

IMPROVING PRE-LAUNCH DIFFUSION FORECASTS: USING SYNTHETIC NETWORKS AS SIMULATED PRIORS

Michael Trusov, William Rand and Yogesh Joshi
Robert H. Smith School of Business
University of Maryland

ABSTRACT

While the role of social networks and consumer interactions in new product diffusion is widely acknowledged, such networks and interactions are often unobservable to the firm/researcher. What may be observable, instead, are aggregate diffusion patterns for past products adopted within a particular social network. We propose an approach for identifying systematic conditions that are stable *across* diffusions, and hence are “transferrable” to new product introductions within a given network. Using Facebook apps data, we show that incorporation of such systematic conditions improves pre-launch forecasts. Our research bridges the gap between the disciplines of Bayesian statistics and agent-based modeling by demonstrating how stochastic relationships simulated within complex systems can be used as meaningful inputs for Bayesian inference models.

Keywords: Agent based models, complex systems, Bayesian inference, consumer networks, diffusion, pre-launch forecasts.

INTRODUCTION

Predicting sales dynamics prior to a new product launch is a critical challenge facing many marketers. Existing approaches leverage historical data by identifying a similar product(s) and assuming that the new product may follow the same sales pattern, and/or by combining characteristics of diffusion curves across multiple products while taking into account product attributes, market conditions, as well as uncertainty in parameter estimates (Lenk and Rao 1990, Lilien and Rangaswamy 2004). We contribute to the pre-launch forecasting literature by leveraging yet another predictive dimension: the underlying consumer network structure. Specifically, we propose that aggregate diffusion dynamics carry a trace of the underlying network of consumer interactions, which for a given period of time can be stable, and hence distinctively affect all diffusions observed in the market. Identifying this trace across multiple diffusions can lead to more informed decisions regarding the dynamics of future adoptions.

While the history of research on the role of social influence in product adoption is extensive (Ryan and Gross 1943, Rogers 1962, Bass 1969), recently there has been a renewed emphasis on understanding how the network of social interactions affects diffusion, and hence a firm's marketing strategy (Valente 1995, Goldenberg et al. 2001). Consumers' social interactions are often depicted as a network of nodes representing consumers and links representing interactions. The structure of such interaction networks has been shown to play a role in targeting (Kempe et al. 2003, Haenlein and Libai 2013), in predicting churn (Dasgupta et al. 2008, Nitzan and Libai 2011), in understanding customer level revenue distribution (Haenlein 2011), in identifying influential consumers (Iacobucci and Hopkins 1992, Iyengar et al. 2011), and in forecasting new product adoption (Toubia et al. 2008). Researchers have also shown that different social network structures generally give rise to different rates of adoption over time

(e.g., Van den Bulte and Joshi 2007, Van den Bulte and Wuyts 2007, Rahmandad and Sterman 2008). Hence, knowing the underlying network characteristics of a market could help improve future sales predictions in that market.

Research on consumer networks has explored restricted network spaces such as grid structures (Goldenberg et al. 2001, 2002, Moldovan and Goldenberg 2003, Libai et al. 2005, 2009) as well as more intricate network structures such as small-world networks (Watts and Strogatz 1998) and scale-free networks (Barabási and Albert 1999, Rand and Rust 2011). Watts (2002), Watts and Dodds (2007) and Goldenberg et al. (2009) have examined the role of influentials in product adoptions, and arrived at differing results in a variety of network structures, while Shaikh et al. (2005) showed that different networks create different adoption curves. Our work differs from the above by inverting the discussion on how networks shape diffusion curves to ask: can some network characteristics be inferred from aggregate diffusion curves, and then used in forecasting?

A key hurdle to incorporating network characteristics in forecasting models has been a lack of network data. Some firms overcome this problem by creating their own online communities around their products and services¹ (Manchanda et al. 2012), providing the firm with information about consumer interactions, which can be used for marketing purposes (Apple's Ping², Nike's Nike+³). For networked products, a firm may already have access to information about its consumer network structure. For instance, telecommunication companies keep logs of phone communications among subscribers and therefore can make some inferences

¹ “Fancy building your own social network?”, BBC, 12/01/2011, <http://www.bbc.co.uk/news/business-15976135>

² “Apple Ping: Social Network for Music”, Washington Post, 09/01/2010, <http://www.washingtonpost.com/wp-dyn/content/article/2010/09/01/AR2010090104538.html>

³ “How Nike’s Social Network Sells to Runners”, BusinessWeek, 11/06/2008, http://www.businessweek.com/magazine/content/08_46/b4108074443945.htm

about such networks (e.g., Hill et al. 2006, Braun and Bonfrer 2011). Social networking websites like Facebook or Google+ can observe the network structure of supported communities (Trusov et al. 2010). Network information can also be obtained through surveys or interviews, however, such direct methods can get expensive for large networks, be hard to administer, are sensitive to less-than-perfect response rates, and are subject to self-report biases (Van den Bulte 2010).

Existing approaches also incorporate network effects into aggregate diffusion models using network primitives (e.g., Toubia et al. 2008, Dover et al. 2012). Common to these approaches is the calibration of a network component of the model using a limited set of early observations, which are then used to improve forecast accuracy at later stages of adoption. Specifically, Dover et al. (2012) explore differences in adoption rate during the growth and decline stages of the diffusion process to make inferences about the degree distribution of the *active* network. These methods improve forecast performance and are particularly useful after product launch, when market specific conditions (not explicitly accounted for in the model) have been reflected in the early diffusion data via the network component and other model parameters.

By and large, however, most firms know little about the social interactions among their current and potential customers. Though firms often lack access to the underlying consumer social networks, they often have detailed information regarding when adoptions occur. The question then becomes: is the unobserved underlying consumer network structure informative with respect to the observed diffusion dynamics? If so, can one learn something, albeit probabilistically, about this unobserved network of interactions from observed diffusion data, and use it to generate better forecasts? An empirical approach to answering this question may involve collecting aggregate diffusion data on a *variety* of networks with *known* characteristics of interest (e.g., density and topology), developing a model that describes the relation between

the diffusion process and the network structure, estimating these parameters, and using them in generating forecasts, whose efficacy can be assessed.

We propose an approach that relies on the observed patterns across a set of products being adopted in a market, to improve a researcher's understanding of the potential underlying network structure within this market. We differ from prior work in that we explore systematic patterns *across* diffusions, rather than *within* an instance of diffusion. The key proposition of our approach is that each individual diffusion curve carries information about the underlying network of consumer interactions. While this information might be difficult (if not impossible) to capture from a single diffusion curve, we argue that if the network imposes a systematic effect on the diffusion processes (i.e. leaves a network footprint), it might be possible to estimate such an effect when observing a set of diffusions over the *same* network. In essence, we learn about a network's footprint by matching the distribution of historical diffusions with a set of synthetic footprints obtained through simulations of known network types. The outcome of this matching procedure is a probability vector representing the likelihood of the observed data being generated from each of the underlying synthetic networks. We combine the synthetic footprints using these probabilities as weights to create an informative prior regarding the underlying network structure. We show that by using this synthetic prior in Bayesian inference, we can improve forecasts of future diffusions.

To validate our method, we first show that it demonstrates strong performance on simulated data. Second, we also show that the general characteristics of the synthetic network recovered from a set of diffusions of Facebook applications closely resemble characteristics of the actual Facebook network. Finally, we illustrate how the inferred latent network characteristics can assist managers in forecasting diffusion of potentially different products

across subsets of the same network. Consistent with prior research (e.g., Trajtenberg and Yitzhaki 1989, Van den Bulte and Lilien 1997, Lilien et al. 2000), we determine market penetration exogenously, and use our method to forecast the diffusion dynamics for future products. We then compare the performance of our approach against three established models for generating pre-launch forecasts that do not incorporate network structure. This demonstrated improvement in forecasting using the latent network structure is the core contribution of our work. From a methodological perspective, we show how stochastic relationships simulated within complex systems can serve as informative priors in Bayesian inference. While our procedure maps the pattern of observed diffusions to the space of synthetic footprints generated for a number of selected network types, the approach is not assumed to converge to the characteristics of the actual network. If the actual network is one of the pre-specified types and our representation of the adoption process is accurate then the characteristics of the identified synthetic network may indeed match the actual network. More generally, the result should be seen as a best approximation or projection of the actual network into a given space of synthetic networks.

The rest of this paper proceeds as follows. We begin with a brief overview of the intuition behind our proposed approach, followed by the model specification. We apply the proposed approach to a large set of online applications (“apps”) collected from Facebook, and compare the forecasting performance of our method with alternative models for pre-launch forecasting, which do not account for latent network structure. We conclude with a discussion of the limitations and the managerial implications of our work.

INTUITION BEHIND OUR PROPOSED APPROACH

Consider the example of an online publisher (e.g., www.nytimes.com) releasing new content on a regular basis and observing adoption in terms of page views. Some consumers find this content on their own while browsing, using search engines, via RSS, etc. Others may be referred, by friends or co-workers. Thus, the publisher observes product consumption dynamics driven in part by consumers' independent behavior and in part by interactions among consumers. Information about this network of interactions, if available, could help the publisher in forecasting the adoption of future products.

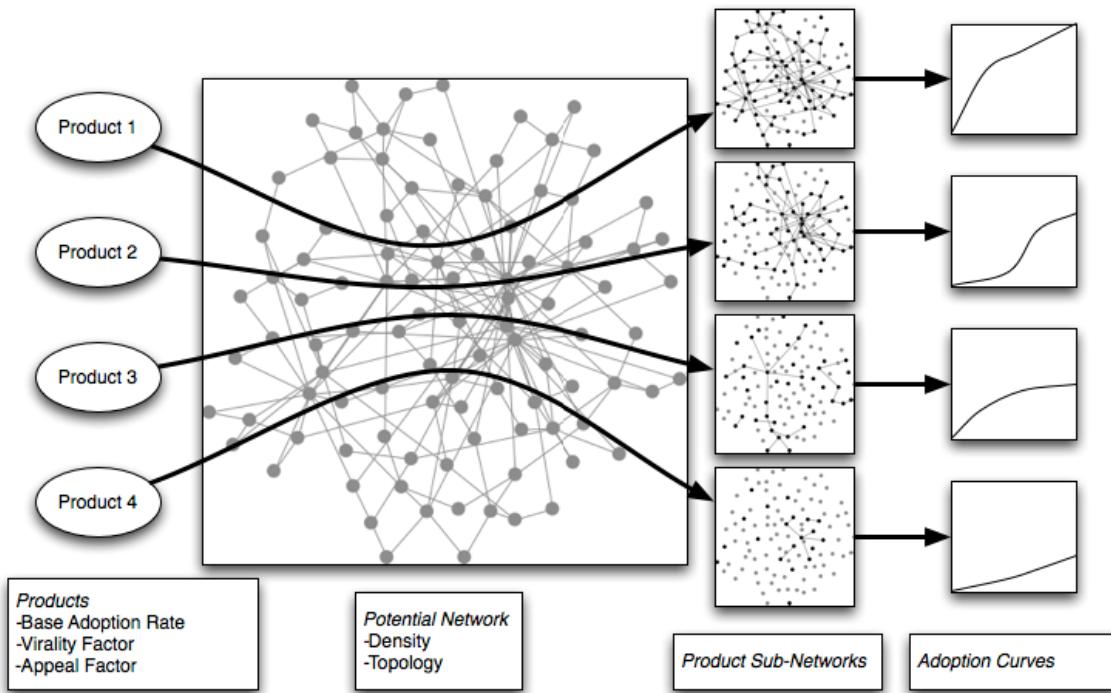


Figure 1. A Potential Interaction Network and Individual Adoption Curves.

What can the publisher learn about the characteristics of the network of consumer interactions on the basis of a single adoption curve? We know from network-based theories of diffusion that diffusion may be faster over a well-connected (high density) network compared to a poorly connected network. Hence the time to takeoff could be informative about network

density. This could be a valid inference if we know that the only difference between the two diffusion processes is the underlying network density. However, such a conclusion cannot be drawn if it is hard to rule out other unobserved factors - for instance, heterogeneity in customers' visits, appeal of the story (e.g., interesting headline), marketing communications, content ranking in search engines, off-line media coverage (e.g., TV or radio) – which can all have an impact on the adoption dynamics that could be misattributed to network effects.

But what if the publisher observes the diffusion of not just one, but several products? Across products, we may think of all existing customers as members of a bigger *potential network of interactions*. We call this the *potential network*, because each particular product potentially appeals to some nonexclusive subset of this network (sub-networks); some members may have different usage patterns than others; and some products may be able to generate more or less buzz on radio and TV. What is common across these instances however is that occasionally some customers share information about content they find on the publisher's site with other members (e.g., by e-mail, tweeting, talking at the water cooler). The superset of all of these communication networks for these different products represents a potential network of interactions that may contribute to adoption (Figure 1). Therefore, while a single diffusion may not necessarily be very informative about either the underlying potential network or the sub-network activated by a particular product, we propose that a set of observed diffusions can help characterize the space of possible underlying potential networks. A sample of observed diffusions over a common potential network can be used to separate unobservable factors that vary from diffusion to diffusion (non-systematic effects) from properties that can be assumed to be reasonably constant across diffusions, such as characteristics of the potential network (systematic effects).

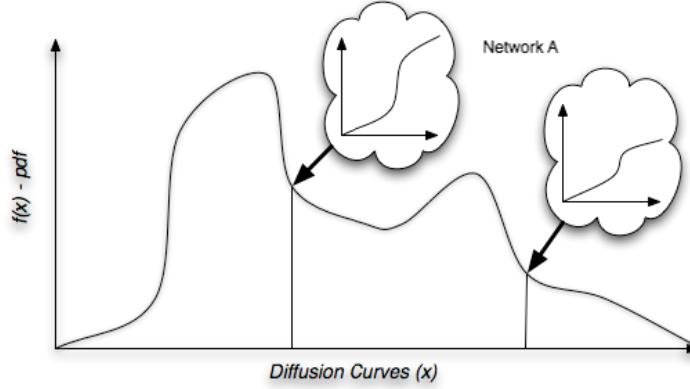


Figure 2. Distribution of Diffusion Curves (X) Observed over Network A.

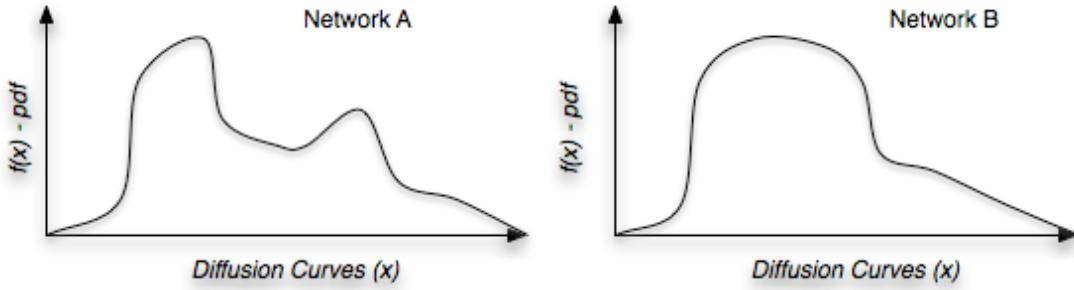


Figure 3. Differences in Diffusion Curve Parameter PDFs for Two Networks.

We now illustrate our approach for characterizing the underlying network. For brevity, instead of looking at a complete time series of the aggregated diffusion curve, we parameterize the curve by X , which for simplicity is assumed to fully characterize the entire diffusion curve. (e.g., the Bass (1969) model characterizes the diffusion curve using three parameters: p , q , and m). Given X , Figure 2 plots a probability density function of observing X over a potential underlying network A. Each point on the x-axis corresponds to a diffusion curve of a distinct shape and can arise, for instance, for different products under different conditions. Similar plots can be generated by observing a set of diffusions on another potential underlying network B, along the same lines as above (Figure 3).

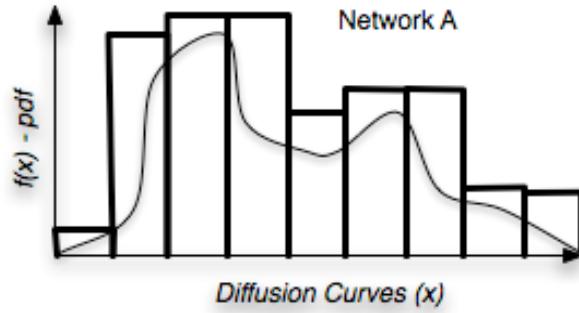


Figure 4. Matching a Footprint with Empirical Distribution.

If this density function, which takes into account different products and conditions, is not identical for different networks after marginalizing all non-network factors, then we can conclude that network characteristics leave a distinct footprint on the distribution of diffusion curves. This information can now be used to make some inferences about the underlying network type by matching the empirical distribution of diffusion curves observed in the data with the footprints for a variety of networks of known types (Figure 4).

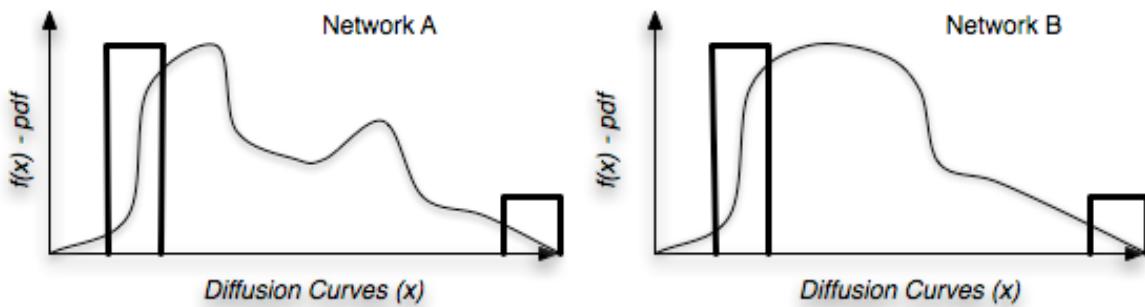


Figure 5. Difficulty in Identification for a Smaller Sample.

Figure 5 illustrates that two density functions may have similar shapes for a subset of values of X, hence network identification for a small number of observed diffusions may be difficult. As shown, it is hard to establish which one of the two networks gave rise to the empirical distribution illustrated by the bold rectangles. However, for a larger sample as shown in Figure 4, a network footprint becomes more pronounced. In subsequent sections we discuss

how the network footprints can be identified and used in conjunction with empirical data.

MODEL FORMULATION

To understand the role of the underlying (unobservable) potential network in generating sets of diffusion curves, assume that the observable and practitioner relevant properties of a diffusion process of product i over a potential network j are fully characterized by vector Y_{ij} , the underlying network characteristics are specified by vector Z_j (e.g., network topology and density) and all other factors (both observable and unobservable) which may influence the diffusion process are given by vector X_i (e.g., product appeal, marketing mix). Then, observing a number of diffusions Y_{ij} under conditions X_i on a set of networks Z_j allows us to explore the relationship:

$$Y_{ij} = f^T(X_i, Z_j) + \varepsilon_{ij}^T, \quad (1)$$

where $f^T(X_i, Z_j)$ represents the true relationship between product/market characteristics, network characteristics, and diffusion; and ε_{ij}^T is a random factor assumed to be independent from X_i and Z_j that captures the stochastic nature of the process. To infer Z given Y , we first learn $f(X, Z)$ and then integrate out the unobservable elements of X . From an empirical perspective, given a reasonably large sample and variations in Y , Z and X , the task of learning $f(X, Z)$ is feasible – different parametric and non-parametric forms of $f(X, Z)$ can be considered and the best model found. The major challenge is that in practice Z (i.e., network structure) is rarely observed, and therefore there is typically no empirical data from which to learn this relationship. The same limitation might apply to at least some of the attributes in X . To address this issue, we hypothesize a relationship, $f^*(X, Z)$, such that:

$$Y_{ij} = f^*(X_i, Z_j) + \varepsilon_{ij} \quad (2)$$

By integrating out X in (2) we arrive at $g(Z|Y)$.⁴ While it might be tempting to start directly with specifying $f^*(Z)$ (therefore skipping the X integration step) we are not aware of any theory describing this relationship reasonably well. On the other hand, several individual level adoption models can be used to construct $f^*(X, Z)$, which we discuss in the next section.

Equipped with a relationship $g(Z|Y)$ and given a set of observed diffusions Y , one can generate some inferences about the latent network Z given Y . How well the latent network Z represents the actual network depends on several factors, among which are the “quality” (or robustness) of the adopted model for f^* (Toubia et al. 2008), the observed sample Y , and the characteristics of the conditional probability density function $g(Z|Y)$. For example, if $g(Z|Y)$ is “flat” or multi-modal then a larger sample might be needed to reduce prediction uncertainty.

Linking the Underlying Network Characteristics with Aggregate Diffusion Dynamics. Among the individual level adoption heuristics used in empirical studies of social networks, most prominent are variants of the SIR model (Anderson and May 1992, Goldenberg et al. 2009, 2010) and the threshold model (Watts 2002, Watts and Dodds 2007). As noted by Van den Bulte and Lilien (2001, p. 1418-19), standard discrete time hazard models used to study network effects are also threshold models. SIR models consider adoption as a combination of two forces –individual innovation and peer imitation. In the threshold model, each individual observes the fraction of her neighbors having adopted, and adopts if this fraction exceeds a certain threshold. These two models, though different, are sub-models of the same general model, and can produce qualitatively similar results (Dodds and Watts 2005, Watts and Dodds 2007), and any differences between the two can be offset by the selection of appropriate model parameters (Kempe et al. 2003). Hence in our study, we focus on a single adoption heuristic (SIR-like). In principle, our

⁴ We may also consider a set of competing f^* , so that $g(Z|Y)$ reflects the result of integration across multiple f^* .

approach can also be used to select between individual level adoption heuristics by treating them as additional unobservable parameters in the model.

Ideally, for the selected heuristic, a closed-form expression for $f^*(X, Z)$ is desirable, given its benefits such as ease of integration and posterior inference. Under restrictive assumptions, closed-form solutions have been proposed (Keeling 1999, Newman and Watts 1999, Shaikh et al. 2005, Toubia et al. 2009). However, to accommodate reasonably general network structures, product features and consumer characteristics, closed form solutions are difficult, if not impossible, to find. Given our objective of exploring multiple realistic underlying network structures, we rely on a simulation-based approach to construct $f^*(X, Z)$.

The Individual Level Adoption Model. We characterize adoption at the individual level via the unobserved product specific vector $X_i = \{A_i, \lambda_{1i}, \lambda_{2i}\}$, where A_i is the fraction of the potential network that product i appeals to, and $(\lambda_{1i}, \lambda_{2i})$ are parameters of the SIR adoption heuristic. λ_{1i} is the independent rate of adoption, i.e., how likely is an individual to adopt a product without any social influence; and λ_{2i} captures the effects of social contagion.⁵

Potential Networks. While our approach is flexible enough to accommodate many different network structures, here we demonstrate its application by considering 16 distinct underlying potential networks (vector Z). To select these networks, we use two criteria. First, we consider four distinct topologies: *lattice / regular*, *random*, *small world*, and *preferential attachment*.⁶ Second, we consider various edge densities, where *edge density* is defined as the fraction of existing edges (i.e., connections) out of all possible edges in a network. Edge density is related to average degree. While the number of neighbors of an individual node is its *degree*, the average

⁵ In principle, X can be expanded to include additional unobserved parameters such as consumer-level heterogeneity (e.g., Van den Bulte and Joshi 2007, Toubia et al. 2008, 2009).

⁶ Preferential attachment is a specific mechanism for generating a scale-free network (Barabási and Albert 1999).

of the degrees of all nodes in a network is its *average degree*.⁷ Let V be the number of nodes in the network. Consequently, edge density can be defined as $(\text{average degree} * V/2) / (V(V-1)/2) = (\text{average degree})/(V-1)$ (Wasserman and Faust 2009, p. 102). For each topology, we examine four comparable classes of edge density.⁸

Characteristics	Lattice/Regular	Random	Small World	Preferential Attachment
Illustration				
Key parameters	d (degree)	p (probability of an edge)	d (average degree), B (rewiring probability)	k (minimal degree)
Degree Distribution	point	Poisson	Varies, from point to Poisson	scale-free
Characteristic Path Length	Relatively long	Relatively short	Relatively short	Shortest of all
Average Clustering Coefficient	Relatively high	Relatively low	Relatively high	Relatively low, dependent on V
Edge Density	$d*V / (V(V-1))$	$p*V*V / (V(V-1))$	$d*V / (V(V-1))$	$\sim 2k(V-1) / (V(V-1))$

Table 1: Comparison of the Four Network Topologies (V = number of nodes).

The above selection of topologies and densities is guided by past research, to capture a range of diffusion patterns across these combinations. Watts and Strogatz (1998) showed that the extent of clustering (the fraction of friends who are also friends) in a network greatly affects the speed of diffusion. Their “small-world” network has the unique properties of having relatively high clustering and low average path length (average of the shortest paths between all possible pairs of individuals in the network). Barabási and Albert (1999) showed that clustering or topology were not sufficient to describe some networks, and that certain degree distributions, such as the

⁷ When simulating preferential attachment networks, we use edge densities comparable to the other topologies.

⁸ The specific edge densities considered correspond to the average densities specified in Table 2.

scale-free distribution, have a dramatic effect on the speed of diffusion. As a result, it is common practice in the network science literature to analyze these four network structures, as they vary both clustering rates and degree density. The key characteristics of the four network topologies are summarized in Table 1.⁹

Simulation Procedure. To learn $f^*(X, Z)$ we run a number of simulations by varying parameter values for X and Z and observing the resulting aggregate diffusion curves Y . We generate ten instances for each potential network, with the number of nodes, V , set to 1000. For each instance of a potential network, the pertinent sub-network is selected using an appeal fraction A , which varies from 0.1 to 1.0 at 0.1 increments. For example, an appeal fraction of 0.5 implies 500 nodes within the network will be randomly selected to constitute the potential adoption sub-network, and the rest are assumed to never adopt this product. While topology and density are explicitly defined for the *potential* network, they need not always match with those for the selected sub-network.¹⁰ The SIR-based individual-level network model uses the activation function, or the likelihood of adoption, equal to $[\lambda_1 + \lambda_2 (v_{it}/V_{it})]$, where V_{it} is the number of neighbors of i , and v_{it} is the number of neighbors who have already adopted the product (Rand and Rust 2011). We vary λ_1 from 0.0007 to 0.03 at 11 equally spaced increments, and varied λ_2 from 0.38 to 0.53 at 11 equally spaced increments. This selected range of *agent-level* parameters results in aggregate diffusion curves that are within the range commonly reported in the empirical marketing literature for the Bass model (e.g., Van den Bulte and Stremersch 2004,

⁹ For a more detailed discussion of these networks, please refer to Watts and Strogatz (1998), Lewis (2009), Wasserman and Faust (2009), as well as the online Appendix A.

¹⁰ Though the simulated potential networks are always connected, since active nodes are chosen at random, the sub-network may frequently be disconnected. Since there exists a non-zero probability for each individual to adopt on her own, relaxing the assumption that the potential network be connected should not dramatically affect our results. In addition, appeal may not always be random, and may be impacted by product attributes. We thank an anonymous reviewer for these observations.

Chandrasekaran and Tellis 2007).

Table 2 summarizes our design. Since both the individual level adoption model and network generation are stochastic, we repeat the simulation of the diffusion process multiple times for each of the 193,600 parameter combinations. Consequently, we observed close to one million adoption curves, each associated with a particular set of micro-level adoption parameters, a network topology and density.

<i>Network Parameters (Z)</i>		
Topologies	Lattice, Random, Small-World	Preferential Attachment
Avg. Degree <i>K</i>	2, 4, 6, 8 n/a	n/a 1,2,3,4
<i>Product Parameters (X)</i>		
	Range	Increment
λ_1	[0.0007, 0.03]	0.00293
λ_2	[0.38, 0.53]	0.015
A	[0.1, 1.0]	0.1

Table 2. Simulation Setup.

From diffusion curves and simulation parameters to a multivariate stochastic function. Each simulation run results in a series of data points representing the number of adopters at time t . We reduce the dimensionality of the simulated data by collapsing each time series into a few descriptive parameters – vector Y in equation (2). To do so, for instance, one could use principal component analysis/ functional data analysis to extract the most prominent dimensions in the data that differentiate between curves (e.g., Bradlow 2002, Ramsay and Silverman 2005, Sood et al. 2009). Alternatively, a generalized functional form could be fitted to the data and the estimated functional parameters could be used as descriptive characteristics (e.g., Stremersch and Lemmens 2009, Lemmens et al. 2012). In this study, we chose to fit a standard Bass model, which is uniquely described by three parameters p , q , and m , estimated using the non-linear least squares approach (Srinivasan and Mason 1986). A Bass model shows remarkably good fits to the simulated data, with an average R^2 of 0.98 (variance 0.002). In the simulations, the level of

product appeal in the potential network (A_i) essentially determines m . While we do estimate m from the simulated data, we focus our attention on p and q . Hence in the subsequent discussion, Y_{ij} represents $\{p_{ij}, q_{ij}\}$.

With each simulation outcome Y_{ij} paired with a unique set of parameters X_i and Z_j , we need to construct $f^*(X, Z)$. An exploratory analysis of the simulated multidimensional space of X , Y and Z showed complex, non-linear relationships between these parameters. We attempted several parametric approaches to formulate $f^*(X, Z)$, however, we did not obtain reasonably good fits with the data. Therefore we resorted to non-parametric methods. Since our ultimate interest is the relationship between Z and Y , at this stage we integrate over X to arrive at $g(Z|Y)$. To do so, we select the simplest form of non-parametric density estimators – a histogram. We discretize the two-dimensional space Y into bins, and for each bin count the number of simulation instances for each Z (i.e., a unique network density and topology pair). We experimented with different bin widths on both dimensions of Y , and ended up with 11 bins for p and 10 bins for q .

The result of this procedure is a four-dimensional frequency table with $4 \times 4 \times 11 \times 10$ cells, which we refer to as a *diffusion hypercube*. If each cell in this table is divided by the total number of simulation runs, then the diffusion hypercube can be viewed as a joint probability density function of Y and Z , and the conditional distribution $g(Z|Y)$ follows naturally. Since visualization of all dimensions of this cube simultaneously is problematic, in Figure 6 we present the distribution of p and q for the 16 underlying synthetic networks.¹¹ While these distributions look different, they show a significant overlap in the function domain. This supports our earlier contention that due to unobserved factors in X and stochasticity in the diffusion process there is a non-trivial probability of different networks giving rise to similar (if not identical) diffusion

¹¹ For additional figures representing slices of this hypercube, please refer to the online Appendix C.

curves. Therefore, to reliably identify properties of the underlying network (i.e., establish which of these 16 cells best describes the underlying network) multiple observations are necessary.

Using this hypercube, in the next section we develop a Bayesian inference model, and subsequently investigate the identification issue empirically using simulated data.

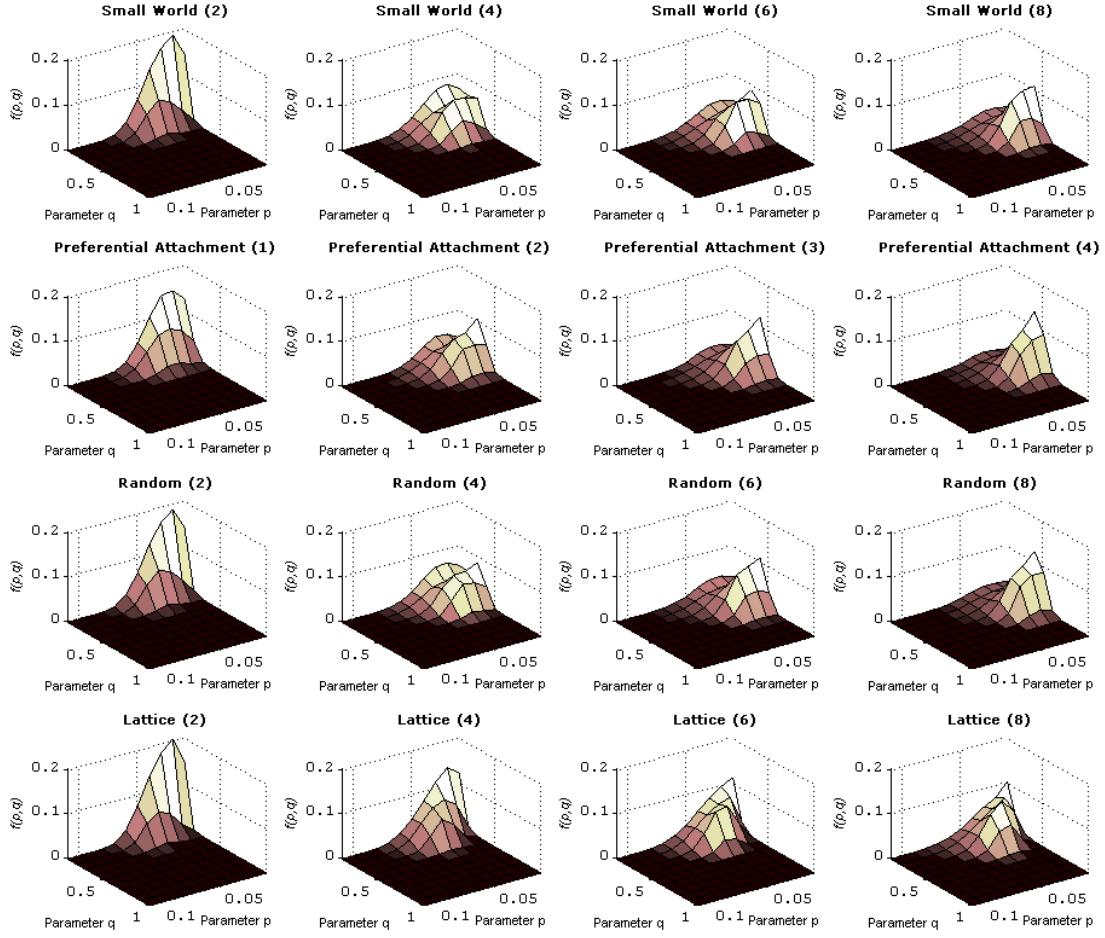


Figure 6. The Diffusion Hypercube: Diffusion Parameter Distribution over each Network.¹²

Bayesian Inference Model. Following a standard Bayesian model selection approach (Chipman et al. 2001), assume that there exists a set of models, $M = \{M_1, M_2, \dots, M_K\}$, $K=16$, that can give a rise to an observed set of diffusion curves represented by Y . M_k corresponds to a potential

¹² The numbers in brackets correspond to average degree for small world/random/lattice networks and edge density for preferential attachment networks.

network k , of a given topology and density. Let $\psi(Y | \theta_k)$ be a probability density function for the distribution of diffusion curves Y , operationalized as a multinomial distribution where Y is a vector of length 110 with each element corresponding to a p -and- q bin. Hence for a single observation (i.e., one product diffusion) Y is a binary vector with all but one elements set to zero $Y = \{0, 0, 0, \dots, 0, 1, 0, 0\}$. Across multiple observations, the elements of vector Y are counts of p -and- q pairs, which fall into the corresponding bins (e.g., $Y = \{1, 0, 3, \dots, 2, 1, 8, 0\}$). Accordingly, θ_k is a 110-element probability vector. Let $\varphi(\theta_k | M_k)$ be a prior probability distribution of θ_k given model M_k . Finally, let $p(M_k)$ be a prior probability of model M_k . The proposed prior formulation allows us to write a joint distribution as:

$$f(Y, \theta_k, M_k) = \psi(Y | \theta_k) \times \varphi(\theta_k | M_k) \times p(M_k). \quad (3)$$

We use the distribution in equation (3) to (a) infer the general characteristics of the unknown underlying diffusion network (i.e., to identify likelihood of M_k), and (b) predict the probability distributions for future product diffusions. The former is achieved by inspecting the posterior distribution of $p(M_k | Y)$. Specifically, assuming a 0/1 loss function, we are interested in the models of the highest posterior probability as defined by equation (4):

$$p(M_k | Y) = \frac{p(Y | M_k) \times p(M_k)}{\sum_i p(Y | M_i) \times p(M_i)}, \text{ where} \quad (4)$$

$$p(Y | M_k) = \int_{\theta_k} \psi(Y | \theta_k) \times \varphi(\theta_k | M_k) d\theta_k. \quad (5)$$

To address the latter, consider a predictive distribution $p(Y^* | Y)$, where Y^* characterizes a distribution of future diffusion curves, by integrating out θ_k and M_k in equation (3). Assuming that Y^* and Y are independent conditional on θ_k , the predictive distribution can be written as:

$$p(Y^* | Y) = \sum_k p(M_k | Y) \times \int_{\theta_k} \psi(Y^* | \theta_k) \times \phi(\theta_k | M_k, Y) d\theta_k, \text{ where} \quad (6)$$

$$\phi(\theta_k | M_k, Y) = \frac{\psi(Y | \theta_k) \times \varphi(\theta_k | M_k)}{p(Y | M_k)} \text{ is the posterior distribution of } \theta_k. \quad (7)$$

Under the mean squared error loss prediction criterion the optimal forecast is the mean of $p(Y^* | Y)$.

Priors. To complete our specification, we define priors for θ and M . The natural conjugate prior for the multinomial distribution $\psi(Y | \theta)$ is the Dirichlet distribution parameterized by vector $\alpha_M = \alpha_{M_1}, \alpha_{M_2}, \dots, \alpha_{M_J}$ where $J=110$ is the number of mass points in the distribution $\psi(Y | \theta)$:

$$\theta | M_k \sim Dirichlet(\alpha_{M_k}). \quad (8)$$

In the proposed framework $\varphi(\theta | M_k)$ is an *informative* prior: for each model M_k , elements of vector α_M come from the corresponding cells of the diffusion hypercube.¹³ For the Dirichlet distribution, the variance of θ_k (which can be interpreted as a probability mass concentrated around a single point) is inversely proportional to $\sum_i \alpha_{M_k,i}$ and hence decreases in the number of simulation runs N_{runs} used to generate the hypercube. Since the main objective in choosing the number of simulation runs N_{runs} is to fully explore the underlying parameter space, $\sum_i \alpha_{M_k,i} \sim N_{runs}$ is large. Therefore, an additional step is needed to scale down the level of *informativeness* that should be given to the prior in equation (7) (i.e. the prior's weight in the posterior distribution). For the multinomial-Dirichlet model, the weight of the prior can be varied by multiplying

¹³ For the informative prior distribution $\varphi(\theta | M_k)$ to be proper, all of the parameters $\alpha_{M_k,j}$ must be greater than zero. In practice this condition may be not satisfied for some combinations of the curve characteristics p and q . Indeed, Figure 6 shows that for several cells there is a zero count of observations in the simulated data. To address the problem of zero $\alpha_{M_k,j}$ we adopt the shrinkage estimator idea in the spirit of Fienberg and Holland (1973). We calculate: $\alpha_{M_k,j} = \alpha_{FH} \times \frac{N_{runs}}{J} + (1 - \alpha_{FH}) \times n_j$, where $\alpha_{FH} = \frac{N_{runs} - Z}{N_{runs} + (N_{runs} - 1) \times Z}$ and $Z = \sum_j \left(n_j - \frac{N_{runs}}{J} \right)^2 / \left(\frac{(J-1) \times N_{runs}}{J} \right)$.

elements of vector α_M by some scalar W^{14} . We argue that in our framework the choice of W is an empirical question and therefore, we determine W by cross-validation using a calibration sample. Finally, we assume a uniform prior for M , $p(M_k) = 1/K$. The various steps involved in our proposed approach are summarized in Figure 7.

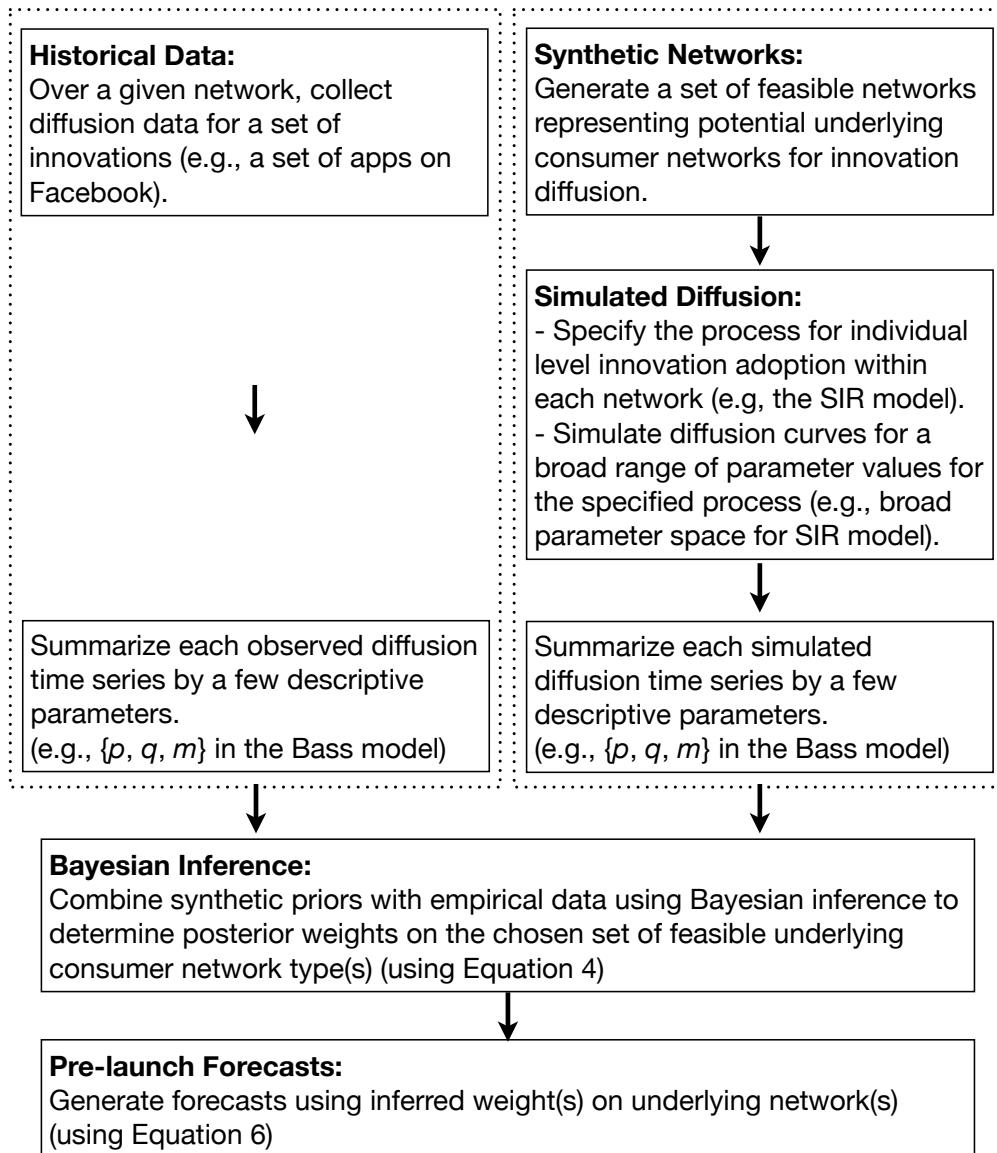


Figure 7: A Summary of Steps in Our Proposed Approach.

¹⁴ The scalar W used to scale Dirichlet parameters in the prior distribution is sometimes called the equivalent sample size parameter.

ILLUSTRATING NETWORK INFERENCE WITH SIMULATED DATA

We now analyze the performance of our proposed approach on simulated data. First, we assess the extent of recovery of the true density and topology of the unknown underlying potential network. Then, we analyze the sensitivity of this recovery to available sample size.

We first generate a set of test networks for each underlying topology and density combination. Next, we randomly draw values for λ_1 , λ_2 and A from a uniform distribution over the domain of the hypercube, so that these values may differ from those used in generating the hypercube. We generate 20 values for each parameter, resulting in 8,000 unique test product scenarios. For each scenario on each of the 16 test networks, we simulate individual level adoption, which we then aggregate to create 128,000 diffusion curves. For each curve, we estimate the Bass (1969) model using the Srinivasan and Mason (1986) technique. The resulting set of p and q values are then binned into 110 intervals (i.e., simulated Y data).¹⁵ To test the performance of the proposed methodology we use the following algorithm:

1. Randomly select N diffusion curves and create 16 vectors Y_{Nk} , each corresponding to a unique density and topology combination representing the randomly selected curves.
2. Using Equation (4), calculate the posterior likelihood for each model M_k and classify the network according to the highest probability model.
3. Calculate the fraction of correctly identified networks.
4. Repeat steps 1-3, 100 times, to calculate mean and variance of correctly classified shares.
5. Increment sample size N and repeat from step 1.

We also consider an alternative (naïve) approach that ignores the condition that all

¹⁵ Though not used, we also estimate m for each diffusion curve.

observed diffusions originate from the same underlying potential network (i.e., share the same density and topology of the potential network). The main difference lies in how the vector \mathbf{Y} is constructed in each case: under the naïve approach we perform classification for each of N selected networks individually without imposing a condition of a shared potential network.

Simulation Results. These are shown in Figure 8. The x-axis is the number of diffusions considered for classification and the y-axis is the fraction of correctly identified networks. As Figure 8 depicts, the naïve model performs better than the random model (i.e., prediction by chance), approximately tripling its performance. But, as expected, it is not sensitive to sample size, and therefore, additional information does not improve its performance.¹⁶ The proposed model, which strongly outperforms the naïve model, gains significantly from additional data. With ~ 100 curves, the underlying network is correctly identified by the proposed method 80% of the time, and the performance levels off at about a 90% accuracy rate with a sample size of 250+. Thus, even under very favorable conditions (i.e., large sample size, finite and known set of potential networks, use of the same data/network generating procedure to create a diffusion hypercube and to simulate test diffusions) the misclassification rate is not trivial. The vast majority of classification errors occur between low density (2) “Random” and “Small World” networks.¹⁷ In other words, these two networks leave nearly undistinguishable footprints on the distribution of diffusion curves (see Figure 6). While this has no negative implications for the accuracy of forecasts (discussed in the next section), it suggests that even under “ideal” conditions it might be difficult to distinguish between some network types, thus highlighting the

¹⁶ The key difference between the naïve model and the proposed model is that the naïve model ignores the common potential network assumption.

¹⁷ To understand (mis)classification, we create a “confusion” matrix, which lists the rate at which the underlying networks are (in)correctly classified. A close observation of this matrix reveals that low density (2) small world networks are frequently confused with low density (2) random networks. This confusion matrix is available upon request from the authors.

difficulty of uniquely identifying the underlying network structure.¹⁸

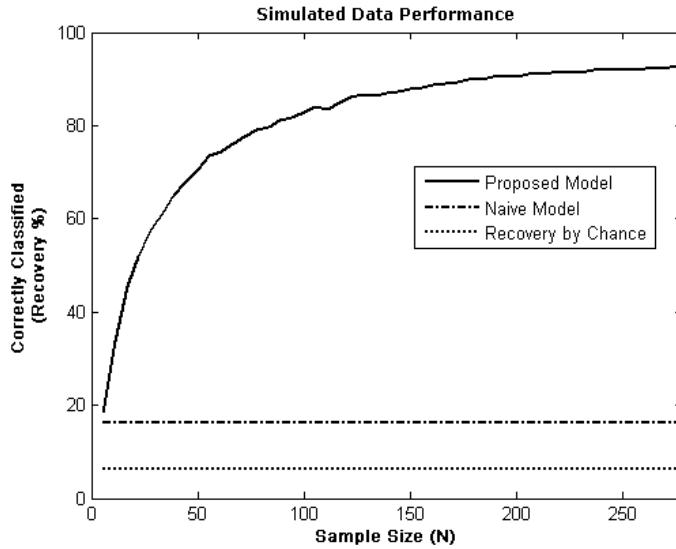


Figure 8. Network Recovery on Simulated Data.

APPLICATION OF THE METHODOLOGY TO FIELD DATA

We now discuss how our approach can be applied in practice. Over the past years, a new digital phenomenon has emerged around “applications” (or “apps”). Differing from traditional software applications, these small software programs are designed for use on mobile devices (e.g., iPhone, iPad) or on social networking sites such as Facebook. For example, Burger King in their “Whopper Sacrifice” campaign, used a Facebook app to generate a free coupon for a whopper, only after the user “de-friended” ten Facebook friends. Social networking website Flixster is using a Facebook app to draw attention to its own website by providing a limited set of the functionality of its full website within its Facebook app.

The diffusion of these applications on Facebook makes it a great testing platform for the proposed methodology for several reasons. First, the social component significantly contributes

¹⁸ For the impact of truncation on network recovery, please see online Appendix E.

to the adoption process for many of these apps. Second, we have access to adoption data for a large number of apps, diffusing over subsets of the same underlying potential network of Facebook users. Third, Facebook apps vary considerably in terms of appeal and virality, facilitating a good exploration of the unobserved parameter space. Finally, while Facebook does not reveal detailed characteristics of its member network, several studies offer some evidence about its actual structure, allowing a test of the face validity of our results.

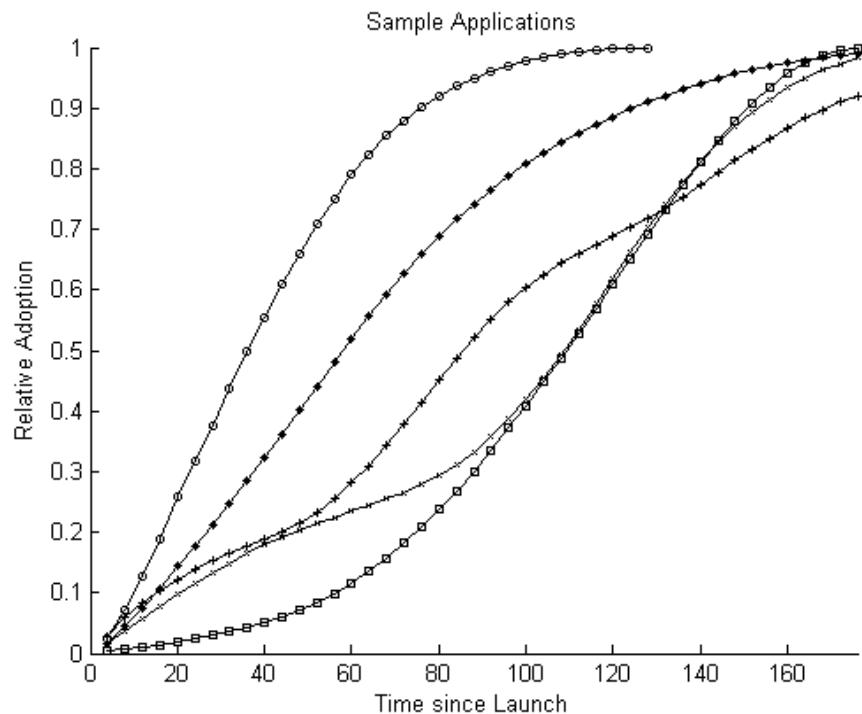


Figure 9. Illustrative Cumulative Adoptions for Five Facebook Apps.

Data. Our data consists of daily cumulative installs, for a sample of about 900 Facebook applications, between June 25, 2007 and February 23, 2008. These apps showed at least four consecutive months of daily performance data (installs). The time series were pre-processed for missing observations and irregularities related to sampling. Final sample data windows have a mean of 158.96 (with a standard deviation of 18.2). Applications belong to various categories, with the two most popular categories being “Games” (36%) and “Utilities” (27%). Total cumulative number of installs varied between 1,400 and 20,720,000 with a mean of 381,959 and

median of 16,183. Figure 9 shows the normalized cumulative adoptions for five sample apps.

For each app, we consider its adoption curve based on the time series of installs and using this data estimate the Bass model (a la Srinivasan and Mason 1986) to obtain p , q and m . We then bin p -and- q pairs into 110 intervals to create the outcome variable Y . Using the inference procedure from the previous section, we (1) estimate posterior probabilities for each of 16 possible topology/density combinations to establish the highest probability of the underlying network, and (2) use a calibration sample to forecast a distribution of diffusions in a holdout sample.

Inferring Facebook Apps Network Characteristics. Based on the entire sample of 900 Facebook apps, our proposed method provides very strong support ($p(M_k | Y) \rightarrow 1$) for the underlying potential network being a low density, preferential attachment topology. This is an encouraging outcome, since recent studies using Facebook apps data and having access to the actual network of adopters report similar results (Nazir et al. 2008). Moreover, many social networks have been shown to exhibit preferential attachment characteristics (Barabási and Albert 1999, Stephen and Toubia 2009). While a low network density (low degree distribution) may first appear contradictory to the Facebook reported statistics of average number of friends (connections) being 130 (Facebook.com 2009), our inference applies to the potential network of app users, rather than to the potential network of Facebook friends. Using the analogy of an online publisher, we are looking at the potential network of readership (i.e., Facebook app users) and not at the entire population of the U.S. (i.e., all Facebook members). Our results are consistent with existing studies showing that only a small fraction of all friendship connections on some major social networking sites can be considered “important” or “influential” (e.g., Trusov et al. 2010, Aral and Walker 2012, Goel et al. 2012).

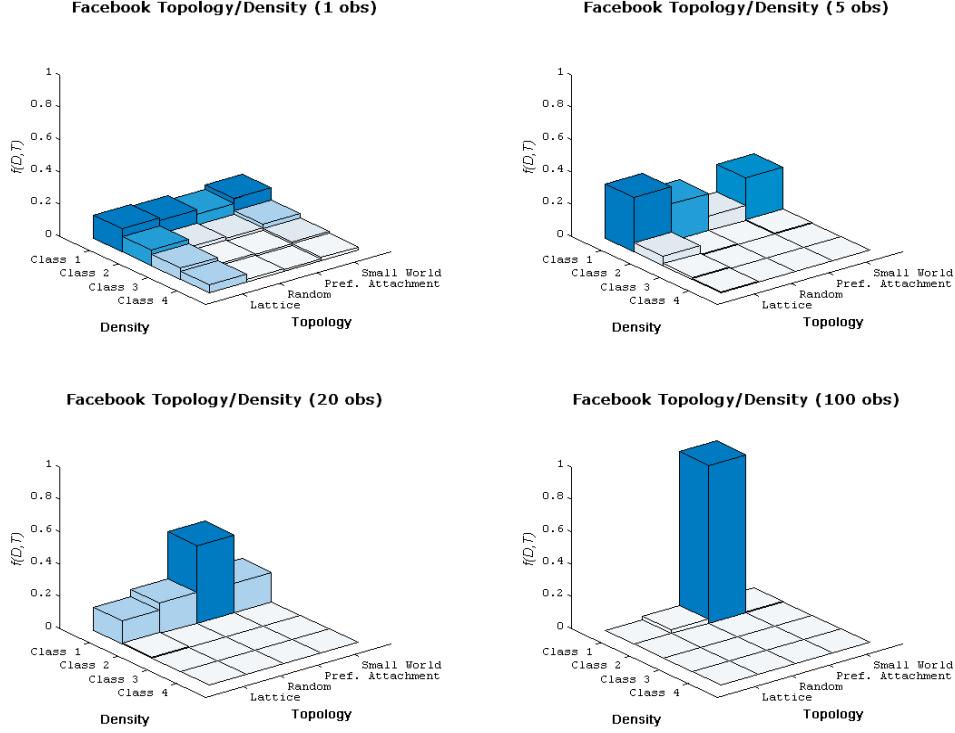


Figure 10. Inferring the Potential Network Structure for Facebook Apps.

While we observe strong convergence ($p(M_k | Y) \rightarrow 1$) to a single network structure with a sample of 900 applications, results for smaller samples are less definite. Figure 10 depicts how posterior probabilities for the 16 topology/density combinations evolve with increasing sample size. Low density class becomes a dominating outcome with as many as just 5 observations; however the network topology does not become pronounced until after 20 observations. In other words, randomly drawing from a sample of 900 products could generate different posterior distributions, but distributions are increasingly consistent as we increase sample size. This further supports our proposition that due to a number of unobserved factors influencing the diffusion process, prediction based on a small number of observations produces noisy results.

While our inferred network structure is congruent with the actual Facebook network reported in previous studies, we reemphasize that the proposed method is *not* meant to converge to the characteristics of the actual network. More generally, this result should be seen as a best

approximation or projection of the actual network into the given space of synthetic networks. As a next step, we examine how our method performs with out-of-sample predictions.

Predicting a Distribution of Diffusion Curves on a Holdout Sample. To predict a distribution of future diffusion curves over the same underlying potential network, we use equation (6). In keeping with the Bayesian tradition, we want our predictions to be based on both the observed data and prior beliefs about the underlying network structure, which in our case are informative and driven by an individual level adoption model. As competing models, we consider two alternatives. The first model, which we call the calibration model, is simply an empirical distribution (histogram of Bass model p and q) in the calibration sample. It is well known in statistics that as the number of observations in a calibration sample grows, the empirical distribution converges to the parent distribution (Glivenko-Cantelli theorem). Hence we expect this first benchmark to do reasonably well on large samples, but show an inferior performance on smaller samples. The second model, which we call the naïve model, is a simplified version of the proposed model. It assigns equal weights to all 16 possible density/topology combinations in equation (6). Hence, for the naïve model, equation (6) becomes:

$$p(Y^* | Y) = \frac{1}{16} \times \sum_k \int_{\theta_k} \psi(Y^* | \theta_k) \times \phi(\theta_k | M_k, Y) d\theta_k. \quad (9)$$

A performance comparison between our proposed model and the naïve model is important, since the difference between the two models allows us to test whether the inferred network footprint (i.e., $\chi(Z | Y)$) adds any value to the prediction. In other words, the existence of such a difference validates our approach as a method of identifying systematic patterns across diffusions, which can be attributed to network effects. Specifically, Equation (9) combines the empirical distribution of the calibration sample with a synthetic distribution of diffusion parameters where the network characteristics are essentially ignored (or integrated out). In

contrast, the proposed model assigns higher weights to the synthetic priors that correspond to the network density/topology combinations identified as more probable given the calibration data.

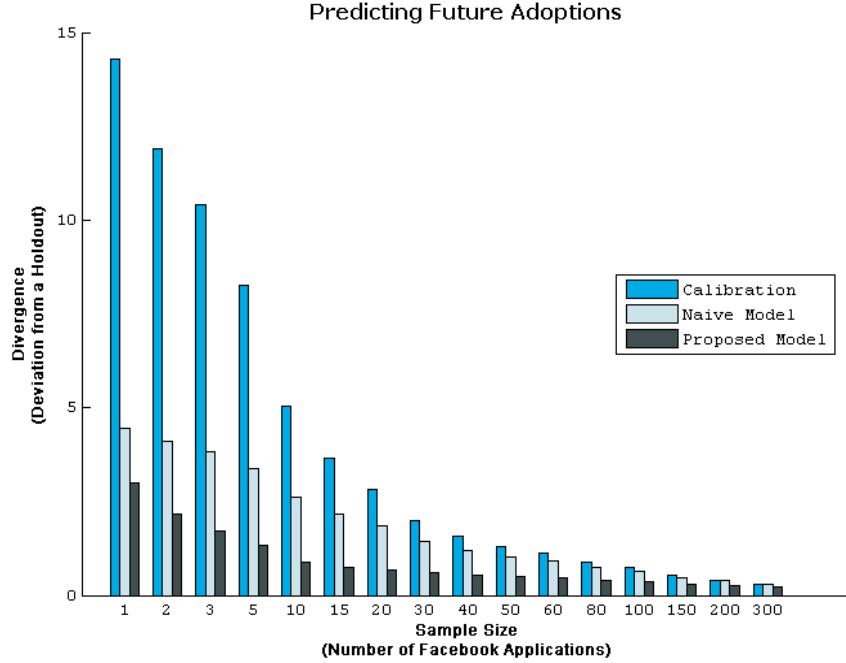


Figure 11. Forecasting Facebook Diffusions – Model Comparison.

Since the outcome of this prediction task is not a point estimate, but rather a probability distribution, we adopt a metric which allows comparison of functional data. Common to such applications is the Kullback-Leibler (KL) symmetric divergence measure. We calculate the distance between the predicted probability density function derived by the alternative models and the empirical probability density from the holdout sample. We use diffusion parameters of 600 randomly selected applications as a holdout sample, and draw random calibration sets of varying size from the remaining 300 applications. Finally, we average KL measures across multiple draws of the same sample size. From Figure 11, the proposed model considerably outperforms the other two models. Both the naïve and the proposed model perform significantly better than the calibration model for all sample sizes, with the difference in performance gradually decaying

as sample size approaches 300.¹⁹

Comparison of the Proposed Approach with Alternative Models for Pre-Launch Forecasting. We now test the performance of our proposed approach against three popular diffusion models: (i) the Bass model (Bass 1969), (ii) the Gamma/Shifted Gompertz model (Bemmaor 1994, Bemmaor and Lee 2002), and (iii) the Weibull-Gamma model (Moe and Fader 2002). In generating pre-launch forecasts, on the one hand, estimating market size (m) is an important managerial concern alongside forecasting the dynamics of diffusion. On the other hand, past research suggests that when forecasting before launch, it is better to estimate m outside the model (e.g., Trajtenberg and Yitzhaki 1989, Van den Bulte and Lilien 1997, Lilien et al. 2000). Mahajan et al. (1993, p. 391) review several past studies and conclude “... applications of diffusion models report better forecasting results by using exogenous sources of information (such as market surveys, secondary sources, management judgments, or other analytical models) for estimating m .” Further, from their research on assessing estimation bias, Van den Bulte and Lilien (1997, p. 349) “caution against estimating market potential using aggregate level diffusion models.”

In light of these issues, we first generate pre-launch forecasts when m is estimated within the proposed and benchmark models and then generate pre-launch forecasts when m is estimated outside these models.²⁰ For the benchmark models, we use intercept-based forecasts to predict diffusion parameters in the holdout (Lenk and Rao 1990). For our proposed approach, the inference procedure is analogous to that presented in the previous section, and uses Equation (6)

¹⁹ For detailed fit statistics comparing these three models, please refer to online Appendix G.

²⁰ We briefly explored the possibility of including product attributes for generating forecasts for market size. To do so, we followed the approach in Hahn et al. (1994). Unfortunately, these initial explorations revealed poor model fits, suggesting a more careful identification of relevant attributes was necessary. For the details of this study and our findings, please see online Appendix B. For analysis within genres, please see online Appendix H.

to generate an informative (synthetic) prior distribution for p and q that we combine with diffusion parameters estimated from the calibration sample. Posterior means of p , q and m are used to generate forecasted diffusion curve. We use 150 randomly selected applications as a holdout sample, and draw random calibration sets of varying size from the remaining applications. From each calibration sample we generate four forecasts corresponding to the proposed model and three benchmark models. We compare our forecasted values with the actual values observed in the holdout sample and calculate standard performance statistics such as MAD, MAPE and MSE for each curve. Since the results of this analysis are sensitive to sample composition (Lenk and Rao 1990), we repeat the above steps 300 times and report average MAPE for the four models in Table 3, with (left) and without (right) prediction of market size.²¹

From Table 3, when market size is estimated from historical data, model fits are quite poor across the board. For example, MAPE varies between 1280% and 1397.1% when the calibration sample is of size one. We attribute these large errors to an extreme heterogeneity in market penetration across apps. For example, when an application with 5 million installations is used as a calibration sample to predict market potential for applications that never reach 10K downloads the forecast errors can get quite dramatic. This is not surprising, as similar issues with data homogeneity have been reported and discussed in previous research (Lenk and Rao 1990). The forecast errors go down as calibration sample size increases, and we are able to get a more accurate estimation of average market size. Nevertheless, on a comparative basis, our proposed approach performs somewhat better when sample sizes are moderate. Our analysis also shows that accounting for only the average degree (density) or only topology provides approximately 2/3rd of the benefit of using both, as measured in terms of improvement in MAPE over the

²¹ For the average MAD and MSE scores, please refer to the online Appendix D.

baseline Bass model forecast.²²

Calibration Sample Size	MAPE - m estimated				MAPE - m known			
	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma
1	1391.2%	1364.0%	1397.1%	1280.0%	29.6%	37.9%	37.8%	49.2%
3	438.5%	460.0%	509.6%	749.6%	26.7%	33.0%	35.9%	47.9%
5	283.1%	306.2%	280.4%	543.7%	25.4%	30.5%	35.3%	46.9%
7	201.8%	217.7%	224.1%	310.0%	25.3%	29.6%	35.0%	44.4%
9	171.0%	186.6%	181.6%	233.1%	24.9%	28.8%	32.9%	42.0%
11	138.3%	147.5%	157.6%	209.7%	24.7%	28.3%	32.4%	38.4%
13	120.2%	128.7%	132.7%	186.6%	24.6%	28.1%	31.7%	37.6%
15	120.2%	128.5%	134.3%	163.2%	24.5%	27.8%	31.7%	37.2%
21	105.1%	111.3%	120.2%	149.1%	24.3%	27.4%	29.3%	34.4%
35	98.3%	102.2%	104.9%	129.2%	24.1%	26.8%	26.9%	30.5%
50	96.2%	99.0%	101.5%	125.3%	24.0%	26.7%	25.7%	29.3%
75	96.0%	98.1%	100.0%	118.2%	23.9%	26.4%	25.1%	28.8%

Table 3: Comparing Model Performance Against Benchmarks.

Given that market size is the likely culprit contributing most to forecast errors, we now compare performance of the four models when market size is exogenous to model estimation.²³

From a comparison of the right side columns of Table 3, the results strongly suggest that our proposed approach using a synthetic prior based on latent network characteristics improves forecast accuracy. Our proposed approach notably outperforms alternative models when the calibration sample size is low, showcasing the value added by incorporating network footprint in generating forecasts.²⁴

²² Please refer to online Appendix I for the details regarding this analysis.

²³ Specifically, for each model, we set market size equal to maximum observed value. While our data does not suggest existence of truncated calibration samples, recent work by Van den Bulte and Iyengar (2011) shows that the accuracy of parameter estimates can be further improved by appropriately accounting for truncation, when it exists.

²⁴ From Table 3, our model performance converges at ~ 25 apps. This is in contrast to the analysis summarized in Figure 7, where we assess the extent to which we correctly identify the underlying network (that task imposes intensive data requirements: ~100 apps for ~80% recovery in simulation). However, data requirements are not as intense when our objective is to predict a future diffusion over potential network types, as in Table 3. This is

Figure 12 provides graphical evidence of the forecasting ability of our proposed approach. It shows the forecasted curve (solid line) using our proposed approach on a calibration sample of 5 apps, for a holdout sample of 300 apps. Also shown is the median curve in the holdout sample, and 10/90 percentile error bounds.²⁵

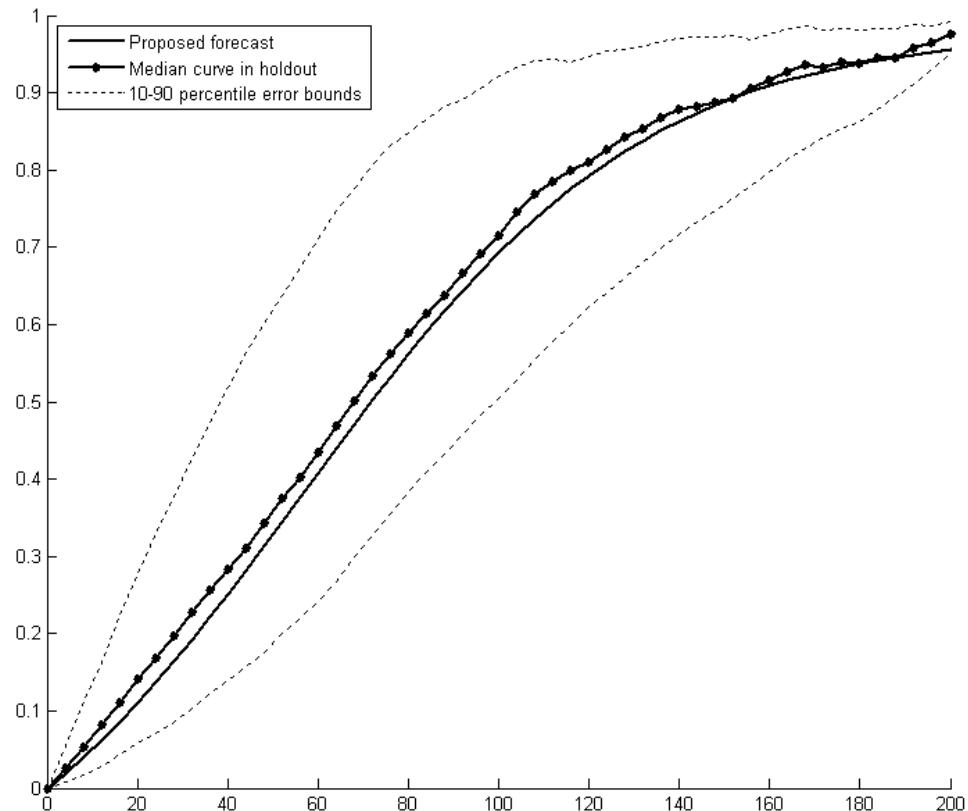


Figure 12. Graphical Illustration of Pre-Launch Forecasts using the Proposed Approach.

DISCUSSION AND CONCLUSIONS

While it is generally acknowledged that social interactions play an important role in product adoption, incorporation of the underlying consumer network information into forecasting

because we leverage posterior beliefs across all possible underlying network types (i.e., all 16 combinations of topologies and densities we consider) to generate predictions. We thank an anonymous reviewer for pointing out this interesting difference.

²⁵ Please see online Appendix F for a graphical illustration of forecasts across benchmark models.

models is still a rather uncommon practice. With the objective of generating improved pre-launch forecasts, in this paper we argue that the product diffusion process is affected by a number of factors both systematic and non-systematic, and hence a model calibrated on a single diffusion instance may face transferability issues when applied to a product that is introduced to the market under different non-systematic conditions. Hence, we propose an approach that focuses on identifying market conditions that are relatively stable (or common to all products) by exploring systematic patterns *across* diffusions rather than *within* a diffusion. We extract this systematic effect from historical data by matching the distribution of historical diffusions with a set of synthetic footprints obtained through simulation on a set of networks of known types.

Our primary contribution is to demonstrate how incorporating network structure can assist managers by improving predictions of future diffusion dynamics of potentially very different products across subsets of the same network. We do so by empirically mapping observed prior diffusions to a selected space of synthetic network footprints. Methodologically, we show how stochastic relationships simulated in complex systems can serve as informative priors for Bayesian inference. Our approach neither confirms or disconfirms the presence of *network* effects in the data. Rather, we show that the systematic component extracted from historical data *according* to network theory helps us improve pre-launch forecasts. Hence, whatever systematic component we identify, it mirrors the impact of an underlying network structure as simulated from a pre-specified set of networks. If the actual network is one of the pre-specified types and our representation of the adoption process is accurate, then the characteristics of the identified synthetic network may indeed match the actual network. However, more generally, the result should be seen as the best projection of the actual network into a given space of synthetic networks. We hope that future research will explore other

potential candidates for systematic effects (e.g., consumer preference heterogeneity) and their application in forecasting tasks. The original approach developed in this paper may serve as a platform for such research inquiries.

The novelty of our method is that it employs an informative prior constructed solely using a pre-specified individual level adoption model. To our knowledge this is the first paper to propose the use of synthetic priors generated through agent-based modeling to inform Bayesian inference of model selection. We have illustrated our method using historical diffusion data for apps on the Facebook network. Our proposed method performs well in predicting future diffusion when the market potential is known, and compares favorably against a number of competing models. It relies on an improved inference of the underlying network for improving forecasts. Recent work (Salganik et al. 2006) suggests that the ability to predict diffusion success decreases with the level of virality, which in turn is likely to be impacted by the underlying network structures that are central to our work. Consequently, the comparative advantage of our approach could potentially vary across categories that differ in terms of their virality. Indeed, this advantage might be stronger for categories that are more viral; we leave the exploration of this conjecture as a potential avenue for future research.

In generating pre-launch forecasts, along with diffusion dynamics it is important to generate estimates for final penetration. We achieve reasonable forecasting performance with respect to diffusion dynamics only if we rely on exogenous information of final penetration. While this is consistent with existing research (e.g., Trajtenberg and Yitzhaki 1989, Van den Bulte and Lilien 1997, Lilien et al. 2000), from a practical viewpoint, this is an important first limitation, and a limitation that our approach shares in common with existing forecasting approaches. To overcome this limitation, just as we construct a hypercube for shape parameters,

relying on theory driven arguments for how the underlying network determines market potential one can extend our approach to include market size as an additional parameter in the hypercube. We leave this extension as an interesting direction to explore in future research. The benefit of keeping penetration exogenous is that our approach can be used across distinct products that could vary significantly in market potential but diffuse across networks having similar underlying characteristics.

In our analysis, we have assumed homogenous adoption propensities for a given product across individuals within synthetic networks. Future research can consider heterogeneity across consumers as well as products. Given the findings of Iyengar et al. (2011), we conjecture that heterogeneity in consumer adoption rates combined with variation in their potential locations within adoption networks can provide a richer prior of synthetic networks, perhaps enhancing forecasting abilities. Along the same lines, different products could generate interest from distinct segments of the entire network, hence we conjecture that investigating the link between characteristics of the sub-networks participating in the diffusion process and characteristics of new products can serve as an alternate avenue for improving forecasts.

The accuracy of our inference depends on the representativeness of the empirical sample. For example, if there is overrepresentation in the empirical sample of diffusion under certain market conditions or products of certain characteristics (dimensions which are being integrated out in the simulated data) then the inference about network characteristics can be biased. This shortcoming can be addressed by introducing additional data to the model that corrects for the unbalanced sample or can be incorporated into the theoretical model used to generate the simulation-based prior. With increasing model complexity, and hence with a larger number of unobservable factors needing to be integrated out, it is likely that the variance of the conditional

distribution $g(Z|Y)$ may increase, which in turn may result in the need for a larger sample size in order to maintain accuracy.

Another limitation of our approach is that in its current form, it does not account for product characteristics. In some contexts, product characteristics can play an important role in determining overall adoption (e.g., Hahn et al. 1994, Berger and Milkman 2012), including network formation (e.g., people who like certain products might form networks of certain structure). In principle, our proposed approach is flexible enough to permit a dependence between X and Z . However, since we do not have an a-priori theoretical basis for hypothesizing how such interactions might unfold in our empirical context, we leave the exploration of this interesting possibility as an area for future research.

REFERENCES

- Anderson, Roy M. and Robert M. May (1992), *Infectious Diseases of Humans: Dynamics and Control*. New York: Oxford University Press.
- Aral, Sinan and Dylan Walker (2012), "Identifying Influential and Susceptible Members of Social Networks," *Science*, 337, 337-341.
- Barabási, Albert-László and Réka Albert (1999), "Emergence of scaling in random networks," *Science*, 286, 509-512.
- Bass, Frank M. (1969), "A New Product Growth Model for Consumer Durables," *Management Science*, 36 (9), 1057-1079.
- Bemmaor, Albert C. (1994), "Modeling the Diffusion of New Durable Goods: Word-of-Mouth Effect Versus Consumer Heterogeneity", in *Research Traditions in Marketing*, Gilles Laurent, Gary L. Lilien, and Bernard Pras, eds. Boston: Kluwer Academic Publishers, 201–223.
- and Janghyuk Lee (2002), "The Impact of Heterogeneity and Ill-conditioning on Diffusion Model Parameter Estimates," *Marketing Science*, 21(2), 209–220.
- Berger, Jonah and Katherine L. Milkman (2012), "What Makes Online Content Viral," *Journal of Marketing Research*, 49(2), 192-205.
- Bradlow, Eric T. (2002), "Exploring repeated measures data sets for key features using Principal Components Analysis," *International Journal of Research in Marketing*, 19(2), 167-179.
- Braun, Michael and André Bonfrer, (2011), "Scalable inference of customer similarities from interactions data using Dirichlet processes," *Marketing Science*, 30(3), 513-531.
- Chandrasekaran, Deepa and Gerard J. Tellis (2007), "A Critical Review of Marketing Research on Diffusion of New Products", in *Review of Marketing Research*, Volume 3, Naresh K. Malhotra ed. Armonk: Emerald Group Publishing, 39-80.
- Chipman, Hugh, Edward I. George, and Robert E. McCulloch (2001), "The Practical Implementation of Bayesian Model Selection," in *IMS Lecture Notes – Monograph Series: Model Selection, Vol 38*, P. Lahiri ed. Beachwood: Institute of Mathematical Statistics, 65-116.
- Dasgupta, Koustuv, Rahul Singh, Balaji Viswanathan, Dipanjan Chakraborty, Sougata Mukherjea, Amit A. Nanavati, and Anupam Joshi (2008) "Social ties and their relevance to churn in mobile telecom networks." In *Proceedings of the 11th international conference on Extending database technology: Advances in database technology*, ACM 668-677.
- Dodds, Peter S. and Duncan J. Watts (2005), "A generalized model of social and biological contagion," *Journal of Theoretical Biology*, 232(4), 587-604.
- Dover, Yaniv, Jacob Goldenberg, and Daniel Shapira (2012), "Network Traces on Penetration: Uncovering Degree Distribution from Adoption Data," *Marketing Science*, 31(4), 689-712.
- Facebook.com (2009), "Press Room: Statistics," <http://www.facebook.com/press/info.php?statistics>, accessed on December 4, 2009.
- Fienberg, Stephen. E. and Paul W. Holland (1973), "Simultaneous Estimation of Multinomial Cell Probabilities," *Journal of the American Statistical Association*, 68(343), 683-691.

- Goel, Sharad, Duncan J. Watts, and Daniel G. Goldstein (2012), "The Structure of Online Diffusion Networks," In *Proceedings of the 13th ACM Conference on Electronic Commerce*, 623-638.
- Goldenberg, Jacob, Barak Libai, and Eitan Muller (2001), "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Marketing Letters*, 12(3), 211-223.
- , --- and --- (2002), "Riding the Saddle, How cross-Market Communications Creates a Major Slump in Sales," *Journal of Marketing*, 66(April), 1-16.
- , Sangman Han, Donald R. Lehmann, and Jae Hong (2009), "The Role of Hubs in the Adoption Process," *Journal of Marketing*, 73 (March), 1-13.
- Hahn, Minhi, Sehoon Park, Lakshman Krishnamurthi, and Andris A. Zoltners (1994), "Analysis of new product diffusion using a four-segment trial-repeat model," *Marketing Science*, 13(3) 224-247.
- Haenlein, Michael (2011), "A Social Network Analysis of Customer-level Revenue Distribution," *Marketing Letters*, 22(1), 15-29.
- and Barak Libai (2013), "Targeting Revenue Leaders for a New Product," *Journal of Marketing*, 77(May), 65-80.
- Hill, Shawndra, Foster Provost, and Chris Volinsky (2006), "Network-Based Marketing: Identifying Likely Adopters via Consumer Networks," *Statistical Science*, 22(2) 256-275.
- Iacobucci, Dawn, and Nigel Hopkins (1992), "Modeling Dyadic Interactions and Networks in Marketing," *Journal of Marketing Research*, 29(1), 5-17.
- Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente (2011), "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30(2), 195-212.
- Keeling, Matthew J. (1999), "The Effects of Local Spatial Structure on Epidemiological Invasions," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1421) 859-867.
- Kempe, David, Jon Kleinberg, and Eva Tardos (2003), "Maximizing the Spread of Influence Through a Social Network," In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 137-146.
- Lemmens, Aurelie, Christophe Croux, and Stefan Stremersch (2012), "Dynamics in international market segmentation of new product growth," *International Journal of Research in Marketing*, 29(1), 81-92.
- Lenk, Peter J. and Ambar G. Rao (1990), "New Models from Old: Forecasting Product Adoption by Hierarchical Bayes Procedures," *Marketing Science*, 9, 42-53.
- Lewis, Ted G. (2011), *Network Science: Theory and Applications*. Hoboken: John Wiley & Sons.
- Libai, Barak, Eitan Muller and Renana Peres (2005), "The Role of Seeding in Multi-Market Entry," *International Journal of Research in Marketing*, 22(4), 375-393.
- , --- and --- (2009), "The Role of Within-Brand and Cross-Brand Communications in Competitive Growth," *Journal of Marketing*, 73(3), 19-34.
- Lilien, Gary L., Arvind Rangaswamy and Christophe Van den Bulte (2000), "Diffusion Models: Managerial Applications and Software," in *New Product Diffusion Models*, Vijay Mahajan, Eitan Muller, and Yoram Wind, eds. Norwell: Kluwer Academic Publishers, 295-311.

--- and --- (2004), *Marketing Engineering: Computer-Assisted Marketing Analysis and Planning*. Victoria: Trafford Publishing.

Mahajan, Vijay, Eitan Muller, and Frank M. Bass (1993), "New Product Diffusion Models," in *Handbooks in OR & MS, Vol. 5*, Jehoshua Eliashberg and Gary L. Lilien, eds. Amsterdam: Elsevier Science Publishers B.V., 349–408.

Manchanda, Puneet, Grant Packard, and Adithya Pattabhiramaiah (2012), "Social Dollars: The Economic Impact of Customer Participation in a Firm-sponsored Online Community," *Working Paper*, Available at SSRN 1984350.

Moe, Wendy W. and Peter S. Fader (2002), "Using Advance Purchase Orders to Forecast New Product Sales," *Marketing Science*, 21, 347–364.

Moldovan, Sarit and Jacob Goldenberg (2003), "Cellular Automata Modeling of Resistance to Innovations: Effects and Solutions," *Technological Forecasting and Social Change*, 71(5), 425-442.

Nazir, Atif, Saqib Raza, and Chen-Nee Chuah (2008), "Unveiling Facebook: A Measurement Study of Social Network Based Applications," In *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, ACM, 43-56.

Newman Mark E. J. and Duncan J. Watts (1999), "Scaling and percolation in the small-world network model," *Physical Review E*, 60(6) 7332-7342.

Nitzan, Irit and Barak Libai (2011), "Social Effects on Customer Retention," *Journal of Marketing*, 75(6), 24-38.

Rahmandad, Hazhir and John Sterman (2008), "Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models," *Management Science*, 54, 998-1014.

Rand, William M. and Roland T. Rust (2011), "Agent-based modeling in Marketing: Guidelines for Rigor," *International Journal of Research in Marketing*, 28(3), 181-193.

Ramsay, Jim O. and Bernard W. Silverman (2005), *Functional Data Analysis*. New York: Springer.

Rogers, Everett M. (1962), *Diffusion of Innovations*. New York: The Free Press.

Ryan, Bryce and Neal C. Gross (1943), "The diffusion of hybrid seed corn in two Iowa communities," *Rural Sociology*, 8(1), 15-24.

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

Shaikh, Nazrul I., Arvind Rangaswamy, and Anant Balakrishnan (2005), "Modeling the diffusion of innovations using small-world networks," *Working Paper*, Pennsylvania State University.

Sood, Ashish, Gareth M. James, and Gerard J. Tellis (2009), "Functional Regression: A New Model for Predicting Market Penetration of New Products," *Marketing Science*, 28(1), 36-51.

Srinivasan, V., and Charlotte H. Mason (1986), "Technical Note - Nonlinear Least Squares Estimation of New Product Diffusion Models," *Marketing Science*, 5(2), 169-178.

Stephen, Andrew T., and Olivier Toubia (2009), "Explaining the Power-Law Degree Distribution in a Social Commerce Network," *Social Networks*, 31(4), 262-270.

Stremersch, Stefan, and Aurelie Lemmens (2009), “Sales Growth of New Pharmaceuticals Across the Globe: The Role of Regulatory Regimes,” *Marketing Science*, 28(4), 690-708.

Trusov, Michael, Anand Bodapati, and Randolph E. Bucklin (2010), “Determining Influential Users in Internet Social Networks,” *Journal of Marketing Research*, 47(4), 643-658.

Toubia, Olivier, Jacob Goldenberg, and Rosanna Garcia (2008), “A New Approach to Modeling the Adoption of New Products: Aggregated Diffusion Models,” *Marketing Science Institute Working Paper Series*, No. 08-001.

---, --- and --- (2009), “Diffusion Forecast Using Social Interactions Data,” *Marketing Science Institute Working Paper Series*, No. 09-210.

Trajtenberg, Manuel and Shlomo Yitzhaki (1989), “The Diffusion of Innovations: A Methodological Reappraisal,” *Journal of Business & Economic Statistics*, 7(1) 35-47.

Valente, Thomas W. (1995), *Network Models of the Diffusion of Innovations*. Creskill: Hampton Press.

Van den Bulte, Christophe (2010), “Opportunities and challenges in studying customer networks,” in *The Connected Customer: The Changing Nature of Consumer and Business Markets*. Stefan H. K. Wuyts, Marnik G. Dekimpe, Els Gijsbrechts, and F. G. M. Rik Pieters, eds. London: Routledge, 7-35.

--- and Gary L. Lilien (1997), “Bias and Systematic Change in the Parameter Estimates of Macro-level Diffusion Models,” *Marketing Science*, 16, 338–353.

--- and --- (2001), “Medical Innovation Revisited: Social Contagion versus Marketing Effort,” *American Journal of Sociology*, 106(5), 1409-35.

--- and Stefan Stremersch (2004), “Social contagion and income heterogeneity in new product diffusion: A meta-analytic test,” *Marketing Science*, 23, 530-544.

--- and Yogesh V. Joshi (2007), “New Product Diffusion with Influentials and Imitators,” *Marketing Science*, 26(3), 400-421.

--- and Stefan H. K. Wuyts (2007), *Social Networks and Marketing*. Cambridge: MSI.

--- and Raghu Ram Iyengar (2011), “Tricked by Truncation: Spurious Duration Dependence and Social Contagion in Hazard Models,” *Marketing Science*, 30(2), 233-248.

Wasserman, Stanley and Katherine Faust (2009), *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

Watts, Duncan J. (2002), “A simple model of global cascades on random networks,” *Proceedings of the National Academy of Science*, 99(9), 5766-5771.

--- and Peter S. Dodds (2007), “Influentials, Networks and Public Opinion Formation,” *Journal of Consumer Research*, 34(4), 441-458.

--- and Steven H. Strogatz (1998), “Collective dynamics of ‘small-world’ networks,” *Nature*, 393(6684), 440-442.

Online Appendix
for
IMPROVING PRE-LAUNCH NEW PRODUCT FORECASTS:
USING SYNTHETIC NETWORKS AS SIMULATED PRIORS

Contents:

Appendix A – Social Network Models

Appendix B – The Role of App Characteristics in Forecasting

Appendix C – Additional Diagrams of the Diffusion Hypercube

Appendix D – Fit Statistics for Pre-launch Forecasting

Appendix E – Network Recovery with Simulated Truncated Data

Appendix F – Graphical Evidence of Forecasting Ability

Appendix G – Comparing Proposed Approach to Naïve and Bass Model

Appendix H – Impact of Product Heterogeneity

Appendix I – Impact of Specific Network Characteristics

Appendix A: Social Network Models

The four networks analyzed in this paper: *lattice*, *random*, *small world*, and *preferential attachment*, are popular candidates in the network science and social network analysis literature for the purposes of investigating several network-based phenomena (e.g., Erdos and Renyi 1960, Watts and Strogatz 1998, Barabasi and Albert 1999, Kiesling et al. 2012). In a recent review of the literature on diffusion within networks, Kiesling et al (2012) report a majority of the past research using one or more of the four networks we consider, with small world and lattice networks being the most represented. The widespread popularity of these four networks stems from the fact that they span the spectrum of properties that in general characterize social networks. In this Appendix, we (i) describe these properties; (ii) discuss each network structure in the context of these properties (qualitatively and quantitatively); and (iii) compare diffusion across networks given differences in these properties.

A.1 Network Properties

Many properties influence diffusion in a social network (e.g., Wasserman and Faust 1994, Scott 2000, Van den Bulte and Wuyts 2007). We describe four key properties below:

- **degree (d_i):** the number of direct connections of node i . The distribution of d_i across the nodes affects the diffusion process. Intuitively, higher the average degree of the network, \hat{d} , faster the spread across the network.
- **edge density (e):** the proportion of existing edges, relative to all possible edges. On an undirected network, $e = \frac{\hat{d}}{(N-1)}$, where N denotes number of nodes in the network.

- **path length** ($l_{i,j}$): the number of steps on the shortest path between nodes i and j . Average path length, also known as characteristic path length is $\hat{l} = \frac{1}{N(N-1)} \sum_{i,j} l_{i,j}$. Note, $l_{i,j} = 0$ if $i=j$. Intuitively, shorter the average path length, quicker the diffusion across the network.
- **clustering coefficient** (c_i): the ratio of existing links connecting the node's neighbors to each other, to the maximum possible number of such links. For nodes with fewer than two neighbors the clustering coefficient is undefined. Informally, this measures the extent to which friends of a node's friends are that node's friends. $c_i = \frac{2e_i}{d_i(d_i-1)}$, where e_i is the number of edges between neighbors of i . The clustering coefficient for the entire network is the average of the clustering coefficients of all nodes in the network.

Past research provides evidence that these properties can play a key role in understanding the diffusion within networks. For instance, Watts and Strogatz (1998) show that a combination of clustering coefficient and path length is necessary to understand the spread of disease across a network. Stonedahl, Rand and Wilensky (2010) show that degree alone is insufficient for maximizing diffusion in real-world social networks, and improvements are observed once clustering coefficient is taken in to account. Given that these properties do play a role, next we discuss in detail the four networks we analyzed in the context of these properties.

A.2 Network Structures

For each network, we describe its formation, characterize its properties, provide a typical example, and illustrate it visually. The numerical values for average properties for all topologies, densities, and levels of appeal used in the analysis in the paper are reported in Tables A1-A4. The formation of each network begins with specifying N, the number of nodes

in the network.

Lattice / Regular Networks: In this network, each individual is connected to exactly d other individuals; and $\{d, N\}$ fully specify the network. The average degree of the network is exactly the degree of any individual, and degree variance is 0. The average path length is large, relative to other topologies (Watts and Strogatz, 1998). The average clustering coefficient is high relative to other topologies, and independent of system size (Watts and Strogatz, 1998). An example of a lattice network is a neighborhood of families where each neighbor has exactly the same number of neighbors as any other neighbor. In our analysis, $d = 2, 4, 6$, and 8 . Figure A1 visually illustrates a lattice network of $N=50$ and $d = 8$ on the left, with a zoomed in view of local network structure on the right.

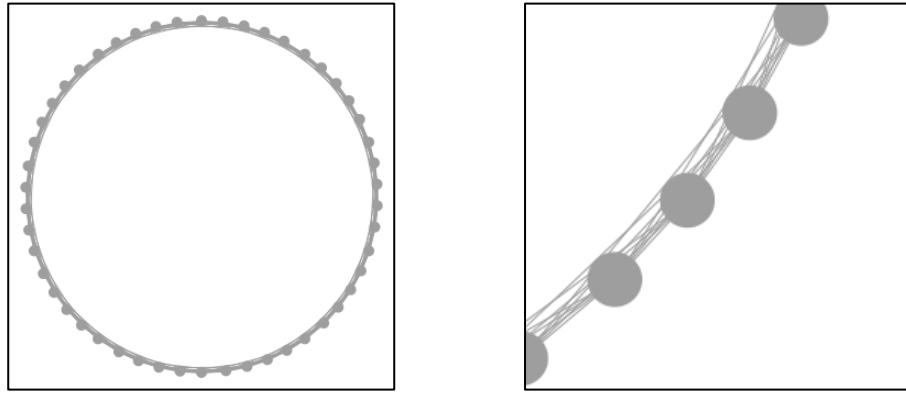


Figure A1: Example of a Lattice/Regular Network Structure.

Random Networks: In this network, an edge exists between any two nodes with probability p , and $\{p, N\}$ fully specify the network. The degree distribution is Poisson for large N . The average path length and the average clustering coefficient is low relative to the other topologies (Erdos and Renyi, 1960). The random model is a uniform mixing model that assumes everyone has an equal probability (p) of meeting everyone else. An example of a random network is connections that might arise at a cocktail party if no one knew anyone else

beforehand and everyone was equally likely to talk to everyone else. In our analysis, $d = 2, 4, 6$, and 8 . Figure A2 visually illustrates a random network of $N=50$ and $d = 8$ on the left, with a zoomed in view of local network structure on the right.

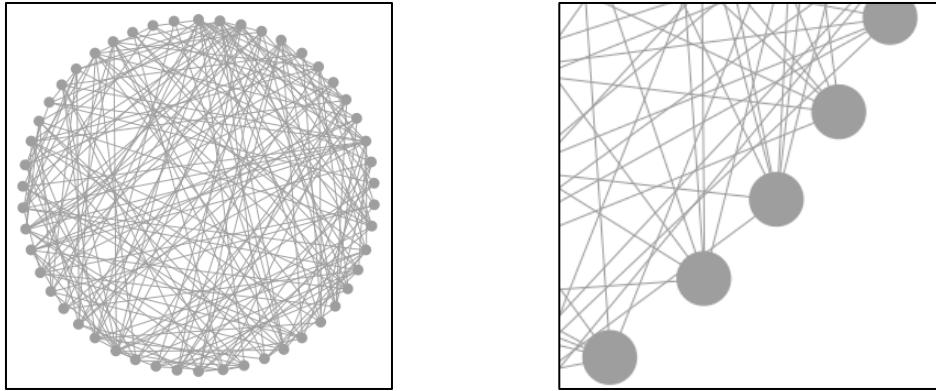


Figure A2: Example of a Random Network Structure.

Small World Networks: This network is first initialized as a lattice. Next, with probability B , links are randomly rewired, i.e., a connection between nodes i and j is replaced with a connection between node i and another random node k (Watts and Strogatz, 1998). This network has the unique property that if $B = 0$ it collapses to the lattice network, and if $B = 1$ it becomes the random network. Thus, $\{d, N, B\}$ fully specify the network. The small world network has a high clustering coefficient, but a relatively low average path length and the degree distribution is similar to that of the random graph especially as B increases close to 1. Examples of small world networks include the network of film actors who have starred together in a movie (Watts and Strogatz, 1998); the network of email connections given a target (Dodds et al., 2003); and connections between file sharing users (Iamnitchi et al., 2004). In our analysis, $d = 2, 4, 6$, and 8 ; and $B=0.3$.²⁶ Figure A3 visually illustrates a small world

²⁶ We choose $B = 0.3$ because at this value, average path lengths in our small world networks are similar to those in random networks of the same density, while clustering coefficients are similar to those of lattice networks of

network of $N=50$, $d = 8$, and $B=0.3$ on the left, with a zoomed in view of local network structure on the right.

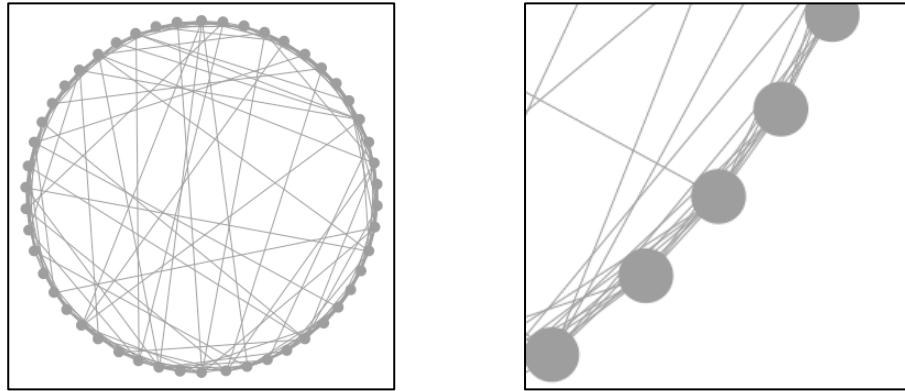


Figure A3: Example of a Small World Network Structure.

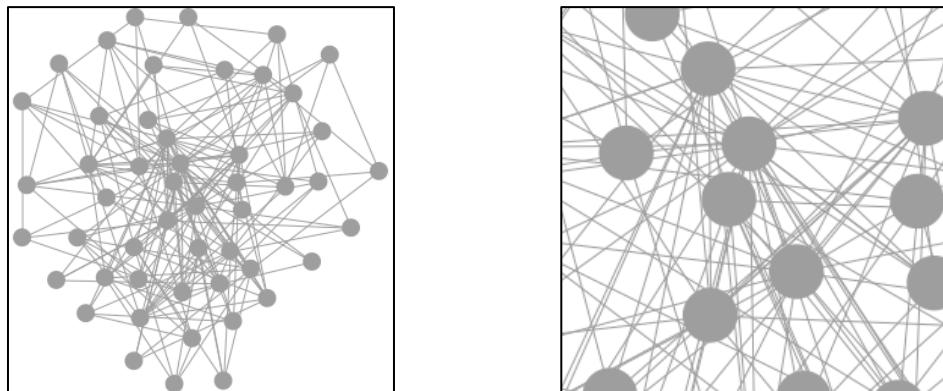


Figure A4: Example of a Preferential Attachment Network Structure.

Preferential Attachment Networks: This network is initialized as a fully connected graph of $k+1$ nodes. Each subsequent node added to the network forms k connections within the already existing nodes, where an edge is formed with node i with probability that is proportional to the number of links that the existing nodes already have (Barabasi and Albert, 1999). Thus, this network is fully specified by $\{k, N\}$. The degree distribution is scale free, in the power law

the same density. This allows our small world networks to be representative of the small world phenomenon, i.e., individuals have tight groups of friends, but at the same time everyone is close to everyone else in the entire network.

form, $\sim d^3$. The average path length is shorter than a random network, and the clustering coefficient is higher than the random network but lower than all other networks. Examples of preferential attachment networks include affiliation of websites on the world wide web (Barabasi and Albert 1999), the network of sexual contacts (Lilijeros et al., 2001), and the network of instant messaging connections with a target (Leskovec and Horvitz, 2007). In our analysis, $k = 1, 2, 3$, and 4 . Figure A4 visually illustrates a preferential attachment network of $N=50$ and $k = 1$ on the left, with a zoomed in view of local network structure on the right.

A.3 Comparing Diffusion Across Network Structures

The above discussion of network properties suggests diffusion can indeed be affected by the underlying network topology. We now discuss past research studying these differences. Under a wide variety of conditions, Rahmandad and Sterman (2008) show that the four networks we consider can greatly alter the diffusion process. Watts and Dodds (2007) examine the role that network topology plays in the effect of influentials and show that the effect varies depending on the underlying topology, including a random influence network, a hyper-influential network which is similar to the preferential attachment network, and a small group network which is similar to the small world network. Goldenberg et al. (2009b) show that hubs, a feature present only in the preferential attachment network, can greatly speed up the diffusion process. Delre et al. (2010) have explored how small world topologies affect diffusion in the presence of promotional activities, heterogeneous consumers, and different types of social influence. Goldenberg and Efroni (2001) and Goldenberg et al. (2001, 2007, 2009a, 2010) have used lattice-like networks to explore the emergence of innovations, word-of-mouth, new product growth and network externalities. Jansen and Jager (2001, 2002, 2003) have used a small

world and preferential attachment-like network to explore the diffusion of fashions, green trends, and the interaction between consumer psychology and network topologies. Kuandykov and Sokolov (2010) used a random network, a preferential attachment-like network and a more clustered small world network to explore the effect of network topology on S-curves of innovation. van Eck et al. (2011) explore how opinion leaders affect diffusion using a preferential attachment-like network. Centola (2010) show that for some behaviors tightly clustered networks, such as the lattice network, promote diffusion, while loosely clustered networks, such as the preferential attachment network and random network, inhibit diffusion, depending on the adoption process.

A Note on the Metrics Calculated in Tables A1-A4: Tables A1-A4 report numerical values for the metrics described above, for each network topology and density/average degree analyzed in the main paper. It is important to note that the above-defined properties apply to a network which is weakly connected, i.e., one that has no isolated nodes. In our numerical analysis where we sample sub-networks from the entire network based on the appeal fraction, A , it is possible that when $A < 1$ the network may not be always weakly connected. While we compute the numerical values for these metrics when $A < 1$ as well, note that average path length is not well defined in the case where there are nodes that are not connected to the network. To address this situation, the value that is shown is the average path length of the largest connected component (also referred to as the giant component) of the network. In the table, we report what fraction of the whole network is represented by this giant component. In addition, since the selection of nodes that participate in the appeal sub-network is random, the results presented are the average of 100 draws from this process. Moreover, since for random, small world, and preferential attachment networks the network generation routine is stochastic, we

average over 10 draws. Thus for each entry in this table, for $A < 1$ for three networks we average over 1,000 networks instances. For the lattice network, for $A < 1$, since the network generation is deterministic we only average over different values of appeal, A , and so each entry is averaged over 100 network instances.

REFERENCES

- Barabási, Albert-László, and Réka Albert (1999), "Emergence of scaling in random networks," *Science* 286(5439), 509-512.
- Centola, Damon (2010), "The spread of behavior in an online social network experiment," *Science* 329(5996), 1194-1197.
- Delre, Sebastiano A., Wander Jager, Tammo H.A. Bijmolt, and Marco A. Janssen (2010), "Will it spread or not? The effects of social influences and network topology on innovation diffusion," *Journal of Product Innovation Management*. 27(2), 267-282.
- Dodds, Peter Sheridan, Roby Muhamad, and Duncan J. Watts (2003), "An experimental study of search in global social networks," *Science* 301(5634), 827-829.
- Erdős, Paul, and Alfréd Rényi (1960), "On the evolution of random graphs," *Magyar Tud. Akad. Mat. Kutató Int. Közl* 5, 17-61.
- Goldenberg, Jacob, and Sol Efroni (2001), "Using cellular automata modeling of the emergence of innovations," *Technological Forecasting and Social Change* 68(3), 293-308.
- , Barak Libai, and Eitan Muller (2001), "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Marketing letters* 12(3) 211-223.
- , ---, Sarit Moldovan, and Eitan Muller (2007), "The NPV of bad news," *International Journal of Research in Marketing*, 24(3), 186-200.
- , Oded Lowengart, and Daniel Shapira (2009a), "Zooming in: Self-emergence of movements in new product growth," *Marketing Science* 28(2), 274-292.
- , Sangman Han, Donald Lehmann, and Jae Hong (2009b), "The role of hubs in the adoption processes," *Journal of Marketing*, 73(2), 1-13.
- , Barak Libai, and Eitan Muller (2010), "The chilling effects of network externalities," *International Journal of Research in Marketing*, 27(1), 4-15.
- Iamnitchi, Adriana, Matei Ripeanu, and Ian Foster (2004), "Small-world file-sharing communities," In *INFOCOM 2004, Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, 2, 952-963.
- Janssen, Marco A., and Wander Jager (2001), "Fashions, habits and changing preferences: Simulation of psychological factors affecting market dynamics," *Journal of Economic Psychology*, 22(6) 745-772.
- and --- (2002), "Stimulating diffusion of green products," *Journal of Evolutionary Economics* 12(3), 283-306.
- and --- (2003), "Simulating market dynamics: Interactions between consumer psychology and social

networks," *Artificial Life*, 9(4), 343-356.

Kiesling, Elmar, Markus Günther, Christian Stummer, and Lea M. Wakolbinger (2012), "Agent-based simulation of innovation diffusion: a review," *Central European Journal of Operations Research*, 20(2), 183-230.

Kuandykov, Lev, and Maxim Sokolov (2010), "Impact of social neighborhood on diffusion of innovation S-curve," *Decision Support Systems*, 48(4), 531-535.

Leskovec, Jure, and Eric Horvitz (2008), "Planetary-scale views on a large instant-messaging network," In *Proceeding of the 17th international conference on World Wide Web*, ACM, 915-924.

Liljeros, Fredrik, Christofer R. Edling, Luis A. Nunes Amaral, H. Eugene Stanley, and Yvonne Åberg (2001), "The web of human sexual contacts," *Nature*, 411(6840), 907-908.

Rahmandad, Hazhir and John Sterman (2008), "Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models," *Management Science*, 54, 998-1014.

Scott, John (2000), *Social Network Analysis: A Handbook*. Thousand Oaks: SAGE Publications.

Stonedahl, Forrest, William Rand, and Uri Wilensky (2010), "Evolving viral marketing strategies," In *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, ACM, 1195-1202.

Van Eck, Peter S., Wander Jager, and Peter SH Leeflang (2011), "Opinion leaders' role in innovation diffusion: A simulation study," *Journal of Product Innovation Management*, 28(2), 187-203.

Van den Bulte, Christophe and Stefan H. K. Wuyts (2007), *Social Networks and Marketing*. Cambridge: MSI.

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

Watts, Duncan J., and Peter S. Dodds (2007), "Influentials, Networks and Public Opinion Formation," *Journal of Consumer Research*, 34(4), 441-458.

--- and Steven H. Strogatz (1998), "Collective dynamics of 'small-world' networks," *Nature*, 393(6684), 440-442.

**Table A1: Average Values for Metrics used in the Hypercube Creation:
Lattice/Regular Networks**

Lattice Topology										
Appeal	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
Density Class 2										
<i>Avg. Degree (d)</i>	0.20	0.40	0.59	0.79	1.00	1.20	1.40	1.60	1.80	2.00
<i>Density (e)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Path Length (l)</i>	1.26	1.69	2.14	2.69	3.40	4.49	6.24	9.25	16.35	250.25
<i>Largest Component Fraction</i>	0.03	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.05	1.00
<i>Clustering Coefficient (c)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Density Class 4										
<i>Avg. Degree (d)</i>	0.38	0.79	1.20	1.60	2.00	2.40	2.80	3.20	3.60	4.00
<i>Density (e)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Path Length (l)</i>	1.36	2.03	2.77	3.63	5.41	7.38	11.56	20.83	56.71	125.38
<i>Largest Component Fraction</i>	0.03	0.03	0.03	0.03	0.04	0.05	0.07	0.12	0.33	1.00
<i>Clustering Coefficient (c)</i>	0.03	0.09	0.17	0.26	0.35	0.41	0.46	0.49	0.50	0.50
Density Class 6										
<i>Avg. Degree (d)</i>	0.60	1.20	1.78	2.39	3.01	3.61	4.19	4.80	5.40	6.00
<i>Density (e)</i>	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
<i>Path Length (l)</i>	1.48	2.36	3.40	5.03	7.72	12.67	23.09	48.60	94.58	83.75

<i>Largest Component Fraction</i>	0.04	0.04	0.04	0.05	0.07	0.11	0.20	0.41	0.93	1.00
<i>Clustering Coefficient (c)</i>	0.07	0.21	0.35	0.46	0.54	0.58	0.59	0.60	0.60	0.60
Density Class 8										
<i>Avg. Degree (d)</i>	0.80	1.60	2.40	3.21	3.99	4.80	5.60	6.40	7.20	8.00
<i>Density (e)</i>	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
<i>Path Length (l)</i>	1.67	2.58	4.16	6.64	11.11	19.61	40.09	72.70	65.33	62.94
<i>Largest Component Fraction</i>	0.05	0.05	0.06	0.08	0.13	0.22	0.45	0.87	1.00	1.00
<i>Clustering Coefficient (c)</i>	0.12	0.32	0.48	0.58	0.62	0.64	0.64	0.64	0.64	0.64

**Table A2: Average Values for Metrics used in the Hypercube Creation:
Random Networks**

Random Topology										
Appeal	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
Density Class 2										
<i>Avg. Degree (d)</i>	0.20	0.40	0.60	0.80	1.00	1.20	1.40	1.60	1.80	2.00
<i>Density (e)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Path Length (l)</i>	1.30	1.82	2.38	3.07	3.92	5.06	6.72	9.67	15.96	34.70
<i>Largest Component Fraction</i>	0.03	0.02	0.02	0.03	0.03	0.04	0.06	0.10	0.21	0.84
<i>Clustering Coefficient (c)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Density Class 4										
<i>Avg. Degree (d)</i>	0.39	0.79	1.20	1.60	2.00	2.40	2.80	3.20	3.60	4.00
<i>Density (e)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Path Length (l)</i>	1.76	3.40	7.63	12.33	9.61	7.92	6.89	6.19	5.70	5.33
<i>Largest Component Fraction</i>	0.05	0.06	0.18	0.60	0.83	0.93	0.97	0.99	1.00	1.00
<i>Clustering Coefficient (c)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Density Class 6										
<i>Avg. Degree (d)</i>	0.60	1.20	1.79	2.40	2.99	3.60	4.20	4.80	5.40	6.00
<i>Density (e)</i>	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
<i>Path Length (l)</i>	2.33	7.04	9.06	6.97	5.90	5.24	4.81	4.51	4.28	4.10

**Table A3: Average Values for Metrics used in the Hypercube Creation:
Small World Networks**

Small World Topology										
Appeal	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
Density Class 2										
<i>Avg. Degree (d)</i>	0.20	0.40	0.60	0.80	1.00	1.20	1.40	1.60	1.80	2.00
<i>Density (e)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Path Length (l)</i>	1.27	1.77	2.28	2.92	3.77	4.97	6.86	9.93	17.11	42.17
<i>Largest Component Fraction</i>	0.03	0.02	0.02	0.02	0.03	0.03	0.05	0.07	0.17	0.88
<i>Clustering Coefficient (c)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Density Class 4										
<i>Avg. Degree (d)</i>	0.40	0.80	1.20	1.60	2.00	2.40	2.80	3.20	3.60	4.00
<i>Density (e)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Path Length (l)</i>	1.62	2.88	5.46	11.89	13.32	10.30	8.58	7.47	6.72	6.18
<i>Largest Component Fraction</i>	0.04	0.05	0.09	0.31	0.72	0.90	0.97	0.99	1.00	1.00
<i>Clustering Coefficient (c)</i>	0.01	0.03	0.06	0.09	0.11	0.14	0.16	0.17	0.18	0.18
Density Class 6										
<i>Avg. Degree (d)</i>	0.59	1.20	1.80	2.40	3.00	3.60	4.20	4.80	5.40	6.00
<i>Density (e)</i>	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
<i>Path Length (l)</i>	2.00	4.82	11.42	9.28	7.39	6.32	5.66	5.22	4.90	4.66

<i>Largest Component Fraction</i>	0.06	0.12	0.51	0.86	0.96	0.99	1.00	1.00	1.00	1.00
<i>Clustering Coefficient (c)</i>	0.02	0.07	0.12	0.16	0.18	0.20	0.21	0.21	0.21	0.21
Density Class 8										
<i>Avg. Degree (d)</i>	0.79	1.60	2.39	3.20	3.99	4.80	5.60	6.40	7.20	8.00
<i>Density (e)</i>	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
<i>Path Length (l)</i>	2.45	8.04	8.60	6.59	5.59	5.00	4.62	4.35	4.15	4.00
<i>Largest Component Fraction</i>	0.08	0.33	0.84	0.96	0.99	1.00	1.00	1.00	1.00	1.00
<i>Clustering Coefficient (c)</i>	0.04	0.11	0.16	0.20	0.22	0.23	0.23	0.23	0.23	0.23

**Table A4: Average Values for Metrics used in the Hypercube Creation:
Preferential Attachment Networks**

Preferential Attachment Topology										
Appeal	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
Density Class 1										
<i>Avg. Degree (d)</i>	0.20	0.40	0.60	0.80	1.00	1.20	1.40	1.60	1.80	2.00
<i>Density (e)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Path Length (l)</i>	1.51	2.16	2.71	3.27	3.78	4.36	4.92	5.55	6.19	6.75
<i>Largest Component Fraction</i>	0.04	0.06	0.08	0.13	0.18	0.25	0.36	0.50	0.72	1.00
<i>Clustering Coefficient (c)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Density Class 2										
<i>Avg. Degree (d)</i>	0.40	0.80	1.19	1.58	1.99	2.39	2.79	3.20	3.59	3.99
<i>Density (e)</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>Path Length (l)</i>	2.08	4.00	5.86	6.11	5.59	5.17	4.81	4.52	4.29	4.06
<i>Largest Component Fraction</i>	0.08	0.20	0.41	0.62	0.78	0.87	0.94	0.98	0.99	1.00
<i>Clustering Coefficient (c)</i>	0.00	0.00	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.03
Density Class 3										
<i>Avg. Degree (d)</i>	0.59	1.18	1.80	2.40	2.98	3.60	4.18	4.78	5.39	5.98
<i>Density (e)</i>	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
<i>Path Length (l)</i>	2.70	5.36	5.41	4.90	4.53	4.20	3.97	3.77	3.62	3.50

<i>Largest Component Fraction</i>	0.14	0.40	0.69	0.84	0.92	0.97	0.99	1.00	1.00	1.00
<i>Clustering Coefficient (c)</i>	0.00	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.03	0.03
Density Class 4										
<i>Avg. Degree (d)</i>	0.79	1.58	2.39	3.20	3.97	4.78	5.58	6.38	7.17	7.97
<i>Density (e)</i>	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
<i>Path Length (l)</i>	3.35	5.36	4.73	4.26	3.95	3.70	3.52	3.38	3.27	3.18
<i>Largest Component Fraction</i>	0.21	0.60	0.82	0.92	0.97	0.99	1.00	1.00	1.00	1.00
<i>Clustering Coefficient (c)</i>	0.01	0.01	0.02	0.03	0.03	0.03	0.03	0.04	0.04	0.04

Appendix B – The Role of App Characteristics in Forecasting

In the context of our work, to understand the role of observable product characteristics as data that may help in the prediction process, we conducted the following analysis. First, we conducted a survey with undergraduate students, an appropriate target market for Facebook apps. 92 subjects from a major mid-Atlantic public university took part in this study. Since the apps we analyze are part of the Facebook platform, for the purposes of this survey, we created dummy Facebook accounts, where we installed a preset number of apps to allow subjects quick access to these apps, in a systematic fashion. During the data collection, each subject was provided instructions on how to access Facebook using the provided dummy accounts, and then asked to briefly familiarize themselves with the apps installed within those dummy accounts. Each subject was offered to evaluate up to 15 apps. Subjects were instructed to skip the app if they have used or seen it before. The detailed instructions provided to the subjects are listed in Table B1.

Once each subject was familiar with the provided applications, they were instructed to then rate each app on predefined attributes common across apps. Doing so allowed us to collect product attribute information for a subset of 284 apps. Categories for attributes we considered include app design and functionality, app entertainment, social and interactive value, and intentions about future usage. The specific set of questions asked is provided in Table B2.

To analyze the effectiveness of product characteristics in predicting parameters of the diffusion process, we regressed app characteristics on the diffusion process parameters estimated separately for each of these apps (e.g., Hahn et al. 1994). We consider the same three diffusion models analyzed earlier, viz. the Bass (1969) model, the Gamma/Shifted-Gompertz model

(Bemmaor and Lee, 2002) and the Weibull-Gamma model (Moe and Fader, 2002). We run a regression where the dependent variables are the log-transformed values of diffusion parameter estimates, and independent variables are the (average) scores on the eleven attribute related questions asked via the survey, genre level fixed effects and the extent of competition within (defined as number of other apps of the same genre available at that time) and across the category²⁷. The results of this exercise are presented in Table B3.

The results in Table B3 show that the predictive ability of app attributes is not very high. Adjusted-R² values from this analysis are quite low across the three models (0.16 and below), and comparatively are even lower for explaining p and q (0.1 and below). In fact, given the low predictive ability of these attributes, we find that pre-launch forecasts that exclude app characteristics perform better than those that include it.

²⁷ We thank an anonymous reviewer for this suggestion.

Table B1: Instructions Provided to Subjects in the Facebook Apps Study.

Social Media marketing - Facebook Applications

Today we will be evaluating some applications in Facebook. The process is as follows:

1. Go to the Facebook website and log in with the details below:

UserName (Email) : <**user@gmail.com**> Password: <**UserPassword**>

2. Once you are logged in, click on “App Center” and then on “My Apps”:



3. Once you are there, you should see the following applications. Start with the first application on this list. If you are familiar with this application, skip it and move on to the next application which is new to you.

- 1. Cliquey
- 2. Birthday Card
- 3. Spanish Audio Word of the Day
- 4. Top Friends
- 5. Pool
- 6. Beer!
- 7. Bubble Bobble
- 8. Texas HoldEm Poker
- 9. Blackjack
- 10. DVDs
- 11. Beer Pong
- 12. dNeero
- 13. Lego Character Creator
- 14. Star Wars Fan
- 15. Battle Ship

4. Spend a minute getting to know this application. You may want to try to play/interact with it, click “About“ link, or do whatever you want to explore the application. You should not provide your personal information though (e.g. links to your real facebook friends, your real name/e-mail address, etc).

5. Once you got a basic idea about the application answer the questions posted on:

<http://-----edu/>

You will find a survey under “User Menu” under “Facebook Applications.”

6. Once you respond to all questions on the survey for this application, please pick the next new application from the list above and repeat steps 4 onwards.

Table B2: Specific Questions Answered by Subjects in the Facebook Apps Study.

- 1 Have you seen this application before?
- 2 What do you like about this application?
- 3 What do you dislike about this application?

Please indicate the extent to which you agree with the statement
(from 1 being Strongly Disagree to 7 being Strongly Agree).

- 4 It is easy to learn how to use this application.
- 5 This application has been designed attractively and professionally.
- 6 This application is new and creative as compared to other applications.
- 7 Some applications are created just for fun and others have commercial purposes. This application is commercial in nature.
- 8 This application provides plenty of entertainment of its own.
- 9 This application provides relevant information.
- 10 This application allows me to express myself to others.
- 11 This application enables me to interact with others.
- 12 I am very likely to include this application in my profile.
- 13 I am very likely to continue using this application 2 weeks from now.
- 14 I expect this application to be very popular among other users.

Table B3: Predicting Diffusion Process Parameters Using App Characteristics.

		Bass Model			G/SG Model (Bemmaor and Lee 2002)				Weibull-gamma Model (Moe and Fader 2002)			
		m*	p	q	m	p	q	alpha	m	a	r	c
Intercept		9.90	-4.64	-3.52	9.76	-6.60	-3.70	0.63	7.61	17.18	13.24	0.42
		(12.45)**	(-14.88)	(-11.02)	(12.23)	(-8.39)	(-3.51)	(0.61)	(4.33)	(7.58)	(4.41)	(3.29)
Attributes	easy to learn	0.07	0.04	0.03	0.09	0.12	-0.03	0.03	0.03	-0.01	0.15	-0.01
		(1.04)	(1.30)	(1.01)	(1.23)	(1.77)	(-0.29)	(0.34)	(0.19)	(-0.05)	(0.57)	(-0.89)
	designed attractively and professionally	0.11	-0.06	0.00	0.11	-0.02	-0.06	0.09	0.13	0.16	-0.09	0.02
		(1.31)	(-2.00)	(-0.09)	(1.32)	(-0.21)	(-0.58)	(0.85)	(0.74)	(0.69)	(-0.31)	(1.76)
	new and creative	-0.10	0.01	-0.02	-0.11	0.00	-0.03	0.03	-0.11	-0.21	-0.25	0.01
		(-1.26)	(0.26)	(-0.63)	(-1.32)	(-0.03)	(-0.27)	(0.30)	(-0.60)	(-0.91)	(-0.81)	(0.67)
	commercial	-0.01	-0.02	0.02	-0.01	0.03	-0.01	0.03	0.20	-0.02	-0.20	0.00
		(-0.19)	(-0.72)	(0.88)	(-0.19)	(0.46)	(-0.12)	(0.35)	(1.65)	(-0.13)	(-0.95)	(-0.08)
	plenty of entertainment	0.07	-0.02	0.05	0.07	0.01	0.06	-0.04	-0.01	0.45	0.57	0.00
		(0.95)	(-0.78)	(1.76)	(0.94)	(0.09)	(0.62)	(-0.39)	(-0.05)	(2.25)	(2.19)	(-0.05)
	relevant information	-0.13	-0.02	-0.03	-0.15	-0.16	0.10	-0.12	0.01	0.19	0.03	-0.01
		(-1.87)	(-0.92)	(-1.02)	(-2.14)	(-2.38)	(1.12)	(-1.34)	(0.09)	(0.98)	(0.14)	(-0.51)
	allows to express myself	0.10	0.01	-0.03	0.10	-0.13	0.04	-0.07	0.26	-0.13	-0.19	-0.01
		(1.40)	(0.23)	(-1.14)	(1.34)	(-1.91)	(0.45)	(-0.80)	(1.64)	(-0.66)	(-0.70)	(-1.32)
	enables to interact with others	0.08	-0.01	0.00	0.09	0.14	-0.08	0.09	0.13	0.01	-0.09	0.01
		(1.13)	(-0.20)	(0.15)	(1.24)	(2.10)	(-0.92)	(0.97)	(0.84)	(0.04)	(-0.37)	(0.51)
	likely to include in my profile	0.07	-0.01	0.01	0.06	-0.05	0.00	0.06	-0.12	0.18	0.16	0.02
		(0.52)	(-0.09)	(0.14)	(0.45)	(-0.35)	(0.02)	(0.31)	(-0.38)	(0.45)	(0.31)	(0.74)
	likely to continue using this application 2 weeks from now	-0.15	0.01	-0.02	-0.15	-0.03	0.07	-0.14	-0.11	-0.24	-0.12	-0.02
		(-1.07)	(0.18)	(-0.32)	(-1.02)	(-0.21)	(0.38)	(-0.77)	(-0.33)	(-0.58)	(-0.22)	(-0.75)
	expect to be popular among others	0.09	0.05	-0.02	0.10	0.11	-0.06	0.05	0.02	-0.22	-0.08	-0.01
		(1.23)	(-0.63)	(-0.70)	(1.28)	(1.51)	(-0.61)	(0.48)	(0.12)	(-1.00)	(-0.27)	(-0.84)
Category	Games	0.69	-0.40	0.07	0.73	0.24	0.14	-0.05	0.14	0.98	0.34	0.18
		(2.00)	(-3.00)	(0.54)	(2.12)	(0.71)	(0.32)	(-0.12)	(0.19)	(0.99)	(0.26)	(3.22)
	Utilities	0.51	-0.13	-0.33	0.53	0.40	0.12	-0.35	0.55	-0.38	-0.92	-0.04
		(0.98)	(-0.64)	(-1.59)	(1.01)	(0.79)	(0.18)	(-0.52)	(0.48)	(-0.26)	(-0.47)	(-0.45)
	Lifestyle	0.11	0.99	-1.23	-0.03	-0.90	2.40	-3.33	1.46	-3.67	-1.60	-0.59
		(0.09)	(2.10)	(-2.53)	(-0.03)	(-0.76)	(1.50)	(2.12)	(0.55)	(-1.07)	(-0.35)	(-3.07)
	Entertainment	2.24	0.02	-0.52	2.14	-0.69	1.41	-1.87	3.37	-2.73	-2.57	-0.28
		(2.12)	(0.05)	(-1.23)	(2.02)	(-0.66)	(1.01)	(-1.36)	(1.44)	(-0.91)	(-0.64)	(-1.68)
	Education	-0.29	0.62	-1.36	-0.44	-0.88	0.81	-1.86	3.08	-5.94	-6.29	-0.59
		(-0.21)	(1.14)	(-2.45)	(-0.31)	(-0.65)	(0.44)	(-1.03)	(1.01)	(-1.51)	(-1.21)	(-2.69)
	Business	-0.82	0.73	-0.69	-0.90	-0.27	0.33	-1.01	-0.46	-5.61	-2.79	-0.50
		(-0.52)	(1.17)	(-1.08)	(-0.57)	(-0.17)	(0.16)	(-0.49)	(-0.13)	(-1.24)	(-0.47)	(-1.98)
	Sports	0.06	1.27	-1.34	-0.17	-2.45	2.93	-4.10	0.57	-0.38	3.89	-0.89
		(0.03)	(1.64)	(-1.69)	(-0.09)	(-1.25)	(1.12)	(-1.58)	(0.13)	(-0.07)	(0.52)	(-2.81)
	Others	-1.19	1.07	-1.76	-1.41	-1.17	2.42	-3.69	0.42	-4.84	-2.34	-0.81
		(-0.84)	(1.92)	(-3.09)	(-0.99)	(-0.83)	(1.29)	(-1.99)	(0.13)	(-1.20)	(-0.44)	(-3.57)
Competition	Overall	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
		(-0.79)	(-0.35)	(3.35)	(-0.60)	(1.81)	(-1.14)	(1.90)	(-0.31)	(-0.02)	(-0.54)	(3.05)
	Within Category	0.00	0.01	-0.01	0.00	-0.01	0.02	-0.02	0.01	-0.03	-0.02	-0.01
		(-0.30)	(2.03)	(-2.16)	(-0.44)	(-0.87)	(1.15)	(-1.71)	(0.52)	(-1.19)	(-0.51)	(-3.22)
	R-squared	0.20	0.15	0.10	0.21	0.08	0.05	0.07	0.08	0.07	0.08	0.10
	Adj. R-squared	0.16	0.10	0.05	0.16	0.03	0.00	0.01	0.03	0.01	0.02	0.05

* - log transformed dependent variables are used

** - statistically significant parameters are shown in bold

REFERENCES

- Bass, Frank M. (1969), "A New Product Growth Model for Consumer Durables," *Management Science*, 36 (9), 1057-1079.
- Bemmaor, Albert C., and Janghyuk Lee (2002), "The Impact of Heterogeneity and Ill-conditioning on Diffusion Model Parameter Estimates," *Marketing Science*, 21(2), 209–220.
- Hahn, Minhi, Sehoon Park, Lakshman Krishnamurthi, and Andris A. Zoltners (1994), "Analysis of new product diffusion using a four-segment trial-repeat model," *Marketing Science*, 13(3) 224–247.
- Moe, Wendy W. and Peter S. Fader (2002), "Using Advance Purchase Orders to Forecast New Product Sales," *Marketing Science*, 21, 347–364.

Appendix C – Additional Diagrams of the Diffusion Hypercube

These figures complement the discussion in the main paper related to Figure 6.

Figure C1 shows a distribution of p and q parameters when density and topology dimensions are integrated out.

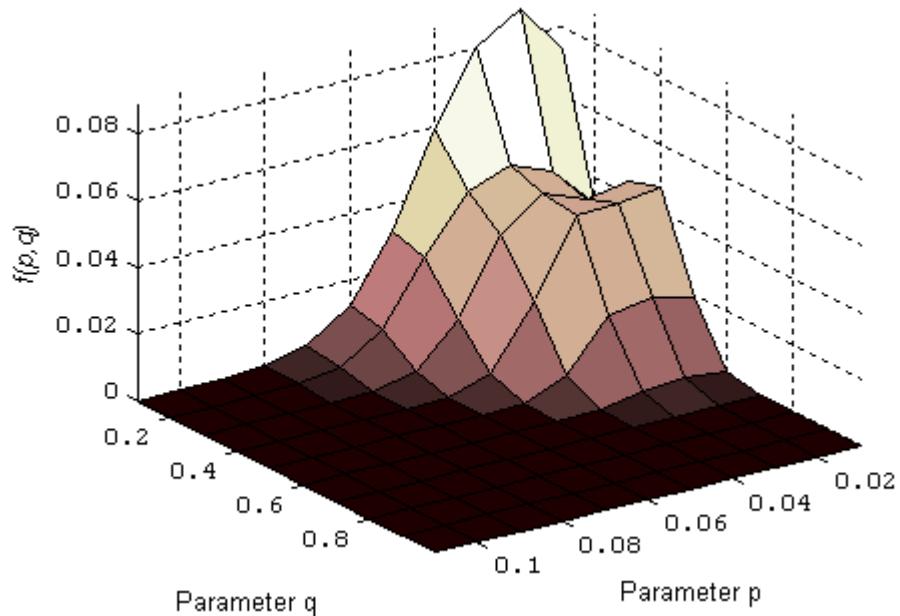


Figure C1. Diffusion Hypercube: Integration over Network Density and Topology.

Figure C2 shows the results of integration over network density.

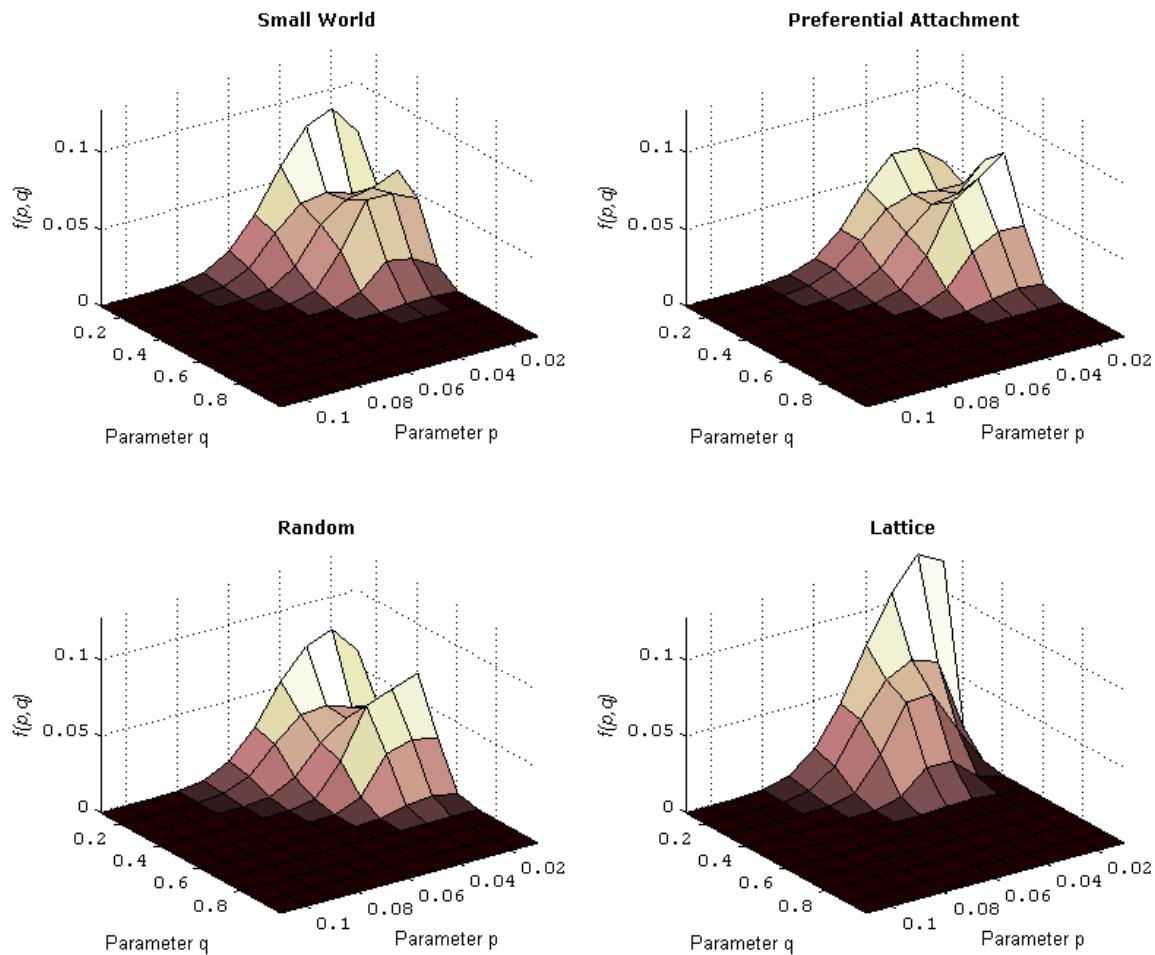


Figure C2. Diffusion Hypercube: Integration over Network Density.

Figure C3 summarizes the results of integration over network topology.

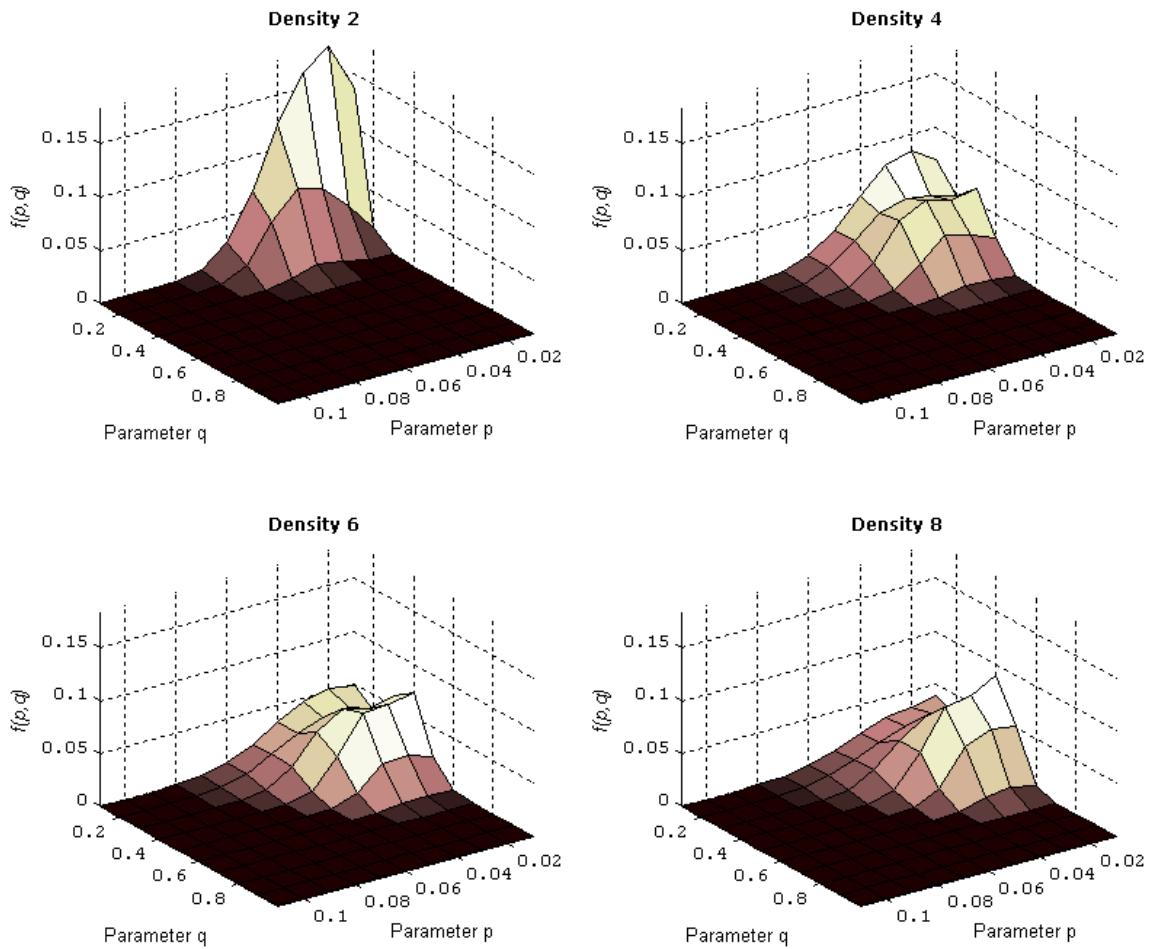


Figure C3. Diffusion Hypercube: Integration over Network Topology.

Appendix D – Fit Statistics for Pre-launch Forecasting

The tables below report on additional performance statistics, including MAD and MSE, to supplement the discussion in the main paper related to Table 3.

Calibration Sample Size	MAPE				MAD				MSE			
	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma
1	1391.2%	1364.0%	1397.1%	1280.0%	95,442	95,224	95,638	92,878	102,159,506,469	116,449,026,312	118,281,475,298	112,653,913,079
3	438.5%	460.0%	509.6%	749.6%	28,879	30,777	32,333	48,811	8,610,617,529	10,053,207,489	11,367,048,948	43,144,764,246
5	283.1%	306.2%	280.4%	543.7%	19,132	20,918	18,945	35,855	3,738,661,234	4,317,421,268	2,566,270,618	16,404,061,924
7	201.8%	217.7%	224.1%	310.0%	13,828	15,186	15,334	21,657	1,377,886,393	1,596,374,375	1,801,267,239	6,083,048,845
9	171.0%	186.6%	181.6%	233.1%	11,758	13,080	12,597	16,684	826,606,542	1,029,383,988	1,031,870,514	2,868,920,191
11	138.3%	147.5%	157.6%	209.7%	9,725	10,718	10,961	15,137	445,592,434	539,203,006	716,725,412	1,179,706,506
13	120.2%	128.7%	132.7%	186.6%	8,421	9,393	9,577	13,883	249,610,828	336,010,687	339,338,816	711,684,920
15	120.2%	128.5%	134.3%	163.2%	8,353	9,298	9,441	12,506	318,063,350	423,167,524	492,633,873	512,268,628
21	105.1%	111.3%	120.2%	149.1%	7,363	8,250	8,497	11,964	122,572,862	162,048,920	206,584,848	353,932,889
35	98.3%	102.2%	104.9%	129.2%	6,606	7,463	7,608	10,670	63,419,113	83,371,159	88,587,716	183,332,505
50	96.2%	99.0%	101.5%	125.3%	6,490	7,353	7,542	10,705	58,329,559	76,799,039	81,072,947	171,349,274
75	96.0%	98.1%	100.0%	118.2%	6,251	7,044	7,288	10,223	53,278,667	68,893,324	73,183,705	148,268,054

Table D1: Comparing Model Performance with Market Size from Historical Data.

Calibration Sample Size	MAPE				MAD				MSE			
	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma
1	29.6%	37.9%	37.8%	49.2%	1071.22	1557.04	1475.71	2339.85	1,636,180	3,801,171	3,521,643	10,256,035
3	26.7%	33.0%	35.9%	47.9%	998.89	1360.23	1459.70	2253.82	1,347,884	2,740,006	3,436,689	9,653,821
5	25.4%	30.5%	35.3%	46.9%	969.00	1254.56	1418.33	2194.09	1,248,635	2,233,334	3,312,259	9,216,717
7	25.3%	29.6%	35.0%	44.4%	961.77	1211.11	1425.25	2095.17	1,227,032	2,097,776	3,588,439	8,532,195
9	24.9%	28.8%	32.9%	42.0%	949.44	1180.58	1319.06	1958.98	1,196,890	1,959,771	2,869,308	7,456,481
11	24.7%	28.3%	32.4%	38.4%	947.12	1147.17	1305.40	1752.33	1,186,398	1,818,293	3,025,073	5,921,850
13	24.6%	28.1%	31.7%	37.6%	940.44	1126.72	1255.23	1729.50	1,167,164	1,743,915	2,530,325	5,815,145
15	24.5%	27.8%	31.7%	37.2%	939.63	1119.77	1266.67	1684.20	1,162,973	1,712,991	2,687,593	5,321,722
21	24.3%	27.4%	29.3%	34.4%	931.21	1093.67	1144.89	1531.36	1,139,783	1,609,323	2,069,819	4,134,197
35	24.1%	26.8%	26.9%	30.5%	926.61	1066.12	1029.62	1315.65	1,118,757	1,499,428	1,486,757	2,642,465
50	24.0%	26.7%	25.7%	29.3%	926.82	1064.79	985.15	1262.86	1,129,496	1,493,024	1,290,847	2,289,784
75	23.9%	26.4%	25.1%	28.8%	928.30	1046.84	954.18	1226.73	1,128,987	1,424,172	1,198,104	2,097,110

Table D2: Comparing Model Performance when Market Size is Exogenous.

Appendix E – Network Recovery with Simulated Truncated Data

To assess the extent to which the recovery of the underlying networks is affected due to right truncation, we conduct the following analysis. Using simulated data, we test recovery of the underlying network for four levels: with no truncation, i.e., with 100% of the data being considered for estimating parameters, with 75% data being considered, with 50% data being considered, and finally, with 25% data being considered. The results from this analysis are reported below in Figure E (similar to Figure 8 in the main paper).

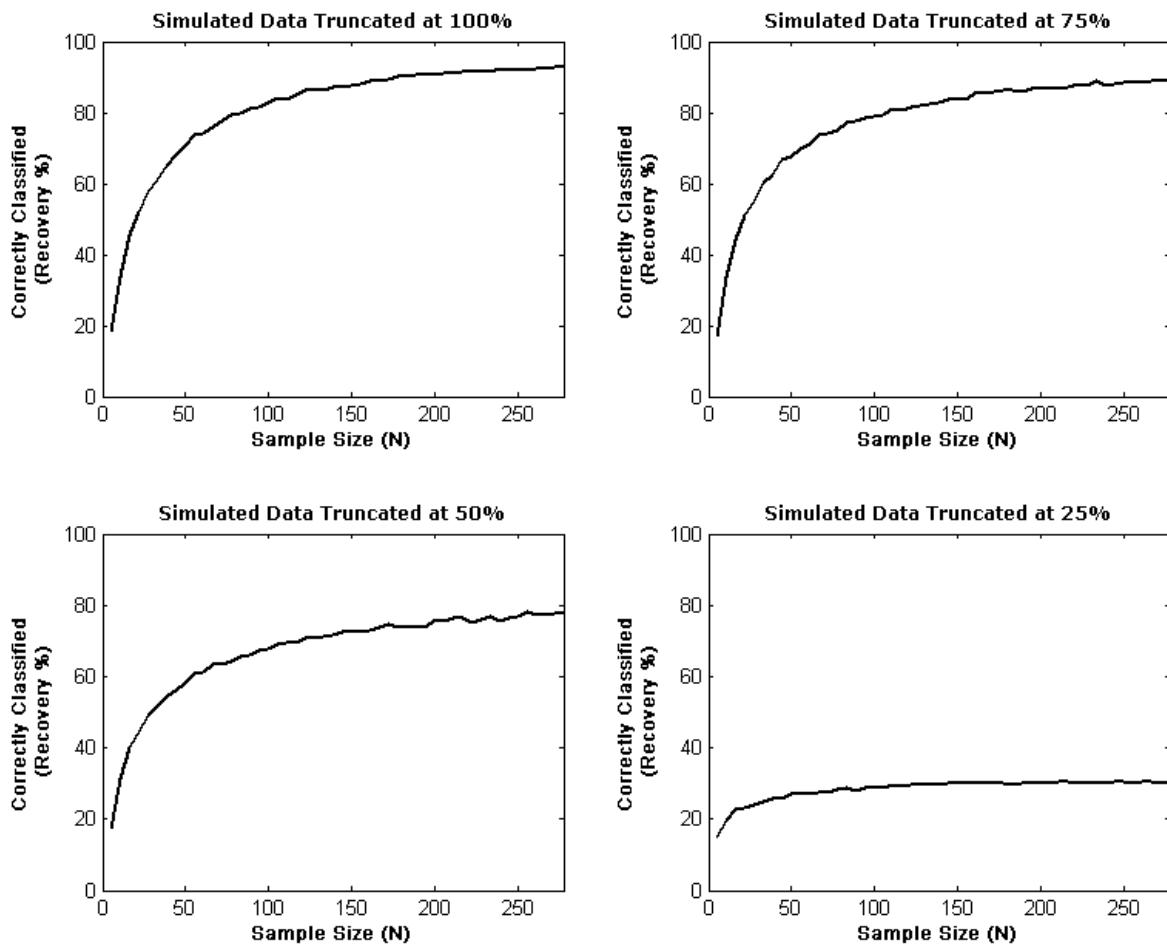


Figure E: Method Performance on Simulated and Truncated Data.

We observe is that while there is a small drop in recovery with 75% data, the drop with 50% data is quite big and recovery is really poor with only 25% data being used. This suggests that in applying the proposed approach, one needs to be careful regarding the extent of truncation, and anticipating its consequent effects on deciphering the underlying network footprint and hence enhancing forecasts.

Appendix F – Graphical Evidence of Forecasting Ability

This appendix provides graphical evidence for the models analyzed in the main paper for the forecasting exercise in Figure F below. For a sample application, there are 4 charts: forecasts generated when market size is known and unknown, as well as forecasts for actual and relative levels. It is important to keep in mind that the figure below is only illustrative, and that the entire data set contains 900+ apps, so forecasting ability varies substantially across apps, as evidenced by the results in Table 3 and Appendix D.

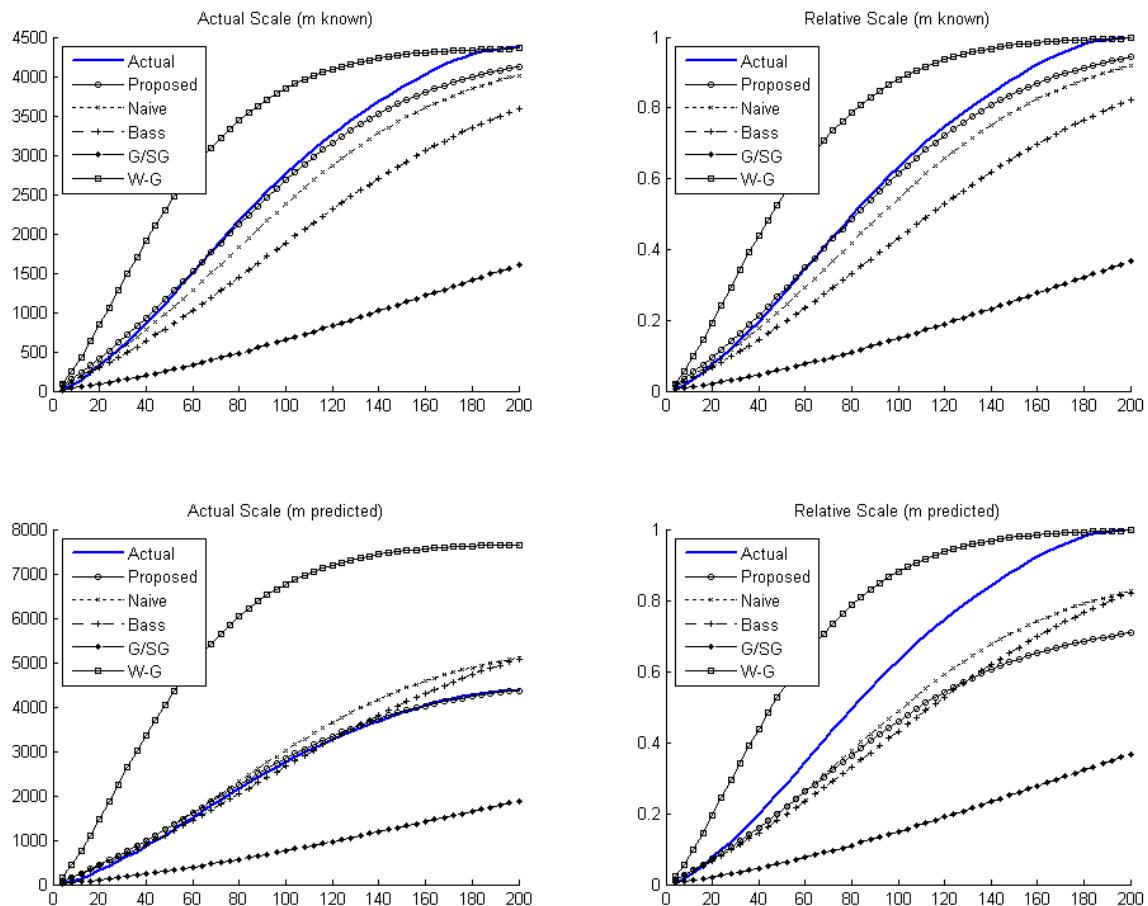


Figure F: Sample Graphical Evidence of Forecasting Ability.

Appendix G – Comparing Proposed Approach to Naïve and Bass Model

Table G below provides fit statistics for the proposed model, the naïve model (which assigns equal weights to each of the 16 underlying synthetic networks), and the Bass model, on the Facebook data (KL divergence measures for these models are plotted in Figure 10 in the main paper). The benefit of the proposed approach is highest at low calibration samples, and model fits converge as the calibration sample size increases.

	MAPE			MAD			MSE		
	Proposed	Naïve	Bass	Proposed	Naïve	Bass	Proposed	Naïve	Bass
Calibration Sample Size									
1	30.8%	36.6%	38.1%	1186	1534	1635	2,052,652	3,893,314	4,514,036
3	27.1%	31.1%	32.4%	1074	1308	1396	1,572,061	2,532,807	2,964,023
5	26.0%	29.3%	30.4%	1036	1220	1296	1,452,313	2,128,708	2,474,947
7	25.4%	28.5%	29.7%	1024	1195	1267	1,397,507	2,003,307	2,312,827
9	25.0%	27.7%	28.8%	1007	1147	1219	1,351,997	1,805,598	2,089,358
11	24.9%	27.3%	28.4%	1000	1132	1202	1,329,109	1,740,271	1,993,535
13	24.7%	27.0%	28.1%	997	1118	1187	1,315,809	1,680,917	1,926,849
15	24.6%	26.8%	27.8%	995	1106	1165	1,311,436	1,633,311	1,851,382
21	24.4%	26.3%	27.3%	993	1083	1134	1,304,274	1,552,040	1,729,298
35	24.3%	26.1%	27.1%	998	1075	1133	1,313,786	1,512,344	1,685,860
50	24.2%	25.9%	26.9%	991	1062	1116	1,289,212	1,471,716	1,638,557
75	24.1%	25.6%	26.6%	994	1058	1106	1,305,072	1,457,769	1,604,329

Table G: Comparing Model Performance when Market Size is Exogenous.

Appendix H – Impact of Product Heterogeneity

In this Appendix, we provide the results of additional analysis we conducted to assess the impact of product heterogeneity on improving pre-launch forecasts. We report forecasts when market size is estimated from past data (Table H1) as well as when market size is estimated exogenously (Table H2). To do so, we select the two largest app categories observed in our data, viz. “Games” and “Utilities,” and analyze each category separately. For comparison, we also report the analysis when “All” categories are included. We report the same performance measures: MAPE, MAD and MSE, for the proposed model, the Bass model, the Gamma-Shifted Gompertz model and the Weibull-Gamma model.

From Table H1, compared to “All” categories, MAPE improves somewhat for “Utilities,” but does not differ much for “Games”. Looking at the results in Table H2, where market size is estimated exogenously, we do not observe much difference in MAPE “All”, “Games”, and “Utilities.” This lack of improvement in forecasts perhaps suggests that there is inherently enough variation within each category; hence forming segments at the category level might not be informative and it may be essential to conduct enhanced segmentation or rely on alternative attributes in order to gain significant improvement in the levels of pre-launch forecasts.

All																
Calibration Sample Size	MAPE				MAD				MSE				Gamma	Shifted	Weibull Gamma	
	Proposed	Bass	Gamma	Welbull	Proposed	Bass	Gamma	Welbull	Proposed	Bass	Gamma	Welbull				
	Gompertz	Gamma	Shifted	Gompertz	Gompertz	Gamma	Shifted	Gompertz	Gompertz	Bass	Gamma	Shifted	Gompertz	Shifted	Weibull	Gamma
estimated market size	1	1391.2%	1364.0%	1397.1%	1280.0%	95,442	95,224	95,638	92,878	102,159,506,469	116,449,026,312	118,281,475,298	112,653,913,079			
	3	438.5%	460.0%	509.6%	749.6%	28,879	30,777	32,333	48,811	8,610,617,529	10,053,207,489	11,367,048,948	43,144,764,246			
	5	283.1%	306.2%	280.4%	543.7%	19,132	20,918	18,945	35,855	3,738,661,234	4,317,421,268	2,566,270,618	16,404,061,924			
	7	201.8%	217.7%	224.1%	310.0%	13,828	15,186	15,334	21,657	1,377,886,393	1,596,374,375	1,801,267,239	6,083,048,845			
	9	171.0%	186.6%	181.6%	233.1%	11,758	13,080	12,597	16,684	826,606,542	1,029,383,988	1,031,870,514	2,868,920,191			
	11	138.3%	147.5%	157.6%	209.7%	9,725	10,718	10,961	15,137	445,592,434	539,203,006	716,725,412	1,179,706,506			
	13	120.2%	128.7%	132.7%	186.6%	8,421	9,393	9,577	13,883	249,610,828	336,010,687	339,338,816	711,684,920			
	15	120.2%	128.5%	134.3%	163.2%	8,353	9,298	9,441	12,506	318,063,350	423,167,524	492,633,873	512,268,628			
	21	105.1%	111.3%	120.2%	149.1%	7,363	8,250	8,497	11,964	122,572,862	162,048,920	206,584,848	353,932,889			
	35	98.3%	102.2%	104.9%	129.2%	6,606	7,463	7,608	10,670	63,419,113	83,371,159	88,587,716	183,332,505			
	50	96.2%	99.0%	101.5%	125.3%	6,490	7,353	7,542	10,705	58,329,559	76,799,039	81,072,947	171,349,274			
	75	96.0%	98.1%	100.0%	118.2%	6,251	7,044	7,288	10,223	53,278,667	68,893,324	73,183,705	148,268,054			
Games																
Calibration Sample Size	MAPE				MAD				MSE				Gamma	Shifted	Weibull Gamma	
	Proposed	Bass	Gamma	Welbull	Proposed	Bass	Gamma	Welbull	Proposed	Bass	Gamma	Welbull				
	Gompertz	Gamma	Shifted	Gompertz	Gompertz	Gamma	Shifted	Gompertz	Gompertz	Bass	Gamma	Shifted	Gompertz	Shifted	Weibull	Gamma
estimated market size	1	1523.4%	1567.9%	1600.7%	1480.7%	101,026	104,942	105,067	102,500	90,138,780,559	103,202,732,684	103,929,328,576	99,363,794,526			
	3	572.4%	619.1%	608.4%	790.0%	37,809	41,188	40,783	51,930	12,348,432,171	14,106,182,822	15,136,184,920	24,792,653,955			
	5	367.3%	423.2%	397.0%	496.1%	24,290	28,098	26,538	32,970	4,013,034,468	5,707,800,254	4,977,756,380	9,887,644,711			
	7	249.9%	289.1%	263.2%	410.6%	16,976	19,665	18,566	27,634	1,565,163,600	2,240,148,996	1,956,807,262	6,472,877,928			
	9	190.7%	217.3%	209.4%	270.9%	13,319	15,372	15,075	18,687	891,105,105	1,233,474,632	1,250,447,515	1,979,673,564			
	11	162.4%	184.5%	189.2%	234.2%	11,358	13,120	13,208	16,954	457,349,575	666,410,516	736,402,260	1,433,321,189			
	13	161.1%	180.8%	185.0%	219.9%	11,144	12,841	12,990	16,000	449,037,387	639,360,039	685,066,048	1,185,991,769			
	15	133.2%	148.0%	151.1%	182.9%	9,278	10,650	10,703	13,541	196,102,271	275,225,899	293,260,764	612,871,518			
	21	110.9%	120.8%	122.6%	143.3%	7,792	8,918	9,022	11,106	100,738,032	139,379,077	143,202,457	230,939,136			
	35	103.9%	110.3%	112.0%	132.2%	7,060	8,065	8,177	10,601	73,902,247	99,997,627	103,146,511	180,196,005			
	50	100.6%	105.8%	106.6%	119.9%	6,774	7,784	7,852	10,114	65,079,464	88,600,484	89,892,277	149,550,493			
	75	97.1%	100.6%	101.5%	111.3%	6,508	7,490	7,588	9,786	55,889,748	75,553,150	77,676,515	128,872,643			
Utilities																
Calibration Sample Size	MAPE				MAD				MSE				Gamma	Shifted	Weibull Gamma	
	Proposed	Bass	Gamma	Welbull	Proposed	Bass	Gamma	Welbull	Proposed	Bass	Gamma	Welbull				
	Gompertz	Gamma	Shifted	Gompertz	Gompertz	Gamma	Shifted	Gompertz	Gompertz	Bass	Gamma	Shifted	Gompertz	Shifted	Weibull	Gamma
estimated market size	1	605.8%	568.9%	584.5%	537.4%	46,176	42,664	42,727	41,400	78,989,920,151	95,281,991,926	97,055,314,574	91,065,275,376			
	3	130.5%	134.0%	136.4%	255.8%	9,753	10,131	10,013	19,187	290,975,367	306,517,090	365,540,636	10,150,442,720			
	5	111.3%	115.9%	117.7%	169.3%	8,328	8,824	8,645	12,510	174,860,461	196,807,669	175,672,756	735,987,226			
	7	99.5%	104.7%	106.6%	136.3%	7,336	7,873	7,841	10,248	94,994,565	116,481,891	113,855,520	252,651,770			
	9	93.1%	97.0%	98.3%	119.4%	6,787	7,355	7,283	9,164	77,173,884	95,787,644	90,084,292	161,984,918			
	11	90.1%	93.7%	94.0%	114.3%	6,524	7,099	7,036	8,926	68,158,025	85,066,337	79,698,749	145,880,200			
	13	89.4%	92.8%	94.2%	112.6%	6,330	6,954	6,867	8,821	63,404,185	80,109,773	76,372,283	145,635,921			
	15	85.8%	88.1%	88.8%	109.7%	5,975	6,542	6,476	8,642	51,990,630	64,621,497	63,435,760	132,925,567			
	21	84.5%	86.5%	87.2%	105.2%	5,872	6,484	6,409	8,465	48,358,581	60,746,937	60,035,358	115,077,066			
	35	83.1%	84.8%	86.1%	98.9%	5,654	6,302	6,305	8,188	43,336,789	55,172,068	55,652,336	103,579,766			
	50	82.3%	83.5%	84.0%	92.3%	5,525	6,197	6,239	7,823	40,147,648	51,883,662	52,636,639	86,918,543			
	75	81.9%	83.1%	83.6%	89.8%	5,409	6,111	6,184	7,672	37,821,996	49,509,501	50,661,203	80,277,252			

Table H1: Comparing Model Performance with Market Size from Historical Data.

All														
Calibration Sample Size	MAPE				MAD				MSE				Known market size	
	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma		
	1	29.6%	37.9%	37.8%	49.2%	1071.22	1557.04	1475.71	2339.85	1,636,180	3,801,171	3,521,643	10,256,035	
Known market size	3	26.7%	33.0%	35.9%	47.9%	998.89	1360.23	1459.70	2253.82	1,347,884	2,740,006	3,436,689	9,653,821	
	5	25.4%	30.5%	35.3%	46.9%	969.00	1254.56	1418.33	2194.09	1,248,635	2,233,334	3,312,259	9,216,717	
	7	25.3%	29.6%	35.0%	44.4%	961.77	1211.11	1425.25	2095.17	1,227,032	2,097,776	3,588,439	8,532,195	
	9	24.9%	28.8%	32.9%	42.0%	949.44	1180.58	1319.06	1958.98	1,196,890	1,959,771	2,869,308	7,456,481	
	11	24.7%	28.3%	32.4%	38.4%	947.12	1147.17	1305.40	1752.33	1,186,398	1,818,293	3,025,073	5,921,850	
	13	24.6%	28.1%	31.7%	37.6%	940.44	1126.72	1255.23	1729.50	1,167,164	1,743,915	2,530,325	5,815,145	
	15	24.5%	27.8%	31.7%	37.2%	939.63	1119.77	1266.67	1684.20	1,162,973	1,712,991	2,687,593	5,321,722	
	21	24.3%	27.4%	29.3%	34.4%	931.21	1093.67	1144.89	1531.36	1,139,783	1,609,323	2,069,819	4,134,197	
	35	24.1%	26.8%	26.9%	30.5%	926.61	1066.12	1029.62	1315.65	1,118,757	1,499,428	1,486,757	2,642,465	
	50	24.0%	26.7%	25.7%	29.3%	926.83	1064.79	985.15	1262.86	1,129,496	1,493,024	1,290,847	2,289,784	
	75	23.9%	26.4%	25.1%	28.8%	928.30	1046.84	954.18	1226.73	1,128,987	1,424,172	1,198,104	2,097,110	
Games														
Calibration Sample Size	MAPE				MAD				MSE				Known market size	
	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma		
	1	29.6%	37.3%	37.9%	45.6%	1509.98	2047.04	1964.26	2856.49	3,275,366	6,543,016	6,340,997	16,880,725	
Known market size	3	27.4%	32.6%	37.0%	44.5%	1414.85	1810.86	1958.61	2749.99	2,779,313	4,965,500	6,051,898	14,901,314	
	5	26.0%	30.2%	34.5%	42.6%	1369.16	1678.84	1853.65	2552.58	2,537,491	4,114,651	5,881,972	12,962,059	
	7	25.3%	28.8%	32.9%	39.5%	1344.89	1595.50	1745.93	2401.91	2,423,672	3,541,940	4,659,596	11,501,737	
	9	25.0%	28.2%	31.2%	37.3%	1329.38	1563.24	1663.76	2268.54	2,377,054	3,336,187	4,614,677	10,157,144	
	11	24.7%	27.8%	30.8%	36.8%	1320.84	1553.20	1649.32	2236.99	2,355,877	3,268,080	4,079,614	9,788,024	
	13	24.6%	27.7%	29.1%	35.6%	1320.44	1541.83	1554.42	2177.80	2,353,644	3,162,664	3,598,872	9,169,662	
	15	24.6%	27.5%	29.3%	34.8%	1317.63	1534.91	1547.34	2108.56	2,328,859	3,109,719	3,556,031	8,060,096	
	21	24.3%	27.4%	27.4%	32.1%	1314.99	1524.07	1454.87	1919.83	2,306,398	3,027,071	2,822,807	5,948,017	
	35	24.0%	27.1%	25.9%	29.3%	1310.76	1515.44	1377.50	1760.75	2,267,933	2,981,258	2,485,007	4,490,818	
	50	23.8%	26.9%	25.5%	28.6%	1298.95	1483.27	1353.08	1683.27	2,227,144	2,831,959	2,408,255	3,914,541	
	75	23.7%	26.6%	25.1%	28.0%	1295.11	1464.34	1344.22	1649.41	2,214,223	2,747,492	2,362,579	3,731,323	
Utilities														
Calibration Sample Size	MAPE				MAD				MSE				Known market size	
	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma	Proposed	Bass	Gamma Shifted Gompertz	Weibull Gamma		
	1	30.8%	39.0%	38.0%	48.5%	1424.43	2041.94	2047.22	2953.22	2,932,544	6,620,795	7,258,017	16,470,159	
Known market size	3	27.7%	34.3%	37.9%	47.4%	1350.34	1863.71	1924.66	2874.89	2,504,342	5,374,394	6,289,131	15,475,079	
	5	26.4%	32.2%	35.7%	46.8%	1313.18	1770.23	1916.14	2796.32	2,333,299	4,558,039	5,962,526	14,318,781	
	7	25.4%	30.5%	35.0%	41.3%	1290.10	1679.54	1839.70	2498.54	2,216,242	4,038,701	5,579,753	11,389,625	
	9	24.8%	29.1%	33.5%	39.5%	1274.84	1607.12	1779.72	2356.77	2,160,811	3,570,008	5,257,391	10,298,139	
	11	24.5%	28.9%	32.9%	38.4%	1270.13	1593.13	1717.98	2304.20	2,134,797	3,514,094	4,740,577	9,787,578	
	13	24.5%	28.7%	32.8%	37.6%	1265.04	1585.76	1712.48	2265.19	2,102,661	3,454,444	4,585,854	9,457,414	
	15	24.3%	28.6%	32.7%	35.7%	1258.21	1584.71	1696.55	2122.46	2,081,526	3,448,502	4,503,564	8,184,470	
	21	24.2%	28.1%	30.4%	32.2%	1256.27	1564.53	1573.42	1872.75	2,069,379	3,333,492	3,667,382	5,690,190	
	35	23.8%	27.5%	27.6%	29.5%	1238.19	1529.70	1439.90	1723.37	2,005,032	3,146,034	2,907,059	4,365,189	
	50	23.8%	27.4%	26.4%	28.2%	1248.86	1533.78	1378.23	1643.11	2,064,628	3,125,575	2,539,270	3,828,701	
	75	23.8%	27.2%	25.9%	27.5%	1265.62	1537.63	1365.08	1615.45	2,168,100	3,145,714	2,458,116	3,611,127	

Table H2: Comparing Model Performance when Market Size is Exogenous.

Appendix I – Impact of Specific Network Characteristics

To generate insights into the role of specific network characteristics – i.e., topology versus density – in forecasting, we analyzed the simulated data by adding two new models to the existing models. The first model ignores the role of density, and only focuses on the gains from recovering topology, whereas the second model ignores the role of topology and only focuses on the gains from recovering density. The results from this analysis are reported in Table I below.

Calibration Sample Size	MAPE				MAD				MSE						
	Proposed			Proposed (ignoring density)	Proposed			Proposed (ignoring topology)	Proposed			Proposed (ignoring density)	Proposed (ignoring topology)		
	Naïve	Bass			Naïve	Bass			Naïve	Bass					
1	30.8%	36.6%	38.1%	34.9%	33.7%	1186	1534	1635	1420	1351	2,052,652	3,893,314	4,514,036	3,265,380	2,881,167
3	27.1%	31.1%	32.4%	29.7%	29.0%	1074	1308	1396	1220	1187	1,572,061	2,532,807	2,964,023	2,130,316	1,979,162
5	26.0%	29.3%	30.4%	28.0%	27.6%	1036	1220	1296	1144	1126	1,452,313	2,128,708	2,474,947	1,818,709	1,743,451
7	25.4%	28.5%	29.7%	27.2%	27.0%	1024	1195	1267	1116	1106	1,397,507	2,003,307	2,312,827	1,706,406	1,663,295
9	25.0%	27.7%	28.8%	26.5%	26.4%	1007	1147	1219	1077	1076	1,351,997	1,805,598	2,089,358	1,566,661	1,552,328
11	24.9%	27.3%	28.4%	26.1%	26.1%	1000	1132	1202	1066	1070	1,329,109	1,740,271	1,993,535	1,520,606	1,528,795
13	24.7%	27.0%	28.1%	25.8%	26.0%	997	1118	1187	1053	1060	1,315,809	1,680,917	1,926,849	1,461,276	1,486,672
15	24.6%	26.8%	27.8%	25.6%	25.8%	995	1106	1165	1045	1052	1,311,436	1,633,311	1,851,382	1,449,704	1,466,723
21	24.4%	26.3%	27.3%	25.2%	25.4%	993	1083	1134	1034	1044	1,304,274	1,552,040	1,729,298	1,398,961	1,427,315
35	24.3%	26.1%	27.1%	25.0%	25.3%	998	1075	1133	1028	1041	1,313,786	1,512,344	1,685,860	1,382,979	1,415,708
50	24.2%	25.9%	26.9%	24.8%	25.2%	991	1062	1116	1013	1026	1,289,212	1,471,716	1,638,557	1,345,827	1,375,575
75	24.1%	25.6%	26.5%	24.6%	25.0%	994	1058	1106	1009	1023	1,305,072	1,457,769	1,604,329	1,340,529	1,368,640

Table I: Forecasts ignoring one network characteristic, m exogenous.

From Table I, first we see that forecasts using only one dimension (i.e., either topology or density) are still better than a naïve model that assumes any combination of topology and density is equally likely. On average, considering any one dimension gets us $\sim 2/3^{\text{rd}}$ of the way there, e.g., in terms of improvement in MAPE over the baseline Bass model forecast. Further, when calibration sample size is low, we lose less in forecasts from ignoring density; but when calibration sample size is high, we lose less by ignoring topology.