1

Deriving Value from Social Commerce Networks

Andrew T. Stephen and Olivier Toubia*

Andrew T. Stephen

Doctoral Candidate in Marketing

Columbia University, Graduate School of Business

3022 Broadway, Uris 311, New York, NY 10027

Email ats2107@columbia.edu

Tel. (646) 709-4675, Fax. (734) 758-2012

Olivier Toubia

David W. Zalaznick Associate Professor of Business

Columbia University, Graduate School of Business

3022 Broadway, Uris 522, New York, NY 10027

Email ot2107@columbia.edu

Tel. (212) 854-8243, Fax. (212) 854-7647

November 6, 2008

Conditionally accepted at *Journal of Marketing Research*

Author Note

Deriving Value from Social Commerce Networks

ABSTRACT

Social commerce is an emerging trend in which *sellers* are connected in online social networks, and where sellers are individuals instead of firms. This paper 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: (i) Does allowing sellers to connect to one another create value (i.e., increase sales), (ii) what are the mechanisms through which this value is created, (iii) 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 it benefits or suffers from the network? We find that: (i) allowing sellers to connect generates considerable economic value; (ii) the network's value lies primarily in making shops more accessible to customers browsing the marketplace (the network creates a "virtual shopping mall"); and (iii) the sellers 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.

INTRODUCTION

Social commerce and social shopping communities are growing in number and in size. Broadly defined, social commerce and social shopping are forms of Internet-based "social media" that allow people to actively participate in the marketing and selling of products and services in online marketplaces and communities. One way to think of these applications is that they 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*. The roles played by consumers vary across websites or platforms, and can range from generating content (e.g., product reviews and recommendations, known as "consumer-generated media," on websites like 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* reported that Internet traffic to social commerce and social shopping websites grew by more than 500% between early 2007 and early 2008 (Palmer 2008), the *New York Times* reported that a number of social commerce firms are attracting substantial venture capital financing (Tedeschi 2006), and further growth and investment in this online retailing segment is expected.

Social shopping revolves around online word-of-mouth, and has recently been studied academically. For example, Chevalier and Mayzlin (2006), Godes and Mayzlin (2004) and Liu (2006) studied word-of-mouth and the influence of consumer-generated media on business outcomes, and Salganik, Dodds, and Watts (2006) and Watts and Dodds (2007) studied some 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 (Yang and Allenby 2003; Zhang and Netzer 2008), and other issues

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

Social commerce, on the other hand, is a more recent phenomenon and has not been studied as extensively. Social commerce marketplaces have four defining characteristics: (i) sellers (or shopkeepers) are individuals instead of firms, (ii) sellers create product assortments organized as personalized online shops, (iii) sellers' can create hyperlinks between their personalized shops, and (iv) 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) based on 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 individuals who participate as sellers in these marketplaces (by increasing sales). Issues related to connecting *sellers* have not been studied extensively (one exception is the shopping center literature that we review below). The value implications of networks have been studied in other contexts, such as inter- and intra-firm networks of a formal or an informal nature (e.g., Ingram and Roberts 2000; Rindfleisch and Moorman 2001; Tsai and Ghoshal 1998), 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; Granovetter 2005), and illegal collusive networks between firms have been studied in sociology (Baker and Faulkner 1993). However, very 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: (i) does allowing sellers to connect to one another create economic value (i.e., increase sales), (ii) what are the mechanisms through which this value is created, (iii) 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 it benefits or suffers from the network?

We address these questions using a novel dataset from an online marketplace which, after hosting a set of independent, consumer-generated online shops for about 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 website). Our dataset covers both a pre-network and a post-network period (therefore allowing 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 (therefore allowing 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). The paper is organized as follows. First, we review relevant literatures on shopping centers and social networks. Second, we describe our dataset. 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 future 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 individuals who participate as sellers in social commerce marketplaces and who earn commissions on sales that 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 (although 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, co-location, and spatial dependence between grouped shops (e.g., Eppli and Benjamin 1994; Eppli and Shilling 1996; Gerbich 1994; Lee and Pace 2005; Martineau 1958). Marketing researchers have also contributed to this literature (e.g., Anderson 1985; 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; Bronnenberg, Mahajan, and Vanhonacker 2000; Bronnenberg and Sismeiro 2002).

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 due to the presence of attractive "anchor" tenants, such as department stores,

supermarkets, or super-stores (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 multi-purpose or "one

stop" shopping (Eppli and Benjamin 1994). These effects are generally empirically well

supported (e.g., Anderson 1985; Brueckner 1993; Eppli and Benjamin 1994; Eppli and Shilling

1996; Martineau 1958; 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),

although any shop in a shopping center is more accessible than a standalone shop outside of the

center in most cases.[1]

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. Hence, 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 since the number of sales is simply equal to the number of visitors multiplied by

the shop's conversion rate.

However, while 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 a customer browses in a shopping center they do not visit (i.e., go into) every shop that 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 mentioned earlier, one 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.

As a result, while 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, networks that are more connected (i.e.,

with a larger number of links) generally tend to better 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 equally contribute to a network's browsability (this is related to the concept of network-based "search;" cf. Watts, Dodds, and Newman 2002). In particular, creating new links sometimes actually *hurts* the browsability of the network. For instance, in Figure 1 a simple network evolves from time 1 to time 2 with the addition of a new link. In 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) then 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).

[INSERT FIGURE 1 ABOUT HERE]

The accessibility of a website (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 indegree centrality (number of incoming links) should benefit more from the network compared to shops with lower indegree centrality because more incoming links equates to a higher chance of customer traffic in one's shop. We also expect that shops with higher incoming proximity (can be reached from more shops in fewer steps) will benefit more from the network compared to shops with lower incoming proximity, because shops that are accessible from fewer other shops and that 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 outdegree centrality (many outgoing links) or a higher outgoing proximity (can

lead to more shops in fewer steps), should themselves benefit less from the network. Also, shops that are connected to by shops that are themselves highly interconnected will be less accessible since the likelihood of browsing customers reaching such a shop will be low unless their browsing path starts in its ego-network. We therefore anticipate that shops with lower incoming clustering coefficients tend to benefit more from the network. The concepts of indegree centrality, outdegree centrality, incoming proximity, outgoing proximity, incoming clustering coefficients and outgoing clustering coefficients are defined formally in Table 1. (Note that hub centrality and authority centrality, which are also defined in Table 1, are addressed later in the paper.)

[INSERT TABLE 1 ABOUT HERE]

Of course, other features will influence a shop's likelihood of generating sales, which can generally be referred 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 of products, and uniqueness of its products vis-à-vis other shops in the marketplace) 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 who provide the products set prices), intensified competition between sellers cannot force them to lower prices.[3]

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 dataset

covers the French marketplace, which is the largest and was the first that this company launched. The company leverages online retailers' "affiliate" selling programs whereby websites 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 over 4 million products across many categories, with these products coming from over 100 vendor retailers such as Amazon.fr, Apple, Buy.com, Dell, and the Gap. Each seller has complete control over their shop's product assortment. Importantly, sellers are individual people, as opposed to 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 who 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 where Roger lists a range of books on Bayesian estimation. If Mark purchases a particular book, he is taken to the corresponding vendor's website (say Amazon), pays the vendor with his credit card, and a few days later receives the book shipped from (or through) Amazon. Since Mark purchased a book from Roger's shop, Amazon pays the marketplace a commission on that sale, and Roger in turn 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 who 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) prior to 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 homepage. 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 email). Therefore, this network is directed.

Our dataset includes the entire French population of shops that were created anytime between the first day and the 781st day of this marketplace's life (the last day of our dataset). 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 were growing quickly: an average of 180 new shops had been created in the marketplace each day, and once the network was born an average of 107 shops had joined the network each day, with an average of 421 new links created each day. Shops in this marketplace are generally 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; although 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 sales transactions and €388,970 in commission revenues (from vendors) had been generated by the end of the observation window.

We first analyze 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 modeling to examine the value created by the network, as well as the relation between this value and some aggregate characteristics of the network. We then analyze the data at the individual shop level in order 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

Modeling Approach

Our analysis of marketplace-level data uses autoregression with exogenous variables (ARX) time series models (these models are also called autoregressive distributed lag [ARDL] models in the econometrics literature; Greene 2003). This type of model, as well as its multivariate version (vector autoregression with exogenous variables or VARX) is increasingly popular in marketing dynamics research (e.g., Pauwels et al. 2004; Trusov, Bucklin, and Pauwels 2008), and is appropriate here because it allows for an endogenous series (daily marketplace-level commission revenues; i.e., how much the marketplace receives from vendors each day for sales that were made that day) to be modeled as a function of p-period lags of itself (i.e., the pth-order vector autoregression component) and as a function of m covariates or exogenous variables, which can also be lagged (up to r periods). The covariates we use dynamically track certain aggregate characteristics of the marketplace and the network within it. Specific variables are described below. Note that an alternative specification would be a VAR or VARX model (with the difference being that VARX includes exogenous regressors whereas VAR is a fully endogenous dynamic system). Whether this is appropriate is ultimately an empirical question,

and is the subject of tests which we describe below. For now, consider the generic ARX/ARDL model:

$$y_t = \beta_0 + \sum_{l=1}^{p} \phi_l \cdot y_{t-l} + \sum_{j=0}^{r} \Lambda_j \cdot \mathbf{x}_{t-j} + \varepsilon_t \qquad (1)$$

Where y_t is a dependent (endogenous) variable at period t, β_0 is an intercept, p is the number of autoregressive lags on the endogenous variable (selected using an appropriate information criterion such as the Bayesian Information Criterion [BIC]), ϕ_l is the effect of the l-period lagged endogenous variable on its current value, Λ_j is an m-dimensional row vector with the effects of the j-period lagged exogenous variables (m is the number of exogenous variables in the model), $\mathbf{x}_{t-j}$ is an m-dimensional column vector of the j-period lagged exogenous variables, r is the number of lags on the exogenous variables (also selected using information criteria), and ε_t is a random disturbance. Ordinary least squares or maximum likelihood estimation techniques are used to estimate these models. This specification is appropriate provided that Granger causality tests confirm that the variables in $\mathbf{x}$ are indeed exogenous (i.e., not caused by lags of the dependent variable y) (Granger 1969; Greene 2003; Hanssens et al. 2001; Trusov et al. 2008). If one or more variables in $\mathbf{x}$ are found to not be exogenous then a VAR specification is needed.

Variables

We now introduce the variables that will be the focus of our time series analysis. Data are available on each of these variables for each of the 781 days in our dataset, covering both pre- and post-network birth periods.

- *Commission_revenues$_t$*: the commissions in € 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;

- *Marketplace_size$_t$*: 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$_t$*: total number of links in the shop network on day t;

- *Dead_ends$_t$*: the total number of dead-end shops in the shop network on day t. See Figure 1.

Results

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, as 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:

$$Commission\_revenues_t = \beta_0 + \sum_{l=1}^{p} \phi_l \cdot Commission\_revenues_{t-l} \qquad (2)$$
$$+ \lambda_0 \cdot Network_t + \lambda_1 \cdot \Delta Marketplace\_size_t + \varepsilon_t$$

The best-fitting model (i.e., with the lowest BIC) had autoregressive lag of $p = 6$.[4] *Commission_revenues$_t$*, which we defined in its *daily* (differenced) form, was found to be stable and not evolving using a Dickey-Fuller unit root test (with a null hypothesis of non-stationarity; $p < .001$). *Marketplace_size$_t$*, defined above in its total or cumulative form was found non-stationary based on a 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 *Marketplace_size$_t$* (i.e., the daily change) and not the cumulative level in our model

(Dekimpe and Hanssens 2004). 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 VAR model was supported by a series of Granger causality tests (Granger 1969; Hanssens et al. 2001; Trusov et al. 2008) which confirmed the exogeneity of $\Delta Marketplace\_size_t$ and $Network_t$. Since 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 be more sure that the results apply at any lag and not just the best-fitting lag for the model (Hanssens 1980; Trusov et al. 2008). $\Delta Marketplace\_size$ 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 $\Delta Marketplace\_size$ 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, 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 controlling 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:

$$Commission\_revenues_t = \beta_0 + \sum_{l=1}^{p} \phi_l \cdot Commission\_revenues_{t-l} + \sum_{j=0}^{r} \Lambda_j \cdot \begin{bmatrix} \Delta Marketplace\_size_{t-j} \\ \Delta Network\_links_{t-j} \\ \Delta Dead\_ends_{t-j} \end{bmatrix} + \varepsilon_t \quad (3)$$

All 781 days in the dataset were used to estimate this model (which makes for a more conservative test of the network's effects given that the corresponding exogenous variables do not vary until day 583).

In this model we more directly examine the effects of evolution in the marketplace's size and network structure on marketplace commission revenues. As before, $\Delta Marketplace\_size_t$ is the number of new shops to join the marketplace on day t. $\Delta Network\_links_t$ and $\Delta Dead\_ends_t$ 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 since 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.

Per our above discussion about modeling time series in levels or differences, an alternative specification of these variables would of course be to use cumulative levels (e.g., cumulative number of links created by the end of day t for $Network\_links_t$). Although conceptually plausible, cumulative versions of these variables pose problems because they are all non-stationary (even after controlling 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 autoregressive lags of the dependent variable contain information about previous days' levels of each of the exogenous variables, thus ensuring that these variables'

histories are accounted for. 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; cf. Dekimpe and Hanssens 2004). The ARX specification in (3) (and not a VAR model) was confirmed by Granger causality tests: (1) $\Delta Marketplace\_size$ was not Granger caused by *Commission_revenues*, $\Delta Network\_links$, or $\Delta Dead\_ends$ ($\chi^2[90] = 60.49$, $p = .99$), (2) $\Delta Network\_links$ was not Granger caused by *Commission_revenues*, $\Delta Marketplace\_size$, or $\Delta Dead\_ends$ ($\chi^2[90] = 88.34$, $p = .53$), and (3) $\Delta Dead\_ends$ was not Granger caused by *Commission_revenues*, $\Delta Marketplace\_size$, or $\Delta Network\_links$ ($\chi^2[90] = 81.20$, $p = .74$).

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, 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 ($\chi^2[2] = 6.77$, $p < .05$). Results are reported in Table 2, and Figure 2 plots the actual and fitted daily commission revenues time series to illustrate the model's fit.

[INSERT TABLE 2 AND FIGURE 2 ABOUT HERE]

There is a positive and significant effect of growth in marketplace size ($\lambda_1 = .35$, $t = 4.31$, $p < .001$), 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 ($\lambda_2 = .25$, $t = 2.03$, $p < .05$), consistent with the idea that more connected networks tend to improve the overall accessibility of their members. Also as predicted, growth in the number of dead-end shops has a negative effect on marketplace performance ($\lambda_3 = -3.10$, $t = -2.54$, $p < .05$). All

autoregressive parameters (effects of lagged daily commission revenues) were also positive and significant (except for the fourth day lag).

Long-Run Impacts. Persistence modeling can be used to examine the long-run effects of the exogenous variables in (3). This approach estimates how a "shock" in an "impulse" variable (e.g., adding a new link) affects a "response" variable over time (e.g., daily commission revenue). This approach has been used in the marketing literature to study the long-run impact of changes in marketing variables on performance variables such as sales (e.g., Dekimpe and Hanssens 2004, 2005) or service adoptions (e.g., Trusov et al. 2008).

Specifically, we use impulse-response functions (IRFs).[6] An IRF simulates the impact over time of a change in one variable (in our case one of the exogenous variables) on the full dynamic system (Bronnenberg, Mahajan, and Vanhonacker 2000). Here the IRFs are based on the estimated coefficients from the ARX model (3) above. The cumulative IRFs (sometimes referred to as total short-run effects) are plotted in Figure 3. These plots show how much commission revenue is generated by a one unit increase in each of the exogenous variables and how long the "shock" lasts in the system (the effect propagates through the lags of the dependent variable). 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. Of course, these values are based on the marketplace during its first 781 days and may not apply when it is more mature. Nevertheless, the cumulative IRFs indicate that growth in the marketplace *and* in the general accessibility of shops in the network have somewhat persistent revenue-enhancing effects.

[INSERT FIGURE 3 ABOUT HERE]

Moreover, these results illustrate the role of the network links in generating value by facilitating browsing, particularly showing that if links *hurt* the network's browsability by sending customers to dead-ends then such links are very costly to the marketplace owner. 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 a net gain of one new dead-end shop in the network. Interestingly, holding 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, hence 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 network adds economic value, after controlling 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. Having 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 as 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, in order to further investigate the role of accessibility, and to assess how the value created by the network is distributed across its members.

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 generated by the network is distributed across its members. We measure *Performance$_i$* as shop *i*'s total commissions earned during the seventh (last) month of our dataset. Performance is modeled as a function of network- and assortment-related variables. All independent variables are measured at the end of the sixth (second-to-last) month of this period.

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 based on 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), Scott (2000), and Van den Bulte and Wuyts (2007), and come from sociology and graph theory. In addition to centrality measures we use other node-level measures that help to describe a node's (shop's) position in the network relative to others. Each measure is defined in Table 1:

- *Indegree centrality*: number of incoming links received by a shop from other shops;

- *Outdegree centrality*: number of outgoing links from a shop to other shops;

- *Incoming proximity* (Faust and Wasserman 1992; Lin 1976; de Nooy et al. 2005): this measures the reachability of shop *i* from other shops in the network. 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 (shop *i*'s "indomain"), and inversely proportional to the mean geodesic distance (shortest path length) from these shops to *i*. Thus, incoming proximity is highest for shops that are accessible from a large number of shops in the network in only few steps. Note that this is a directed graph analog of Freeman's (1979) standard "closeness" metric for undirected graphs;

- *Outgoing proximity* (Faust and Wasserman 1992; Lin 1976; de Nooy et al. 2005): this measures the reachability of other shops in the network from shop *i*. 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 (shop *i*'s "outdomain"), and inversely proportional to the mean geodesic distance (shortest path length) from shop *i* to these shops. Thus, outcoming proximity is highest for shops from which a large number of shops in the network are accessible in only few steps.

- *Incoming clustering coefficient* (Watts and Strogatz 1998; Zhou 2002): the incoming clustering coefficient of shop *i* is the degree of interconnectedness among the shops that link to shop *i*. Specifically, it is the proportion of possible links that exist among the shops in shop *i*'s incoming ego-network. The higher a shop's incoming clustering coefficient, the more densely interconnected its incoming ego-network is, i.e., that shop is connected to by other shops that are themselves highly interconnected (as opposed to being connected to by shops that are more dispersed throughout the network). Note that

clustering is not a centrality measure, but rather a measure of how dense each shop's ego-network is;

- *Outgoing clustering coefficient* (Watts and Strogatz 1998; Zhou 2002): the outgoing clustering coefficient of shop i is the degree of interconnectedness among the shops that shop i links to, and thus is the proportion of possible links that exist among the shops in shop i's outgoing ego-network. The higher a shop's outgoing clustering coefficient, the more densely interconnected its outgoing ego-network is, i.e., that shop connects to other shops that are themselves highly interconnected;

- *Hub centrality* and *authority centrality* (Kleinberg 1999): these measures of centrality are directed graph analogs of eigenvector centrality (Bonacich 1987) for outgoing links (hub) and incoming links (authority). The basic concept of eigenvector centrality is that a shop is more prominent in the network if it is well-connected to other well-connected shops. A shop with a high hub score links to many shops with high authority scores, and a shop with a high authority score is linked to by many shops with high hub scores (these metrics are based on eigenvector decompositions of the network's adjacency matrix; see Kleinberg 1999 for derivations). Because these centrality measures are not directly related to accessibility, we do not expect them to have a significant impact on shop performance.

We use the following variables to describe a shop's product assortment, also computed at the end of the sixth (second-to-last) month of the dataset:

- *Number of products listed* by shop i;

- *Average popularity* of the products listed in shop i, based on how many other shops in the marketplace feature the same products.[7]

In addition to these product assortment control variables, we also 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 indegree and outdegree 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).

A potential problem, however, 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 instance, a seller's overall unobserved ability, which makes them a better seller in the marketplace as well as helps them 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 (cf. 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 used by Hui, Bradlow, and Fader (2007) to model "category attractiveness" in their shopping path model for the movements of customers around grocery stores.

Our model is as follows:

$$Network\,Position_{i,j} = \gamma_{0,j} + \gamma_{1,j}Ability_i + \delta_{i,j} \tag{4}$$

$$Product\,Assortment_{i,k} = \alpha_{0,k} + \alpha_{1,k}Ability_i + \zeta_{i,k} \tag{5}$$

$$Performance_i^* =$$

$$\beta_0 + \beta_1 Ability_i + \beta_2 Age_i + \sum_{j=1}^{J}\beta_{3,j}\delta_{i,j} + \sum_{k=1}^{K}\beta_{4,k}\zeta_{i,k} + \sum_{l=1}^{L}\beta_{5,l}\delta_{i,j}^2 + \sum_{j',k'}\beta_{6,j'k'}\delta_{i,j'}\zeta_{i,k'} + \varepsilon_i \tag{6}$$

$$Performance_i = \begin{cases} 0 & \text{if } Performance_i^* = 0 \\ Performance_i^* & \text{if } Performance_i^* > 0 \end{cases} \tag{7}$$

Where j indexes the network position measures ($j = 1, \ldots, J$) taken at the end of the sixth month,

k indexes the product assortment measures ($k = 1, \ldots, K$) (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 $j' = 1, \ldots, 8$ and $k' = 1$), the γ's

are the network equation parameters, the α's 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), the

β's 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 (7) 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 (6) would be inappropriate as it

would give rise to biased and inconsistent network- and assortment-related estimates.[8] Instead,

the residuals from the network (4) and assortment (5) equations (δ_i and ζ_i, respectively) are used

as regressors in the performance equation (6). 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 may be induced by a common latent variable.[9]

A hierarchical Bayesian procedure was used to estimate the parameters in this model. Technical details are provided in the appendix. The prior on latent ability was

$Ability_i \sim N(-1, \eta^2)$; diffuse priors were used for β_0, β_1, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$, $\boldsymbol{\beta}_4$, $\boldsymbol{\beta}_5$, $\boldsymbol{\alpha}_0$, $\boldsymbol{\alpha}_1$, $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$; the

priors on σ, η, and $\boldsymbol{\Lambda}$ were $\sigma^2 \sim InverseGamma\left(\frac{r_0}{2}, \frac{s_0}{2}\right)$, where $r_0 = s_0 = 1$,

$\eta^2 \sim InverseGamma\left(\frac{r_0}{2}, \frac{s_0}{2}\right)$, where $r_0 = s_0 = 1$, and $\boldsymbol{\Lambda} \sim InverseWishart(n_0, n_0\boldsymbol{\Lambda}_0)$, $n_0 = J + K + $

3, and $\boldsymbol{\Lambda}_0 = \mathbf{I}$.[10]

All network position- and product assortment-related variables were standardized (mean = 0, standard deviation = 1) before running the model. All shops in the dataset 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) were included in the shop-level analysis. These criteria resulted in a set of 85,708 shops for estimating this model. Means, standard deviations, and correlations for the unstandardized variables are reported in Table 3. Estimation was based on 20,000 MCMC iterations with the first 10,000 as burn-in. All parameters mixed well and convergence was fast.

[INSERT TABLE 3 ABOUT HERE]

Results

Model Fit and Validation. First, in-sample fit was checked 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 dataset. Fit may be assessed using posterior checks (Gelman, Meng, and Stern 1996). In Figure 4 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 €0.1782, extremely close to the actual value of €0.1783. For each draw we also computed the

median (across shops) of the absolute deviation (MAD) between actual and predicted

performance. The mean MAD over all draws was reasonably small (€0.056), keeping in mind

that our dependent variable is a financial performance variable, which are 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.

[INSERT FIGURE 4 ABOUT HERE]

Second, we compared the full model in (4) to (7) to a nested model without network

position effects; i.e., in (6) restricting $\beta_3 = \beta_5 = \beta_6 = 0$. Following Newton and Raftery (1994) and

Rossi, Allenby, and McCulloch (2005) we computed the log marginal densities for the two

models using the harmonic means of the respective models' likelihoods across posterior draws

(every tenth draw after burn-in, for computational reasons). The full model had a better fit

$(-2 \log \hat{p}(y \mid M_{product-only}) = 2.334 \times 10^{-5}$ versus $-2 \log \hat{p}(y \mid M_{full}) = 2.308 \times 10^{-5}$; smaller is

better), and a large log Bayes factor based on these log marginal densities ($\log BF_{full\,vs\,product-only} =$

1289; note that $\log BF_{AvsB} > 5$ is "strong evidence" in favoring model A over model B), which

provided very strong evidence in favor of the full model over the product-only restricted model.

The mean correlation between actual and predicted commissions (over MCMC draws) for this

restricted model was clearly inferior to the full model (.07 versus .22).

Third, we performed another nested model comparison, this time comparing the full

model in (4) to (7) to a restricted model with $\alpha = 0$, $\gamma = 0$ and $\beta_1 = 0$, which was a simpler Tobit

model without the latent ability variable and entering the network position and product

assortment variables directly as regressors instead of their ability-adjusted residuals. This simpler

model's fit and log Bayes factor were worse ($-2 \log \hat{p}(y \mid M_{simple}) = 2.311 \times 10^{-5}$ versus

$-2 \log \hat{p}(y \mid M_{full}) = 2.308 \times 10^{-5}$; $\log BF_{full\,vs\,simple} = 139$), and the correlation between actual and

predicted commissions was also worse than for the full model (.15 versus .22). We therefore base our findings on the full model with the latent "ability" specification. Parameter estimates for the simpler Tobit model are reported in Table 5 for sake of comparison and as a robustness check. All effects that we report as significant in the full model below are confirmed by this simpler specification, and the magnitudes of the effects are very comparable. A few effects, such as that of outgoing clustering coefficient and authority, are not significant in the full model but are significant in the simpler model. Since the former fits better than the latter we do not focus heavily on these points of difference. Indeed, basing our substantive findings on the full model with fewer significant effects is conservative.

Fourth, out-of-sample fit was checked. The common approach of randomly splitting the data into an estimation sample and a hold-out validation sample is inappropriate here since our latent ability variable needs to be estimated for each shop, and predicting shop performance in a validation sample would require estimating latent ability for each shop in that sample, which would require using the validation data for estimation. Instead, we re-estimated the model with the network and product assortment variables measured at the end of month 5 (instead of the end of month 6) and the commission revenues measured during month 6 (instead of during month 7; i.e., "month 5/6" data instead of "month 6/7" data). As an indicator of robustness we found no qualitative differences between the results reported below from analysis of the month 6/7 data and the results from this month 5/6 data (details available from the authors). We then assessed out-of-sample fit by using these month 5/6 parameter estimates (including shop-level latent ability estimates) to predict the commission revenues earned during month 7 as a function of the network and assortment variables at the end of month 6 (i.e., using the month 5/6 data for calibration and the month 6/7 data for validation). The average predicted month 7 commission

(mean taken across MCMC draws of the parameter estimates) was €0.1947, close to the actual mean month 7 commission of €0.1783 (and the actual mean lied well within the distribution of the predicted means from the MCMC draws). The MAD across draws was €0.077, which was slightly larger than for in-sample fit but still reasonable, and the mean correlation between actual and predicted commissions was .21.

Parameter estimates are reported in Tables 4A, 4B and 4C (full model) and Table 5 (simpler model without latent ability). Note that only one of the interaction terms between the network 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 where appropriate.

[INSERT TABLES 4A, 4B, 4C AND 5 ABOUT HERE]

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 (4, 5) were not significant either. Hence, the residuals (δ, ζ) in the performance equation (6) are very similar to the original standardized independent variables. Note that the fact that the full latent variable model was favored over the simpler non-latent model based on the Bayes factor suggests that modeling latent ability is the more appropriate specification, despite the non-significance of these effects.

Degree Centrality Effects. A shop's position in this network affects its commission revenues. The largest effects on shop's commissions were associated with degree centrality, that is, the number of ties going into and out of each shop. Indegree centrality had a positive effect on commissions, and outdegree 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 indegree and outdegree effects are nonlinear given the significant quadratic terms (negative for indegree, positive for outdegree). Thus, the positive (negative) impact of a new incoming (outcoming) link diminishes as the number of existing links increases.[11]

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 indegree means that a shop is more likely to be found by a browsing customer. A shop with a higher outdegree, on the other hand, makes it easier for customers to leave that shop, which hurts performance. The asymmetry between incoming and outcoming links highlights one key difference between bricks-and-mortar shopping centers and social commerce marketplaces: such asymmetries are not possible in the former since links in bricks-and-mortar shopping centers are undirected.

These effects also raise some interesting game theoretic issues because they imply that sellers have an incentive to try to attract others to connect to their shops, but have a disincentive to connect their shops to others' shops. This also raises the interesting 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 indegree centrality is greater than that of outdegree 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 indegree and outdegree by 1, leading to a positive net effect). The posterior means for the standardized effects of indegree and outdegree centrality are .923 and -.503, respectively. The corresponding unstandardized effects are .272 and -.114,[12] and the proportion of links that were reciprocated in our dataset is 61.7%. Supposing 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 × .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 inproximity and outproximity effects were significant, and respectively positive and negative. The positive effect of inproximity 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 having to pass through fewer other shops) benefit more from the network compared to shops that are less easily reached. It also shows that not only direct incoming links are important; rather, direct and indirect paths into a shop are 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 outcloseness complements this, and indicates that being positioned in the network such that many other shops can be reached from one's shop in relatively few steps hurts a shop's performance. The inproximity × number of products interaction was negative, suggesting that the positive effect of a shop's incloseness 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 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 instance, if most of shop A's incoming links come from a set of shops that are themselves highly interconnected then the chance that a customer entering the network at a randomly chosen shop in the network will browse to shop A is smaller 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, where clustering provides the benefit of reduced customers' traveling costs. Stores in bricks-and-mortar shopping centers are generally at least moderately clustered, and this clustering or grouping does not appear to have negative effects in the offline context (Eppli and Benjamin 1994). The effect of outgoing clustering, on the other hand, was not strong (it was marginally significant and positive). Indeed, outgoing clustering only has a weak and indirect effect on the accessibility of *other* shops outside of a shop's ego-network.

Hub and Authority Effects. The prominence (eigenvector centrality) effects—hub and authority—were both non-significant, as 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 that we did however find 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 played by 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. 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, we find that 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 indegree centrality), few outgoing ties (negative effect of outdegree centrality), are easily reachable from other shops (positive effect of inproximity centrality), cannot easily reach other shops (negative effect of outproximity centrality), and are connected to by shops that are not densely interconnected (negative effect of incoming clustering coefficient).

Our findings are broadly consistent with discussions of shop accessibility in the offline shopping center literature, although 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 indegree and outdegree, inproximity centrality and outproximity 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 compared to bricks-and-mortar shopping centers.

Finally, the non-significance of the prominence effects and the product assortment size effects suggest that it is not necessary to be prominently positioned in a social commerce network in order to benefit from the network, and that having a larger shop is not necessarily helpful.

GENERAL DISCUSSION

Despite the rapid growth in online social networks, and the recent emergence of online social commerce marketplaces where opportunities for social interactions in online retailing and e-commerce contexts are provided, extremely little is known about social networks between sellers. The findings reported here represent a first step in understanding the role that social

networks play in e-commerce and online retailing. Critically, this paper shows how networks can help to generate economic value for social commerce marketplace owners and for the individuals who participate in such marketplaces.

Overall, our 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, therefore 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" that customers cannot browse away from. 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 instance, sellers may wish 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 connecting to shops that are many steps away from one's shop (i.e., a path-shortening effect, similar to the cross-cutting paths that are 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 while 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 future 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, while social networks naturally tend to evolve towards clustered groups (e.g., if A→B and A→C then it is more likely than chance that the B→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.com, MySpace.com) start introducing 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, on the other hand, appears to help here as receiving an incoming link as a result of reciprocity appears 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.

This research is not without its limitations. First, our 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. Future research might explore

variations on this business model, and consider other marketplaces with different retailing concepts (e.g., regular shops versus auction sellers). Second, while our shop-level model captures interdependence between shops through a set of node-level network position measures, future research may also capture interdependence through the error structure, based for example on the work by Hoff (2003) in statistics and the spatial econometrics literature (e.g., Anselin 2006), or by examining dynamic spillover effects across shops.[13] For example, future research may explore how shops' commission revenues and product assortments influence the commission revenues earned by shops that either 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 future research will lead to the development of statistical models that are compatible with today's large network datasets, and that allow capturing a wide range of effects such as strategic behavior, interdependence between nodes, and time dynamics.

REFERENCES

Anderson, P. M. (1985), "Association of Shopping Center Anchors with Performance of a Non-anchor Specialty Chain's Stores," *Journal of Retailing*, 61, 61-74.

Anselin, Luc (2006) "Spatial Econometrics." In T.C. Mills and K. Patterson (Eds.), *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*. Basingstoke, Palgrave Macmillan.

Baker, Wayne E. and Robert R. Faulkner (1993), "The Social Organization of Conspiracy: Illegal Networks in the Heavy Electrical Equipment Industry," *American Sociological Review*, 58 (6), 837-860.

Bonacich, Phillip (1987), "Power and Centrality: A Family of Measures," *American Journal of Sociology*, 92 (5), 1170-1182.

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), 284-299.

———, ———, and Wilfred Vanhonacker (2000), "The Emergence of Market Structure in New Repeat-Purchase Categories: A Dynamic Approach and an Empirical Application," *Journal of Marketing Research*, 37 (1), 16-31.

——— and Catarina Sismeiro (2002), "Using Multimarket Data to Predict Brand Performance in Markets for Which No or Poor Data Exist," *Journal of Marketing Research*, 39 (1), 1-17.

Brown, Stephen (1992), *Retail Location: A Micro-scale Perspective*, Brookfield, VT: Avebury.

Brueckner, J. K. (1993), "Inter-Store Externalities and Space Allocation in Shopping Centers," *Journal of Real Estate Finance and Economics*, 7, 5-16.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345-354.

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

——— and ——— (2005), "The Persistence of Marketing Effects on Sales," *Marketing Science*, 14 (1), 1-21.

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.

——— and J. D. Shilling (1996). "How Critical is a Good Location to a Regional Shopping Center?" *Journal of Real Estate Research*, 12, 459-468.

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, 215-239.

Freeman, Linton C., D. Roeder, R.R. Mulholland (1980), "Centrality in Social Networks: II. Experimental Results," *Social Networks*, 2, 119-141.

Gelman, Andrew, Xiao-Li Meng, and Hal Stern (1996), "Posterior Predictive Assessment of Model Fitness Via Realized Discrepancies," *Statistica Sinica*, 6, 733-807.

Gerbich, Marcus (1994), "Shopping Center Rentals: An Empirical Analysis of the Retail Tenant Mix," *Journal of Real Estate Research*, 15 (3), 283-296.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication, *Marketing Science*, 23 (4), 545-560.

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

Granovetter, Mark S. (2005), "The Impact of Social Structure on Economic Outcomes," *Journal of Economic Perspectives*, 19 (1), 33-50.

Greene, William H. (2003), *Econometric Analysis*, 5th ed., New York: Prentice Hall.

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.

Hartmann, Wesley R., Puneet Manchanda, Harikesh Nair, Matthew Bothner, Peter Dodds, David Godes, Kartik Hosanagar and Catherine Tucker (2008), "Modeling Social Interactions: Identification, Empirical Methods and Policy Implications," *Marketing Letters*, forthcoming.

Hoff, Peter D. (2003), "Random Effects Models for Network Data," Working paper, University of Washington.

Hui, Sam, Eric Bradlow, and Peter Fader (2007), "An Integrated Model of Grocery Store Shopping Path and Purchase Behavior," Working Paper, University of Pennsylvania.

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

Ingram, Paul, and Peter W. Roberts (2000), "Friendships among Competitors in the Sydney Hotel Industry," *American Journal of Sociology*, 106 (2), 387–423.

Iyengar, Raghuram, Thomas Valente, and Christophe Van den Bulte (2008), "Opinion Leadership and Social Contagion in New Product Diffusion," Working Paper, University of Pennsylvania.

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

———, Peter Pal Zubcsek, and Miklos Sarvary (2008), "Joining the Network: Personal Influences as Determinants of Diffusion," Working Paper, INSEAD.

Kleinberg, Jon M. (1999), "Authoritative Sources in a Hyperlinked Environment," *Journal of the Association for Computing Machinery*, 46 (5), 604-632.

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

Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70 (July), 74-89.

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

Martineau, P. (1958), "The Personality of the Retail Store," *Harvard Business Review*, 36, 47-55.

Mason, Charlotte H., and William D. Perreault, Jr., "Collinearity, Power, and Interpretation of Multiple Regression Analysis," *Journal of Marketing Research*, Vol. 28, No. 3 (Aug., 1991), pp. 268-280.

Montgomery, Alan L., Shibo Li, Kannan Srinivasan, and John C. Liechty (2004), "Predicting Online Purchase Conversion Using Web Path Analysis," *Marketing Science*, 23 (4), 579-595.

Nair, Harikesh, Puneet Manchanda, and Tulikaa Bhatia (2006), "Asymmetric Peer Effects in Prescription Behavior: The Role of Opinion Leaders," Working Paper, Stanford University.

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

Newton, Michael A. and Adrian E. Raftery (1994), "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society – Series B*, 56 (1), 3-48.

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

Palmer, Maija (2008), "Shoppers Find Goods Well Recommended," *Financial Times*, Jan. 20.

Park , Young-Hoon and Peter S. Fader (2004), "Modeling Browsing Behavior at Multiple Websites," *Marketing Science*, 23 (3), 280-303.

Pauwels, Koen, Imran Currim, Marnik G. Dekimpe, Eric Ghysels, Dominique M. Hanssens, Natalie Mizik, and Prasad Naik (2005), "Modeling Marketing Dynamics by Time Series Econometrics," *Marketing Letters*, 15 (4), 167-183.

——— (2004), "How Dynamic Consumer Response, Competitor Response, Company Support, and Company Inertia Shape Long-Term Marketing Effectiveness," *Marketing Science*, 23 (4), 596-610.

Richardson, Sylvia (1996), "Measurement Error." In *Markov Chain Monte Carlo in Practice*, W.R. Gilks, S. Richardson, and D.J. Spiegelhalter (eds), London, UK: Chapman & Hall, 401-418.

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(2), 1-18.

Rossi, Peter E., Greg M. Allenby, and Rob McCulloch (2005), *Bayesian Statistics and Marketing*, New York, NY: Wiley.

Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts (2006), "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market," *Science*, 311 (5762), 854-856.

Scott, John (2000), *Social Network Analysis: A Handbook*, 2nd ed., Thousand Oaks, CA: Sage.

Snijders, Tom A. B. (2006), "Statistical Models for Network Dynamics." In: S. R. Luchini et al. (eds.), *Proceedings of the XLIII Scientific Meeting, Italian Statistical Society*, pp. 281--296. Padova: CLEUP.

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," *New York Times*, September 11.

Trusov, Michael, Randolph E. Bucklin, and Koen Pauwels (2008), "Estimating the Dynamic Effects of Online Word-of-Mouth on Member Growth of a Social Network Site," Working Paper, University of Maryland.

Tsai, Wenpin and Sumantra Ghoshal (1998), "Social Capital and Value Creation: The Role of Intrafirm Networks," *Academy of Management Journal*, 41 (4), 464-476.

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), 1409-35.

——— 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 (056104).

Vitorino, Maria Ana (2008), Empirical Entry Games with Complementarities: An Application to the Shopping Center Industry," Working Paper, 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), 441-458.

———, ———, and Mark E. J. Newman (2002), "Identity and Search in Social Networks," *Science*, 296 (5571), 1302-1305.

——— and Steven H. Strogatz (1998), "Collective Dynamics of Small-World Networks," *Nature*, 393 (6684), 440-442.

Yang, Sha and Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research*, 40 (3), 282-294.

Zhang, Jonathan Z. and Oded Netzer (2008), "Dynamics in Interdependent Consumer Preferences," Working Paper, Columbia University.

Zhou, Haijun (2002), "Scaling Exponents and Clustering Coefficients of a Growing Random Network," *Physical Review E*, 016125.

FOOTNOTES

(1)   Similar logic applies to retailer co-location in local districts or zones (e.g., the diamond district in midtown Manhattan, second-hand book stores lining the Seine in Paris, and so-called "Auto Malls" with multiple car dealerships lining major streets in many American and European cities.) More generally, retailer co-location is a micro type of the concept of industrial districts introduced by Marshall (1890/1961; see also Brown 1992).

(2)   Nodes with incoming links but zero outgoing links are also called "sinks" in graph theory and network analysis (de Nooy, Mrvar, and Batagelj 2005).

(3)   Note that sellers may, however, "under-invest" in service quality, for example by spending less time making their shops visually attractive, in response to increased competition (Stern, El-Ansary, and Coughlan 1996).

(4)   We considered adding a time trend and monthly seasonality effects in Equations (2) and (3), however in both cases Wald tests could not reject the null hypothesis that the time trend and all month effects were zero, thus suggesting that time and seasonality effects were not needed.

(5)   Our argument is that allowing shops to be connected in a network makes the marketplace more "sticky" and therefore increases the probability that a visitor to the marketplace will make a purchase. An alternative explanation is that overall visitor traffic to the marketplace increased between the pre- and post-network birth periods. Although the company did not provide traffic data, we were able to collect approximate aggregate (at the level of the website domain, e.g., *website*.com) traffic data using Google's *Trends for Websites* product. In the two months on either side of the birth of the network the marketplace received approximately 80,000 unique daily visitors, and this gradually *decreased* over the seven months from the birth of the network to the end of the dataset to approximately 65,000 unique visitors per day. Given that traffic actually declined between the pre- and post-network birth periods, it cannot be the case that the positive effect of the network's presence on revenues was due to increased overall traffic to the marketplace.

(6)   Readers are referred to Dekimpe and Hanssens (2004, 2005) for a more detailed explanation of impulse-response functions, and Pauwels (2004) for an example of their use in modeling the impact of marketing actions or changes in marketing variables on dynamic systems. More typical uses of IRFs in marketing (e.g., stock return models) attempt to quantify long-run effects of marketing mix variables that last for months or even years. In our case, given the rapid growth of this marketplace and its social network, our time frame is much shorter.

(7)   The popularity variable is an indicator of how popular a shop's products are relative to other shops in the marketplace. This is measured 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 on shops A and B, thus giving it a count of 2). The mean of this count within each shop and over all of the shop's products is then taken as the measure of average popularity, where a shop with a higher average popularity has, on average, a less unique product assortment than a shop with a lower average popularity.

(8)   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.

(9)   Note however that this approach may not address the collinearity between these measures that follows from the fact that there are linked by definition. With the large number of observations used to estimate this model, collinearity is less likely to be a problem (Mason and Perreault 1991). The largest correlation between any of the 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 reasonably low (.24). The largest correlation was between authority centrality and hub centrality (.92), both of which did not significantly affect *Performance$_i$*. The model was re-estimated without hub centrality and the results did not change (the posterior means were very similar to those reported for the full model - details available from the authors). The next-largest correlation was between indegree and outdegree (.80). Since both have large effects on commission revenues it is not meaningful to remove them from the model. Instead, we decomposed indegree and outdegree into unreciprocated indegree (number of incoming links that shop *i* has not reciprocated), unreciprocated outdegree (number of outgoing links that have not been reciprocated), and reciprocated degree (number of reciprocated, two-way links). The highest correlation between these three variables was .56 (between

reciprocated degree and unreciprocated outdegree). We then re-estimated the model with these three variables replacing indegree and outdegree. The other effects remained unchanged, and we found significant positive unreciprocated indegree (posterior mean = .46) and reciprocated degree effects (posterior mean = .26) with the unreciprocated outdegree effect marginally significant (posterior mean = -.08).

(10)   The model is identified by (i) fixing the prior mean on $Ability_i$, which gives the latent variable a "scale," and (ii) because $Ability$ varies across shops but not network position and product assortment measures, while the α and γ parameters vary across measures but not across shops.

(11)   The maxima for the quadratic indegree effect is approximately 27 standard deviations above the mean indegree (indegree of about 93, which is less than the maximum observed indegree of 184), and the minima for the quadratic outdegree effect is approximately 36 standard deviations above the mean outdegree (outdegree of about 159, which is above the maximum observed outdegree of 100).

(12)   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 indegree and outdegree centrality, respectively.

(13)   However, tractability becomes an issue with large networks, and complex error structures may not always be necessary: on this dataset preliminary models with interdependence entered into error structures (e.g., similar to the model used by Yang and Allenby 2003) did not suggest that sellers' commissions were highly interdependent. Endogenous distributed effects that are entered directly as additional ability-adjusted regressors and not through the model's error structure (e.g., neighbors' commissions affecting own commissions), however, were significant in a preliminary endogenous spillover model that included effects of neighbors' commission revenues and neighbors' shop sizes (numbers of products offered).

TABLE 1
NETWORK POSITION MEASURES

Measure	Formula	Definition and Intuition
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 $N \times N$ adjacency matrix, $\mathbf{X}$, with diagonal elements $x_{ii} = 0$ (for all $i = 1, \ldots, N$), and off-diagonal elements $x_{ij} = 1$ if there is a link from i to j (and 0 otherwise).		
Indegree centrality	$(\mathbf{X' \cdot 1})_i = k_i^{in}$ where $\mathbf{1}$ is an N-dimensional vector of ones.	The number of links from other shops that go to shop i.
Outdegree centrality	$(\mathbf{X \cdot 1})_i = k_i^{out}$ where $\mathbf{1}$ is an N-dimensional vector of ones.	The number of links from shop i that go to other shops.
Incoming proximity	$[n_i^{in}/(N-1)] \times [n_i^{in} / \sum_{j \in in_i} d(j,i)]$, 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	$[n_i^{out}/(N-1)] \times [n_i^{out} / \sum_{j \in out_i} d(i,j)]$, 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 indegree of shop i.	A measure of the density of 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 outdegree of shop i.	A measure of the density of 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 $\mathbf{X}^T\mathbf{X}$.	

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)*

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

TABLE 3

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

	Mean (st. dev.)	Correlations (between unstandardized variables)									
Indegree centrality	.76 (3.40)										
Outdegree 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					
Inproximity centrality	.01 (.03)	.60	.45	.48	.45	.21	.17				
Outproximity 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 in this table are computed on the sample of shops used in the estimation of the shop-level model, not all shops.

TABLE 4A

FULL LATENT VARIABLE TOBIT MODEL: EFFECTS OF NETWORK POSITION, PRODUCT ASSORTMENT, AND LATENT ABILITY ON SHOP-LEVEL COMMISSION REVENUES (EQUATION 6)

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			
Indegree centrality ($\beta_{3,1}$)	.923***	.0006	.82, 1.05
Outdegree 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
Inproximity (incloseness) centrality ($\beta_{3,7}$)	.176***	.0004	.10, .24
Outproximity (outcloseness) 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 significant interaction effects[a]			
Indegree^2 ($\beta_{5,1}$)	-.017***	.00002	-.02, -.01
Outdegree^2 ($\beta_{5,2}$)	.007***	.00002	.003, .01
Authority × number of products ($\beta_{6,5}$)	.094*	.0006	-.03, .20
Inproximity × number of products ($\beta_{6,7}$)	-.041**	.0002	-.09, -.002

*   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: (1) the error standard deviation (σ) has posterior mean = 3.226, standard error = .0004, and 95% credible interval (3.10, 3.27). (2) All network position- and product assortment-related variables were standardized (mean = 0, standard deviation = 1) before running the model.

TABLE 4B

FULL LATENT VARIABLE TOBIT MODEL: EFFECTS OF LATENT ABILITY ON

NETWORK POSITION MEASURES (EQUATION 4)

Parameter	Posterior Mean ($\times 10^{-3}$)	Posterior Standard Error	95% Credible Interval
Network effects: intercepts $\gamma_{0,j}$			
Indegree centrality ($\gamma_{0,1}$)	-.491	.0003	-.07, .06
Outdegree 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
Inproximity (incloseness) centrality ($\gamma_{0,7}$)	-.437	.0003	-.06, .06
Outproximity (outcloseness) centrality ($\gamma_{0,8}$)	-.614	.0003	-.07, .07
Network effects: slopes $\gamma_{1,j}$			
Indegree centrality ($\gamma_{0,1}$)	.505	.0003	-.06, .07
Outdegree centrality ($\gamma_{0,2}$)	.447	.0003	-.06, .07
Incoming clustering coefficient ($\gamma_{0,3}$)	.422	.0003	-.06, .06
Outgoing clustering coefficient ($\gamma_{0,4}$)	.496	.0003	-.06, .07
Authority (in eigenvector centrality) ($\gamma_{0,5}$)	.524	.0003	-.07, .06
Hub (out eigenvector centrality) ($\gamma_{0,6}$)	.565	.0003	-.06, .07
Inproximity (incloseness) centrality ($\gamma_{0,7}$)	.404	.0003	-.06, .07
Outproximity (outcloseness) centrality ($\gamma_{0,8}$)	.573	.0003	-.06, .07

TABLE 4C

FULL LATENT VARIABLE TOBIT MODEL: EFFECTS OF LATENT ABILITY ON

PRODUCT ASSORTMENT MEASURES (EQUATION 5)

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

TABLE 5

SIMPLER TOBIT MODEL: EFFECTS OF NETWORK POSITION AND PRODUCT

ASSORTMENT ON SHOP-LEVEL COMMISSION REVENUES

Parameter	Posterior Mean	Posterior Standard Error	95% Credible Interval
Intercept (β_0)	.189***	.0001	.17, .21
Latent ability (β_1)	0[a]	n/a	n/a
Shop age (β_2)	.018	.0002	-.004, .04
Network effects			
Indegree centrality ($\beta_{3,1}$)	.914***	.0004	.84, .98
Outdegree centrality ($\beta_{3,2}$)	-.498***	.0005	-.59, -.41
Incoming clustering coefficient ($\beta_{3,3}$)	-.152***	.0002	-.19, -.12
Outgoing clustering coefficient ($\beta_{3,4}$)	.050***	.0002	.02, .09
Authority (in eigenvector centrality) ($\beta_{3,5}$)	-.084***	.0003	-.15, -.02
Hub (out eigenvector centrality) ($\beta_{3,6}$)	.038	.0005	-.02, .10
Inproximity (incloseness) centrality ($\beta_{3,7}$)	.178***	.0002	.14, .22
Outproximity (outcloseness) centrality ($\beta_{3,8}$)	-.071***	.0002	-.11, -.03
Product assortment effects			
Number of products ($\beta_{4,1}$)	.013	.0001	-.01, .04
Average product popularity ($\beta_{4,2}$)	-.015	.0002	-.04, .01
Quadratic and significant interaction effect[b]			
Indegree^2 ($\beta_{5,1}$)	-.017***	.00001	-.02, -.01
Outdegree^2 ($\beta_{5,2}$)	.007***	.00002	.003, .01
Authority × number of products ($\beta_{6,5}$)	.100**	.0004	.02, .18
Inproximity × number of products ($\beta_{6,7}$)	-.036**	.0002	-.07, -.01
Outproximity × number of products ($\beta_{6,8}$)	.021*	.0001	-.004, .05

*   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]   Fixed to zero.
[b]   None of the other interaction effects are close to being significantly different from zero.
Notes: (1) These estimates are based on a non-hierarchical Tobit model without a latent ability variable. Estimation involved 20,000 MCMC draws (10,000 burn-in). The model mixed well and converged quickly. (2) The error standard deviation (σ) has posterior mean = 3.260, standard error = .0001, and 95% credible interval (3.24, 3.28). All network position- and product assortment-related variables were standardized (mean = 0, standard deviation = 1) before running the model. Thus, posterior means in this table can be compared as "standardized" coefficients.

FIGURE 1

ILLUSTRATION OF DEAD-ENDS

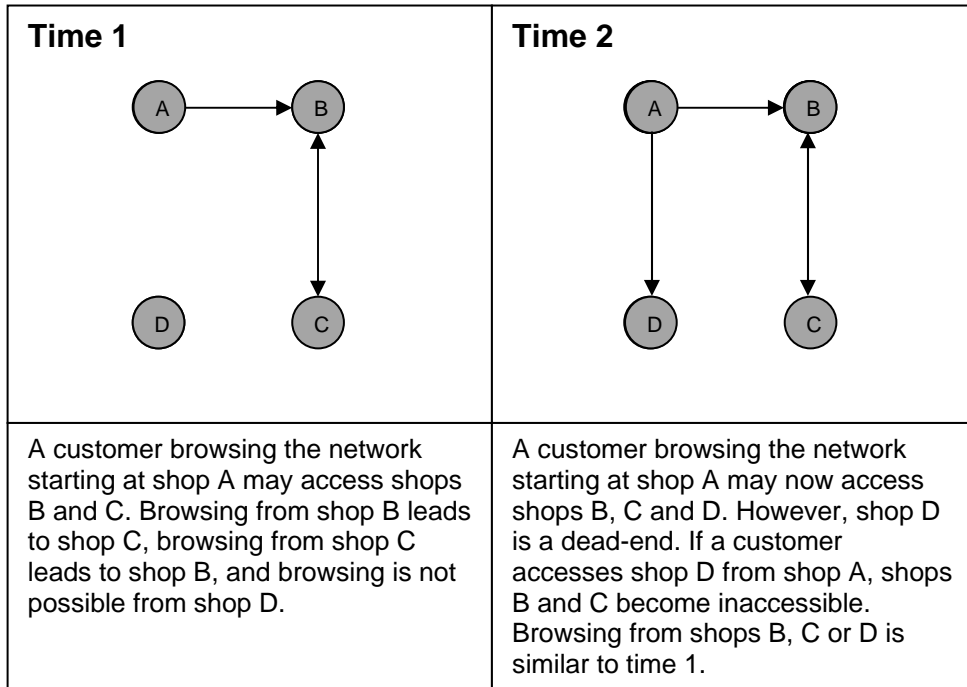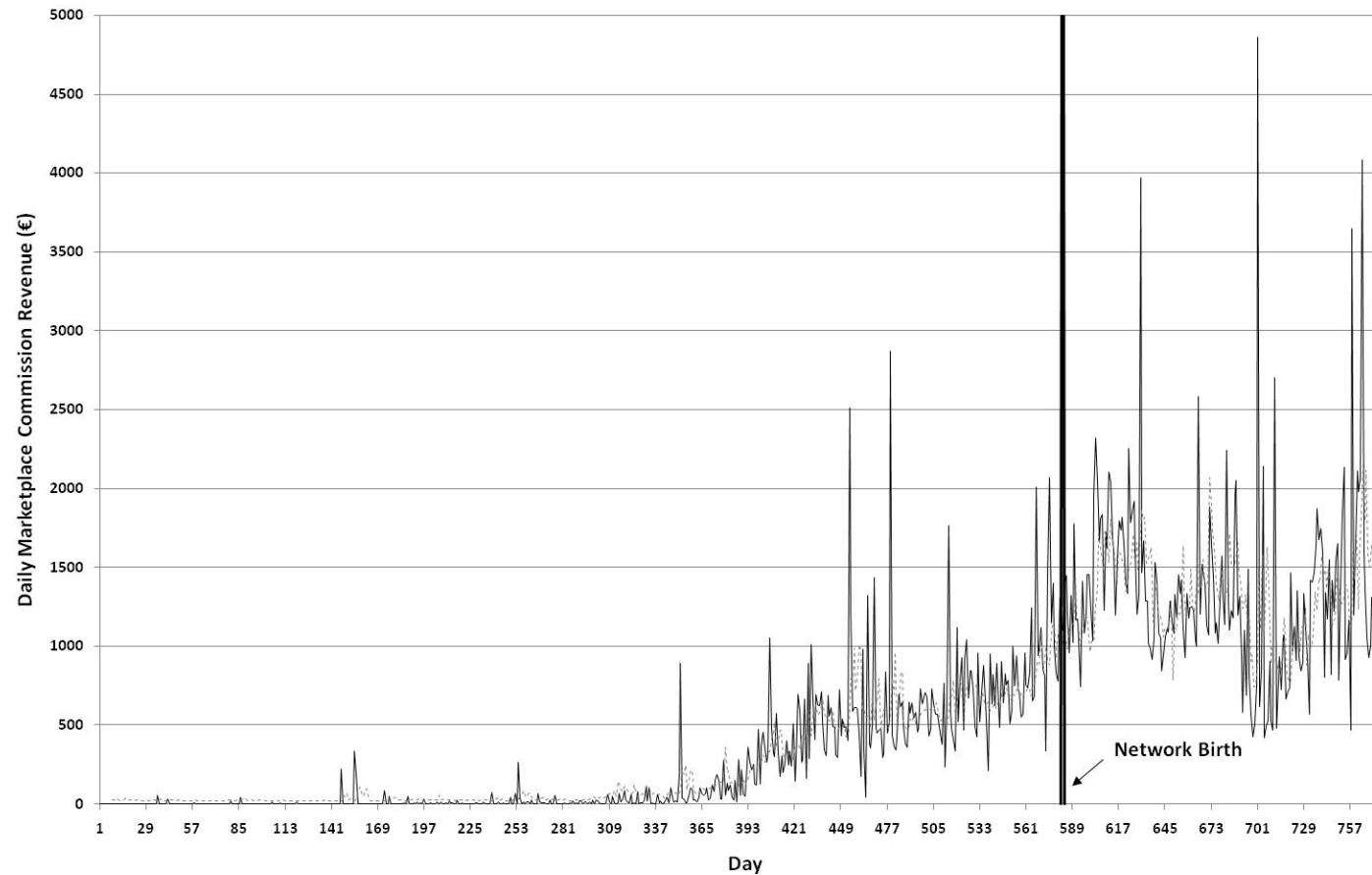| Time 1	Time 2
A customer browsing the network starting at shop A may access shops B and C. Browsing from shop B leads to shop C, browsing from shop C leads to shop B, and browsing is not possible from shop D.	A customer browsing the network starting at shop A may now access shops B, C and D. However, 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.

FIGURE 2

DAILY MARKETPLACE COMMISSION REVENUES

Note: the solid black line is the actual daily marketplace commission revenues, and the broken grey 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.

54

FIGURE 3
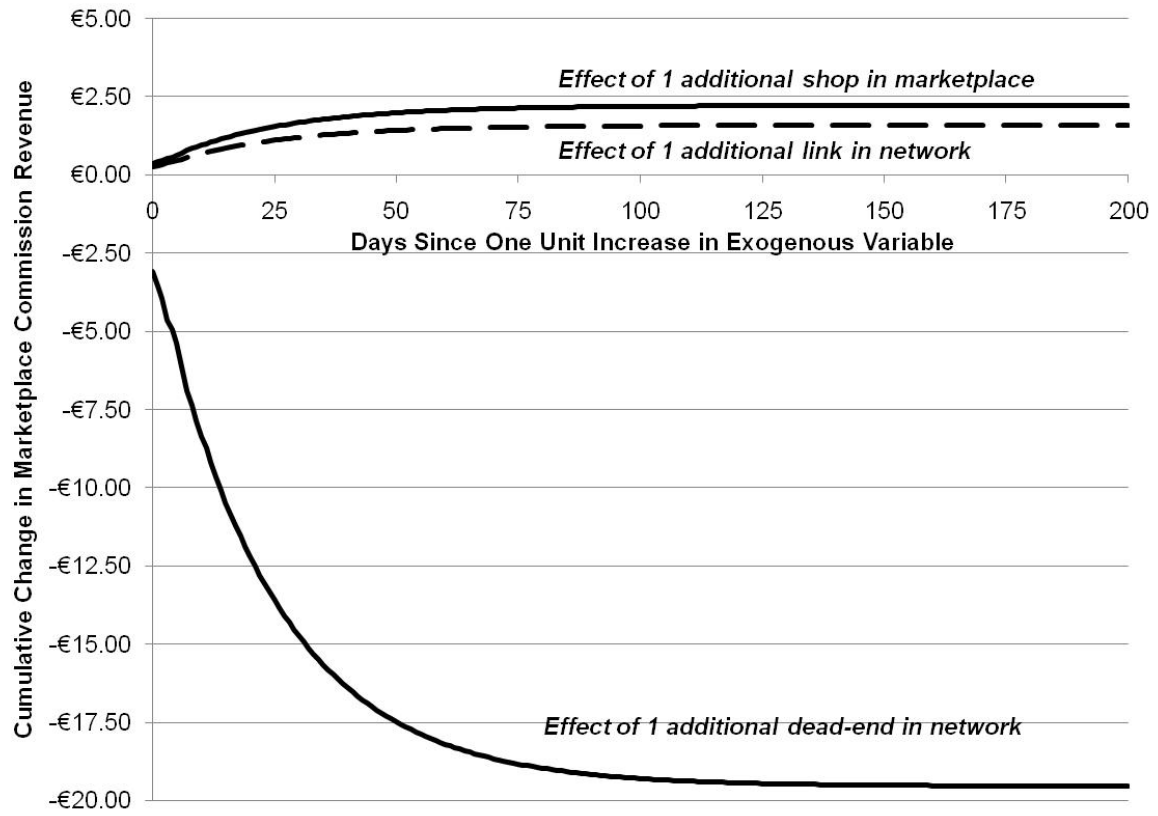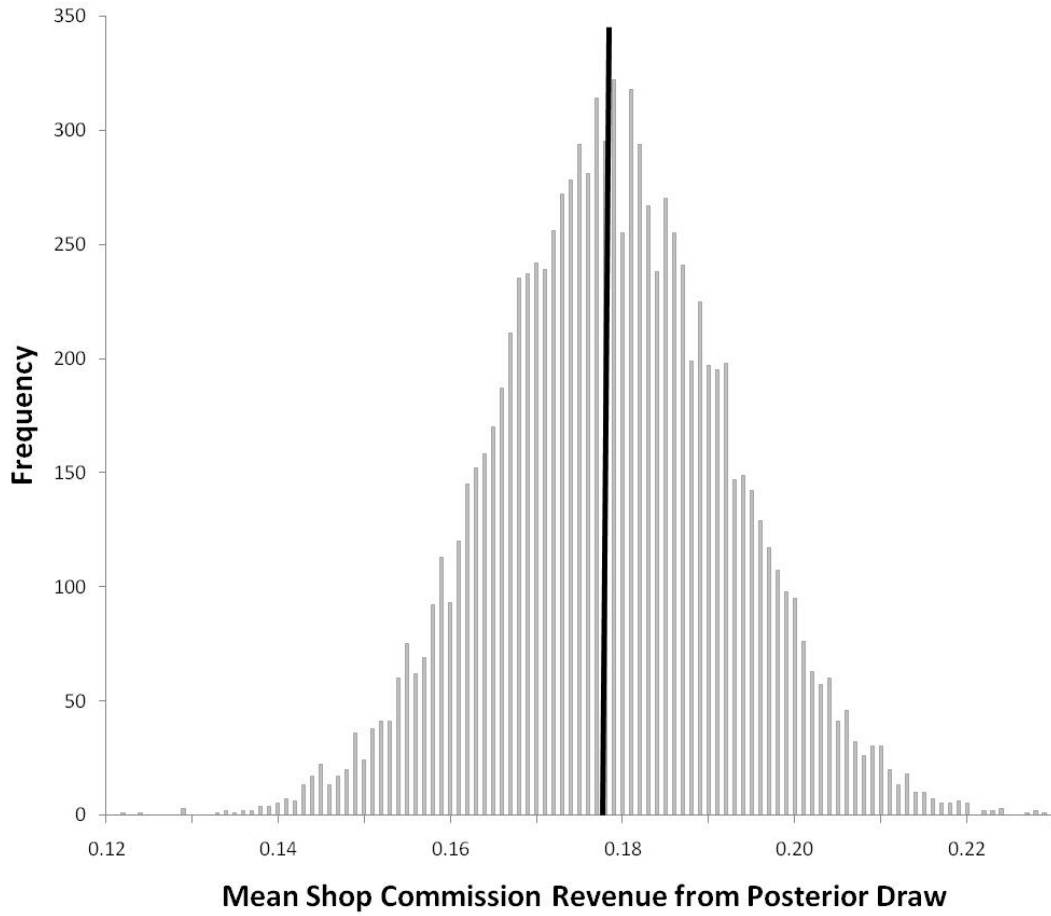**CUMULATIVE IMPULSE-RESPONSE FUNCTIONS**

FIGURE 4

**POSTERIOR CHECKS: POSTERIOR DISTRIBUTION OF PREDICTED MEAN
COMMISSION REVENUES**

Note: the histogram shows the posterior distribution of the fitted mean commission revenues, the solid vertical line represents the actual value (€0.1783).

APPENDIX: BAYESIAN ESTIMATION PROCEDURE FOR SHOP-LEVEL MODEL

Priors

- $Ability_i \sim N(-1, \eta^2)$.

- Diffuse on β_0, β_1, β_2, $\boldsymbol{\beta}_3$, $\boldsymbol{\beta}_4$, $\boldsymbol{\beta}_5$, $\boldsymbol{\beta}_6$, $\boldsymbol{\alpha}_0$, $\boldsymbol{\alpha}_1$, $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$.

- $\sigma^2 \sim InverseGamma\left(\dfrac{r_0}{2}, \dfrac{s_0}{2}\right)$, where $r_0 = s_0 = 1$.

- $\eta^2 \sim InverseGamma\left(\dfrac{r_0}{2}, \dfrac{s_0}{2}\right)$, where $r_0 = s_0 = 1$.

- $\boldsymbol{\Lambda} \sim InverseWishart(n_0, n_0\boldsymbol{\Delta_0})$, where $n_0 = p + q + 3$, and $\boldsymbol{\Delta_0} = \mathbf{I}$, with p the number of network-related variables (7) and q the number of assortment-related variables (3).

Markov Chain Monte Carlo Simulation Steps

Step 1

$L(Ability_i \mid Performance_i, Performance_i^*, \boldsymbol{\beta}, \boldsymbol{\delta}_i, \boldsymbol{\zeta}_i, \sigma^2, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \sim Normal(m_i, V_i)$, where

$$m_i = V_i \cdot \left(\frac{Performance_i^* - \beta_0 - \beta_2 Age_i - \sum_{j=1}^{J} \beta_{3,j}\delta_j - \sum_{k=1}^{K} \beta_{4,k}\zeta_k - \sum_{l=1}^{L} \beta_{5,l}\delta_{i,j}^2 - \sum_{j',k'} \beta_{6,j'k'}\delta_{i,j'}\zeta_{i,k'}}{\beta_1} \cdot \frac{\beta_1^2}{\sigma^2} + [\boldsymbol{\alpha}_1, \boldsymbol{\gamma}_1]\boldsymbol{\Lambda}^{-1}[(Network_i - \boldsymbol{\gamma}_0)'; (Assortment_i - \boldsymbol{\alpha}_0)'] + \frac{1}{\eta^2} \right)$$

$$V_i = \left(\frac{\beta_1^2}{\sigma^2} + [\boldsymbol{\alpha}_1, \boldsymbol{\gamma}_1]\boldsymbol{\Lambda}^{-1}[\boldsymbol{\alpha}_1, \boldsymbol{\gamma}_1]' + \frac{1}{\eta^2} \right)^{-1}$$

Step 2

$L(Performance_i^* \mid Performance_i, Ability_i, \boldsymbol{\beta}, \boldsymbol{\delta}_i, \boldsymbol{\zeta}_i, \sigma^2, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \sim$

$TruncatedNormal(\beta_0 + \beta_1 Ability_i + \boldsymbol{\beta}_2\boldsymbol{\delta}_i + \boldsymbol{\beta}_3\boldsymbol{\zeta}_i, \sigma^2)$

if $Performance_i < 0$. Otherwise $Performance_i^* = Performance_i$.

Step 3

$$L(\eta^2 \mid Performance_i, Performance_i^*, Ability_i, \boldsymbol{\beta}, \boldsymbol{\delta}_i, \boldsymbol{\zeta}_i, \sigma^2, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \sim$$

$$InverseGamma\left(\frac{r_0 + \sum_{i=1}^{N}(Ability_i + 1)^2}{2}, \frac{s_0 + N}{2}\right)$$

Step 4

$$L(\sigma^2 \mid Performance_i, Performance_i^*, Ability_i, \boldsymbol{\beta}, \boldsymbol{\delta}_i, \boldsymbol{\zeta}_i, \sigma^2, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \sim$$

$$InverseGamma\left(\frac{r_0 + \sum_{i=1}^{N}(Performance_i^* - (\beta_0 + \beta_1 Ability_i + \beta_2 Age_i + \boldsymbol{\beta}_3\boldsymbol{\delta}_i + \boldsymbol{\beta}_4\boldsymbol{\zeta}_i + \boldsymbol{\beta}_5\boldsymbol{\delta}_i^2 + \boldsymbol{\beta}_6\boldsymbol{\delta}_i\boldsymbol{\zeta}_i))^2}{2}, \frac{s_0 + N}{2}\right)$$

Step 5

$$L(\boldsymbol{\Lambda} \mid Performance_i, Performance_i^*, Ability_i, \boldsymbol{\beta}, \boldsymbol{\delta}_i, \boldsymbol{\zeta}_i, \sigma^2, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \sim$$

$$InverseWishart(n_0 + N, n_0\boldsymbol{\Delta}_0 + \sum_{i=1}^{N}[\boldsymbol{\delta}_i'; \boldsymbol{\zeta}_i']'[\boldsymbol{\delta}_i, \boldsymbol{\zeta}_i])$$

Step 6

$$L([\beta_0; \beta_1; \beta_2; \boldsymbol{\beta}_3; \boldsymbol{\beta}_4; \boldsymbol{\beta}_5; \boldsymbol{\beta}_6] \mid Performance_i, Performance_i^*, Ability_i, \boldsymbol{\delta}_i, \boldsymbol{\zeta}_i, \sigma^2, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \sim$$

$$Normal((\mathbf{X'X})^{-1}\mathbf{X'y}^*, \sigma^2(\mathbf{X'X})^{-1})$$

where $\mathbf{X}_i = [1, Ability_i, Age_i, \boldsymbol{\delta}_i, \boldsymbol{\zeta}_i, \boldsymbol{\delta}_i^2, \boldsymbol{\delta}_i\boldsymbol{\zeta}_i]$

Step 7

For all j:

$$L([\gamma_{0,j}; \gamma_{1,j}] \mid Performance_i, Performance_i^*, Ability_i, \boldsymbol{\beta}, \boldsymbol{\delta}_i, \boldsymbol{\zeta}_i, \sigma^2, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \gamma_{0,(l\neq j)}, \gamma_{1,(l\neq j)}) \sim$$

$$Normal\left((\mathbf{X'X})^{-1}\mathbf{X'W}, \boldsymbol{\Lambda}_{j,j}(\mathbf{X'X})^{-1}\right)$$

where $\mathbf{X}_i = [1, Ability_i]$ and $\boldsymbol{W}_i = [\boldsymbol{Network}_{i,j}]$, and similarly for all k:

$$L([\alpha_{0,k}; \alpha_{1,k}] \mid Performance_i, Performance_i^* Ability_i, \boldsymbol{\beta}, \boldsymbol{\delta}_i, \boldsymbol{\zeta}_i, \sigma^2, \alpha_{0,(l\neq k)}, \alpha_{1,(l\neq k)}, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) \sim$$

$$Normal\left((\mathbf{X'X})^{-1}\mathbf{X'W}, \boldsymbol{\Lambda}_{J+k,J+k}(\mathbf{X'X})^{-1}\right)$$

Note that we did not update $[\gamma_{0j}; \gamma_{1j}]$ and $[\alpha_{0j}; \alpha_{1j}]$ for all j's and k's simultaneously for tractability reasons.