NET Institute*

www.NETinst.org

Working Paper #11-29

October 2011

Assessing Value in Product Networks

Eyal Carmi	Gal Oestreicher	Barak Libai	Ohad Yassin
Tel Aviv Univ.	Tel Aviv Univ.	Arison School of Business	Tel Aviv Univ.

* The Networks, Electronic Commerce, and Telecommunications ("NET") Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, "virtual networks" comprised of computers that share the same technical standard or operating system, and on network issues in general.

ASSESSING VALUE IN PRODUCT NETWORKS

Eyal Carmi*, Gal Oestreicher-Singer*, Barak Libai**, Ohad Yassin*

ABSTRACT

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 eCommerce environments, where products are often presented as a collection of webpages linked by recommendation hyperlinks, creating a large-scale product network. Here we present the first attempt to use a systematic approach to estimate products' true value to a firm in such a product network. Our approach, which is in the spirit of the PageRank algorithm, uses easily available data from large-scale electronic 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.

We apply this approach to data collected from Amazon.com and from BarnesAndNoble.com. Focusing on one domain of interest, we find that if products are evaluated according to their direct revenue alone, without taking their network value into account, the true value of the "long tail" of electronic commerce may be underestimated, whereas that of bestsellers might be overestimated[1].

Keywords: product value, cross-selling, electronic commerce, recommendation networks, social networks, long tail

JEL Classification Numbers: D4

---

[1] The authors thank the NET Institute (www.NETinst.org) for financial support.

* The Leon Recanati Graduate School of Business Administration, Tel Aviv University, Israel
** The Arison School of Business, Interdisciplinary Center, Israel

INTRODUCTION

A fundamental insight of research on products and brands is that the demand for different products can be interrelated. It is widely recognized that purchases across categories are correlated among consumer goods that are complements or substitutes for one another (Seetharaman et al. 2005). This interconnection of product demand is a widespread phenomenon that has been observed in a variety of settings. For example, a “loss leader” drives purchases for other products (Hess and Gerstner 1987), the availability 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, Muller, and Peres 2009), and the demand for a sub-brand can affect the consumption of other members of the brand portfolio (Aaker 2004). Such inter-product dependencies are of much interest to marketers since they can affect issues such as optimal pricing decisions (Niraj, Padmanabhan, and Seetharaman 2008.), predicting the sales of new products (Sriram, Chintagunta, and Agarwal 2010), assessing cross-selling opportunities (Li, Sun, and Wilcox 2005), or understanding competitive dynamics (Wedel and Zhang 2004).

To date, the investigation of inter-product associations has generally focused on dyads or on small numbers of entities; however, a larger picture often exists. For example, while pricing of bacon has been shown to affect the demand for eggs and vice versa (Niraj, Padmanabhan, and Seetharaman 2008), the demand for these two categories may be correlated with demand for other categories, which are further related to additional categories. Similarly, if, as proposed, the adoption of Apple’s iPod affected the sales of the company’s Mac computers (Economist 2006), it may have affected other products as well, which in turn affected additional products.

One can think of the connections among products as forming a *product network*. To understand how products in a product network influence one another, it is helpful to consider the analogy of word of mouth in a social network. In a social network, a person's decision to adopt a certain behavior is often influenced by "recommendations" made by her social connections, and this person can in turn affect others in the network, in a contagion-like process (Van den Bulte and Wuyts 2007). Similarly, in a product network, the purchase of a given product is influenced by the purchase of neighboring products, and in turn can encourage the purchase of other connected products. Thus, the purchase of a product in the network can be seen as a "recommendation" for the purchase of a neighboring product. By applying network analysis tools to large-scale online social network databases, researchers have been able to better understand the dynamics of social influence in networks. The dynamics and implications of large-scale product networks are yet to be fully explored.

An important development towards that end is the emergence of hyperlinked online product networks, based on inter-product connections formed among products sold on the web. Such products are often presented as a collection of connected webpages (which are nodes in the product network), each offering one product. If one imagines the process of browsing an eCommerce site as being analogous to walking the aisles of a physical store, then the aisle structure of an online retailer consists not of a built physical system of shelves, but instead of the structure of interconnected hyperlinks. Such structures are often generated by advanced collaborative filtering algorithms (Herlocker et al. 2004). The placement of a product in this graph thus 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 approach is increasingly used by diverse sellers such as Zappos.com, Hotels.com and Walmart.com to cross-sell products to their customers, and thus enables them to more fully exploit the inherent relationships among products.

Here we present a novel approach for the analysis of large-scale online product networks, aiming to investigate the value of a product to the firm, taking into account the product's position in the hyperlinked network. Understanding the full value of a product (a category or a brand) can help managers to recognize which products to offer (or to stop offering) to customers, which products to promote, and how to better price different products. Traditionally, the value of a product or brand has been assessed 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, considering both the revenue that an item generates by directing traffic to other items, and the revenues that an item is not "entitled" to due to traffic directed to it by other items.

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 webpages. First, for each item, we differentiate between the *intrinsic value* portion of its revenue, which is self-generated by the item, and the *extrinsic value* portion of its revenue, which is driven by the recommendation links pointing from other items to the focal item. We can therefore attribute the extrinsic value of each product to the items that point to it. The product's *network value* is the sum of its intrinsic value and the value it generates for its neighbors, which we label *generated value*.

The approach we present here is applicable to large-scale databases and can be implemented in a relatively straightforward way, relying on easily observable data. This is of considerable importance given that a product's 'full value' in relation to that of other products is of interest not only to the retailer, who has access to all information regarding the purchase processes that take place on its site, but also to external parties whose access to such data may be limited. For example, most product networks managed by online retailers include multiple items made by different manufacturers. A manufacturer naturally has an interest in assessing not only the revenue of its products, but the full value of those products given their network connectivity. However, unlike the retailer, the manufacturer can obtain only limited information regarding the network, and must take the network structure as given. Our approach, which uses data that can be accessed by any observer, should thus be of interest not only to retailers but also to manufacturers. Of course, retailers who manage the product network may use their comprehensive access to data to further extend the approach presented here, to better calibrate the parameters used to assess product value, and possibly to develop ways to influence the network structure to achieve optimal connectivity of products.

We apply our approach to large datasets collected from Amazon.com and BarnesAndNoble.com to demonstrate how it can be used to study the differential drivers of network value in such environments. Similarly to PageRank, our model can be applied iteratively to account for the effect of the entire network on the value of specific products. Specifically, we examine how the difference between network value and revenue in these environments can lead to an underestimation of the value of the "long tail" of low-selling books.

This study contributes to current literature in a number of ways. First, it broadens the analysis of inter-product 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 owing to the complexity associated with larger-scale investigation (Blattberg, Kim, and Neslin 2008). It also has implications regarding the composition of the optimal assortment a retailer should carry and the consequent relationships with the manufacturers of the various products (Dukes, Geylani, and Srinivasan 2009). A product made by a certain manufacturer may receive extrinsic value and provide generated value to products of other manufacturers. This can create a discrepancy between the product's value to the manufacturer and its value to the retailer, and it may affect optimal pricing and balance of power. Lacking techniques with which to quantify product network value, researchers have not explored these issues to date.

Second, this study highlights the need to consider the different means by which a product generates value for a firm. In recent years there is a growing understanding that a customer's value to a firm stems not only from his purchases, but also from his social influence, e.g., through word of mouth and imitation (Hogan, Lemon, and Libai 2003; Kumar et al. 2010). Other research has begun to examine the value created by networks in cases of interconnected sellers (Stephen and Toubia 2010), interconnected blogs (Mayzlin and Yoganarasimhan 2008), or content networks (Dellarocas, Katona, and Rand 2011). In the same spirit we aim to explicitly distinguish between a product's independent value and its contribution to the value of other products. This enables us to suggest a new look at a firm's product portfolio, based on the types of value that each product contributes and receives.

Third, we are able to demonstrate the usefulness of the product network value approach by looking at the empirical issue of the "long tail" of demand distribution in electronic commerce (Anderson 2008; Brynjolfsson, Hu, and Smith 2006; Hinz, Eckert, and Skiera 2011).

The existence of the long tail phenomenon, its drivers, and its magnitude have been at the center of a recent ongoing discussion, yet the value that products contribute to and take from the network has not been considered in this regard. Specifically, there have been claims that recommendation systems can help drive the long tail phenomenon, but the effects that the long-tail products produce through the recommendation system have not been investigated fully.

In a product network, we expect bestsellers to receive a larger number of incoming links, which in our model will result in a higher extrinsic value for these products, as compared with less popular products. However, at the same time, high selling products are assumed to drive more traffic to neighboring products, resulting in higher generated value. These two opposing forces make the overall effect of the network value approach non-trivial. Here, we are able to study the sources of network value for different revenue tiers in outlets such as Amazon.com and BarnesAndNoble.com. We find that when a product's value is assessed on the basis of its revenue alone, the network value of long-tail products may be underestimated on average, while the network value of bestsellers may be overestimated. This issue should be taken into account by managers who decide on the placement of low-selling products in their portfolios.

The paper continues as follows: We first introduce our theoretical model for computation of the network value of products. We then apply our model to data from Amazon.com (and further, as a robustness check, to data from BarnesAndNoble.com), and specifically, we examine its implications for the estimated value of the long tail. We then discuss the implications of our results and directions for future work.

MODELING NETWORK VALUE OF A PRODUCT

The Setting

We consider a large-scale network of interlinked products. The *outdegree* of a product *u* represents the number of links that originate from product *u* and point to other products, while the *indegree* is the number of links that point to *u* from other products. To help us demonstrate our approach, we will 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 creates based on co-purchases of books.

The problem we analyze is of a firm that wants to understand the actual value contribution and the types of value generated by each product in the database. Note that our aim is not to analyze the optimal policy of the firm in shaping the network. This issue is highly complex and is beyond the scope of this paper. Rather, we accept the structure of the network and the overall sum of revenue of all items as given: What we examine is how to redistribute this sum. Our approach is therefore applicable not only to retailers but also to external parties, such as product manufacturers (e.g., publishers of books in Amazon), who are not able to affect the links in the product network and must accept the network as given.

We assume that the firm (manufacturer or retailer) can observe only the following information for each product: the product's indegree and outdegree, quantity sold, and price. This is indeed the only information available to outside observers of product networks. Retailers are likely to have access to additional data for each product, which they could exploit to reveal additional effects. Yet, even for a retailer such detailed analysis over an entire large scale network is likely to require a heavy investment of time and computational resources, which makes the approach presented here appealing.

We divide the revenue of a product into two parts, which we define as follows: (1) The *intrinsic value* portion of the revenue is self-generated by the item. One can think of it as the revenue that the product would be expected to yield if it were not connected to others. (2) A product's *extrinsic value* is driven by the recommendation links that point to that product from other products. Thus, for product *u:*

$$Revenue(u) = ExtrinsicValue(u) + Intrinsic\ Value\ (u)$$

Our focus here is on the actual contribution of any focal product to the firm, which we label *network value*. Network value of a product stems from two sources, one of which is its intrinsic value. The other is the contribution of this product to the extrinsic values of products it recommends, which we label the *generated value* of the focal product.

Network Value (u) = Intrinsic Value (u) + Generated Value (u)

This view is consistent with previous work aiming to assess the value of customers in a network by distinguishing between 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 re-allocate 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 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 webpages. 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 webpage in the hyperlinked structure of the web, based on the following model:

(1)

$$PageRank(u) = \sum_{v \in In(u)} \frac{PageRank(v)}{OutDegree(v)}$$

where $In(u)$ is the set of webpages (nodes) linking to node *u*, and *OutDegree (v)* is the number of outgoing links from node *v*. 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[1]. 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. The ranking of a page thus ends up being the stationary probability that a random surfer who starts at a random page will visit the specific page. Therefore, a page can gain a high ranking by either having many pages that point to it or having a few highly ranked pages point to it. While 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 hence its demand).

PageRank has been followed by other algorithms such as HITS (Kleinberg 1999), which gives a hub and authority score, TrustRank (Gyöngyi, Garcia-Molina, and Pedersen 2004), which focuses on separating useful pages from spam, and Hilltop (Bharat and Mihaila 2000), which focuses on the most authoritative pages. Some properties of our model are inspired by these ranking techniques as well.

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 <u>creates</u> for other products, not only on the traffic it receives. Furthermore, similarly to PageRank, we want to take into account the fact 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, a link from Dan Brown's bestseller *The Da Vinci Code* is likely to be a more fruitful recommendation compared with one from a lower-selling book.

Consider a product *u* that receives certain traffic to its webpage. We define <u>*u*'s</u> *InTraffic* as the traffic that flows into product *u*'s page from its neighbors. However, not every link exposure leads to a purchase. We therefore define α to be the Recommendation Conversion Rate (RCR), which represents the probability that a link exposure will result in a purchase. In the case of books on Amazon, this probability is a combination of the probability that a link will be clicked on, and the probability that the user's visit to the next page will result in a purchase. For simplicity, in the following formulation we use a fixed α value for all products, which can be based on the average α of the specific product network. A retailer who has better information may be able to assess a specific expected α_u for product *u*, or even a conversion rate α_{vu} for the specific recommendation link from product *v* to product u. We further discuss the estimation of α in the empirical section.

We can now define the intrinsic value of the product, that is, the sales of the product that are not attributed back to the network, as:

(2)

$$Intrinsic\ Value\ (u) = [Q(u) - \alpha * InTraffic\ (u)] * P(u)$$

where $P(u)$ is the price of product *u*. Note that the greater the volume of traffic directed to the product from neighboring products (i.e., the greater its $InTraffic$), the lower the fraction of revenue that should be attributed to the intrinsic value of the product. For example, the intrinsic value of a book on Amazon that is recommended by many bestsellers should be lower 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 *extrinsic value* (that is, the revenue portion that is generated by incoming links from other items in the network):

(3)

$$ExtrinsicValue(u) = Revenue(u) - IntrinsicValue(u)$$

$$= \alpha \cdot InTraffic(u) \cdot P(u)$$

We can now attribute the respective portions of the extrinsic value to the items that originated these revenues. We define product *v*'s share in the extrinsic value of item *u* as the following ratio:

(4)

$$M(v, u) = \frac{Traffic(v \to u)}{InTraffic(u)}$$

which is the amount of traffic that item *v* has generated for item *u*, relative to the overall traffic that item *u* has received from all its incoming links (and assuming a constant α). Note that by construction, $0 \leq M(v,u) \leq 1$ and $\sum_{v \in In(u)} M(v,u) = 1$. Thus, M can be viewed as a transition probability matrix.

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

(5)

$$GeneratedValue(v) = \sum_{u \in Out(v)} M(v,u) ExtrinsicValue(u)$$

$$= \sum_{u \in Out(v)} ExtrinsicValue(u) \frac{Traffic(v \rightarrow u)}{InTraffic(u)}$$

Adding the intrinsic value and the generated value of item *v*, we obtain an expression for *v*'s network value:

(6)

$$NetworkValue(v) = IntrinsicValue(v) + GeneratedValue(v)$$

$$= [Q(v) - \text{á} * InTraffic(v)] \cdot P(v)$$

$$+ \sum_{u \in Out(v)} ExtrinsicValue(u) \frac{Traffic(v \rightarrow u)}{InTraffic(u)}$$

An illustration of the application of the model using a simple example is presented in Web Appendix A.

Iterations and Convergence

When applied over a given graph once, our model pushes extrinsic revenues back to the originating items located one link away from the recommended products.

Take, for example, a product network where item A recommends item B, which recommends item C, which recommends item D. Consider item B. 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.

However, the picture may be more complicated. Some of the revenue from sales of item C that is attributed to item B should in fact be attributed backwards to item A, which generated part of item B's traffic to begin with. In fact, 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, since some of C's value comes from the extrinsic value driven by B. In other words, an item is entitled to a share of another item's network value, not just its revenue. Hence, to calculate the network value for the complete chain, we need to apply the model iteratively, so that for iteration n ≥ 1:

$$NetworkValue_n(v) = IntrinsicValue_n(v) + GeneratedValue_n(v)$$

When considering iterations one should take into account that as we get further from the focal product we can expect decay in the effect. Because the average consumer's shopping basket is typically limited, it is reasonable to expect that the probability of further browsing and purchasing decreases with the number of purchased items. Hence we use a decay rate β that accounts for the diminishing effectiveness of the links as the number of links in the path grows (we discuss the implementation of β in the empirical section of this paper).

For convenience of calculation, we can rewrite the intrinsic value in terms of revenue as follows:

(7)

$$IntrinsicValue(u) = \left(1 - \frac{\text{á} \cdot InTraffic(u)}{Q(u)}\right) Revenue(u)$$

Thus, considering product *v*:

(8)

$$GeneratedValue_n(v) = GeneratedValue_{n-1}(v) -$$

$$\text{â}\left(\frac{\text{á} * InTraffic(v)}{Q(v)}\right) \cdot GeneratedValue_{n-1}(v)$$

$$+ \ \text{â} \sum_{u \in Out(v)} M(v,u) \left(\frac{\text{á} * InTraffic(v)}{Q(v)}\right) GeneratedValue_{n-1}(v)$$

In each iteration of the model, the network value of a given product is reduced by the portion of its extrinsic value that is given back to the previous item, yet it is increased by the product’s contribution to the network value of the items it recommends.

To run the iteration process we need a functional form for the decay rate β. A critical criterion is the distance after which we can expect the effect to essentially disappear. Carmi and colleagues (2009) found that, on average, the effect of a shock that travels through recommendation links on Amazon is limited to a close neighborhood of up to three links around the source. This number is consistent with findings from the social network literature that show influence to be 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 and BarnesAndNoble.com contains fewer than three items (de los Santos 2008). Consistent with previous research, we use an exponential function to model the rate of decay at which influence diminishes across the network (Carmi, Oestreicher-Singer, and Sundararajan 2009; Deschatres and Sornette 2005). In the datasets we examined, by setting β at $1/10^{n-1}$, convergence is reached after four iterations[2]. In this case the implication is that after the third iteration, the effect is very close to zero.

APPLYING THE NETWORK VALUE ASSESSMENT APPROACH: THE ISSUE OF PROFITABILITY TIERS

The Long Tail of Demand

While there are multiple dimensions in which assessing the network value of products can bring value to marketers, here we focus on one application: the use of product network value to assess the real value of what is often known as the *long tail*.

The long tail of demand in electronic commerce is a well-documented business phenomenon of recent years. It suggests that electronic commerce is composed of a relatively large proportion of sales of low-selling and even very-low-selling items, many of which are not sold in traditional stores (Anderson 2008). Previous literature has suggested that supply-side factors, such as broader product variety (Brynjolfsson, Hu, and Smith 2003; Clemons, Gao, and Hitt 2006; Hinz, Eckert, and Skiera 2011), contribute to the emergence of the long tail. Others have focused on demand-side factors, such as reduced search costs (Brynjolfsson, Hu, and Smith 2006; Cachon, Terwiesch, and Xu 2008) and preference isolation (Choi, Hui, and Bell 2010). Additional work has shown that selling niche products can help Internet firms avoid competition from mainstream retailers (Brynjolfsson, Hu, and Rahman 2009), and that exposure to niche products can drive consumers to develop a taste for more niche products (Brynjolfsson, Hu, and Smith 2009). Yet, different studies have shown that easier search and observational learning effects can also increase the power of "superstars" in overall sales, and create in addition a "steep tail" (Elberse and Oberholzer-Gee 2007; Tucker and Zhang 2007). While there is evidence that the proportion of the long tail has changed in recent years (though there are conflicting findings as to the direction of the change; Brynjolfsson, Hu and Smith 2009; Tan and Netessine 2010), it is still considered a notable phenomenon of online commerce that should be exploited.

Researchers have associated the long tail phenomenon with the emergence of recommendation systems that enable consumers to learn from the choices of other buyers (Brynjolfsson, Hu, and Smith 2006); such systems form the basis of the type of product network analyzed here. However, the exact relationship between the two phenomena is not yet clear. On one hand, recommendation networks could increase the demand for niche products by making items that consumers might otherwise not have been aware of visible to them (Anderson 2008; Hervas-Drane 2009). On the other hand, recommendations based on sales and ratings may reinforce the popularity of already popular products; however, this depends on the specific mechanism of the recommendation system (Fleder and Hosanagar 2009). A recent suggestion is that the exact effect may depend on the product category and the nature of the relationship between products (Oestreicher-Singer and Sundararajan 2011).

The product network value analysis presented here enables us to shed new light on the contribution of long-tail products to an organization. Contrary to current literature that examines how recommendation systems may contribute to the demand for niche products and thus to the emergence of the long tail, we want to examine the network contribution of long-tail products compared to that of other products. Specifically, we investigate whether the ‘true’ value of products in the tail, when calculated using a full network value approach, differs from the value that is currently attributed to these products. To do so we will apply our model to a large-scale network for which the long tail has been extensively examined: that of Amazon.

The Amazon Co-purchase Network

We created a database of product data including pricing, demand, rating and co-purchase network information for over 900,000 books sold on Amazon.com on a particular day in 2010. While Sales Rank is not an exact measure of sales, 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[3] data generated by Amazon and following a log-linear conversion model suggested by Chevalier and Goolsbee (2003) and by Brynjolfsson, Hu and Smith (2009) with the correction shown by Gabaix and Ibragimov (2009).

Amazon's book recommendation system is probably the best known among electronic retailers and has been widely used to demonstrate the role of recommender systems in general (Brynjolfsson, Hu, and Smith 2003; Fleder and Hosanagar 2009; Linden, Smith, and York 2003). Each product on Amazon.com has an associated webpage containing a set of "co-purchase links", which are hyperlinks to products that were co-purchased most frequently with that product on Amazon.com. The co-purchase set for each webpage is limited to five items and is listed under the heading, "Customers who bought this item also bought …"[4]. The network was collected using a snowball sampling method, which started from a number of seed books and resulted in a large connected component.

Web Appendix A presents the key network parameters for the Amazon network. These parameters include the size of the network (number of nodes), average indegree, average outdegree, density, and average clustering coefficient, which represents the tendency to form clustered groups of connected items (CC1 in Newman 2003). The network has a single major component, that is, there are no isolated units, or isolated clusters, and the degree distribution follows a power law shape consistent with prior empirical work.

Adjustments Made for the Empirical Dataset

In order to apply the algorithm to the Amazon data we need to make several assumptions, given that we do not observe the full click-through data.

The level of Intraffic. *InTraffic* reflects the traffic that flows into product *u*'s page from its neighbors. As we do not have these data, we use the following proxy, in the spirit of PageRank:

(2)

$$InTraffic(u) = \sum_{v \in In(u)} \frac{Q(v)}{OutDegree(v)}$$

where *Q(v)* is the quantity of purchases of a given neighbor *v*, and serves as a proxy for the traffic to *v*'s page. The equation is based on the assumption that the amount of traffic to *u* that stems from *v* is a function of *Q(v)*, since the more traffic *v* receives (i.e., the more units *v* sells), the greater the chance of traffic being redirected from *v*'s page to *u*'s page. Hence, we focus here on the network value that stems from purchases, and not only page views. Another issue that follows from this formulation is that, similarly to PageRank, we divide the outgoing traffic equally among the target books.

The value of the recommendation conversion rate (RCR). A key parameter value needed to apply the model above is that of α – the RCR, or the probability that a recommendation link will convert into sales of a second product. Here we use a single average RCR for the empirical analysis, and examine the effect of this level on the network value of products.

Regarding the RCR, we note first that while the term "conversion rate" is often used to describe the percentage of overall page views or clicks that turn into sales (Moe and Fader 2004), in our model, we focus on purchases that originate from products that the consumer has already purchased. Thus, the RCR is related to a case in which an actual shopping decision has been made, which means that RCR is likely to be much higher compared to a case of just browsing the Internet. A seller such as Amazon can follow the links and obtain the true RCR, and in fact can even assess a separate RCR for each dyad of books. However, unlike the traditional conversion rate (percent of visits that turn into a purchase), which is regularly published for many eCommerce websites, RCR is not published by Amazon or other sites.

In appendix B we describe a number of ways that helped us assess the range of RCR values for the Amazon data. In appendix A we present the analysis of the Amazon network for a range of RCRs, from 1% to 20%, but we use a mid-range value of 10% to demonstrate the main results.

Endogeneity issues in the Amazon network. Before we present the results, it is necessary to acknowledge endogeneity issues, which present challenges to the study of social networks (Manski 2000), and are also present in the case of product networks. One notable source of potential endogeneity is the fact that the product networks for retailers such as Amazon are created via the recommendation systems. Two items are connected if they were frequently purchased together in the past, yet this link guides new consumers to make additional purchases. Hence, network position is a function of past sales, which biases the study of the network's influence on subsequent sales. Thus, it makes it more difficult to determine whether an Amazon customer who buys a product after clicking on a recommendation link would have bought that product anyway, even in the absence of a link—which would suggest that the real recommendation conversion rate α might be lower than the one determined on the basis of straightforward analysis of clickstream data. To partially account for this potential bias, we repeated our estimation for a range of values of α, and found that the results are directionally the same for any given value. One should acknowledge, however, that the real level of α may be lower than what is assessed using simple aggregate-level analysis.

Moreover, this bias may vary for different products and even for different dyads of books. For example, recommendations for more popular books or books written by well known authors may be associated with higher conversion rates. This bias is not easily controlled for, as is the measurement problem of endogeneity in social networks in general. One approach may be to use natural experiments such as external shock to the product network (Carmi, Oestreicher-Singer, and Sundararajan 2009) to examine how demand for a connected product changes. Retailers who have access to comprehensive click-stream data can also plan or design experiments in which links among products are varied in a controlled way to understand to what extent such bias exists in the simple measurement of the RCR and how this may be affected by item-specific characteristics. Doing so on a large scale is far from trivial, and is a clear source of interest for further research.

Still, in the analysis we present here, the issue of bias in the measurement of a product-specific α is less of a problem, since we do not aim to estimate the RCR for each dyad, but analyze the consequences of different values of α in general. We elaborate on the direction of such potential biases when discussing our findings.

Basic Results for Amazon

We ran the iterative network value algorithm on the Amazon data, generating, for each book, measures of intrinsic value, extrinsic value and generated value. Consequently, we could compute the network value for each item. Table 1 shows summary statistics of our network value estimations for an RCR of 10% on the Amazon dataset, binned according to revenue. Note that column 2 (Network value) is the sum of column 4 (Intrinsic value) and column 5 (Generated value).

Table 1 about here

A number of observations emerge from Table 1. First, books' extrinsic value is considerably lower than their intrinsic value—most value comes from the book itself and not from recommendations. However, the ratio between the two values varies across books in different revenue tiers. In the lowest selling tier—the bottom 20% in terms of revenue ("low sellers")[5]—books' average extrinsic value is about 6% of the their intrinsic value. For the top selling tier—the top 20% in revenue ("bestsellers")—this ratio is close to 11%. That is, the extrinsic value is relatively higher for bestsellers.

In terms of the generated value, i.e, the value books generate for other books, the picture is somewhat different. Like extrinsic value, generated value is considerably lower than intrinsic value, yet in this case we observe a different trend. Looking at the ratio of the average generated value to revenue across tiers, we find that for low sellers the proportion of the generated value (12%) is higher than for other tiers. This proportion monotonically decreases as the revenue tier increases, culminating in a value of 8.3% for the bestsellers.

These two competing effects influence the network value of different tiers in opposite directions. High selling products seem to benefit more from the network (and as a result a larger fraction of their revenue should be attributed back to the network), but they also contribute more to other products in the network (and as a result should be attributed some of those products' demand). Hence, the overall effect of the network value approach should be carefully analyzed. A direct comparison between network value and revenue (column 6 of Table 1) shows the overall variation across revenue tiers. Among low sellers, network value is 13.3% higher than revenue. This percentage monotonically decreases as revenue tier increases, and for the bestsellers, the network value is nearly 0.8% lower than revenue. That is, if one assesses a book's value according to its revenue only, then the value generated by books in the tail may be underestimated, while value generated by books in the head may be overestimated. Using the demand conversions over our sample, we find that the value of the head of the distribution (top 20%) is overestimated by $300,000 a week, of which $70,000 are attributed to the tail of the distribution (lowest 20%). Of course, 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 these books are underestimated by the conventional revenue valuation.

To further study the source of this finding, consider Table 2, which presents information about the books that create extrinsic value (indegree) for the different profitability tiers. Recall that the outdegree (the books that the focal book recommends) is limited in size, owing to the way that co-purchased products are presented on Amazon. The indegree, in contrast, varies substantially, and we see that that low selling books receive half the number of recommendations that bestselling books receive (column 1). This may not be not surprising given that popular books are co-sold with more other books; yet this implies that low sellers may benefit less from the network.

It is also interesting to observe the types of books that the products in each tier are connected to. We see that books that form the indegree network for low sellers sold 2.3 units on average (column 2), whereas books in bestsellers' indegree networks sold 55 units on average. Similarly, we see that books in a given revenue tier receive a large percentage of their recommendations from books in that same tier: 41% of recommendations for low selling books come from low sellers, and 45% of the recommendations for bestsellers come from bestsellers. 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.

These phenomena can explain the relatively low network value of bestsellers in product networks such as Amazon. Compared with low sellers, bestsellers are recommended more frequently, and the recommendations they receive are from higher selling books. This results in a relatively high extrinsic value for high selling products. While such products also generate more value, this value is not enough to "compensate" for the extrinsic value. For low selling books, the opposite is true.

Table 2 about here

Note that in the above calculation we did not make any assumptions regarding the RCRs of low sellers and bestsellers: the calculations assumed a similar RCR across revenue tiers. The results may change if there are consistent differences between the various tiers. While there is no clear empirical evidence as to the magnitude of the potential differences in RCR for different products, one can expect that for the most part, recommendations to high selling products will yield a higher conversion rate (see also Oestreicher-Singer and Sundararajan, 2011). This can be due to the popularity of the author or genre or unobserved higher intrinsic quality. If this is indeed the case, the extrinsic value of high selling products should be even higher, resulting in an even lower network value.

Results for Different RCRs

Since RCRs may vary among product networks, it is interesting to examine how a change in the RCR may affect the patterns we saw above. We repeated our analysis using several other values for the average RCR, ranging from 1% to 20%. We present the full tables for the different RCR values in appendix A. Although the magnitude of the results changes with the level of RCR, the patterns discussed above are remarkably consistent across RCR levels. The results that relate to the main parameter of interest—the difference between network value and revenue—are presented in Table 3.

Table 3 about here

Looking at the tables in appendix A, the following picture emerges for the effect of the RCR. The higher the RCR: (a) the larger the extrinsic value; (b) the larger the generated value; and (c) the lower the intrinsic value. This pattern is observed 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, which reflects more traffic through the network. We see that as information passes more easily in the network, items increasingly affect others and are affected more. Thus, the role of the intrinsic value of the item decreases. It should be noted, however, that even with an RCR value of 20%, intrinsic value is by far greater than extrinsic or generated value.

The network value of each revenue tier is also affected by the RCR. Table 3 shows the relative difference between network value and revenue for different levels of RCR. We see that the larger the RCR, the greater the contribution of the low sellers to the network, i.e., the higher their network value in comparison to their revenue. For bestsellers the opposite is true: the larger the RCR, the lower the network value compared to revenue. These results demonstrate how a "stronger" product network in terms of interconnectivity among books has a differential effect across revenue tiers – it increases the network value of the lower selling, "long-tail" products, and it decreases the network value of bestsellers. It should be noted that despite the differences in relative contributions across different values of RCR, the absolute generated value the bestsellers give the network is always considerably higher than that of the low sellers.

Applying the Model to BarnesandNoble.Com Data

To examine our results on another product network, we replicated the analysis on a second dataset of 257,000 books taken from the eCommerce website of BarnesAndNoble.com. (For a discussion of the collection method and descriptive statistics see Web appendix A.)

We find that the results for the Barnes and Noble data are similar to those for the Amazon data (see bottom part of Table 1). The ratio of extrinsic value to intrinsic value increases as the revenue tier increases, and the ratio of network value to revenue decreases as the revenue tier increases. The overall difference between network value and revenue shows a similar pattern to that of Amazon. Low sellers provide the network with the highest ratio of network value to revenue, while the bestsellers receive more value from the network than they contribute through recommendations. However, the variation is not as strong as in the case of Amazon. One reason may be that we were not able to collect as many books from the Barnes and Noble website. (This is due to differences in the websites as well as in the data collection algorithm.) This can lead to a smaller proportion of niche products that are historically low sellers. With a "thinner" tail, we witness a smaller redistribution effect.

DISCUSSION

The emergence of online recommendation systems makes product networks a vivid reality for firms, and highlights the question of how managers can take advantage of product networks to enhance profitability. Here we focused on how product valuation can take into account the effects of the product network. We presented an approach that enables marketers to adopt a network value view of products, and we demonstrated its applicability to large-scale databases. We have further shown that the product network value approach enables a better understanding of the contribution of different product revenue tiers to the firm. Next, we briefly summarize our results and their implications and discuss directions for further research on this issue.

The sources of product network value. We separate value that stems from a product's revenue into the *intrinsic value*, which is the value independent of the influence of other products, and *extrinsic value*, which stems from recommendations from other products. The *network value* of a product is the true value generated by a product that is part of a network – it comprises the product's intrinsic value in addition to its *generated value*, i.e., the value the product creates for its neighbors. In the datasets we examined, while the network certainly plays a role, intrinsic value still creates the vast majority of profits and is on average much greater than extrinsic or generated values.

The network value of bestsellers and low sellers. The overall effect of the network value approach is non-trivial: On the one hand, higher-selling products benefit from a larger number of incoming links, which in our model translates to a higher extrinsic value. On the other hand, higher-selling products are likely to drive more traffic to neighboring products, resulting in higher generated value. Our analysis of data from Amazon.com and from BarnesAndNoble.com shows that on average, the ratio of network value to revenue is lower among bestsellers than among low sellers. Furthermore, the network value of high selling products is lower than their revenue, suggesting that the true value of these products is consistently overestimated. Similarly, the average network value of low selling products is higher than their revenue, suggesting that the true value of low sellers is consistently underestimated.

Two issues are notable here. First, network value is an unexplored source of value of the long tail of electronic commerce. Whereas previous research examined whether and how recommendation systems affect the size of the long tail and that of the bestseller category, we show how incorporating the product network structure into value assessments changes the relative value of each segment. Second, it is interesting to see that the relative change in value is largely due to incoming recommendations. Whereas in the traditional social network literature the social value of an individual has largely been measured in terms of outgoing effects (e.g., word of mouth to others), for the product networks we analyze, we see that incoming ties (the number of recommendations an item receives from others) are a main driver of value.

The effects of the RCR. The RCR is an important factor in the assessment of network value. A higher RCR indicates that the product network is more active and that products tend to "recommend" each other more effectively on average. If we examine the effect of the RCR level, a few conclusions emerge. First, the basic results presented above are generally consistent across a range of RCRs, yet the magnitude of the effects changes. Second, a higher RCR increases extrinsic and generated values and decreases intrinsic value for all revenue tiers, which is an expected consequence of a situation in which items influence one another in a stronger way.

However, the RCR's effect on products' network value differs across revenue tiers. A larger RCR increases the network value of the low sellers and decreases the network value of the bestsellers. Naturally, sellers aim to develop better recommendation systems that will increase the RCR (Bodapati 2008). We find that if that happens, the relative role of the "long tail" in the network value created to the firm will increase.

Managerial Implications

The full value of products. In recent years, managers have moved from a limited view of customer value—stemming solely from his or her purchases—to a broader view that also takes into account the customer's effect on others via word of mouth (Kumar et al. 2010; Libai et al 2010). Similarly, managers should take into account the full value of a product that is part of a product network. Indeed, firms increasingly aim to optimize the number and types of products and brands they hold in their portfolio, using products' value to the firm as a key criterion in determining which products to carry (Aaker 2004). The long-tail analysis presented here illustrates the need to take a product's full network value into consideration, rather than its revenue alone. Some products bring in little revenue through sales, yet their relative contribution to the network is much greater. Therefore, it might be a mistake to eliminate these products from the portfolio.

A retailer vs. a manufacturer. From another angle, the product network approach can be used to better understand channel relationships in the optimal management of product assortment. Recent research in this area has focused on the relationship between manufacturers and retailers and considered the optimal assortment the retailer should carry, and its price (Draganska, Mazzeo, and Seim 2009; Dukes, Geylani, and Srinivasan 2009). Literature on slotting allowance, for example, has discussed the information asymmetry between the manufacturer and the retailer with regard to the true quality (and hence value) of the product (e.g., see Bloom, Gundlach, and Cannon 2000). Yet, none of these streams of literature has taken product network issues into account. A product made by one manufacturer can generate recommendations for and receive recommendations from other products, which creates a discrepancy 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, pricing, and in more general terms power in channels. Its consequences are yet to be explored.

Marketing mix. Generally, marketing mix decisions should take into account the effect of the network value of products. Through the effect on network value one can better assess how pricing changes or promotions for one product can have a system-wide effect. Furthermore, one can also better compare the effectiveness of selling a product via an online channel that allows a hyperlinked product network, versus offline channels where display options are more limited. While such marketing mix analysis can be done for cases of two interconnected products in the supermarket (Niraj, Padmanabhan, and Seetharaman 2008; Bezawada et. al 2009), the advantage of the approach presented here is the ability to take into account effects among multiple products in a large-scale product network.

Building the product network. Finally, the approach presented here can help managers to think about the best ways of building the product network. In online environments the firm can influence the connectivity among products in the product network, and the question is how to do so to maximize the system-wide profitability. The network value approach described here presents marketers with a tool to assess the implications of different recommendation systems that customers can see.

Future Research and Limitations

The analysis of value created in product networks raises many questions that are beyond the scope of this paper, and can be explored in further research. Here are just a few:

Richer specification. We used just a few variables to create the mechanism that determines the different dimensions of value. An advantage of this approach is the ability to make inferences given limited data, which makes our model more applicable and simpler to run in large-scale electronic commerce systems. However, the model can be extended to take into account the rich environment in which product networks operate. For example, in the empirical analysis we used the estimated sales of product v as a proxy for the amount of traffic it sends to product u. Given richer data sets, the number of views of product v's web page can be used to assess its effect on u. The use of page-view data can enable researchers to take additional processes into account. For instance, one can envision a case in which a person viewing product v clicks on a recommendation link for product u and ultimately chooses to purchase product u rather than product v. In this case, the recommendation for u negatively affects the sales of v, an option that does not exist in the model we used.

Heterogeneous RCR. In the empirical analysis we used a constant RCR in a range of values that was consistent with prior empirical findings in this area. Retailers like Amazon, however, can aim to obtain a specific RCR for each dyad of books, which may, of course, enhance the assessment of the network value for specific items. Data on product-specific recommendations can also enable researchers to study how a specific item's characteristics and network position can affect its network value. Furthermore, researchers might study how to minimize the endogeneity issue associated with recommendation networks by studying how changes in product characteristics affect the value of α.

Note, however, that the use of a constant RCR has its advantages. First, assessing an item-specific RCR from transaction data demands careful analysis and may become resource-consuming in evaluating a product network of millions of products. An average RCR may be sufficient for gaining various insights as demonstrated above. Second, while the retailer has access to detailed item-specific information, manufacturers that sell though product networks might not, and they may have to use aggregate-level measurements to determine the network value of their products.

Implications for optimal behavior in the market. Our analysis accepted the network as given and aimed to derive consequences in terms of products' network value. Retailers can theoretically affect the nature of the recommendation system, and thus use the method presented here to examine the overall change that may increase profits. It should be noted, however, that firms such as Amazon have indicated in the past that the recommendations are not presented in a strategic way, but reflect customers' true purchase behavior.

Other types of networks. While online recommendation systems are natural candidates for product network analysis, product networks exists in various forms in many consumption situations, including offline ones. Items in a supermarket, products in a catalog, or stores in a mall can also be examples of product networks. Figuring out the associations among products in offline settings may, however, 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 towards building product networks based on product associations.

Conclusions

Assessing the value of a product is central to informed marketing management. It forms the basis of well-planned advertising, brand portfolio planning, channel placement, cross-selling initiatives, pricing, and compensation of marketing personnel. The incorporation of network value is thus of essential importance to marketers. As the share of online purchases increases, the ability of firms to measure and affect network value will rise. This study should therefore be only a first step towards a better understanding of this important concept.

REFERENCES

Aaker, David A. (2004), *Brand Portfolio Strategy*. New-York: Free Press.

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.

Bezawada, Ram, S. Balachander, P. K. Kannan, and Venkatesh Shankar (2009), "Cross-category Effects of Aisle and Display Placements: A Spatial Modeling Approach and Insights," *The Journal of Marketing*, 73 (3), 99–112.

Bharat, Krishna and George A. Mihaila (2000), "Hilltop: A Search Engine Based on Expert Documents," in *Proceedings of the 9th International World Wide Web Conference*. New York: Elsevier Science.

Binken, Jeroen L. G. and Stefan Stremersch (2009), "The Effect of Superstar Software on Hardware Sales in System Markets," *The Journal of Marketing*, 73 (2), 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," *The Journal of Marketing*, 64 (2), 92–108.

Bodapati, Anand V. (2008), "Recommendation Systems with Purchase Data," *Journal of Marketing Research*, 45 (1), 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 Mohammad Saifur Rahman (2009), "Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition," *Management Science*, 55 (11), 1755–1765.

———, ———, 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), 1580–1596.

———, ———, and ——— (2006), “From Niches to Riches: Anatomy of the Long Tail,” *MIT Sloan Management Review*, 47 (4), 67.

———, ———, 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, 14–15 December.

Cachon, Gérard P., Christian Terwiesch, and Yi Xu (2008), “On the Effects of Consumer Search and Firm Entry in a Multiproduct Competitive Market,” *Marketing Science*, 27 (3), 461–473.

Carmi, Eyal, Gal Oestreicher-Singer, and Arun Sundararajan (2009), “Spreading the Oprah Effect: The Diffusion of Demand Shocks in a Recommendation Network,” *Proceedings of the Twenty-Ninth Annual International Conference on Information Systems (ICIS).*

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.

Choi, Jeonghye, Sam K. Hui, and David R. Bell (2010), “Spatiotemporal Analysis of Imitation Behavior Across New Buyers at an Online Grocery Retailer,” *Journal of Marketing Research*, 47 (1), 75–89.

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.

Clemons, Eric K., Guodong Gordon Gao, and Lorin M. Hitt (2006), “When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry,” *Journal of Management Information Systems*, 23 (2), 149–171.

Clifton, Rita, John Simmons and Sameena Ahmad (2009), *Brands and Branding.* Princeton, NJ: Bloomberg Press.

de los Santos, Babur (2008), “Consumer Search on the Internet,” NET Institute Working Paper No. 08-15.

Dellarocas, Chrysanthos, Zsolt Katona, and William M. Rand (2011), “Media, Aggregators and the Link Economy: Strategic Hyperlink Formation in Content Networks,” NET Institute Working Paper No. 10-13.

Deschatres, F. and D. 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,” *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining*. 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

Dukes, Anthony J., Tansev Geylani, and Kannan Srinivasan (2009), “Strategic Assortment Reduction by a Dominant Retailer,” *Marketing Science*, 28 (2), 309–319.

Economist (2006), “Podtastic: The iPod has Turned Apple into a Superbrand,” Jan 12, print edition.

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,” *Journal of Business and Economic Statistics*, Published online October 19.

Gyöngyi, Zoltán, Hector Garcia-Molina, and Jan Pedersen (2004), “Combating Web Spam with Trustrank,” in *Proceedings of the Thirtieth International Conference on Very Large Data Bases*. VLDB Endowment, 587.

Herlocker, Jonathan L., Joseph A. Konstan, Loren G. Terveen, and John T. Riedl (2004), “Evaluating Collaborative Filtering Recommender Systems,” *ACM Transactions on Information Systems*, 22 (1), 5–53.

Hervas-Drane, Andres (2009), "Word of Mouth and Taste Matching: A Theory of the Long Tail," Working paper.

Hess, James D. and Eitan Gerstner (1987), "Loss Leader Pricing and Rain Check Policy," *Marketing Science*, 6 (4), 358–374.

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.

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.

Katona, Zsolt and Miklos Sarvary (2008), "Network Formation and the Structure of the Commercial World Wide Web," *Marketing Science*, 27 (5), 764.

Kleinberg, Jon M. (1999), "Authoritative Sources in a Hyperlinked Environment," *Journal of the ACM*, 46 (5), 604–632.

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.

Li, Shibo, Baohong Sun, and Ronald T. Wilcox (2005), "Cross-selling Sequentially Ordered Products: An Application to Consumer Banking Services," *Journal of Marketing Research*, 42 (2), 233–239.

Libai, Barak, Ruth Bolton, Marnix S. Büegel, Ko de Ruyter, Oliver Götz, Hans Risselada, and Andrew T. Stephen (2010), "Customer-to-Customer Interactions: Broadening the Scope of Word of Mouth Research," *Journal of Service Research*, 13 (3), 267–282.

Libai, Barak, Eitan Muller, and Renana Peres (2009), "The Role of Within-Brand and Cross-Brand Communications in Competitive Growth," *The Journal of Marketing*, 73 (3), 19–34.

Linden, Greg, Brent Smith, and Jeremy York (2003), "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Computing*, 7 (1), 76–80.

Manski, Charles F. (2000), "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, 14 (3), 115–136.

Mayzlin, Dina and Hema Yoganarasimhan (2008), "Link to Success: How Blogs Build an Audience by Promoting Rivals," Working paper, Yale School of Management.

Moe, Wendy W. and Peter S. Fader (2004), "Dynamic Conversion Behavior at E-commerce Sites," *Management Science*, 50 (3), 326–335.

Newman, M. E. J. (2002), "Assortative mixing in networks," *Physical Review Letters*, 89 (20), 208701-4.

——— (2003), "The Structure and Function of Complex Networks," *SIAM Review*, 45 (2), 167–256.

Nicholls, Charles (2010), *Lessons Learned from the Top 10 Converting Websites*. Andover, MA: SeeWhy.

Niraj, Rakesh, V. Padmanabhan, and P. B. Seetharaman (2008), "A Cross-Category Model of Households' Incidence and Quantity Decisions," *Marketing Science*, 27 (2), 225–235.

Oestreicher-Singer, Gal and Arun Sundararajan (2011), "Recommendation Networks and the Long Tail of Electronic Commerce," *MIS Quarterly*, Forthcoming.

Seetharaman, P. B., Siddhartha Chib, Andrew Ainslie, Peter Boatwright, Tat Chan, Sachin Gupta, Nitin Mehta, Vithala Rao, and Andrei Strijnev (2005), "Models of Multi-Category Choice Behavior," *Marketing Letters*, 16 (3), 239–254.

Senecal, Sylvain and Jacques Nantel (2004), "The Influence of Online Product Recommendations on Consumers' Online Choices," *Journal of Retailing*, 80 (2), 159–169.

Smith, Donnavieve, Satya Menon, and K. Sivakumar (2005), "Online Peer and Editorial Recommendations, Trust, and Choice in Virtual Markets," *Journal of Interactive Marketing*, 19 (3), 15–37.

Sriram, S., Pradeep K. Chintagunta, and Manoj K. Agarwal (2010), "Investigating Consumer Purchase Behavior in Related Technology Product Categories," *Marketing Science*, 29 (2), 291–314.

Stephen, Andrew T. and Oliver Toubia (2010), "Deriving Value from Social Commerce Networks," *Journal of Marketing Research*, 47 (2), 215–228.

Tan, Tom F. and Serguei Netessine (2010), “Is Tom Cruise Threatened? Using Netflix Prize Data to Examine the Long Tail of Electronic Commerce,” Working paper, University of Pennsylvania.

Tucker, Catherine and Juanjuan Zhang (2007), “Long Tail or Steep Tail? A Field Investigation into How Online Popularity Information Affects the Distribution of Customer Choices,” Working paper, Massachusetts Institute of Technology (MIT), Sloan School of Management.

Van den Bulte, Christophe and Stefan Wuyts (2007), *Social Networks and Marketing*. Cambridge, MA: Marketing Science Institute.

Wedel, Michel and Jie Zhang (2004), “Analyzing Brand Competition across Subcategories,” *Journal of Marketing Research*, 41 (4), 448–456.

FOOTNOTES

1. 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 $PageRank(u) = \frac{(1-d)}{N} + d\sum_{v\in In(u)}\frac{PageRank(v)}{OutDegree(v)}$.
2. We also experimented with alternative forms such as $1/n$, making convergence a bit longer; results were robust to these specifications.
3. The Sales Rank is a number associated with each product on Amazon.com, which measures its demand of relative to other products. The lower the number is, the higher the sales of that particular product.
4. Currently Amazon.com provides a list of more than five items in each co-purchase network. Due to screen size limitations, users are initially exposed to the top five, and can then click to view the next five products, and so on.
5. The binning of 20% follows the conversion when discussing demand distribution, and specifically the tail of the distribution.

TABLES

	1	2	3	4	5	6
Revenue Percentile	Average Revenue ($)	Average Network value ($)	Average Extrinsic value ($)	Average Intrinsic value ($)	Average Generated value ($)	% of difference (NV-R)/R
Amazon.com						
0-20%	5.83	6.210	0.35	5.48	0.73	13.30%
20-40%	11.20	11.60	0.79	10.40	1.17	3.43%
40-60%	17.30	17.60	1.38	15.90	1.66	1.72%
60-80%	28.80	29.00	2.39	26.40	2.62	0.84%
80-100%	105.00	104.00	10.30	94.90	8.71	-0.83%
BarnesNoble.com						
0-20%	10.201	10.836	0.694	9.507	1.333	6.79%
20-40%	18.752	19.509	1.286	17.466	2.048	4.08%
40-60%	28.070	28.852	2.078	25.992	2.866	2.82%
60-80%	44.637	45.315	3.637	40.999	4.320	1.60%
80-100%	150.850	147.940	15.261	135.590	12.329	-0.75%

Table 1 – Network value estimation results for Amazon and BarnesNoble.com using an RCR of 10%.

	1	2	3
Revenue Percentile	**Indegree size**	**Average units indegree**	**% same group indegree**
0-20%	3.28	2.33	0.45
20-40%	3.87	3.59	0.28
40-60%	4.41	5.25	0.24
60-80%	5.06	8.30	0.25
80-100%	7.49	55.32	0.41

Table 2 – Revenue tiers and indegree statistics for Amazon

Revenue Percentile	(NV-R)/R				
	RCR=1%	RCR=5%	RCR=10%	RCR=15%	RCR=20%
0-20%	1.33%	6.65%	13.30%	20.00%	26.60%
20-40%	0.34%	1.72%	3.43%	5.15%	6.87%
40-60%	0.17%	0.86%	1.72%	2.58%	3.44%
60-80%	0.08%	0.42%	0.84%	1.26%	1.68%
80-100%	-0.08%	-0.42%	-0.83%	-1.25%	-1.66%

Table 3 – Results of network value estimations using different RCR values (Amazon).

APPENDIX A – RESULTS FOR DIFFERENT RCRS

	1	2	3	4	5
Revenue Percentile	Average Revenue ($)	Average Network Value ($)	Average Extrinsic Value ($)	Average Intrinsic Value ($)	Average Generated Value ($)
0-20%	5.827	5.865	0.035	5.792	0.073
20-40%	11.196	11.234	0.079	11.117	0.117
40-60%	17.278	17.307	0.138	17.141	0.166
60-80%	28.781	28.804	0.239	28.542	0.262
80-100%	105.180	105.020	1.030	104.150	0.872

Table A1 – Network value estimation results for Amazon using an RCR of 1%

	1	2	3	4	5
Revenue Percentile	Average Revenue ($)	Average Network Value ($)	Average Extrinsic Value ($)	Average Intrinsic Value ($)	Average Generated Value ($)
0-20%	5.827	6.018	0.175	5.652	0.367
20-40%	11.196	11.384	0.397	10.799	0.585
40-60%	17.278	17.422	0.688	16.590	0.832
60-80%	28.781	28.896	1.196	27.585	1.311
80-100%	105.180	104.400	5.148	100.030	4.359

Table A2 – Network value estimation results for Amazon using an RCR of 5%

	1	2	3	4	5
Revenue Percentile	Average Revenue ($)	Average Network Value ($)	Average Extrinsic Value ($)	Average Intrinsic Value ($)	Average Generated Value ($)
0-20%	5.830	6.400	0.525	5.300	1.110
20-40%	11.200	11.800	1.190	10.000	1.760
40-60%	17.300	17.700	2.060	15.200	2.500
60-80%	28.800	29.100	3.590	25.200	3.930
80-100%	105.000	103.000	15.400	89.700	13.100

Table A3 – Network value estimation results for Amazon using an RCR of 15%

	1	2	3	4	5
Revenue Percentile	Average Revenue ($)	Average Network Value ($)	Average Extrinsic Value ($)	Average Intrinsic Value ($)	Average Generated Value ($)
0-20%	5.830	6.590	0.701	5.130	1.480
20-40%	11.200	11.900	1.590	9.610	2.350
40-60%	17.300	17.900	2.750	14.500	3.330
60-80%	28.800	29.200	4.780	24.000	5.250
80-100%	105.000	102.000	20.600	84.600	17.400

Table A4 – Network value estimation results for Amazon using an RCR of 20%

APPENDIX B – THE RCR FOR AMAZON.

While direct data on the recommendation conversion rate (RCR) are not available, there are indications that RCRs for retailers such as Amazon are sizeable. Previous research has shown that many online consumers seek and accept recommendations in order to effectively manage the amount of information available during online search processes, and that consumers prefer such recommendations over other types of effort-reducing cues that might be available during online search (Smith, Menon, and Sivakumar 2005). Research has further shown that the use of a recommendation system can double the probability of consumers' eventual purchase (Senecal and Nantel 2004).

Clearly, the effectiveness of an online recommendation may depend on the type of product and the customer's characteristics (Senecal and Nantel 2004), as well as on the algorithm for the choice offered to consumers (Bodapati 2008). Amazon is considered one of the top websites in conversion of visitors to customers, with reported conversion rate ranges that can reach 14% and 17% (Moe and Fader 2004; Nicholls 2010). It is also reported that Amazon assesses that 30% of its revenues come from product recommendations, on the website and via email (Nicholls 2010). It is therefore not surprising that the average book purchase on Amazon includes more than two books (de los Santos 2008).

Recently, Carmi et al. (2009) provided some insights regarding the RCR by looking at what they labeled the "Oprah Effect" – the impact on sales of an external shock such as a book review featured on the *Oprah Winfrey Show* or in the *New York Times Book Review*. They found that after a review, not only did the sales of the reviewed book increase, but the sales of the books that were connected to it on Amazon rose considerably, as well. The ratio between the increase in the sales of a reviewed book and the increase in the sales of a connected book can give us an indication of the size of the RCR. While the authors reported a large variation in the effect on connected books, the average effect was substantial. After a review in the *New York Times*, for example, the increase in the sales of connected books was, on average, 7.8% of the increase in the sales of the reviewed book. For books recommended by Oprah Winfrey, the percentage was 20.1% on average (the difference might be partly related to the types of books recommended, which differed between the two sources).

In order to further explore the range of values for RCR, we conducted an additional analysis that considers another network, which we label the UPD (Ultimate Purchase Decision) network. Customers who scroll down the webpage of an Amazon book can see a section titled “What Do Customers Ultimately Buy After Viewing This Item?” While its poorer position makes it a less favorable recommendation tool for consumers, this network can still help us to assess the RCR for the recommendation network. In Web Appendix C we describe the method we used to do so, which yielded an average RCR of around 17%.

In practice, the RCR may vary substantially across products and eCommerce sites. In the following empirical analysis we will use an average RCR which is the same for all dyads. We present the analysis of the Amazon network for a range of RCRs, from 1% to 20%, but we use a mid-range value of 10% to demonstrate the main results. This number is consistent with the information above, and also with recent findings that suggest that a well-designed recommendation system can result in a recommendation conversion rate of over 9% (Bodapati 2008).