010. The sales rank is a number associated with each product on Amazon.com, which measures its demand relative to other products. The lower the number is, the higher the sales of that particular product. Although sales rank is not an exact measure of sales, previous research has suggested methods of converting it into a sales measure. Thus, we computed demand on the basis of the sales rank data provided by Amazon.com and by following a log-linear conversion model suggested by Chevalier and Goolsbee (2003) and Brynjolfsson, Hu, and Smith (2009) with the correction Gabaix and Ibragimov (2009) demonstrate.

Amazon.com’s recommendation system is probably the best known among electronic retailers and has been widely used to demonstrate the role of recommender systems in general (Fleder and Hosanagar 2009; Kim, Albuquerque, and Bronnenberg 2011). Each product on Amazon.com has an associated web page containing a set of “co-purchase links,” which are hyperlinks to products that were co-purchased most frequently with that product on Amazon.com (listed under the title “Customers who bought this item also bought ...”). In most online product networks that are based on a recommendation system, the number of recommended items is limited. On Amazon.com, for example, the co-purchase set for each web page was limited to five items until relatively recently. Currently, more entries are allowed, but recommendations are effectively limited to no more than five for most users due to screen size constraints. We collected the network using a snowball sampling method, which started from a number of seed books and resulted in a large connected component. We report the details of the data collection in the section “Data and Data Collection” in the Web Appendix ([www.marketingpower.com/jm\\_webappendix](http://www.marketingpower.com/jm_webappendix)).

*Ultimate purchase decision data.* At the time the data were collected, another source of data was available that enabled us to produce a richer representation of the effects



in the product network (this source is no longer available in the full form described in what follows): Near the bottom of each book's page, Amazon.com presented a list titled "What Do Customers Ultimately Buy after Viewing This Item?" which showed the books purchased by visitors to the page and the percentage of visitors who bought each book (including the focal book). As we elaborate in what follows, this information, which is a synthesis of Amazon.com's click-through data, provided us with a proxy for the recommendation conversion rate for different items. Ultimate purchase decision (UPD) data were available for the majority of the books (764,769 books of 916,944 of the books in the recommendation network), so the final network we used included the connected books for which we have UPD data.

### **Adjustments Made for the Empirical Data Set**

We adapted the data collected from Amazon.com to our model as follows. The number of impressions,  $\text{Impressions}(v)$ , is the number of Amazon.com buyers visiting product  $v$ 's page. Although this information was not directly available, the value of impressions for each book could be calculated according to the UPD data and the demand for the book. In addition to information on the eventual purchase of other books, the UPD network also reports the percentage of book  $v$ 's visitors who actually purchased book  $v$  (labeled  $k_v$ ). As mentioned previously, we computed the demand for book  $v$  (denoted  $\text{Demand}(v)$ ) using the sales rank. Thus, given the demand for book  $v$ , if  $k_v$  percent of eventual buyers who viewed book  $v$  purchased book  $v$ , the number of impressions for book  $v$  is

$$(6) \quad \text{Impressions}(v) = \frac{\text{Demand}(v)}{k_v}.$$

Note that the population in which we are interested and that is included in our data set consists solely of people who eventually purchased something.

*The value of the RCR.* A key parameter value needed to apply the preceding model is that of  $\alpha_{v \rightarrow u}$ : the RCR, or the probability that a consumer exposed to a recommendation link from book  $v$  will purchase book  $u$ . For a link between a given book  $v$  and a given book  $u$  in the recommendation network, we use the percentage of viewers of book  $v$  who ultimately bought book  $u$  as  $\alpha_{v \rightarrow u}$ . These values are available to us from the UPD data described previously. (When recommended book  $u$  was not included in the UPD data, we set the RCR to zero.) For example, if the UPD network indicated that 27% of consumers who visited product  $v$ 's page ultimately bought product  $u$ , we set the value of  $\alpha_{v \rightarrow u}$  to .27.

*Endogeneity issues in the Amazon.com network.* It is important to acknowledge our limitations when demonstrating our approach on the Amazon.com data—in particular, endogeneity issues, which present challenges to the study of social networks (Manski 2000) and also exist in the case of product networks.

A primary concern stems from the fact that product networks for retailers such as Amazon.com are created through the recommendation systems, so network position is a func-

tion of past sales, which biases the study of the network's influence on subsequent sales. Thus, sales (in our case, sales rank) are endogenous on the network. Given this issue, using correlations between purchases, one can easily overestimate the actual strength a link between two items, as a person might have purchased the recommended book even in the absence of a recommendation. In our context, this means that the actual RCR values may be lower than those inferred from a straightforward measurement of purchases using clickstream data.

As in the case of separating homophily from contagion for social networks, estimating the unbiased RCR in product networks is far from trivial, even for the retailer itself, and demands comprehensive data and extensive investigation that typically includes a dynamic data analysis over time for the same item. Recent research investigating online product networks has used such analysis to demonstrate that recommendation networks affect demand beyond alternative sources of correlation (Oestreicher-Singer and Sundararajan 2012b). Although such data are not available to all stakeholders in a product network (e.g., publishers of books listed on Amazon.com), product network organizers such as Amazon.com can collect such data over time and use clickstream analysis to develop a less biased RCR for each pair of connected items.

Unfortunately, we did not have access to the type of data needed for such large-scale analysis. Although we cannot rule out all the possible endogeneity effects, we did take several steps to address the issue (for more discussion, see the section "A Simulated RCR") so that any bias would have a minimal effect on the substantive results we report next. First, we repeated our analysis using simulated RCR values (for each dyad of books) that were exogenously and randomly assigned on the basis of a normal distribution. That is, we assigned RCR values that were not the result of the true sales patterns. In addition, we repeated the analysis using the empirically determined RCR values, but with a random reordering; that is, we randomly assigned each value to a dyad in the network. In this way, we avoided measurement bias stemming from preexisting correlation among books. Second, we carried out additional analyses in which we varied the mean of the distribution of the simulated RCR, providing an indication of how the fundamental results are affected by the mean level of influence in the network. As we discuss subsequently, the substantive findings remain the same under different RCR levels. Finally, we note that in the following analysis, our focus is on the differences among sales tiers and not on measurement per se for specific items. Thus, even if the RCRs that are attributed to the existence of the product network should be lower than the values used here, our essential results will still be relevant as long as the overall value distribution pattern characterizing the different tiers remains the same.

### **Basic Results for Amazon.com**

We ran the iterative network value algorithm on the Amazon.com data, generating for each book measures of intrinsic value, incoming value, and outgoing value. Consequently, we were able to compute the network value for each item.

Columns 2–5 of Table 3 show summary statistics of our network value, incoming value, intrinsic value, and outgoing value binned according to the revenue of the corresponding products. The binning of 20% follows the conversion when discussing demand distribution, specifically the tail of the distribution. Figure W2 in the section “Distribution of Values” in the Web Appendix ([www.marketingpower.com/jm\\_webappendix](http://www.marketingpower.com/jm_webappendix)) illustrates the distribution of values for the intrinsic value, incoming value, and outgoing value.

Several observations emerge from Table 3. The books’ incoming values are considerably lower than their intrinsic values. That is, most value originates from the book itself and not from recommendations. However, the ratio between the incoming value and revenue varies across books in different revenue tiers. In the lowest-selling tier—the bottom 20% in terms of revenue (“low sellers”)—books’ average incoming value is approximately 13.37% of their revenue (see column 8 of Table 3). For the top-selling tier—the top 20% in revenue (“best sellers”)—this ratio is close to 22.74%. That is, the incoming value is relatively higher for best sellers.

In terms of the outgoing value, the picture is somewhat different. Like incoming value, outgoing value is considerably lower than intrinsic value, and yet in this case, we observe a different trend. Looking at the ratio of the average outgoing value to revenue across tiers, we find that for low sellers, the proportion of the outgoing value (48.79%) is higher than for other tiers. This proportion monotonically decreases as the revenue tier increases, culminating in a value of 15.67% for the best sellers (see column 9 of Table 3).

To better observe the results, we define a measure that combines the different types of value for a given book, representing the difference between the value the item contributes to the network and the value it receives. We label this measure “net influence”:

$$(7) \quad \text{Net Influence} = \text{Outgoing Value} - \text{Incoming Value} = \text{Network Value} - \text{Revenue}.$$

It is evident that net influence is also the difference between the network value of an item and its revenue. Because this amount may depend on the sales level of the book, we can also consider the relative net influence of a book: net influ-

ence divided by revenue, which represents the portion of change between network value and revenue. Columns 6 and 7 of Table 3 show summary statistics of the net influence and the relative net influence, binned by revenue tier.

Table 3 shows a clear and monotonic pattern: higher-selling items contribute more to the network than lower-selling items do, and yet they benefit even more from it. This is apparent for the net influence, and the difference among tiers is even more vivid when we consider the relative net influence. Among low sellers (bottom 20%), relative net influence is 35.38%. Relative net influence monotonically decreases as revenue tier increases; for the best sellers, the relative net influence is -7.07%. That is, if we assess a book’s value according to its revenue only, the actual value that the seller derives from books in the tail may be underestimated, whereas the actual value derived from books in the head may be overestimated.

Similar findings are evident in Figure W3 in the section “Distribution of Values” in the Web Appendix ([www.marketingpower.com/jm\\_webappendix](http://www.marketingpower.com/jm_webappendix)), which provides a graphic comparison of total network value versus revenue and illustrates the impact of considering the network when assessing the value of products. Using the demand conversions over our sample, we find that the value of the head of the distribution (top 20%) is overestimated by \$1,166,683 a week, of which \$292,645 is attributed to the tail of the distribution (lowest 20%). This does not mean that the books in the tail generate more absolute value than the books in the head. However, the value that low sellers generate is greater than the direct revenue they bring in, and therefore, the value of these books is underestimated.

### ***The Sales Tier Effect: Why Are Best Sellers Overestimated?***

*The number of incoming links.* Consider Table 4, column 1, which presents information about the average indegree for books from different revenue tiers. Recall that the out-degree (the books that the focal book recommends) is effectively limited in size, owing to the way that Amazon.com presents co-purchased products. The indegree, in contrast, varies substantially and is unlimited; yet we observe that that low-selling books receive half the number of recommendations that best-selling books receive. This may not be

**TABLE 3**  
**Estimation Results for Amazon.com: Average Values per Revenue Tier**

	1	2	3	4	5	6	7	8	9
Revenue Percentile	Revenue (\$)	Network Value (\$) (4 + 5)	Incoming Value (\$)	Intrinsic Value (\$)	Outgoing Value (\$)	Net Influence (\$) (5 – 3)	Relative Net Influence (5 – 3)/1	Incoming Value/ Revenue 3/1	Outgoing Value/ Revenue 5/1
0%–20% (low sellers)	6.21	8.40	.83	5.38	3.03	2.2	35.38%	13.37%	48.79%
20%–40%	11.75	13.91	2.11	9.65	4.26	2.15	18.33%	17.96%	36.26%
40%–60%	18.15	20.00	3.63	14.51	5.49	1.86	10.21%	20.00%	30.25%
60%–80%	30.43	32.23	6.34	24.09	8.14	1.8	5.90%	20.83%	26.75%
80%–100% (high sellers)	113.08	105.08	25.72	87.36	17.72	–8	–7.07%	22.74%	15.67%



**TABLE 4**  
**Revenue Tiers and Statistics About Incoming Links**

Revenue Percentile	1 Average Indegree	2	3
		Average Units Sold by the Incoming Link's Books	Fraction of Incoming Links from Books of the Same Revenue Tier
0%–20%	2.48	.87	.58
20%–40%	3.24	1.03	.33
40%–60%	4.18	1.16	.28
60%–80%	5.38	1.33	.29
80%–100%	10.12	2.16	.44

surprising given that popular books are co-sold with more other books; yet this implies that low sellers may benefit less from the network (i.e., lower incoming value).

*The assortativity of demand.* The next issue involves the types of books to which the products in each tier are connected. Note from Table 4 (column 2) that among books pointing to low sellers, each book sold .87 units on average, whereas among books pointing to best sellers, each book sold 2.16 units on average. Similarly, from column 3, we observe that books in a given revenue tier receive a large percentage of their recommendations from books in that same tier. That is, 58% of recommendations for low-selling books come from low sellers, and 44% of the recommendations for best sellers come from best sellers. Thus, the product network is characterized by a high degree of network assortativity, a phenomenon frequently observed in social networks (Newman 2002), in which nodes in a network tend to be connected to nodes with similar attributes. In our context this means that high-selling books get on average more traffic from each book to which they are connected and thus benefit from a higher incoming value per link.

*The average RCR.* One other source of difference can stem from a differential RCR among books. To understand the patterns in our data, consider Table 5, which presents the average RCR from each tier to each of the others. Note that on average, the conversion rate of recommendations from low sellers to best sellers is 3.37%, whereas the RCR from best sellers to low sellers is 1.48%. Observing the average effects for each tier, two issues arise, which we discuss in the following subsections.

*Incoming RCR.* First, we find that the average RCR of links pointing to best sellers is higher than the RCRs of

links pointing to lower sellers (e.g., 3.01% on average for best sellers vs. 1.97% for low sellers). This result is consistent with the findings of Oestreicher-Singer and Sundararajan (2012b), who show a similar pattern even after controlling for other sources of demand correlation between best sellers and other books. In our case, the implication is that, on average, a link pointing to a best seller contributes more incoming value to the recommended book compared with a link pointing to a lower seller.

*Outgoing RCR.* The second observation is that the outgoing RCR of low sellers is higher than that of best sellers. This increases the outgoing value created by low sellers. That is, people who visit the web page of a best seller are less likely to continue searching and to buy other items. Note, however, that even given this issue, the outgoing value of best sellers is considerably higher than that of low sellers (see Table 3) because of the higher level of traffic to best sellers' pages. Therefore, although this phenomenon moderates the difference between low sellers and best sellers, it does not refute the intuition that links from best sellers will lead to more sales.

Overall, we observe that best sellers are recommended more frequently, the recommendations they receive are from higher-selling books, and the conversion rates of their incoming links are higher. This results in a relatively high incoming value for high-selling products. Although such products also generate more value (i.e., higher outgoing value), this value is not enough to “compensate” for the higher incoming value, and their overall network value is lower than their revenue.

## Robustness Checks

### A Simulated RCR

Because the RCRs we use are based on Amazon.com's co-purchase data, and given the issue of endogeneity discussed previously, a question that arises is what our results would look like if the RCR were independent of the network structure. One way to examine this is to randomly draw the RCR for each dyad from a predefined distribution of RCR values. Selecting RCR values in this way would help avoid any measurement bias that stems from correlation in items' demand levels. (Note that the sales rank is not being manipulated in this exercise.)

Thus, we reran the analysis with the same items and same network structure, but instead of using the empirical proxy

**TABLE 5**  
**Average RCR Among Revenue Tiers**

From Tier	To Tier					Average Outgoing RCR
	0%–20%	20%–40%	40%–60%	60%–80%	80%–100%	
0%–20% (low sellers)	2.43%	3.52%	3.85%	3.87%	3.37%	3.41%
20%–40%	2.16%	2.99%	3.44%	3.60%	3.44%	3.13%
40%–60%	1.97%	2.49%	2.91%	3.27%	3.51%	2.83%
60%–80%	1.83%	2.17%	2.51%	2.94%	3.49%	2.59%
80%–100% (best sellers)	1.48%	1.76%	1.96%	2.23%	3.01%	2.09%
Average incoming RCR	1.97%	1.76%	1.96%	2.23%	3.01%	

for RCR, we drew RCR values from a normal distribution. We repeated this analysis using different mean values for the RCR distribution. Table 6 presents the results for the main parameter of interest: the relative net influence (for the full analysis, equivalent to the one in Table 3, see the section “Results for Different Mean RCR Values” in the Web Appendix at [www.marketingpower.com/jm\\_webappendix](http://www.marketingpower.com/jm_webappendix)).

In addition, we reran the analysis with the same items and same network structure, and using the same RCR values, we obtained using the UPD data, but we reassigned each value to a randomly selected product dyad. That is, we used the exact same RCR distribution but in a random ordering (for the full analysis, equivalent to the one in Table 3, see the section “Results for Randomly Reordered RCR Values” of the Web Appendix at [www.marketingpower.com/jm\\_webappendix](http://www.marketingpower.com/jm_webappendix)).

Note that the substantive picture that emerges regarding the average network value of the revenue tiers is generally similar to that with the empirically based RCR and is consistent across all levels of the simulated average RCR. Low sellers (bottom 20% of books) are still undervalued. As the sales tier increases, the positive difference between network value and revenue becomes smaller, and for best sellers (the top 20%), the difference becomes negative. As in the case of our original analysis (Table 3), best sellers’ network value is lower on average than their revenue. It should be noted, however, that in these robustness analyses, the difference between tiers is not as strong as in the case of the empirically based RCR, in which best sellers are more affected by the network.

These results shed light on the sources of the difference between the revenue tiers. Although the high incoming RCR that we observed for best sellers may have contributed to the low relative net influence of this tier (as compared with lower-selling tiers), this is clearly not the main driver of the effects we observed. Rather, the differences among sales tiers are driven more by differences in product network connectivity. The finding that the RCR distribution is not the main source of the observed effects is especially important given the potential bias in the empirical measurement of RCR.

When we examined our results for analyses using different average RCR values (for the full tables, see “Results for Randomly Reordered RCR Values” in the Web Appendix at [www.marketingpower.com/jm\\_webappendix](http://www.marketingpower.com/jm_webappendix)), the following picture emerged: the higher the average RCR, (1) the larger the incoming value, (2) the larger the outgoing value, and (3) the lower the intrinsic value. We observed

this pattern for all revenue tiers. Recall that a higher RCR value suggests that people are more likely to click on a link on a product page and buy, which reflects more traffic through the network. We find that as consumers use the network more, items increasingly affect others and are affected more. Thus, the role of the intrinsic value of the item decreases. Note, however, that even with an RCR value of 5%, intrinsic value is far greater than incoming or outgoing value.

### ***Applying the Model to Barnesandnoble.com Data***

Although the prominent status of Amazon.com has made the website a source of analysis for numerous academic explorations of electronic commerce, it is unclear to what extent the results we report here will hold in other environments. To examine this point, we replicated the analysis using a second data set of 257,000 books collected from the e-commerce website of Barnesandnoble.com (for a discussion of the collection method and descriptive statistics, see the section “Data and Data Collection” in the Web Appendix at [www.marketingpower.com/jm\\_webappendix](http://www.marketingpower.com/jm_webappendix)). We should note, however, that the data we were able to retrieve for Barnesandnoble.com were far more limited than the data for Amazon.com. Specifically, Barnesandnoble.com did not report data similar to the UPD data at Amazon.com, and we were therefore unable to assess impressions and conversion rates. Thus, in what follows, we use demand (units sold) as a proxy for impressions. We add a similar analysis for Amazon.com so we can compare the two product networks. For each dyad, we chose the RCR from a normal distribution  $N(2\%, .4\%)$ ; we selected this distribution on the basis of the range of RCR values we observed for Amazon.com. Furthermore, to compare the two networks, we focused on a subnetwork of 183,544 books that were present in both samples. Given that this is a subsample and not a complete network, the average outdegree is lower than that of the complete Amazon.com network.

Table 7 presents the results of the analysis for the Barnesandnoble.com data and for the restricted Amazon.com data set; note that they are directionally similar. As we expected, the intrinsic values are similar between the two platforms. The correlation between the computed “intrinsic values” across the two websites is .65. As we mentioned previously, differences in intrinsic value for the same product on different platforms can be due to differences in pricing, promotion, review ratings, and general presentation. Indeed, when we repeated the analysis for products with the same sale price on both networks, we found a correlation of .75 on the intrinsic value of the same items. In percentage

**TABLE 6**  
**Results of Relative Net Influence Estimations Using Different Mean RCR Values (Amazon.com)**

Revenue Percentile	Relative Net Influence				
	RCR = 1%	RCR = 2%	RCR = 3%	RCR = 4%	RCR = 5%
0%–20%	4.96%	9.93%	14.98%	20.08%	25.23%
20%–40%	2.69%	5.36%	8.07%	10.81%	13.54%
40%–60%	1.41%	2.84%	4.26%	5.70%	7.14%
60%–80%	.77%	1.54%	2.32%	3.09%	3.88%
80%–100%	–1.00%	–2.01%	–3.02%	–4.04%	–5.07%

TABLE 7

Network Value Estimation Results for Amazon.com and Barnesandnoble.com Using an Average RCR of 2%

Revenue Percentile	1 Average Revenue (\$)	2 Average Network Value (\$)	3 Average Incoming Value (\$)	4 Average Intrinsic Value (\$)	5 Average Outgoing Value (\$)	6 Average Net Influence
<b>Amazon.com</b>						
0%–20%	5.83	6.19	.33	5.50	.70	6.29%
20%–40%	11.20	11.56	.75	10.45	1.12	3.26%
40%–60%	17.28	17.56	1.31	15.97	1.60	1.66%
60%–80%	28.78	29.02	2.29	26.49	2.53	.82%
80%–100%	105.18	103.64	9.90	95.29	8.36	–1.47%
<b>Barnesandnoble.com</b>						
0%–20%	10.20	10.81	.64	9.56	1.24	5.95%
20%–40%	18.75	19.50	1.17	17.58	1.91	3.97%
40%–60%	28.07	28.86	1.88	26.19	2.68	2.82%
60%–80%	44.64	45.37	3.29	41.35	4.02	1.64%
80%–100%	150.85	148.51	13.71	137.14	11.37	–1.55%

terms, for approximately 80% of books whose prices were the same on both platforms, the books' intrinsic values on the two sites differed by less than 10%; that is, the difference was minor.

In addition, for both product networks, the overall patterns of value distribution are consistent with those we identified previously. The ratio between incoming value and intrinsic value increases as the revenue tier increases, and so the net influence decreases as the revenue tier increases. Furthermore, we observe a discrepancy in network value between best sellers and low sellers similar to that observed for the complete Amazon.com network: low sellers provide the network with the highest ratio of network value to revenue, whereas the best sellers receive more value from the network than they contribute through recommendations.

## Discussion

The emergence of online recommendation systems makes product networks a reality for firms and highlights the question of how managers can take advantage of product networks to enhance profitability. Here, we focus on how product valuation can take the product network into account, present an approach that enables marketers to adopt a network value view of products, and demonstrate its applicability for understanding the relationship between sources of value and revenue tiers in large-scale databases. In what follows, we focus on the implications and applications of the product network value approach.

### Basic Measurement

The key to informed use of a product network is proper measurement of the effects, and especially the RCRs among dyads. There are two kinds of stakeholders we can consider: external users, such as manufacturers and advertising agencies, and internal users, that is, the e-commerce retailers (Amazon.com in our example). The external user (Amazon.com would also be considered an external user if it wants to analyze the data at Barnesandnoble.com) has restricted access to the product network data. Yet it can still assess

network value by using use proxies, based on available online data as we do here, or purchased clickstream data.

With regard to their own networks, retailers such as Amazon.com have access to much richer data compared with external users. For example, the retailer can use actual clickstream data to track how the links between items are activated, and thus, it can more precisely measure the recommendation conversion rate and better deal with endogeneity issues of the types described previously. Although it may not be practical to carry out continuous experiments for millions of items, an analysis of samples can help determine the answer. Approaches such the ones presented in Oestreicher-Singer and Sundararajan (2012b) can help e-commerce retailers develop better capabilities in this regard.

This information can serve the firm in carrying out a dynamic analysis of the product network. We recommend a temporal, dynamic analysis of product networks for two reasons. First, product networks change with time: new products enter, the demand for certain items may saturate as a function of natural diffusion processes, and different trends affect demand. These processes may result in changes to the network value of the products in the network. The magnitude and range of such changes warrant exploration. A second issue is that dynamic changes present an opportunity to explore the cross-selling effects in product networks, for example, to get more accurate estimates for the RCR. This can happen when product networks are first formed (Stephen and Toubia 2010) but also in the natural process in which links are formed and disappear over time. However, such analysis should be done carefully to control for exogenous events that happen during the intervals between observations (e.g., pricing or advertising changes) that can affect the intrinsic value of items and consequently the product network calculations.

A dynamic analysis can also help managers determine how to build optimal product networks, an issue of interest to online retailers. This requires further investigation beyond the scope of this study, which focuses on reallocating existing revenues and not on optimal ways to increase overall revenue. Recall that retailers such as Amazon.com maintain that



they do not try to manipulate the recommendation system in their favor, and thus, any optimization effort should take into account the integrity of the recommendation system.

### **Managerial Use of the Product Network Knowledge**

In recent years, the use of recommendation systems has grown along with efforts to make these systems more effective. Yet much of the effort has focused on effective ways to build the recommendation systems, with little exploration of the profit implications (Hennig-Thurau et al. 2010). Managers can use the value-related measures presented here to better understand marketing in the presence of product networks in several ways, which we outline in the following subsections.

*Which products to keep.* Companies increasingly aim to optimize their product and brand portfolios by considering the value that each product generates and eliminating items that do not provide enough value. Given the possible discrepancy we demonstrate between revenue and network value, firms should look beyond revenue when making such decisions.

Note that this direction is similar to the transition in the customer management literature from viewing a customer's value (and consequently the customer portfolio) as based solely on his or her purchases to a broader view that also takes into account the customer's effect on others through word of mouth (Kumar et al. 2010; Libai et al. 2010). The measure of the value to the firm created as a result of customers' connectivity in the social network has been labeled customer referral value (Kumar, Petersen, and Leone 2010) or customer social value (Libai, Muller, and Peres 2012) and should become of essential managerial importance as customers become more connected through tools such as social media and as marketers' ability to follow this connectivity increases. Likewise, the ubiquity of product networks can make the network value of a product an important measure to help managers make informed product management decisions.

In a broader sense, it is worthwhile to tie these results to literature on the long tail of demand in electronic commerce. This stream of literature is based on the idea that electronic commerce is composed of a relatively large proportion of sales of low-selling and even very-low-selling items that, together, provide an overall high value to sellers (Anderson 2008). Previous literature has suggested that supply-side factors, such as broader product variety, and demand-side factors, such as reduced search costs, contribute to the emergence of the long tail (Brynjolfsson, Hu, and Smith 2003; Hinz, Eckert, and Skiera 2011). Yet different studies have shown that easier search and observational learning effects can also increase the power of "superstars" in overall sales and also create a "steep tail" (Elberse and Oberholzer-Gee 2007). Researchers have argued that recommendation systems can increase the demand for long-tail products by making items that consumers might otherwise not have been aware of visible to them (Anderson 2008), and yet these systems may also reinforce the popularity of already popular products (Fleder and Hosanagar 2009; Oestreicher-Singer and Sundararajan 2012a). The network value of products had not been taken into account in this stream to date. Our empirical analysis with the Amazon.com data suggests a need for further research using click-

stream data to compare the RCRs of best sellers with those of low sellers. Such research would facilitate further assessment of the real value of the long tail.

*Marketing mix.* Marketing-mix decisions should take into account how each product affects the other items connected to it. For example, additional intrinsic value created for an item through a promotion or low pricing may affect the outgoing value for other items that point to the focal product. Such issues are notable for advertising strategies. Many online environments use the pay-per-click (e.g., on search engines such as Google and Bing) or pay-per-purchase (e.g., in affiliate networks) advertising models. Often, keywords related to best sellers have a higher bidding price than keywords related to low sellers, which makes the cost per click for the best sellers' keywords higher than the cost per click for the low sellers' keywords. To take advantage of this difference, the advertiser should optimize his or her advertising spending by investigating the network value of each of product instead of just its revenue. Similarly, in affiliate marketing, a commission is paid for each sale generated. In this case, to increase the number of sales and maximize the total network value, a firm could offer a higher commission percentage on the lower-selling products.

*Managing vertical relationships.* A noteworthy implication of the product network value approach involves the relationship between the e-commerce retailer and the manufacturers (e.g., the publishers of books sold on Amazon.com). A product made by one manufacturer can generate recommendations for and receive recommendations from other products, which creates a divergence between the value of the product to the retailer and the value to the producer. This discrepancy can affect the composition of the optimal product assortment and pricing, and in more general terms, it can affect the balance of power in channels. Recent research in this area has focused on the relationship between manufacturers and retailers and considered the optimal assortment the retailer should carry, its price, and its quality (Bloom, Gundlach, and Cannon 2000; Dragan-ska, Mazzeo, and Seim 2009), and yet this research has not taken product network issues into account.

### **Influencers in Product Networks**

It is worthwhile to compare the results we obtain regarding the flow of value in product networks with those obtained in research on social networks. Numerous studies have been devoted to the role of influencers and, in particular, hubs in social networks (Goldenberg et al. 2009; Iyengar, Van den Bulte, and Valente 2011; Katona, Zubcsek, and Sarvary 2011).

Who (or rather, which products) may be labeled "influencers" in a product network? In most product networks, the number of outgoing links is limited, so a straightforward comparison to social network hubs is not trivial. More essential to the effect on others is the amount of traffic an item can send through its existing links, which we capture here in the number of impressions on the item's page.

To connect influence to revenue tiers, we define influencers as the top 10% in terms of outgoing value. The overall picture that emerges for the Amazon.com data (see also Table W11 in the section "Distribution of Influence" of the Web

Appendix at [www.marketingpower.com/jm\\_webappendix](http://www.marketingpower.com/jm_webappendix)) is that the distribution of influence in the Amazon.com product network is spread across revenue tiers. Although the majority (56%) of influencers are among the best sellers (top 20% of revenue), the effect is still spread across other revenue tiers.

Another issue is that of net influence. Here, the comparison to social networks is limited, given that the research on the contribution of influencers in social networks has focused largely on the role of outgoing links, not incoming ones. If we define influencers as the group of products with the top 10% of net influence, we find that best sellers are again most likely to be influencers, but in this case, they make up a smaller portion of the group (42%) compared with influencers defined on the basis of outgoing value (see also Table W11 in the section “Distribution of Influence” of the Web Appendix at [www.marketingpower.com/jm\\_webappendix](http://www.marketingpower.com/jm_webappendix)). This result is particularly intriguing given that, on average, the net influence of this group is negative (see Table 3). Thus, it seems that there may be large differences among best sellers in terms of net influence, which are driven by incoming value.

### Limitations

We highlight several limitations of the current study in the preceding section and mention some additional ones here. For example, we focus on value created through sales of products, and yet researchers may also explicitly consider models that are based on advertising-based revenue. While sales are still the major source of revenue for online retailers such as Amazon.com, advertising is an increasingly important source of revenue for websites. Note that in an advertising-based model, outgoing value can be created by the mere act of directing traffic to another product’s page and does not necessarily depend on the conversion to sales. Thus, the actual “conversion rate” may be higher. However, the revenues per visitor may be lower. An extension of our work in this direction would be a worthwhile avenue for further research.

Another type of limitation is related to the data used in the empirical analysis. As we discussed previously, rich large-scale clickstream data that include link usage could help produce more accurate assessments of the RCRs between dyads. Such data could also help analyze factors that affect the RCR—for example, category popularity and web page design. Another issue to explore further is the role of product complementarity. Choosing a specific camera may lead to the purchase of a lens, but it is not often that the purchase of a lens will lead to the purchase of a camera. Clickstream

data should be carefully analyzed in such cases to ensure that customers’ use of links actually creates a conversion effect.

Finally, although online recommendation systems are natural candidates for product network analysis, product networks exist in various forms in many consumption situations, including offline ones. Items in a supermarket, products in a catalog, and stores in a mall can also be examples of product networks. A noteworthy observation in this context is that the pattern of interaction between interrelated products on shelves is different from that generated by the interconnected hyperlinks in an online store. The online product network is a complex graph, which can be visualized using systems such as Directed Association Visualization (Hao et al. 2001). Therefore, its structure is very different from that of the three-dimensional brick-and-mortar store. For example, one important difference is that in the online product network links are not necessarily bidirectional. In addition, in the product network, the outdegree can be limited, but there is no limitation on the indegree. As a result, a product’s indegree may be different from its outdegree, and the overall outdegree and indegree distributions may differ. Furthermore, in physical stores, consumers do not have the option to randomly jump to another aisle.

Determining the associations among products in offline settings may be a larger challenge compared with the analysis of online recommendation systems. The market-basket analysis literature is a good reference for building association rules among products based on purchase data (Agrawal, Imieliński, and Swami 1993; Blattberg, Kim, and Neslin 2008) and may be used as a starting point toward building product networks based on product associations.

## Conclusion

The increasing amount of research studying Internet recommendation systems is evidence of the increasing role of recommendation systems in consumers’ online shopping environments. Assessing the value of a product is central to informed marketing, including well-planned advertising, brand portfolio planning, channel placement, cross-selling initiatives, pricing, and compensation of marketing personnel. Understanding the network value of products is thus of essential importance for marketers. As the share of online purchases increases, firms’ ability to measure and affect network value will grow. Therefore, this study should be a significant step toward a better understanding of this important concept.

---

## REFERENCES

- Agrawal, Rakesh, Tomasz Imieliński, and Arun Swami (1993), “Mining Association Rules Between Sets of Items in Large Databases,” *ACM SIGMOD Record*, 22 (2), 207–216.
- Anderson, Chris (2008), *The Long Tail: Why the Future of Business Is Selling Less of More*. New York: Hyperion Books.
- Binken, Jeroen L.G. and Stefan Stremersch (2009), “The Effect of Superstar Software on Hardware Sales in System Markets,” *Journal of Marketing*, 73 (March), 88–104.
- Blattberg, Robert C., Byung-Do Kim, and Scott A. Neslin (2008), *Database Marketing: Analyzing and Managing Customers*. New York: Springer.
- Bloom, Paul N., Gregory T. Gundlach, and Joseph P. Cannon (2000), “Slotting Allowances and Fees: Schools of Thought and the Views of Practicing Managers,” *Journal of Marketing*, 64 (April), 92–108.
- Bodapati, Anand V. (2008), “Recommendation Systems with Purchase Data,” *Journal of Marketing Research*, 45 (February), 77–93.
- Brin, Sergey and Lawrence Page (1998), “The Anatomy of a Large-Scale Hypertextual Web Search Engine,” *Computer Networks and ISDN*, 30 (1–7), 107–117.

- Brynjolfsson, Erik, Yu Jeffrey Hu, and Michael D. Smith (2003), "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science*, 49 (11), 1580–96.
- , ———, and ——— (2009), "A Longer Tail? Estimating the Shape of Amazon's Sales Distribution Curve in 2008," Workshop on Information Systems and Economics (WISE), Phoenix, AZ, (December 14–15).
- Carmi, Eyal, Gal Oestreicher-Singer, and Arun Sundararajan (2009), "Spreading the Oprah Effect: The Diffusion of Demand Shocks in a Recommendation Network," in *Proceedings of the Twenty-Ninth Annual International Conference on Information Systems (ICIS)*. Atlanta: Association for Information Systems.
- Chevalier, Judith and Austan Goolsbee (2003), "Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com," *Quantitative Marketing and Economics*, 1 (2), 203–222.
- Christakis, Nicholas A. and James H. Fowler (2009), *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. New York: Little, Brown and Company.
- Clifton, Rita, John Simmons, and Sameena Ahmad (2009), *Brands and Branding*. Princeton, NJ: Bloomberg Press.
- De, Prabhuddha, Yu (Jeffrey) Hu, and Mohammad Saifur Rahman (2010), "Technology Usage and Online Sales: An Empirical Study," *Management Science*, 56 (11), 1930–45.
- De los Santos, Babur (2008), "Consumer Search on the Internet," NET Institute Working Paper No. 08-15, (September), (accessed February 18, 2013), [available at [http://www.netinst.org/DeLosSantos\\_08-15.pdf](http://www.netinst.org/DeLosSantos_08-15.pdf)].
- Deschatres, Fabrice and Didier Sornette (2005), "Dynamics of Book Sales: Endogenous Versus Exogenous Shocks in Complex Networks," *Physical Review E*, 72 (1), 016112.
- Domingos, Pedro and Mark Richardson (2001), "Mining the Network Value of Customers," in *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 57–66.
- Draganska, Michaela, Michael Mazzeo, and Katja Seim (2009), "Beyond Plain Vanilla: Modeling Joint Product Assortment and Pricing Decisions," *Quantitative Marketing & Economics*, 7 (2), 105–146.
- Elberse, Anita and Felix Oberholzer-Gee (2007), "Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales," *MSI Reports*, 45 (4), 49–72.
- Fleder, Daniel and Kartik Hosanagar (2009), "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity," *Management Science*, 55 (5), 697–712.
- Gabaix, Xavier and Rustam Ibragimov (2009), "Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," NBER Technical Working Paper No. 342.
- Goldenberg, Jacob, Sangman Han, Donald R. Lehmann, and Jae Weon Hong (2009), "The Role of Hubs in the Adoption Process," *Journal of Marketing*, 73 (March), 1–13.
- Hao, Ming C., Umeshwar Dayal, Meichun Hsu, Thomas Sprenger, and Markus H. Gross (2001), "Visualization of Directed Associations in Ecommerce Transaction Data," in *Proceedings of the Joint Eurographics—IEEE TCVG Symposium*, Vol. 1. Aire-la-Ville, Switzerland: Eurographics Association, 185–92.
- Hennig-Thurau, Thorsten, Edward C. Malthouse, Christian Friege, Sonja Gensler, Lara Lobschat, Arvind Rangaswamy, et al. (2010), "The Impact of New Media on Customer Relationships," *Journal of Service Research*, 13 (3), 311–30.
- Hess, James D. and Eitan Gerstner (1987), "Loss Leader Pricing and Rain Check Policy," *Marketing Science*, 6 (4), 358–74.
- Hinz, Oliver, Jochen Eckert, and Bernd Skiera (2011), "Drivers of the Long Tail Phenomenon: An Empirical Analysis," *Journal of Management Information Systems*, 27 (4), 43–70.
- , Bernd Skiera, Christian Barrot, and Jan U. Becker (2011), "Seeding Strategies for Viral Marketing: An Empirical Comparison," *Journal of Marketing*, 75 (November), 55–71.
- Hogan, John E., Katherine N. Lemon, and Barak Libai (2003), "What Is the True Value of a Lost Customer?" *Journal of Service Research*, 5 (3), 196–203.
- Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente (2011), "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30 (2), 195–212.
- Katona, Zsolt and Miklos Sarvary (2008), "Network Formation and the Structure of the Commercial World Wide Web," *Marketing Science*, 27 (5), 764–78.
- , Peter Pal Zubcsek, and Miklos Sarvary (2011), "Network Effects and Personal Influences: The Diffusion of an Online Social Network," *Journal of Marketing Research*, 48 (June), 425–43.
- Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg (2010), "Online Demand Under Limited Consumer Search," *Marketing Science*, 29 (6), 1001–1023.
- , ———, and ——— (2011), "Mapping Online Consumer Search," *Journal of Marketing Research*, 48 (February), 13–27.
- Kumar, V., Lerzan Aksoy, Bas Donkers, Rajkumar Venkatesan, Thorsten Wiesel, and Sebastian Tillmanns (2010), "Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value," *Journal of Service Research*, 13 (3), 297–310.
- , Andrew Petersen, and Robert P. Leone (2010), "Driving Profitability by Encouraging Customer Referrals: Who, When, and How," *Journal of Marketing*, 74 (September), 1–17.
- Libai, Barak, Ruth Bolton, Marnix S. Büegel, Ko de Ruyter, Oliver Götz, Hans Risselada, et al. (2010), "Customer-to-Customer Interactions: Broadening the Scope of Word of Mouth Research," *Journal of Service Research*, 13 (3), 267–82.
- , Eitan Muller, and Renana Peres (2009), "The Role of Within-Brand and Cross-Brand Communications in Competitive Growth," *Journal of Marketing*, 73 (May), 19–34.
- , ———, and ——— (2012), "Decomposing the Value of Word of Mouth Seeding Programs: Acceleration vs. Expansion," *Journal of Marketing Research*, (published electronically December 6), [DOI: 10.1509/jmr.11.0305].
- Manski, Charles F. (2000), "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, 14 (3), 115–36.
- Newman, M.E.J. (2002), "Assortative Mixing in Networks," *Physical Review Letters*, 89 (20), 208701.
- Niraj, Rakesh, V. Padmanabhan, and P.B. Seetharaman (2008), "A Cross-Category Model of Households' Incidence and Quantity Decisions," *Marketing Science*, 27 (2), 225–35.
- Oestreicher-Singer, Gal and Arun Sundararajan (2012a), "Recommendation Networks and the Long Tail of Electronic Commerce," *MIS Quarterly*, 36 (1), 65–83.
- and ——— (2012b), "The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets," *Management Science*, 58 (11), 1963–81.
- Raeder, Troy and Nitesh Chawla (2011), "Market Basket Analysis with Networks," *Social Network Analysis and Mining*, 1 (2), 97–113.
- Seetharaman, P.B., Siddhartha Chib, Andrew Ainslie, Peter Boatwright, Tat Chan, Sachin Gupta, et al. (2005), "Models of Multi-Category Choice Behavior," *Marketing Letters*, 16 (3), 239–54.
- Stephen, Andrew T. and Oliver Toubia (2010), "Deriving Value from Social Commerce Networks," *Journal of Marketing Research*, 47 (April), 215–28.
- Van den Bulte, Christophe and Stefan Wuyts (2007), *Social Networks and Marketing*. Cambridge, MA: Marketing Science Institute.
- Zubcsek, Peter Pal and Miklos Sarvary (2011), "Advertising to a Social Network," *Quantitative Marketing and Economics*, 9 (1), 71–107.