ANDREW T. STEPHEN and OLIVIER TOUBIA*

Social commerce is an emerging trend in which sellers are connected in online social networks and sellers are individuals instead of firms. This article examines the economic value implications of a social network between sellers in a large online social commerce marketplace. In this marketplace, each seller creates his or her own shop, and network ties between sellers are directed hyperlinks between their shops. Three questions are addressed: (1) Does allowing sellers to connect to each other create value (i.e., increase sales)? (2) What are the mechanisms through which this value is created? and (3) How is this value distributed across sellers in the network and how does the position of a seller in the network (e.g., its centrality) influence how much he or she benefits or suffers from the network? The authors find that (1) allowing sellers to connect generates considerable economic value, (2) the network's value lies primarily in making shops more accessible to customers browsing the marketplace (the network creates a "virtual shopping mall"), and (3) the sellers who benefit the most from the network are not necessarily those who are central to the network but rather those whose accessibility is most enhanced by the network.

Keywords: social networks, social commerce, electronic commerce, network analysis, marketing value

Deriving Value from Social Commerce Networks

Social commerce and social shopping communities are growing in number and size. Broadly defined, social commerce and social shopping are forms of Internet-based "social media" that allow people to participate actively in the marketing and selling of products and services in online marketplaces and communities. In other words, these applications merge online shopping and social networking (Tedeschi 2006). The distinction between social shopping and social commerce is that while social shopping connects customers, social commerce connects sellers. Consumers' roles vary across Web sites or platforms and can range from generating content (e.g., product reviews and recommendations, known as "consumer-generated media," on Web sites such as Epinions.com, ThisNext.com, and Yelp.com) to being sellers and curators of online stores (e.g., eBay MyWorld/Neighborhoods, Squidoo.com, and Zlio.com). The *Financial Times* reports that Internet traffic to social commerce and social shopping Web sites grew by more than 500% between early 2007 and early 2008 (Palmer 2008), the *New York Times* reports that several social commerce firms are attracting substantial venture capital financing (Tedeschi 2006), and further growth and investment in this online retailing segment are expected.

Social shopping revolves around online word of mouth and has recently been studied in academia. For example, Chevalier and Mayzlin (2006) and Godes and Mayzlin (2004) study word of mouth and the influence of consumer-generated media on business outcomes, and Watts and Dodds (2007) study marketing-related implications of social contagion from a social networks perspective. Recent research in marketing has also examined related issues, such as consumer interdependence in choice and spatial models

*Andrew T. Stephen is Assistant Professor of Marketing, INSEAD (e-mail: andrew.stephen@insead.edu). Olivier Toubia is David W. Zalaznick Associate Professor of Business, Graduate School of Business, Columbia University (e-mail: ot2107@columbia. edu). The authors thank Asim Ansari; Jacob Goldenberg; Kamel Jedidi; Oded Koenigsberg; Donald Lehmann; Duncan Watts; the three anonymous *JMR* reviewers; and seminar participants at Columbia University, Emory University, INSEAD, London Business School, New York University, the University of California, San Diego, the University of Florida, the University of Maryland, the University of Pittsburgh, Washington University in St. Louis, and the WIMI conference on Modeling Social Network Data for their helpful comments and suggestions. They also thank Jeremie Berrebi, Ilan Abehassera, and David Levy from ZLIO.com for their collaboration. This article is based on Andrew Stephen's dissertation. Dominique Hanssens served as associate editor for this article.

ISSN: 0022-2437 (print), 1547-7193 (electronic)

(Yang and Allenby 2003), and other issues related to social networks in marketing contexts (Iyengar, Valente, and Van den Bulte 2008; Katona and Sarvary 2008; Nair, Manchanda, and Bhatia 2006; Trusov, Bucklin, and Pauwels 2009; Van den Bulte and Joshi 2007; Van den Bulte and Lilien 2001).

Social commerce is a more recent phenomenon and has not been studied as extensively. Social commerce marketplaces have four defining characteristics: (1) Sellers (or shopkeepers) are individuals instead of firms, (2) sellers create product assortments organized as personalized online shops, (3) sellers can create hyperlinks between their personalized shops, and (4) sellers' incentives are based on being paid commissions on sales made by their shops. What emerges is a consumer-driven online marketplace of personalized, individual-curated shops that are connected in a network. Links between sellers' shops in this network are directed, clickable hyperlinks that customers can use to move from shop to shop. In the specific marketplace that we study, which is a large and typical social commerce marketplace created in Europe, the products that sellers add to their shops come from vendors (e.g., Amazon, Apple, Gap) through arrangements made by the marketplace owner. As a result, sellers do not own any inventory and do not set prices; they only manage the product mix.

Our aim is to understand social commerce as a new business concept, focusing on whether and how it generates economic value for marketplace-owning firms and for the people who participate as sellers in these marketplaces (by increasing sales). Issues related to connecting sellers have not been studied extensively (an exception is the shopping center literature that we review subsequently). The value implications of networks have been studied in other contexts, such as inter- and intrafirm networks of a formal or an informal nature (e.g., Rindfleisch and Moorman 2001; Tsai and Ghoshal 1998; Wuyts et al. 2004) and collaborative group networks (e.g., Freeman, Roeder, and Mulholland 1980; Grewal, Lilien, and Mallapragada 2006). The economic implications of social structure have also been discussed recently in economics and sociology (e.g., Goyal 2007; Greif 2006). However, little is understood about whether networks provide some economic value in marketing and retailing contexts. We consider the following questions in relation to social commerce: (1) Does allowing sellers to connect to each other create economic value (i.e., increase sales)? (2) What are the mechanisms through which this value is created? and (3) How is this value distributed across sellers in the network, and how does the position of a seller in the network (e.g., centrality) influence how much the seller benefits or suffers from the network?

We address these questions using a novel data set from an online marketplace that, after hosting a set of independent, consumer-generated online shops for approximately 18 months, became a social commerce marketplace by allowing its sellers to connect their shops and form a shop network (a connection from Shop A to Shop B is represented by a directed hyperlink to Shop B on Shop A's Web site). Our data set covers both a prenetwork and a postnetwork period (enabling us to study the effect of the introduction of the networking feature), and it contains detailed information on the characteristics and performance of each shop (enabling us to explore how the value created by the network is shared across members). We use multiple methods and analyze these data at the marketplace level (using time-series analysis) and at the shop level (using Bayesian statistical analysis). We organize the article as follows: First, we review relevant literature on shopping centers and social networks. Second, we describe our data set. Third, we report the results of our marketplace-level analysis. Fourth, we report the results of our shop-level analysis. Finally, we conclude with a general discussion of the results and suggestions for further research.

BACKGROUND AND THEORY

Network ties between sellers in social commerce marketplaces are links between sellers' shops that customers can use to browse between shops, akin to browsing through a virtual shopping center. For the people who participate as sellers in social commerce marketplaces and who earn commissions on the sales they make, the network can make their shops more accessible and more likely to be discovered by a browsing customer.

Bricks-and-mortar shopping centers are possible analogs to online social commerce marketplaces. Whereas social commerce shops are connected by directed hyperlinks, in offline shopping centers, shops are linked by spatial proximities (though these offline "links" are not inherently social and are usually determined by retail planners). The literature on shopping centers in real estate economics has considered relevant issues, such as tenant mixes and locations, rent setting, customer traffic generation, colocation, and spatial dependence between grouped shops (e.g., Eppli and Benjamin 1994; Lee and Pace 2005). Marketing researchers have also contributed to this literature (e.g., Nevin and Houston 1980), including recent work examining retailers' decisions to enter shopping centers (Vitorino 2008) and work on spatial dependence between marketing variables (e.g., market shares) at the geographic region level (e.g., Bronnenberg and Mahajan 2001).

An important concept in the shopping center literature is that of "retail demand externalities." A positive retail demand externality exists when customers are drawn to a shopping center because of the presence of attractive "anchor" tenants, such as department stores, supermarkets, or superstores (Eppli and Benjamin 1994). Smaller shops benefit from being in the same center as an anchor because anchors increase customer traffic and thus increase smaller shops' chances of attracting customers and making sales (Ingene and Ghosh 1990). The benefits to customers include reduced travel costs and the convenience of multipurpose or "one-stop" shopping (Eppli and Benjamin 1994). In general, these effects are empirically well supported (e.g., Eppli and Benjamin 1994; Nevin and Houston 1980). The spatial dependence stream of the shopping center literature (e.g., Lee and Pace 2005) suggests that shops' locations within offline shopping centers can influence their sales (e.g., being next door to an anchor may boost sales), though any shop in a shopping center is more accessible than a standalone shop outside the center in most cases.[1]

[1]Similar logic applies to retailer colocation in local districts or zones (e.g., the diamond district in midtown Manhattan, secondhand bookstores lining the Seine in Paris). More generally, retailer colocation is a micro type of the concept of industrial districts introduced by Marshall ([1890] 1961).

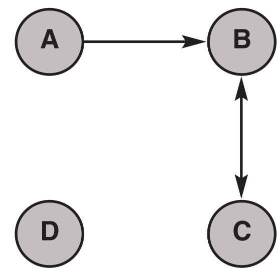
Therefore, bricks-and-mortar shopping centers generate value primarily by making stores more accessible to customers. We argue that social commerce networks act as virtual shopping centers that create economic value through the same basic concept of accessibility. For individual sellers in large social commerce marketplaces, being found by customers can be challenging. Thus, being part of a network and, importantly, being accessible and reachable in that network has a benefit comparable to the benefit offered by bricks-and-mortar shopping centers. Increasing the overall accessibility of the shops in the network makes the marketplace more "sticky" (i.e., helps retain customers within the marketplace for longer). In other words, the network has the potential to affect the number of visitors to any given shop, which has a direct impact on sales because the number of sales is simply equal to the number of visitors multiplied by the shop's conversion rate.

However, though similar to offline shopping centers at a basic level, social commerce marketplaces are not merely online equivalents of shopping centers, thus making social commerce a theoretically and substantively interesting context to study. In particular, the drivers of accessibility are likely to differ between social commerce networks and bricks-and-mortar shopping centers for at least four reasons. First, links between shops in a social commerce network are directed—that is, a hyperlink is from Shop A to Shop B and not necessarily the reverse. In bricks-and-mortar shopping centers, "links" (physical proximities) between shops are obviously undirected, and customers' browsing paths are not as structurally constrained as in an online shop network. Second, when customers browse in a shopping center, they do not visit (i.e., go into) every shop they pass on their way from one shop to another. In the online context, however, browsing requires that the customer visit every shop along a path. For example, on a path A → B → C, going from A to C requires visiting B. Third, traveling costs are lower online (relative to the goods being purchased). As we mentioned previously, a key benefit of grouping bricks-and-mortar retailers to form a shopping center is convenience (i.e., reduced traveling costs for customers). This benefit is likely to be less critical online. Fourth, while the number of neighbors of a bricks-and-mortar shop in a shopping mall is physically constrained, the number of links to and from any given shop in a social commerce network is not constrained.

Thus, although we use the shopping center literature to motivate the concept of accessibility, we turn to the social networks literature to study the drivers of accessibility in social commerce networks. At the marketplace level, in general, networks that are more connected (i.e., with a larger number of links) tend to improve the overall accessibility of their members. However, simply increasing the number of links may not always be beneficial. Improvements in accessibility depend on the browsability of the network, and not all links contribute equally to a network's browsability. In particular, creating new links sometimes actually hurts the browsability of the network. For example, in Figure 1, a simple network evolves from Time 1 to Time 2 with the addition of a new link. At Time 2, Shop D is brought into the network by Shop A with the creation of the A → D link. This makes Shop D accessible from Shop A, but this also makes D a "dead end" (i.e., a shop with at least one incoming link but no outgoing links).[2] If a customer browses to D (starting at A), he or she will not be able to access Shop B or C. Although the network at Time 2 is more connected, the creation of a dead end makes it less browsable (at least from B's and C's perspective).

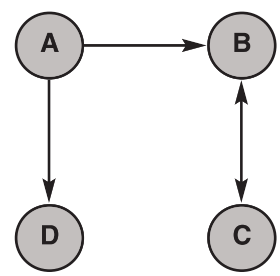
Figure 1
ILLUSTRATION OF DEAD ENDS

A: Time 1



A customer browsing the network starting at Shop A may access Shops B and C. Browsing from Shop B leads to Shop C, browsing from Shop C leads to Shop B, and browsing is not possible from Shop D.

B: Time 2



A customer browsing the network starting at Shop A may now access Shops B, C, and D. However, Shop D is a dead end. If a customer accesses Shop D from Shop A, Shops B and C become inaccessible. Browsing from Shops B, C, or D is similar to Time 1.

The accessibility of a Web site (in our case, a shop) is influenced by the structure of the network to which it belongs and its position in this network relative to other sites (Vázquez 2003). Specifically, shops with higher in-degree centrality (number of incoming links) should benefit more from the network than shops with lower in-degree centrality because more incoming links equate to a greater chance of customer traffic. We also expect that shops with higher incoming proximity (shops than can be reached from more shops in fewer steps) will benefit more from the network than shops with lower incoming proximity because shops that are accessible from fewer other shops and lie further down browsing paths are less likely to be visited. Conversely, shops that provide many opportunities for customers to leave by having a higher out-degree centrality (many outgoing links) or a higher outgoing proximity (can lead to more shops in fewer steps) should themselves benefit

[2]Nodes with incoming links but zero outgoing links are called "sinks" in graph theory and network analysis (De Nooy, Mrvar, and Batagelj 2005). Note that there is a possibility that a browsing customer could simply click the Web browser's "back" button to backtrack out of a dead-end shop. Therefore, the effect of dead ends found in our data set is conservative.

less from the network. In addition, shops that are connected to by shops that are themselves highly interconnected will be less accessible because the likelihood of browsing customers reaching such a shop will be low unless their browsing path starts in the shop's ego network. Therefore, we anticipate that shops with lower incoming clustering coefficients tend to benefit more from the network. We formally define the concepts of in-degree centrality, out-degree centrality, incoming proximity, outgoing proximity, incoming clustering coefficients, and outgoing clustering coefficients in Table 1. (Note that we address hub centrality and authority centrality [also defined in Table 1] subsequently.)

Other features will also influence a shop's likelihood of generating sales, which we refer to as a shop's "attractiveness" to customers. In this context, a shop's attractiveness may be related to its product assortment (e.g., number and uniqueness of its products) and the general ability or skill of the shopkeeper in creating an appealing product assortment. Finally, note that allowing sellers to connect to other shops could also have the effect of intensifying competition. However, because sellers typically do not set prices in such marketplaces (e.g., sellers select merchandise for their shops but do not set prices; vendors that provide the products set prices), intensified competition between sellers cannot force them to lower prices.[3]

[3]However, sellers may underinvest in service quality (e.g., visual attractiveness of their shops) in response to increased competition (Stern, El-Ansary, and Coughlan 1996).

DATA

Our data come from a company that runs popular and rapidly growing social commerce marketplaces in France, Germany, the United Kingdom, and the United States. Our data set covers the French marketplace, which is the largest and the first this company launched. The company leverages online retailers' "affiliate" selling programs, whereby Web sites that refer purchases to these retailers are paid commissions. Individuals ("sellers") join this marketplace and are given tools to create their own online personalized stores or "shops" (each shop has its own URL). Sellers add products to their shops from a database of more than four million products across many categories; these products come from more than 100 vendor retailers, such as Amazon, Apple, Dell, and the Gap. Each seller has complete control over his or her shop's product assortment. Importantly, sellers are individual people instead of companies.

The purchasing process in this marketplace is as follows: When a customer selects a product from a shop, he or she is referred to the corresponding vendor, which then processes the transaction and ships the product (i.e., the marketplace owner and its members hold no inventory). The vendor then pays the marketplace owner a commission for each transaction generated by one of the marketplace's shops, and this commission is shared with the seller whose shop generated the sale. For example, suppose that Mark visits Roger's shop in this marketplace, in which Roger lists a range of books on Bayesian estimation. If Mark purchases a particular book, he is taken to the corresponding vendor's Web site (e.g., Amazon), pays the vendor with his credit card, and a

Table 1
NETWORK POSITION MEASURES

Measure	*Formula*	*Definition and Intuition*
In-degree centrality	$(\mathbf{X'} \times \mathbf{1})_i = k_i^{in}$, where **1** is an N-dimensional vector of ones.	The number of links from other shops that go to shop i.
Out-degree centrality	$(\mathbf{X} \times \mathbf{1})_i = k_i^{out}$, where **1** is an N-dimensional vector of ones.	The number of links from shop i that go to other shops.
Incoming proximity	$[n_i^{in}/(N-1)] \times [n_i^{in}/\Sigma_{j \in in_i} d(j, i)]$, where in_i is the set of shops from which shop i can be reached in a finite number of steps, n_i^{in} is the number of shops in that set (indomain), and d(j, i) is the geodesic distance (shortest path length) from shop j to shop i.	Shop i's incoming proximity is proportional to the proportion of shops in the network other than i that can reach i in a finite number of steps and inversely proportional to the mean geodesic distance (shortest path length) from these shops to i.
Outgoing proximity	$[n_i^{out}/(N-1)] \times [n_i^{out}/\Sigma_{j \in out_i} d(j, i)]$, where out_i is the set of shops that can be reached from shop i in a finite number of steps, n_i^{out} is the number of shops in that set (outdomain), and d(i, j) is the geodesic distance (shortest path length) from shop i to shop j.	Shop i's outgoing proximity is proportional to the proportion of shops in the network other than i that can be reached from i in a finite number of steps and inversely proportional to the mean geodesic distance (shortest path length) from shop i to these shops.
Incoming clustering coefficient	$e_i^{in}/k_i^{in}(k_i^{in}-1)$, where e_i^{in} is the number of directed links between shops that connect to shop i directly (excluding i) and k_i^{in} is the in-degree of shop i.	The proportion of possible links that exist among the shops in shop i's incoming ego network. A shop with a higher incoming clustering coefficient is connected to by more clustered (as opposed to more dispersed) shops.
Outgoing clustering coefficient	$e_i^{out}/k_i^{out}(k_i^{out}-1)$, where e_i^{out} is the number of directed links between shops that shop i connects to directly (excluding i) and k_i^{out} is the out-degree of shop i.	The proportion of possible links that exist among the shops in shop i's outgoing ego network. A shop with a higher outgoing clustering coefficient connects to more clustered (as opposed to more dispersed) shops.
Hub centrality	The hub score for shop i is the ith component of the eigenvector corresponding to the largest eigenvalue of $\mathbf{XX^T}$.	These are both directed network versions of eigenvector centrality, which is an indicator of position-related status (i.e., being well connected to other well-connected shops). Good hubs connect to many good authorities, and good authorities connect to many good hubs.
Authority centrality	The authority score for shop i is the ith component of the eigenvector corresponding to the largest eigenvalue of $\mathbf{X^TX}$.	

Notes: The shop network is comprised of N shops. A directed link from shop i to shop j is denoted by x_{ij}. The network is represented as an N × N adjacency matrix, **X**, with diagonal elements $x_{ii} = 0$ (for all i = 1, …, N) and off-diagonal elements $x_{ij} = 1$ if there is a link from i to j (and 0 if otherwise).

few days later receives the book shipped from (or through) Amazon. Because Mark purchased a book from Roger's shop, Amazon pays the marketplace a commission on that sale, and in turn, Roger earns a portion of this commission. In summary, sellers are individual shopkeepers who do not own any inventory but create online shops that direct customers to online vendors and earn commissions on the sales made by their shops.

In June 2007, approximately 18 months after the marketplace had been established, the firm introduced a new feature that allowed members to link their shops to other shops (at that time, the marketplace had 74,291 shops). Shops were independent (all disconnected) before the introduction of this feature. This feature gave birth to a network with shops as nodes and directed hyperlinks as ties. A link from Shop A to Shop B means that Shop A's owner placed a hyperlink to Shop B on Shop A's home page. Shop B's owner cannot reject the incoming link but is not required to reciprocate this link (Shop B's owner is notified of the incoming link by e-mail). Therefore, this network is directed.

Our data set includes the entire French population of shops that were created anytime between the 1st day and the 781st day of this marketplace's life (the last day of our data set). The network was created (i.e., sellers were given the ability to link their shops to other sellers' shops) on the 583rd day of the marketplace's life; therefore, our data cover approximately the first seven months of the network's life. After 781 days of this marketplace's life in France, 136,774 shops had been created, and 21,373 of these shops (15.6%) were part of the network (i.e., had at least one incoming or outgoing link). By this time, the network had 82,810 directed links (network density, or the proportion of possible directed links that exist, was 1.21×10^{-5}). The marketplace and the network within it had grown quickly: An average of 180 new shops had been created in the marketplace each day, and after the network was born, an average of 107 shops joined the network each day, with an average of 421 new links created each day. In general, shops in this marketplace are small and have limited product assortments (the average shop features nine products). While the average commission revenue generated by each shop was modest (€2.84; though shops that made at least one sale had a higher average commission of €8.36), the aggregate revenue generated by the entire marketplace was nontrivial: 2.3 million transactions and €388,970 in commission revenues (from vendors) had been generated by the end of the observation window.

We begin by analyzing the data at the marketplace level to assess the economic value created by allowing sellers to link their shops to other sellers' shops in a manner observable to customers. We use time-series models to examine the value created by the network and the relationship between this value and some aggregate characteristics of the network. We then analyze the data at the individual shop level to address the issue of how the economic value created by the network is distributed across its members, using a hierarchical Bayesian tobit model with latent variables.

MARKETPLACE-LEVEL ANALYSIS

We first introduce the variables that are the focus of our time-series analysis. Our analysis of marketplace-level data uses autoregression with exogenous variables (ARX) time-series models. Data are available on each of these variables for each of the 781 days in our data set, covering both pre- and postnetwork birth periods:

- $\text{Commission\_revenues}_t$: the commissions, in euros, paid to the marketplace owner by the vendors for the sales made by the shops on day t,
- Network_t: a dummy variable indicating whether the network existed on day t or not,
- $\text{Marketplace\_size}_t$: the total number of shops in the marketplace at the end of day t (includes shops that are in the marketplace but do not have any network links),
- Network_links_t: the total number of links in the shop network on day t, and
- Dead_ends_t: the total number of dead-end shops in the network on day t (see Figure 1).

Impact of Allowing Sellers to Form a Network

Before investigating more specific issues related to network connectivity and marketplace commission revenues, we performed an initial test to address whether adding the networking feature had a positive effect on the commission revenues earned by the marketplace. In other words, did changing the disconnected online marketplace into a social commerce marketplace improve commission revenues? We addressed this question using a regime shift model because introducing the shop network feature on the 583rd day of the marketplace's life was a regime shift for this marketplace. The impact of this regime shift can be modeled with the following ARX model:

$$(1) \quad \text{Commission\_revenues}_t = \beta_0 + \sum_{l=1}^{p} \phi_l \times \text{Commission\_revenues}_{t-l} + \lambda_0 \times \text{Network}_t + \lambda_1 \times \Delta\text{Marketplace\_size}_t + \varepsilon_t.$$

The best-fitting model (i.e., with the lowest Bayesian information criterion [BIC]) had an autoregressive (AR) lag of $p = 6$.[4] $\text{Commission\_revenues}_t$, which we defined in its daily (differenced) form, was stable and not evolving using an augmented Dickey–Fuller unit-root test (with a null hypothesis of nonstationarity; $p < .001$). $\text{Marketplace\_size}_t$, defined in its total or cumulative form, was nonstationary based on an augmented Dickey–Fuller test ($p = .99$); however, when differenced (i.e., the daily change in the number of shops in the marketplace), it was stationary ($p < .001$). Thus, we use the difference in $\text{Marketplace\_size}_t$ (i.e., the daily change) and not the cumulative level in our model (Dekimpe and Hanssens 2004). Indeed, it would be unreasonable to model a stationary series as a function of a nonstationary one. A Stock and Watson common trend test for cointegration found that these series were not cointegrated. The use of an ARX model instead of a vector autoregressive (VAR) model was supported by a series of Granger-causality tests (Granger 1969; Hanssens et al. 2001; Trusov, Bucklin, and Pauwels 2009), which confirmed the exogeneity of

[4]We considered adding a time trend and monthly seasonality effects in Equations 1 and 2; however, in both cases, Wald tests could not reject the null hypothesis that the time trend and all month effects were zero, suggesting that time and seasonality effects were not needed.

ΔMarketplace_size$_t$ and Network$_t$. Because an incorrect choice of the AR lag p can erroneously conclude the absence of Granger causality, we selected a high lag (AR $p = 30$) to ensure that the results apply at any lag and not just the best-fitting lag for the model (Hanssens 1980; Trusov, Bucklin, and Pauwels 2009). As such, ΔMarketplace_size was not "Granger-caused" by either Commission_revenues or Network ($\chi^2[60] = 65.19$, $p = .30$), and Network was not Granger-caused by either ΔMarketplace_size or Commission_revenues ($\chi^2[60] = 40.07$, $p = .98$).

The regime shift model appeared to fit the actual series well ($R^2 = .72$, BIC = 11.77, median absolute deviation [MAD] = €59.78, and median absolute percentage error [MAPE] = 17.94%). The parameter for the network indicator was positive and significant ($\lambda_0 = 112.54$, t = 2.06, $p < .05$), suggesting that shifting to a networked marketplace was a revenue-boosting decision on the marketplace owner's part.[5] The effect of increasing marketplace size was also positive and significant ($\lambda_1 = .36$, t = 4.45, $p < .01$). Overall, these results indicate that, after we control for marketplace size, allowing sellers to network their shops permanently increased the mean daily commission revenues (i.e., the network effect can be interpreted as a small discontinuity or "jump").

Effects of Marketplace and Network Characteristics on Commission Revenues

Given this preliminary evidence indicating that the network's effect on commission revenues is positive, we next examined the influence of the marketplace's and the network's aggregate properties on daily commission revenues with the following model:

$$\text{(2)} \quad \text{Commission\_revenues}_t = \beta_0 + \sum_{l=1}^{p} \phi_l \times \text{Commission\_revenues}_{t-l} + \sum_{j=0}^{r} \Lambda_j \times \begin{bmatrix} \Delta\text{Marketplace\_size}_{t-j} \\ \Delta\text{Network\_links}_{t-j} \\ \Delta\text{Dead\_ends}_{t-j} \end{bmatrix} + \varepsilon_t.$$

We used all 781 days in the data set to estimate this model.

In this model, we examine more directly the effects of evolution in the marketplace's size and network structure on marketplace commission revenues. As previously, ΔMarketplace_size$_t$ is the number of new shops that join the marketplace on day t. In addition, ΔNetwork_links$_t$ and ΔDead_ ends$_t$ capture daily evolution of the network. If the network adds value at the marketplace level by making shops more accessible and by facilitating customer browsing, we should expect new links to have a positive effect on marketplace commission revenues because new links increase browsing opportunities. However, if the network's browsability is adversely affected by dead-end shops, we should expect an increase in the number of dead-end shops in the network to have a negative effect on marketplace commission revenues.

An alternative specification of these variables would be to use cumulative levels (e.g., cumulative number of links created by the end of day t for Network_links$_t$). Though conceptually plausible, cumulative versions of these variables pose problems because they are all nonstationary (even after we control for a time trend; all Dickey–Fuller tests ps > .95, whereas all Dickey–Fuller tests on daily/differenced series ps < .001). This suggests that differences are appropriate. The AR lags of the dependent variable contain information about the previous days' levels of each of the exogenous variables, thus ensuring that we account for these variables' histories. Note that we also found no evidence of cointegration for the daily series, making a VAR or ARX model in differences appropriate (instead of, for example, an error-correction model; see Dekimpe and Hanssens 2004). The ARX specification in Equation 2 (and not a VAR model) was confirmed by Granger-causality tests: (1) ΔMarketplace_size was not Granger-caused by Commission_revenues, ΔNetwork_links, or ΔDead_ends ($\chi^2[90] = 60.49$, $p = .99$); (2) ΔNetwork_ links was not Granger-caused by Commission_revenues, ΔMarketplace_size, or ΔDead_ends ($\chi^2[90] = 88.34$, $p = .53$); and (3) ΔDead_ends was not Granger-caused by Commission_revenues, ΔMarketplace_size, or ΔNetwork_links ($\chi^2[90] = 81.20$, $p = .74$).

The best-fitting model (i.e., with the lowest BIC) had $p = 7$ and r = 0 (i.e., the effects of the exogenous variables appear to be contemporaneous). The model has reasonable fit ($R^2 = .72$, BIC = 11.77, MAD = €49.73, and MAPE = 15.73%). A Wald test for the joint hypothesis that $\Lambda = 0$ (i.e., all exogenous variables' effects are zero) was significant ($\chi^2[3] = 27.44$, $p < .001$), indicating that the exogenous variables have an impact on commission revenues. A Wald test for the joint hypothesis that $\lambda_2 = \lambda_3 = 0$ (i.e., the network-related exogenous variables) was also significant ($\chi^2[2] = 6.77$, $p < .05$). We report the results in Table 2, and

Table 2

EFFECTS OF AGGREGATE MARKETPLACE AND NETWORK CHARACTERISTICS ON MARKETPLACE COMMISSION REVENUES

Variable	*Estimated Effect (t-Value) on Marketplace Commission Revenues*	
Intercept (β_0)	15.36	(.90)
Commission revenues$_{t-1}$ (ϕ_1)	.15	(4.34)***
Commission revenues$_{t-2}$ (ϕ_2)	.10	(2.70)**
Commission revenues$_{t-3}$ (ϕ_3)	.20	(5.61)***
Commission revenues$_{t-4}$ (ϕ_4)	.02	(.44)
Commission revenues$_{t-5}$ (ϕ_5)	.08	(2.10)*
Commission revenues$_{t-6}$ (ϕ_6)	.15	(4.31)***
Commission revenues$_{t-7}$ (ϕ_7)	.15	(4.18)***
Daily increase in marketplace size (λ_1)	.35	(4.31)***
Daily increase in number of "normal" network links (λ_2)	.25	(2.03)*
Daily increase in number of dead-end links (λ_3)	–3.10	(–2.54)*

*$p < .05$.
**$p < .01$.
***$p < .001$.
Notes: $R^2 = .72$, BIC = 11.77, MAD = €49.73, and MAPE = 15.73%.

[5]Alternatively, visitor traffic might have increased between the pre- and postnetwork birth periods. In the two months on either side of the network's birth, the marketplace received 80,000 unique visitors per day, which decreased over the seven months following the network's birth to 65,000 visitors per day. Thus, it cannot be that increased traffic caused the positive effect of the network's presence on revenues.

Figure 2 plots the actual and fitted daily commission revenues time series to illustrate the model's fit.

There is a positive and significant effect of growth in marketplace size (λ_1 = .35, t = 4.31, $p < .001$), meaning that daily marketplace revenues receive a boost from each new shop added. Likewise, growth in the number of links in the network has a positive and significant effect (λ_2 = .25, t = 2.03, $p < .05$), consistent with the idea that more connected networks tend to improve the overall accessibility of their members. As we also predicted, growth in the number of dead-end shops has a negative effect on marketplace performance (λ_3 = –3.10, t = –2.54, $p < .05$). All AR parameters (effects of lagged daily commission revenues) were also positive and significant (except for the fourth day lag).

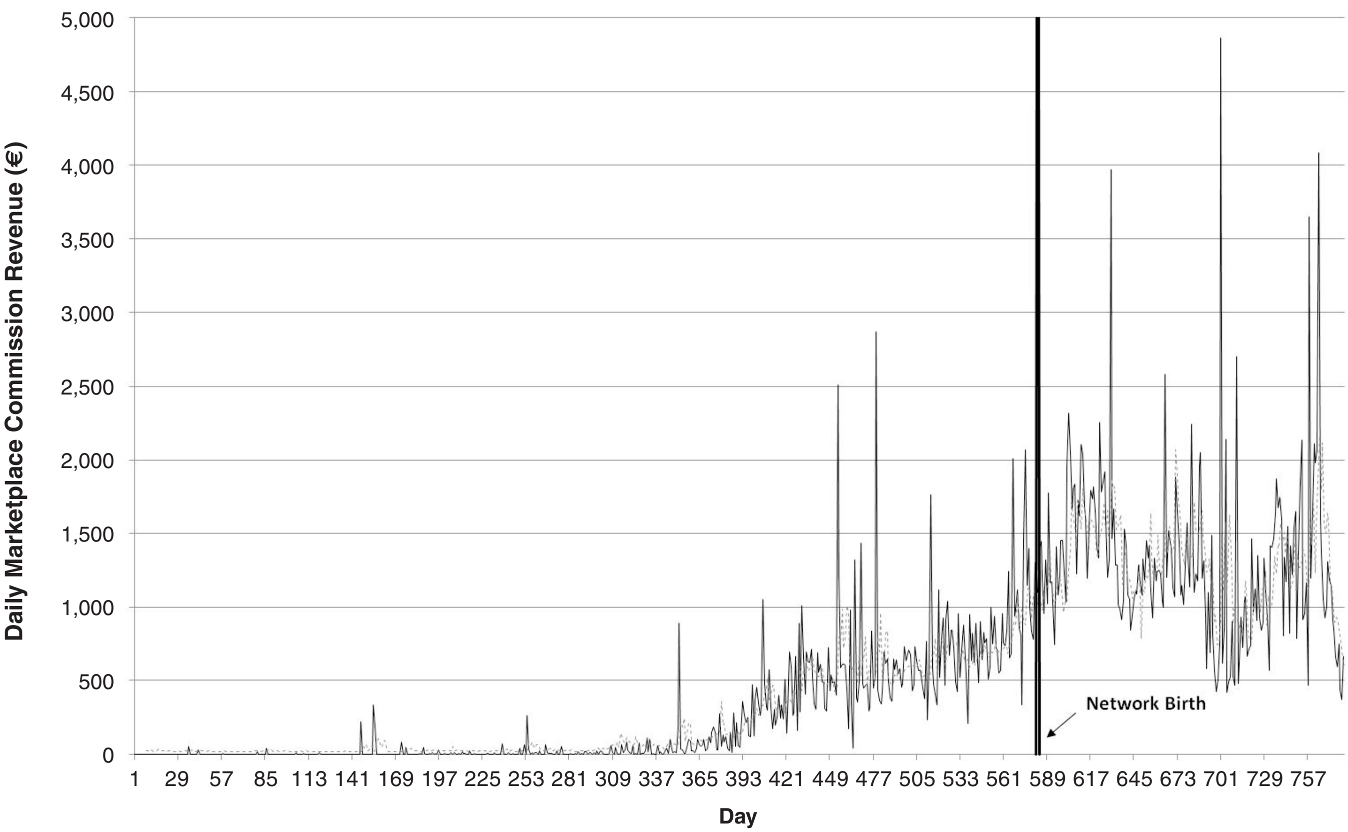
We used impulse response functions to quantify how much commission revenue is generated by a punctual increase of one unit in each of the exogenous variables and how this "shock" affects the system over time (for details, see the Web Appendix at http://www.marketingpower.com/jmrapril10). Adding an extra shop to the marketplace generates €2.22 of additional revenues. Adding an extra new link to the network generates €1.59 of additional revenues. Adding an extra dead end to the network, however, costs €19.54 in lost revenues to the marketplace owner. For each of these impacts, 95% of the total impact is realized by 70 days after the shock. The size of the dead-end effect is surprisingly large; for example, approximately nine new shops in the marketplace or 13 new links in the network would be needed to offset the cost of one new dead-end shop in the network. Notably, if we hold the number of dead ends constant, one new link has approximately 72% of the value of one new shop. Acquiring new sellers is likely to be more costly than encouraging existing sellers to create new links; thus, the value of a link compares favorably with the value of a shop.

Discussion of Marketplace Results

The time-series modeling in this section demonstrates the generally positive effect of the shop networking feature on aggregate marketplace performance: The presence of the network adds economic value, after we control for the growth of the marketplace itself. These marketplace-level findings also provide initial support for our hypothesis that this value is generated by making shops more accessible. A network structure that provides customers with opportunities to move from shop to shop by having a more connected network with few dead ends is valuable because it helps customers browse and find appealing shops and products before abandoning the marketplace. We now turn to a shop-level analysis of the marketplace to further investigate the role of accessibility and to assess how the value created by the network is distributed across its members.

Figure 2

DAILY MARKETPLACE COMMISSION REVENUES



Notes: The solid black line is the actual daily marketplace commission revenues, and the broken gray line is the fitted daily commission revenues using the model in Equation 3. The solid vertical line indicates the birth of the network. The model tracks daily marketplace commission revenues well. Marketplace commission revenue is the total commissions paid to the marketplace owner on day t for all sales that are made by shops in the marketplace on on day t.

SHOP-LEVEL ANALYSIS

Model

In this section, we examine the effect of the network at the level of the individual seller (shop). Our aims here are to further test the role of accessibility as a mechanism through which the network enhances economic value in this marketplace, to further explore similarities and differences between social commerce marketplaces and bricks-and-mortar shopping centers, and to address our question of how the value the network generates is distributed across its members. We measure Performance_i as shop i's total commissions earned during the last month of our data set (i.e., seventh month after the network's birth). We model performance as a function of network- and assortment-related variables. We measure all independent variables at the end of the second-to-last month of this period (i.e., sixth month after the network's birth).

We examine how a shop's position in the network relative to other shops at the end of the sixth month influences its performance in the seventh month. We use several network-related measures to describe a shop's network position relative to other shops. Specifically, we use various traditional measures of a shop's centrality in the network, computed on the basis of the state of the network at the end of the second-to-last (sixth) month of data. The measures included are mostly based on those outlined by Faust and Wasserman (1992), Freeman (1979), De Nooy, Mrvar, and Batagelj (2005), and Van den Bulte and Wuyts (2007) and come from sociology and graph theory. In addition to centrality measures (e.g., degree), we use other node-level measures that help describe a node's (shop's) position in the network relative to others. We define each measure in Table 1 (for an additional description, see the Web Appendix at http://www.marketingpower.com/jmrapril10).

We use two variables to describe a shop's product assortment (as control variables), also computed at the end of the sixth (second-to-last) month of the data set: (1) number of products listed by shop i and (2) average popularity of the products listed in shop i, based on how many other shops in the marketplace feature the same products.[6] In addition to these product assortment control variables, we control for shop i's age (the number of days between the time the shop was created in the marketplace and the last day of the sixth month of data), include in the model quadratic terms for in-degree and out-degree to allow nonlinear effects on performance, and allow for interactions between the network position variables and the number of products listed in a shop (shop assortment size).

However, a potential problem is that a shop's network position, its assortment, and its performance may all be influenced by a common latent variable. This endogeneity could be due to, for example, a seller's overall unobserved ability, which makes the seller better in the marketplace and helps the seller get into a better position in the network and have a better assortment (e.g., akin to a general seller-specific "skill" effect). This type of endogeneity is an important issue when modeling social network–related variables (see Hartmann et al. 2008) because many social network properties may be driven by either unobserved attributes or endogenous attributes that are related to network structure or network position (Handcock, Raftery, and Tantrum 2007; Wasserman and Faust 1994). We control for this using latent variables. We allow each network position and product assortment variable to be a function of a shop's latent ability. Our approach is similar to the data augmentation approach that Hui, Bradlow, and Fader (2007) use to model "category attractiveness" in their shopping path model for the movements of customers around grocery stores.

Our model is as follows:

$$\text{(3)} \qquad \text{NetworkPosition}_{i,j} = \gamma_{0,j} + \gamma_{1,j}\text{Ability}_i + \delta_{i,j},$$

$$\text{(4)} \qquad \text{ProductAssortment}_{i,k} = \alpha_{0,k} + \alpha_{1,k}\text{Ability}_i + \zeta_{i,k},$$

$$\text{(5)} \qquad \begin{aligned}\text{Performance}_i^* = {}& \beta_0 + \beta_1\text{Ability}_i + \beta_2\text{Age}_i + \sum_{j=1}^{J}\beta_{3,j}\delta_{i,j} + \sum_{k=1}^{K}\beta_{4,k}\zeta_{i,k} + \sum_{l=1}^{L}\beta_{5,l}\delta_{i,j}^2 \\ & + \sum_{j',k'}\beta_{6,j'k'}\delta_{i,j'}\zeta_{i,k'} + \varepsilon_i, \text{ and}\end{aligned}$$

$$\text{(6)} \qquad \text{Performance}_i = \begin{cases} 0 & \text{if Performance}_i^* = 0 \\ \text{Performance}_i^* & \text{if Performance}_i^* > 0, \end{cases}$$

where j indexes the network position measures (j = 1, …, J) taken at the end of the sixth month; k indexes the product assortment measures (k = 1, …, K) (in our case, J = 8 and K = 2) taken at the end of the sixth month; l indexes the quadratic terms (in our case, L = 2); j′ and k′ index the position and assortment size interaction components (in our case, j′ = 1, …, 8, and k′ = 1); γs are the network equation parameters; αs are the assortment equation parameters; $[\delta_i, \zeta_i] \sim N(\mathbf{0}, \mathbf{\Lambda})$ ($\mathbf{\Lambda}$ is unconstrained, allowing nonzero covariances between the residuals); βs are the performance equation parameters; ε is a random i.i.d. error with $\varepsilon_i \sim N(0, \sigma^2)$; and Performance_i is the observed performance (i.e., commission revenues) of shop i in the last (seventh) month. We use a tobit specification (Equation 6) because commission revenues in this marketplace cannot be negative.

In addition to influencing shop performance, our model allows the Ability latent variable to influence each shop's network measures and assortment characteristics. Directly entering the network and assortment variables into the performance Equation 5 would be inappropriate because it would give rise to biased and inconsistent network- and assortment-related estimates.[7] Instead, we use the residuals from the network (Equation 3) and assortment (Equation 4) equations (δ_i and ζ_i, respectively) as regressors in the performance Equation 5. These residuals are ability-adjusted network- and assortment-related variables. Our latent variable approach appears to be an appropriate and relatively straightforward technique for dealing with endogeneity issues in these types of models.

This specification may also help us deal with some potential collinearity between the network position measures that a common latent variable may induce. However, this

[6]We measured the popularity variable by first taking a count of the number of times each product is listed in shops in the marketplace (e.g., product X might appear in Shops A and B, thus giving it a count of 2). We then take the mean of this count within each shop and over all the shop's products as the measure of average popularity.

[7]This is correct only under the assumption that there actually is a genuine underlying causal variable at work (ability) and that the ability variable is not simply a latent factor common to the genuinely causal variables of network position and product assortment.

approach may not fully address collinearity between these measures because they are definitionally linked.[8] The largest correlation between unstandardized regressors is .60 among those based on incoming links, .55 among those based on outgoing links, and .92 among pairs of in- versus out-link versions of the same measure. The mean correlation is otherwise low (.24). The largest correlation was between authority and hub centrality (.92), both of which did not affect Performance_i. We reestimated the model without hub centrality, which yielded similar results (details are available on request). The next-largest correlation was between in-degree and out-degree (.80). Because both have large effects on commission revenues, it is not meaningful to remove them from the model. Instead, we decomposed in-degree and out-degree into unreciprocated in-degree (number of in-links that shop i has not reciprocated), unreciprocated out-degree (number of out-links that have not been reciprocated), and reciprocated degree (number of two-way links). The highest correlation between these variables was .56. We reestimated the model with these three variables, replacing in-degree and out-degree (but leaving all other regressors as previously specified), and found significant, positive unreciprocated in-degree (posterior mean = .46) and reciprocated degree (posterior mean = .26) effects, with the unreciprocated out-degree effect marginally significant (posterior mean = –.08).

We used a hierarchical Bayesian procedure to estimate the parameters in this model. Technical details are in the Web Appendix (http://www.marketingpower.com/jmrapril10). The prior on latent ability was $\text{Ability}_i \sim N(-1, \eta^2)$; diffuse priors were used for β_0, β_1, β_2, β_3, β_4, β_5, α_0, α_1, γ_0, and γ_1; the priors on σ, η, and Λ were $\sigma^2 \sim \text{InverseGamma}[(r_0/2)/(s_0/2)]$, where $r_0 = s_0 = 1$; $\eta^2 \sim \text{InverseGamma}[(r_0/2)/(s_0/2)]$, where $r_0 = s_0 = 1$; and $\Lambda \sim \text{InverseWishart}(n_0, n_0\Delta_0)$, $n_0 = J + K + 3$ and $\Delta_0 = I$.

All network position- and product assortment–related variables were standardized (M = 0, SD = 1) before we ran the model. We included all shops in the data set with at least one product at the end of the sixth month of the network's life (i.e., at the time the independent variables were computed) in the shop-level analysis. These criteria resulted in a set of 85,708 shops for estimating this model. We report means, standard deviations, and correlations for the unstandardized variables in Table 3. Estimation was based on 20,000 Markov chain Monte Carlo (MCMC) iterations with the first 10,000 as burn-in. All parameters mixed well, and convergence was fast.

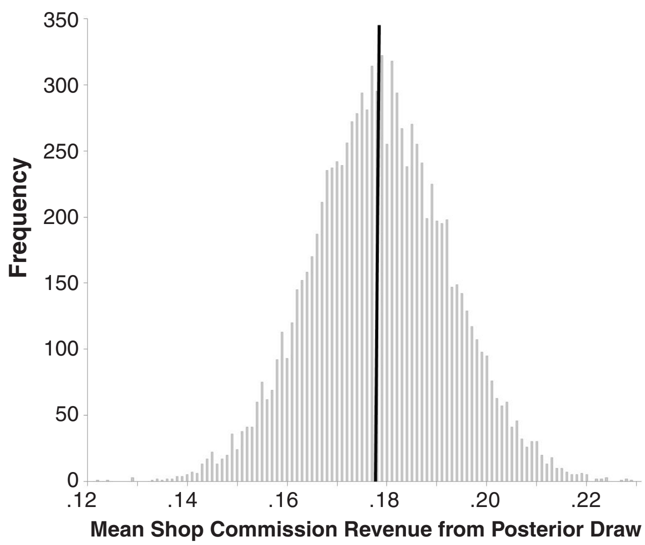
Results

Model fit and validation. We checked in-sample fit by comparing shops' actual commission revenues with those predicted by the model. For each of the 10,000 post-burn-in MCMC draws, we computed the predicted values for the shop performance-dependent variable for each shop in the data set. Fit can be assessed using posterior checks. In Figure 3, we plot the distribution across MCMC draws of the mean (across shops) predicted performance. The average across MCMC draws of the mean predicted performance is €.1782, extremely close to the actual value of €.1783. For each draw, we also computed the MAD (across shops) between actual and predicted performance. The mean MAD over all draws was reasonably small (€.056), keeping in

[8]However, the large number of observations used to estimate our model makes collinearity less of a problem (Mason and Perreault 1991).

Figure 3

POSTERIOR CHECKS: POSTERIOR DISTRIBUTION OF PREDICTED MEAN COMMISSION REVENUES



Notes: The histogram shows the posterior distribution of the fitted mean commission revenues, and the solid vertical line represents the actual value (€.1783).

Table 3

MEANS, STANDARD DEVIATIONS, AND CORRELATION COEFFICIENTS OF THE NETWORK POSITION MEASURES[a]

	M (SD)	*Correlations (Between Unstandardized Variables)*									
In-degree centrality	.76 (3.40)										
Out-degree centrality	.76 (4.41)	.80									
Incoming clustering coefficient	.04 (.15)	.39	.29								
Outgoing clustering coefficient	.04 (.16)	.40	.28	.74							
Authority (in-eigenvector centrality)	.0002 (.003)	.54	.51	.10	.07						
Hub (out-eigenvector centrality)	.0002 (.003)	.49	.54	.09	.06	.92					
In-proximity centrality	.01 (.03)	.60	.45	.48	.45	.21	.17				
Out-proximity centrality	.01 (.04)	.57	.55	.41	.40	.23	.22	.73			
Number of products	43.00 (102.0)	.08	.07	.05	.05	.01	.01	.12	.09		
Average product popularity	315.39 (673.9)	–.08	–.06	–.08	–.08	–.02	–.02	–.12	–.10	–.11	
Commission revenues	.18 (3.30)	.12	.06	.04	.06	.03	.02	.11	.07	.02	–.02

[a]All statistics are computed on the sample of shops used in the estimation of the shop-level model, not all shops.

mind that our dependent variable is a financial performance variable, which is typically difficult to predict with extremely high accuracy. The mean over the MCMC draws of the correlation between the shops' actual and predicted commissions was .22. We also performed a series of nested model comparisons, which provided support for the introduction of network position effects and latent ability into the model. Finally, we assessed out-of-sample fit and found it to be reasonable, albeit slightly worse than in-sample fit. Details of the model comparisons and out-of-sample fit are in the Web Appendix (http://www.marketingpower.com/jmrapril10).

We report parameter estimates in Table 4, Panels A–C. Note that only one of the interaction terms between network

Table 4

FULL LATENT VARIABLE TOBIT MODEL

A: Effects of Network Position, Product Assortment, and Latent Ability on Shop-Level Commission Revenues (Equation 5)

Parameter	*Posterior Mean*	*Posterior Standard Error*	*95% Credible Interval*
Intercept (β_0)	.175	.0067	–1.18, 1.34
Latent ability (β_1)	.015	.0067	–1.14, 1.37
Shop age (β_2)	.015	.0002	–.02, .05
Network Effects			
In-degree centrality ($\beta_{3,1}$)	.923***	.0006	.82, 1.05
Out-degree centrality ($\beta_{3,2}$)	–.503***	.0006	–.62, –.39
Incoming clustering coefficient ($\beta_{3,3}$)	–.155***	.0003	–.22, –.10
Outgoing clustering coefficient ($\beta_{3,4}$)	.049*	.0003	–.01, .11
Authority (in-eigenvector centrality) ($\beta_{3,5}$)	–.087	.0009	–.27, .09
Hub (out-eigenvector centrality) ($\beta_{3,6}$)	.039	.0009	–.14, .22
In-proximity (in-closeness) centrality ($\beta_{3,7}$)	.176***	.0004	.10, .24
Out-proximity (out-closeness) centrality($\beta_{3,8}$)	–.072**	.0003	–.14, –.01
Product Assortment Effects			
Number of products ($\beta_{4,1}$)	.012	.0002	–.02, .04
Average product popularity ($\beta_{4,2}$)	–.019	.0002	–.05, .01
Quadratic and Interaction Effects[a]			
In-degree^2 ($\beta_{5,1}$)	–.017***	.00002	–.02, –.01
Out-degree^2 ($\beta_{5,2}$)	.007***	.00002	.003, .01
Authority × number of products ($\beta_{6,5}$)	.094*	.0006	–.03, .20
In-proximity × number of products ($\beta_{6,7}$)	–.041**	.0002	–.09, –.002

B: Effects of Latent Ability on Network Position Measures (Equation 3)

Parameter	*Posterior Mean ($\times 10^{-3}$)*	*Posterior Standard Error*	*95% Credible Interval*
Network Effects: Intercepts $\gamma_{0,j}$			
In-degree centrality ($\gamma_{0,1}$)	–.491	.0003	–.07, .06
Out-degree centrality ($\gamma_{0,2}$)	–.422	.0003	–.07, .06
Incoming clustering coefficient ($\gamma_{0,3}$)	–.504	.0003	–.07, .06
Outgoing clustering coefficient ($\gamma_{0,4}$)	–.498	.0003	–.07, .07
Authority (in-eigenvector centrality) ($\gamma_{0,5}$)	–.493	.0003	–.07, .06
Hub (out-eigenvector centrality) ($\gamma_{0,6}$)	–.515	.0003	–.07, .07
In-proximity (in-closeness) centrality ($\gamma_{0,7}$)	–.437	.0003	–.06, .06
Out-proximity (out-closeness) centrality ($\gamma_{0,8}$)	–.614	.0003	–.07, .07
Network Effects: Slopes $\gamma_{1,j}$			
In-degree centrality ($\gamma_{1,1}$)	.505	.0003	–.06, .07
Out-degree centrality ($\gamma_{1,2}$)	.447	.0003	–.06, .07
Incoming clustering coefficient ($\gamma_{1,3}$)	.422	.0003	–.06, .06
Outgoing clustering coefficient ($\gamma_{1,4}$)	.496	.0003	–.06, .07
Authority (in-eigenvector centrality) ($\gamma_{1,5}$)	.524	.0003	–.07, .06
Hub (out-eigenvector centrality) ($\gamma_{1,6}$)	.565	.0003	–.06, .07
In-proximity (in-closeness) centrality ($\gamma_{1,7}$)	.404	.0003	–.06, .07
Out-proximity (out-closeness) centrality ($\gamma_{1,8}$)	.573	.0003	–.06, .07

C: Effects of Latent Ability on Product Assortment Measures (Equation 4)

Parameter	*Posterior Mean ($\times 10^{-3}$)*	*Posterior Standard Error*	*95% Credible Interval*
Product Assortment Effects: Intercepts $\alpha_{0,k}$			
Number of products ($\alpha_{0,1}$)	–.448	.0003	–.07, .06
Average product popularity ($\alpha_{0,2}$)	–.415	.0003	–.07, .06
Product Assortment Effects: Slopes $\alpha_{1,k}$			
Number of products ($\alpha_{1,1}$)	.505	.0003	–.06, .07
Average product popularity ($\alpha_{1,2}$)	.412	.0003	–.06, .06

*The 90% credible interval does not contain zero (two-sided).

**The 95% credible interval does not contain zero (two-sided).

***The 99% credible interval does not contain zero (two-sided).

[a]None of the other interaction effects are close to being significantly different from zero.

Notes: In Panel A, the error standard deviation (σ) has posterior mean = 3.226, standard error = .0004, and 95% credible interval (3.10, 3.27). All network position- and product assortment–related variables were standardized (M = 0, SD = 1) before we ran the model.

position measures and shop product assortment size was statistically significant. Therefore, we focus on the main effects and quadratic terms in our discussion and mention the significant interaction when appropriate.

Latent ability effects. The effect of a shop's latent ability on its performance was not significant, and the parameters (intercepts and slopes) in each of the network position and product assortment equations (Equations 3 and 4) were also not significant. Thus, the residuals (δ, ζ) in the performance equation (Equation 5) are similar to the original standardized independent variables. Note that the full latent variable model was favored over a simpler, nonlatent model based on the Bayes factor (see the Web Appendix at http://www.marketingpower.com/jmrapril10), which suggests that modeling latent ability is the more appropriate specification, despite the nonsignificance of these effects.

Degree centrality effects. A shop's position in this network affects its commission revenues. The largest effects on a shop's commissions were associated with degree centrality, that is, the number of ties going into and out of each shop. In-degree centrality had a positive effect on commissions, and out-degree centrality had a (smaller) negative effect. This suggests that shops with more links going into them from other shops and fewer links going out of them to other shops tend to perform better in terms of generating commission revenues for themselves. Both the in-degree and the out-degree effects are nonlinear given the significant quadratic terms (negative for in-degree, positive for out-degree). Thus, the positive (negative) impact of a new incoming (outgoing) link diminishes as the number of existing links increases.[9]

These results support our argument that the value of the network at the individual shop level lies in how the network makes shops more or less accessible to browsing customers. A higher in-degree means that a shop is more likely to be found by a browsing customer. Conversely, a shop with a higher out-degree makes it easier for customers to leave that shop, which hurts performance. The asymmetry between incoming and outcoming links highlights a key difference between bricks-and-mortar shopping centers and social commerce marketplaces: Such asymmetries are not possible in the former, because links in bricks-and-mortar shopping centers are undirected.

These effects also raise some game-theoretic issues because they imply that sellers have an incentive to try to attract others to connect to their shops but a disincentive to connect their shops to others' shops. This also raises the issue of how the marketplace owner could incentivize its sellers to create network links that facilitate browsing. Note that because the absolute effect size of in-degree centrality is greater than that of out-degree centrality, Shop A may still benefit from linking to Shop B as long as Shop B reciprocates this link (in which case Shop A would increase both its in-degree and its out-degree by 1, leading to a positive net effect). The posterior means for the standardized effects of in-degree and out-degree centrality are .923 and –.503, respectively. The corresponding unstandardized effects are .272 and –.114, and the proportion of links that were reciprocated in our data set is 61.7%.[10] If we suppose that when a shop creates an outgoing link to another shop there is a 61.7% chance of receiving a reciprocal incoming link, the expected net effect (unstandardized) on the performance of the shop that created the outgoing link is $-.114 + .617 \times .272 = .0538$. In other words, given the tendency for reciprocity in this network, creating links to other shops is not costly in expectation.

Proximity centrality effects. The in-proximity and out-proximity effects were significant and positive and negative, respectively. The positive effect of in-proximity means that shops that can be reached from a greater proportion of other shops in the network in fewer steps (i.e., with the customer needing to pass through fewer other shops) benefit more from the network than shops that are less easily reached. Thus, not only are direct incoming links important, but direct and indirect paths into a shop are also value relevant. This directly supports our accessibility argument. It also suggests that being closer to the start of potential browsing paths through the network is important. The comparatively weaker negative effect of out-proximity complements this and indicates that being positioned in the network such that many other shops can be reached from a shop in relatively few steps hurts a shop's performance. The in-proximity × number of products interaction was negative, suggesting that the positive effect of a shop's in-closeness centrality on its performance decreases as its assortment size becomes larger.

Clustering effects. A shop's incoming clustering coefficient had a negative effect on commissions, as we expected. It appears to be better for a shop not to be connected to by shops that are themselves highly interconnected. Shops that are connected to by shops that are themselves highly interconnected have relatively poor accessibility from other shops in the network that are not in their ego network. For example, if most of Shop A's incoming links come from a set of shops that are themselves highly interconnected, the chance that a customer entering the network at a randomly chosen shop in the network will browse Shop A is less than if Shop A's incoming links come from a more dispersed, less interconnected set of shops. This result contrasts with bricks-and-mortar shopping centers, for which clustering provides the benefit of reduced traveling costs. In general, stores in bricks-and-mortar shopping centers are at least moderately clustered, and this clustering or grouping does not appear to have negative effects in the offline context (Eppli and Benjamin 1994). Conversely, the effect of outgoing clustering was not strong (it was marginally significant and positive). Indeed, outgoing clustering has only a weak and indirect effect on the accessibility of other shops outside a shop's ego network.

Hub and authority effects. The prominence (eigenvector centrality) effects—hub and authority—were both non-

[9]The maxima for the quadratic in-degree effect is approximately 27 standard deviations above the mean in-degree (in-degree of approximately 93, which is less than the maximum observed in-degree of 184), and the minimum for the quadratic out-degree effect is approximately 36 standard deviations above the mean out-degree (out-degree of approximately 159, which is above the maximum observed out-degree of 100).

[10]To convert a standardized estimate into an unstandardized estimate (which is required for this comparison), the posterior mean can be divided by the unstandardized variable's standard deviation. The unstandardized standard deviations are 3.40 and 4.41 for in-degree and out-degree centrality, respectively.

significant, as we expected. The finding that being in a more prominent position in the network does not affect commission revenues confirms that accessibility, not prominence or "prestige," is the primary driver of how the value created by the network is distributed across shops. Note, however, that we found a weak positive interaction between authority and the number of products in a shop, indicating that larger shops may benefit from being linked into by so-called authority shops.

Discussion of Shop-Level Results

These results provide further support for the importance of networks in social commerce marketplaces and specifically highlight the critical role of the network in making shops more accessible to customers browsing the marketplace. The economic value of the network is distributed across shops according to how accessible they are made by the network. We find that shops that are more accessible from other shops in the network generally enjoy higher commission revenues, after controlling for potential product assortment, shop age, and latent ability effects.

Importantly, depending on how centrality is defined, it may help or hurt accessibility. The shops that benefit the most from the existence of the network are those with many incoming ties (positive effect of in-degree centrality), those with few outgoing ties (negative effect of out-degree centrality), those that are easily reachable from other shops (positive effect of in-proximity centrality), those that cannot easily reach other shops (negative effect of out-proximity centrality), and those that are connected to by shops that are not densely interconnected (negative effect of incoming clustering).

These findings are broadly consistent with discussions of shop accessibility in the offline shopping center literature, though the drivers of accessibility in this social commerce network differ substantially from bricks-and-mortar shopping centers. The directed nature of the links between shops in this marketplace creates asymmetries (between in-degree and out-degree, in-proximity centrality and out-proximity centrality, and incoming clustering and outgoing clustering) that do not exist in offline shopping centers. Moreover, clustering has a different effect in social commerce networks than in bricks-and-mortar shopping centers. Notably, the nonsignificant prominence and product assortment size effects suggest that it is not necessary to be prominently positioned in a social commerce network to benefit from the network and that having a larger shop is not necessarily helpful.

Finally, we considered an extended version of this shop-level model, adding a one-month lag of the dependent variable as an additional regressor (i.e., the commission revenues earned by each shop during the sixth month after the network's birth). The substantive results we reported in the previous sections and in Table 4, Panels A–C, were unchanged. Not surprisingly, the addition of the lagged dependent variable regressor improved the model's fit. For example, the mean MAD over all draws decreased (€.033 versus €.056), and the mean over the MCMC draws of the correlation between the shops' actual and predicted commissions increased (.61 versus .22). (Details are available on request.)

GENERAL DISCUSSION

Despite the rapid growth in online social networks and the recent emergence of online social commerce marketplaces in which opportunities for social interactions in online retailing and e-commerce contexts exist, little is known about social networks between sellers. The findings we report here represent a first step toward understanding the role of social networks in e-commerce and online retailing. Critically, this article shows how networks can help generate economic value for social commerce marketplace owners and for the individuals who participate in such marketplaces.

Overall, the findings suggest that social commerce networks between sellers can play an important economic, value-creating role. A key issue for shops (or individual sellers) in large, online marketplaces is simply being accessible to customers. Social networks between sellers act as "virtual shopping centers" by helping customers browse between shops, thus improving the accessibility of the network's shops. A more connected network tends to improve the overall accessibility of its members, especially if it is structured in a way that minimizes the number of dead ends from which customers cannot browse forward. The shops that benefit the most from the network are not necessarily those that are central to the network but rather those whose accessibility is most enhanced by the network. Network-based notions of centrality need to be carefully considered when examining the relationship between network position and performance outcome (or more generally, any node-level dependent variable) because different measures of a node's centrality can have opposite effects.

The marketplace- and shop-level results suggest some measures that social commerce firms (marketplace owners) and sellers (individual members of this marketplace) could take to enhance their performance. For example, sellers may want to improve their network position so that they receive more incoming links from shops that are dispersed (i.e., not locally clustered). Given the strong effect of reciprocity, this could involve a shop owner connecting to other shops that are many steps away (i.e., a path-shortening effect, similar to the cross-cutting paths discussed in the "small-world" network model; Watts and Strogatz 1998) instead of connecting to nearby shops. At the marketplace level, it may be possible for the marketplace owner to develop mechanisms to encourage sellers to create links (thus improving overall accessibility) while discouraging the creation of dead ends. Note that though these kinds of interventions are possible, we caution that "strategic" attempts to alter a social network's structure can lead to unintended consequences, given the inherent complexity of such systems. We leave more detailed considerations of these issues to further research.

These results also shed light on how network dynamics, which are driven by inherently social processes, influence economic value outcomes. Some of the drivers of network evolution that are typical in directed social networks may not be ideal for driving network-derived economic value. For example, although social networks naturally tend to evolve toward clustered groups (e.g., if $A \rightarrow B$ and $A \rightarrow C$, it is more likely than chance that the $B \rightarrow C$ link will form), we find that the clustering of shops hurts their performance in this marketplace. This is particularly relevant as online social networks that have relatively high levels of clustering (e.g., Facebook, MySpace) begin to introduce e-commerce "marketplace" features. Such social networks are possibly not structurally well suited to being networks of sellers. Although clustering is problematic for sellers, reciprocity appears to help because receiving an incoming link as a result of reciprocity tends to offset the "cost" of the corresponding outgoing link. Reciprocity is also a common driver of link formation in social networks. We encourage further research that explores the appropriateness of different types of network structures and the corresponding network evolution processes for facilitating commercial operations (see, e.g., Stephen and Toubia 2009).

This research is not without its limitations. First, the findings come from the study of a single online social commerce marketplace. Notwithstanding, the marketplace we studied is a pioneer in social commerce and is large and established. Further research might explore variations on this business model and consider other marketplaces with different retailing concepts (e.g., regular shops versus auction sellers). Second, although our shop-level model captures interdependence between shops through a set of node-level network position measures, further research might also capture interdependence through the error structure, based for example on the work by Hoff (2003) in statistics or on spatial econometrics models, or by examining dynamic spillover effects across shops. For example, research might explore how shops' commission revenues and product assortments influence the commission revenues earned by shops either that link to them or to whom they link. Third, we did not capture network dynamics in our shop-level model (Snijders 2006), in part for tractability reasons. We hope that further research will lead to the development of statistical models that are compatible with today's large network data sets and that make it possible to capture a wide range of effects, such as strategic behavior, interdependence between nodes, and time dynamics.

REFERENCES

Bronnenberg, Bart J. and Vijay Mahajan (2001), "Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables," *Marketing Science*, 20 (3), 284–99.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (August), 345–54.

De Nooy, Wouter, Andrej Mrvar, and Vladimir Batagelj (2005), *Exploratory Social Network Analysis with Pajek*. Cambridge, UK: Cambridge University Press.

Dekimpe, Marnik G. and Dominque M. Hanssens (2004), "Persistence Modeling for Assessing Marketing Strategy Performance," in *Assessing Marketing Strategy Performance*, Christine Moorman and Donald R. Lehmann, eds. Cambridge, MA: Marketing Science Institute, 69–94.

Eppli, Mark J. and John D. Benjamin (1994), "The Evolution of Shopping Center Research: A Review and Analysis," *Journal of Real Estate Research*, 9 (1), 5–32.

Faust, Katherine and Stanley Wasserman (1992), "Centrality and Prestige: A Review and Synthesis," *Journal of Quantitative Anthropology*, 4 (1), 23–78.

Freeman, Linton C. (1979), "Centrality in Social Networks: Conceptual Clarification," *Social Networks*, 1 (3), 215–39.

———, Douglas Roeder, and Robert R. Mulholland (1980), "Centrality in Social Networks: II. Experimental Results," *Social Networks*, 2 (2), 119–41.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545–60.

Goyal, Sanjeev (2007), *Connections: An Introduction to the Economics of Networks*. Princeton, NJ: Princeton University Press.

Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37 (3), 424–38.

Greif, Avner (2006), *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge, UK: Cambridge University Press.

Grewal, Rajdeep, Gary L. Lilien, and Girish Mallapragada (2006), "Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems," *Management Science*, 52 (7), 1043–1056.

Handcock, Mark S., Adrian E. Raftery, and Jeremy M. Tantrum (2007), "Model-Based Clustering for Social Networks," *Journal of the Royal Statistical Society: Series A*, 170 (2), 301–354.

Hanssens, Dominique M. (1980), "Market Response, Competitive Behavior, and Time Series Analysis," *Journal of Marketing Research*, 17 (November), 470–85.

———, Leonard J. Parsons, and Randall L. Schultz (2001), *Market Response Models: Econometric and Time Series Analysis*. Norwell, MA: Kluwer.

Hartmann, Wesley R., Puneet Manchanda, Harikesh Nair, Matthew Bothner, Peter Dodds, David Godes, et al. (2008), "Modeling Social Interactions: Identification, Empirical Methods and Policy Implications," *Marketing Letters*, 19 (3), 287–304.

Hoff, Peter (2003), "Random Effects Models for Network Data," working paper, Department of Statistics, University of Washington.

Hui, Sam, Eric Bradlow, and Peter Fader (2007), "An Integrated Model of Grocery Store Shopping Path and Purchase Behavior," working paper, The Wharton School, University of Pennsylvania.

Ingene, C.A. and A. Ghosh (1990), "Consumer and Producer Behavior in a Multipurpose Shopping Environment," *Geographical Analysis*, 22 (1), 70–91.

Iyengar, Raghuram, Thomas Valente, and Christophe Van den Bulte (2008), "Opinion Leadership and Social Contagion in New Product Diffusion," working paper, The Wharton School, University of Pennsylvania.

Katona, Zsolt and Miklos Sarvary (2008), "Network Formation and the Structure of the Commercial World Wide Web," *Marketing Science*, 27 (5), 764–78.

Lee, Ming-Long and R. Kelley Pace (2005), "Spatial Distribution of Retail Sales," *Journal of Real Estate Finance and Economics*, 31 (1), 53–69.

Marshall, Alfred ([1890] 1961), *Principles of Economics*. London: MacMillan.

Mason, Charlotte H. and William D. Perreault Jr. (1991), "Collinearity, Power, and Interpretation of Multiple Regression Analysis," *Journal of Marketing Research*, 28 (August), 268–80.

Nair, Harikesh, Puneet Manchanda, and Tulikaa Bhatia (2006), "Asymmetric Peer Effects in Prescription Behavior: The Role of Opinion Leaders," working paper, Graduate School of Business, Stanford University.

Nevin, J.R. and M.J. Houston (1980), "Image as a Component of Attraction to Intraurban Shopping Areas," *Journal of Retailing*, 56 (1), 77–93.

Palmer, Maija (2008), "Shoppers Find Goods Well Recommended," *Financial Times*, (January 20), (accessed January 18, 2010), [available at http://www.ft.com/cms/s/0/349ca1ba-c7c3-11dc-a0b4-0000779fd2ac.html?nclick_check=1].

Rindfleisch, Aric and Christine Moorman (2001), "The Acquisition and Utilization of Information in New Product Alliances: A Strength-of-Ties Perspective," *Journal of Marketing*, 65 (April), 1–18.

Snijders, Tom A.B. (2006), "Statistical Models for Network Dynamics," in *Proceedings of the XLIII Scientific Meeting, Italian Statistical Society*, S.R. Luchini, ed. Padova, Italy: CLEUP, 281–96.

Stephen, Andrew T. and Olivier Toubia (2009), "Explaining the Power-Law Degree Distribution in a Social Commerce Network," *Social Networks*, 31 (4), 262–70.

Stern, Louis W., Adel I. El-Ansary, and Anne T. Coughlan (1996), *Marketing Channels*. New York: Prentice Hall.

Tedeschi, Bob (2006), "Like Shopping? Social Networking? Try Social Shopping," *The New York Times*, (September 11), (accessed November 17, 2009), [available at http://www.nytimes.com/2006/09/11/technology/11ecom.html].

Trusov, Michael, Randolph E. Bucklin, and Koen Pauwels (2009), "Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing*, 73 (September), 90–102.

Tsai, Wenpin and Sumantra Ghoshal (1998), "Social Capital and Value Creation: The Role of Intrafirm Networks," *Academy of Management Journal*, 41 (4), 464–76.

Van den Bulte, Christophe and Yogesh Joshi (2007), "New Product Diffusion with Influentials and Imitators," *Marketing Science*, 26 (3), 400–421.

——— and Gary L. Lilien (2001), "Medical Innovation Revisited: Social Contagion Versus Marketing Effort," *American Journal of Sociology*, 106 (5), 1409–1435.

——— and Stefan Wuyts (2007), *Social Networks and Marketing*. Cambridge. MA: Marketing Science Institute.

Vázquez, Alexei (2003), "Growing Network with Local Rules: Preferential Attachment, Clustering Hierarchy, and Degree Correlations," *Physical Review E*, 67 (5), 056104.

Vitorino, Maria Ana (2008), "Empirical Entry Games with Complementarities: An Application to the Shopping Center Industry," working paper, The Wharton School, University of Pennsylvania.

Wasserman, Stanley and Katherine Faust (1994), *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.

Watts, Duncan J. and Peter Sheridan Dodds (2007), "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research*, 34 (December), 441–58.

——— and Steven H. Strogatz (1998), "Collective Dynamics of Small-World Networks," *Nature*, 393 (6684), 440–42.

Wuyts, Stefan, Stefan Stremersch, Christophe Van den Bulte, and Philip Hans Franses (2004), "Vertical Marketing Systems for Complex Products: A Triadic Perspective," *Journal of Marketing Research*, 41 (November), 479–87.

Yang, Sha and Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research*, 40 (August), 282–94.