Marketing Science Institute Working Paper Series 2015
Report No. 15-100

How Free Digital Products Grow

Gil Appel, Barak Libai, and Eitan Muller

Report Summary

An intriguing development in the marketing landscape of recent years is the substantial increase in the quantity of digital products that are introduced to the market for free (using different business models to create profit following the free product). The free model has become ubiquitous across electronic platforms with app stores and large online sites offering a large variety of free products.

Recent reports highlight the omnipresence of the phenomenon: More than 90% of the downloaded smartphone applications in 2013 were free, with this percentage expected to continue rising in the near future. In established markets for digital products, a fierce battle is being waged among "freemium" and "premium" business models.

The large scale introduction of free products raises the question of whether our knowledge of the dynamics of conventional product markets can be applied to these new phenomena. A fundamental question regards the pattern of growth for free products. Would it follow the growth pattern recognized for conventional non-free new products? Would the relatively riskless adoption and the ubiquity of offers change the pattern by which customers adopt such products? These issues are essential for prediction, segmentation, and management in such markets. They are of interest also for marketers of traditionally non-digital goods that increasingly introduce free mobile applications and software that helps the brand in terms of promotions, service, and reputation.

Using a database that documents the life cycle of tens of thousands of digital products, authors Appel, Libai, and Muller provide first evidence that freeware's growth pattern departs from the life cycle pattern observed for classic new products—the traditional bell-shaped growth with a cumulative "s-shaped" pattern. They show that freeware products may actually start their market presence with a drop in demand, and identify archetype growth patterns for such free products. They further show that the growth pattern strongly depends on the product's popularity, where highly popular products (which are in fact a minority in the market) follow the bell-shaped growth, yet this tendency declines as popularity decreases. Finally the authors elaborate on how the nature of influence in such markets can explain the patterns observed.

These findings suggest that, when dealing with free digital products, managers may want to use caution in making predictions and marketing plans that are based on the conventional wisdom of traditional products. In particular, the tendency of past research to focus on popular products may bias our understanding of markets where many products are not superstars.

Gil Appel is a doctoral student in marketing at the Glazer Faculty of Business and Management, Ben-Gurion University of the Negev, Israel. Barak Libai is Professor of Marketing, Arison School of Business, Interdisciplinary Center, Herzliya, Israel. Eitan Muller is Professor of Marketing at the Stern School of Business, New York University and at the Arison School of Business, Interdisciplinary Center, Herzliya, Israel.

Acknowledgments
The authors would like to thank Zvi Gilula, Oded Netzer, Jacob Goldenberg and Gal Elidan for their advice and help during the research process.

Introduction

An intriguing development in the marketing landscape of recent years is the substantial increase in the number of new products available for free (Anderson 2009). Occasionally labeled *freeware* for digital products, such offers had been available for a while for computer software products supplied via online platforms, and more recently for smartphones and web applications. Some of this availability stems from the "freemium" business model under which a certain percentage of adopters will eventually upgrade to a less restricted version (Kumar 2014), or purchase in-app byproducts. Yet the increase in freeware also follows other developments such as the rise of open-source software projects, where many users join forces to produce software products that will be free except for technical support (Mallapragada et al. 2012).

The free model has thus become ubiquitous across electronic platforms with app stores and large online sites such as CNET's download.com and SourceForge, offering a huge variety of freeware. Recent reports highlight the omnipresence of the phenomenon: More than 90% of the downloaded smartphone applications in 2013 were free, with this percentage expected to continue rising in the near future (Olson 2013). In established markets such as task management tools and anti-virus programs, a fierce battle is being waged among "freemium" and "premium" business models (Dunn 2011; Woods 2013).

The question we raise here is to what extent the growth of free product follows the classic new product growth pattern observed in the past, such as in durables, pharmaceuticals, or services. Current literature is unambiguous on this issue, suggesting that growth in digital environments in general (Rangaswamy and Gupta 2000) and freeware in particular (Jiang and Sarkar 2010; Lee and Tan 2013; Whitmore et al. 2009; Yogev 2012) should follow the commonly observed s-shaped diffusion patterns – with bell-shaped non-cumulative growth – and can be thus analyzed using traditional diffusion models.

This conventional wisdom can be questioned on at least two fronts. One issue is the change in adoption behavior when no price is charged. The bell-shaped growth of new products had been largely attributed to the risk associated with new products, which gradually decreases with information on others' adoption thereof (Rogers 2003; Bass 1969). In a relatively risk-free environment, this logic may be questioned, as information from others may not be that essential. Observations that people tend to download more freeware applications than they actually use are evidence of the non-conventional decision making that may be associated with freeware (Gupta 2013).

A second issue is the ubiquity of less popular products among freeware. Conventional products are associated with significant R&D costs, as well as costs of manufacturing, marketing, and maintaining market presence. Therefore, firms will invest in screening before market launch, and will be motivated, internally or due to channel pressure, to take a product off the shelf if it seems to fail. The case of freeware differs, in particular due to the low barriers to development and introduction of digital products of that type into the market. The share of small and less experienced suppliers is large, and investments in marketing are relatively low. Thus, it is reported that a large share of freeware products are considered a failure, and eventually do not even cover development costs (Foresman 2012; Rubin 2013). Because typically the costs of keeping the product alive are low, even with low demand, we can expect a high presence of relatively unpopular "long tail" products in the market.

Yet we know little about the growth pattern of unpopular products. In fact, one of the essential concerns with the understanding of the diffusion of innovations is that nearly all knowledge comes from what are considered successful innovations (Greve 2011; Rogers 2003). This is true also for the case of freeware, where the limited evidence we have on patterns of freeware growth is almost entirely based on highly successful examples, and it is not clear how much it represents the actual growth of most freeware products in the market.

We believe these questions provide a strong basis for an in-depth exploration of the growth of free products. The shape of the growth curve is considered "the most important and most widely reported finding about new product diffusion" (Chandrasekaran and Tellis 2007). Studying growth patterns is a fundamental stepping stone to the understanding of markets for new products: It is used to understand the driving forces of new products' success; as a base for modeling and optimizing firm behavior in the context of new product introductions; for decisions of termination or further support for new products; and for segmentation by adoption times (Golder and Tellis 1997; Peres et al. 2010). Given the role of freeware in contemporary digital markets, this issue is of essential theoretical and managerial importance.

Fortunately, digital environments provide an opportunity to conduct a large-scale analysis in a way seldom available to new product growth researchers. Historically, due to the difficulty of obtaining reliable new product growth data, studies were conducted on a relatively small number of product types, with an additional problem of a lack of data on the product's early days (Jiang et al. 2006). In this study, we use data on the monthly growth since launch-day of a large number of software products in a number of categories, with downloads per product ranging from a few hundred to millions, making this one of the largest new product diffusion studies to date. Our main data source is the SourceForge database, which enables us to study the growth of close to 60,000 free products, mostly intended for PCs. We are able to complement this analysis by also looking at data on the growth of more than 7,000 mobile apps.

Our results show that the growth pattern of freeware indeed differs from the commonly studied durables and services. In particular, we see the following:

- Most freeware products start their market presence with a drop in demand.
- One can see three archetypes of growth patterns: *Diffuse* (which is consistent with the bell-shaped curve), *Slide* (an exponential-type constant decline), and *Slide & Diffuse* (S&D), which begins with a long enough Slide, yet turns into a Diffuse at some point.
- Looking at freeware at various levels of popularity (number of downloads), we see a clear correlation between share of pattern archetypes and level of popularity. A larger share of Diffuse patterns and a smaller share of Slide and S&D characterize products that are more popular.
- Among the highly popular products (at least one million downloads), more than 96% exhibit the Diffuse pattern. The tendency of past research to focus on popular products may thus explain how this diverse nature of freeware growth has not been thus far identified.

These findings have essential importance for our understanding of growth in freeware markets, and for attempts to model and optimize growth in such markets. From a broader theoretical sense, it implies that the generalization developed on the product life cycle, its turning points, and its drivers (Golder and Tellis 2004) may need re-examining in the rapidly growing, dynamic world of free products. After we describe the findings, we elaborate on possible explanations for this compound growth pattern observed. In particular, we argue that a phenomenon of an early-on disproportional external influence, which should be expected in markets for freeware, can explain the patterns observed, and we discuss the implications of our findings.

Background

Our study relates to a number of research streams:

Markets for free products: The term *freeware* is used to describe a number of market models under which software is provided to users (at least initially) at no monetary cost. The rationale for such strategies varies. A common case is that of *freemium* under which a subset of consumers pay for a fuller version at a later stage (Kumar 2014); however other models such as profits from advertising, in-game purchases, and complementary paid products are also becoming increasingly popular (Washington and Miller 2010). The ubiquity of freeware is further affected by the increase in use of open-source software, where the aim to make software a freely available good is part of a movement that advocates the development and use of such free products (Kumar et al. 2011).

Research in this area has examined issues such as optimal initial spread of freeware as part of profit maximization in the longer run (Cheng and Liu 2012; Niculescu and Wu 2014), free-riding and competitive dynamics (Haruvy and Prasad 2005; Kumar et al. 2011), and the impact of the creation process on success (Grewal et al. 2006; Mallapragada et al. 2012). Other research has focused on the effect on demand by bestseller ranking and consumer ranking (Carare 2012; Lee and Tan 2013), as well as other factors such as price discount on in-app purchases (Ghose and Han 2014; Jung et al. 2012). We add to this literature by providing the first large-scale analysis of the growth patterns of freeware. This is of particular importance given the assumption that freeware grows in a manner similar to other products, so that freeware growth processes “do not significantly differ from the adoption and the diffusion of other goods described by the Bass diffusion model” (Lee and Tan 2013), which serves as a basis for modeling in this area (e.g., Jiang and Sarkar 2010).

Patterns of innovation growth: In a more general sense, our effort is related to the on-going efforts to study the pattern of new product growth, which spans various disciplines (Rogers 2003). The fact that the adoption rate of successful innovations follows a bell-shaped or logistic-type curve, and a cumulative S-shaped curve, is considered one of the fundamental discoveries of social science and the most important and widely reported finding regarding new product diffusion (Chandrasekaran and Tellis 2007) with the perception across disciplines that “the s-curves are everywhere” (Bejan and Lorente 2012). These observations form the bases of diffusion-of-innovations theory and forecasting new product growth using consistent growth shapes, such as the Bass model, Gompertz, or logistic curves (Meade and Islam 2006).

The dynamic role of social influence among customers in various forms (word of mouth, norms, network externalities) had been a dominant explanation for the s-shape pattern (Peres et al. 2010; Young 2009). In fact, even in markets where a constant decline pattern had been observed (such patterns are sometimes labeled “r-shaped” in contrast to “s-shaped” because the cumulative pattern resembles the shape of a “r”), communications effects were largely used as an explanation. In the case of low-involvement supermarket goods, the explanation is the lack of inter-customer social influence, and a dominant role of external effects such as advertising (Fourt and Woodlock 1960; Gatingon and Robertson 1985). For entrainment products such as movies, it is the concentration of marketing efforts early on, or the fact that the communication process may start well before the observed product launch associated with an r-shaped curve (Moe and Fader 2002; Sawhney and Eliashberg 1996).

There have been also arguments for other explanations for the growth patterns, in particular that heterogeneity among adopters in propensity to adopt, especially considering a decreasing price over time (Chandrasekaran et al. 2013; Van den Bulte and Stremersch 2004), or economic environment and technological cycles (Chandrasekaran and Tellis 2011) drive these patterns. Yet it is notable that these views also largely assume an s-shaped curve growth pattern.

Our research contributes to this stream by focusing on the relatively insufficiently considered growth of free products. Several issues are interesting in this case: One is the low risk associated with freeware adoption, which may render the need for information from other adopters less pressing than it is for higher-priced products. On the other hand, the selection of new freeware offers is extremely wide, which complicates choice and may drive users to rely highly on social media and customer ratings in the adoption decision (Lee and Tan 2013). As we will see next, the patterns that emerge are non-trivial, adding to what we know about turning points in the product life cycle for durables (Golder and Tellis 2004).

In addition, freeware may enable an interesting view of growth that is highly organic. Many factors can affect the adoption pattern of new products, and it may be idiosyncratic to specific categories, locale, and time periods, for example price decline over time, supply and infrastructure issues, and impact of changes in the economic environment. Freeware represents a relatively “organic” type of growth: It is clearly less affected by firms’ marketing strategies in specific markets, price changes, and economic shifts (Yogev 2012), which enabled us to study the fundamental growth process in a way that would be difficult otherwise.

New product success and failure: Finally, this study contributes to research on success and failure of new products. The question of what factors separate new products’ success or failure is naturally of high interest to the business community, and thus has attracted considerable research effort over the years (Evanschitzky et al. 2012; Griffin and Page 1996).

However, there is practically no empirical evidence for the growth pattern of unsuccessful new products, which may be attributable to difficulty in obtaining such information, and possibly to researchers' tendency to follow success stories (Rogers 2003).

Theoretically, one could speculate that an unsuccessful growth pattern will show a smaller-scale s-shaped curve, where the unpopular product then undergoes a diffusion process as the popular product does, yet with a lower effective market potential. However, if "failure" is associated with lack of positive word of mouth, we may expect an r-shaped (constantly declining) curve for a growth that is driven mostly by external influence, similar to low-involvement products (Garber et al. 2004). As part of our exploration, we provide the first large-scale evidence of the relationship between a product's popularity and its growth pattern.

Growth Patterns of SourceForge Freeware

Our primary source of data is SourceForge.net, a large, open-source software (OSS) repository that empowers software developers to control and manage open-source software, and enables users to download these products for free (Madey 2013). As of June 2013, when we downloaded the data, SourceForge had about 400,000 registered projects, with 3.4 million developers and 4 million downloads a day. As such, it is among the largest download sites online, and home to some well-known consumer software products such as VLC media player, eMule, and 7-Zip. In fact, many users may not be aware that products they download from various software download sites are actually being downloaded from SourceForge.

Using the SourceForge database, we retrieved the monthly history of downloads for a large number of products. The number of downloads is largely used to assess the success of open-source products (Daniel et al. 2013; Grewal et al. 2006) and in a broader sense acts as a proxy for the success of free products (Chandrashekaran et al. 1999). While SourceForge contains a large number of products (close to 400,000), many of them are inactive and had zero downloads, and thus are not relevant for our analysis. We focused on the download patterns for the 59,343 products that stood up to the following criteria:

- *Data from five years of growth*. We look at a 60-month window for all products. Naturally, the life cycles of freeware are considerably shorter than the typically analyzed growth of durables (although for some products, the cycle may be longer). Thus, to reduce cases of right censoring and to use a consistent time frame, we considered only products launched before mid-2008. Yet our analysis suggests that we cover the majority of downloads for the various products[1].

- *At least 200 downloads at the five-year window*. This enables us to capture actual growth processes that are not affected much by possible developers' noise over the product life cycle.

The data of the type we use presents a challenge to new product growth researchers. Previous new product growth efforts have typically focused on the analysis of a few to a few dozen products, which enabled the utilization of visual inspection in recognizing patterns of growth (Golder and Tellis 1997). Here we look at close to 60,000 different product growth curves. In addition, we use monthly data, which is much noisier than the yearly data typically used in new product growth research (Chandrasekaran et al. 2013), magnifying the challenge of pattern recognition.

To deal with this challenge, we clustered the various patterns in our data, employing the Functional Data Analysis (FDA) approach, which is a collection of techniques in statistics for the analysis of curves or functions (Foutz and Jank 2010; Sood et al. 2009). Appendix 1 provides some technical details for the use of FDA in our case.

The initial drop

Looking at the resulting clusters (Figure 1), we see that in most clusters, the pattern begins with some drop in downloads. To further understand this phenomenon, we examined the patterns of the individual products, looking for sequences of drops in the beginnings of the patterns. We found that only 38.7% of the projects began with an increase in downloads (meaning that the month 2 download number was higher than the first); and in 61.3% of the products, there was a drop in sales, with an average drop length of 9 months.

Result 1: *The growth pattern of most freeware products begins with a drop in the number of downloads.*

Growth archetypes

We next used the 18 clusters that emerged (Figure 1) to identify three apparent archetypes of non-cumulative growth patterns:

[1] SourceForge data is reported monthly. Because the first month is incomplete and with varying lengths, which can bias the results, we use the first full month for which we have data as the first month.

- *Slide*. Clusters 1, 2, 3, 12, and 13 demonstrate a pattern that begins at a relatively high level of downloads, then drops almost exponentially to a minimal level and remains there. When the share of downloads covered by this drop is more than 90% of the overall, as with the patterns in these clusters, we label the pattern a "Slide" (as it resembles a playground Slide). The average drop at the beginning of the Slide pattern is 11.7 months.
- *Diffuse*. Clusters 9, 15, and 17 demonstrate a clear pattern similar to the classic bell-shaped diffusion curve. Some patterns have a drop (yet considerably smaller compared to Slides, at 3.5 months on average), but otherwise are similar in type as well. When the share of drop is small, less than 10%, as in clusters 4, 7, 11, 14, 16, and 18, we consider them as a classic growth curve and label these patterns "Diffuse".
- *Slide & Diffuse*. Clusters 5, 6, 8, and 10 demonstrate a unique type of pattern: They begin with a drop similar to the "Slide", but after some time become a growth curve similar to the "Diffuse" pattern. In all of these patterns, the share of downloads covered by the drop is larger than 10% but smaller than 90%, and thus these cannot be part of the other two patterns. We will label the patterns in these clusters "Slide & Diffuse" (S&D) patterns. The average initial drop for an S&D archetype is 7.5 months.

Overall, of the 59,343 patterns, 34,710 (58.5%) are associated with a Diffuse pattern, 9,237 (15.6%) are associated with a Slide pattern, and 15,396 (25.9%) with the Slide & Diffuse pattern. Clearly, freeware growth is more complicated than the s-shaped curve assumed by the previous literature.

Result 2: *The growth of freeware can be grouped into three archetypes: s-shaped (Diffuse), r-shaped (Slide), and a combination of the two (Slide & Diffuse). While the s-shape may be the largest group, the other two represent a significant portion of the growth patterns.*

Growth and popularity

Given Result 2, an interesting question is why the evidence we do have on freeware growth highlights only the s-shape pattern (Lakka et al. 2012; Mäkinen et al. 2014; Whitmore et al. 2009; Yogev 2012). Our subsequent analysis of the relationship between popularity and download pattern can help to illuminate this issue.

To look at the effect of popularity, we divided the dataset by download ranking deciles (e.g., top 10%, 11%-20%) and examined the share of each archetype pattern at various deciles (see Table 1 for deciles statistics, Tables follow references throughout). Figure 2a shows the relationship between decile membership and the percentage of each archetype in every popularity decile. We see a clear "funnel" with a monotonic increase in the share of Diffuse patterns from low-popularity products to high-popularity ones, and nearly monotonic decrease in the share of Slide and S&D patterns as products become more and more popular. Diffuse is the clear majority in the top 10% (83%), yet is less than half of the cases (45%) in the bottom 10% of popularity.

Thus, the reason that only s-shaped had been considered to date in this category may be a function of product type selection bias by researchers. As in the case with other new products, freeware researchers (and software growth researchers in general) centered on the popular categories and products to analyze, sometimes focusing on the case of a single popular product. In fact, looking closely at our data, we see that with the most successful products, the dominance of the Diffuse pattern is especially strong: Over 95% of the million-plus downloaded products exhibit a Diffuse pattern. Therefore, when such successful products are analyzed, it is no wonder that an s-shaped pattern is observed. However, these are just a minority of the freeware growth patterns that exist in the market; the larger picture contains products with differing levels of popularity.

Result 3: *There is a robust relationship between freeware popularity and growth pattern ubiquity. Less popular products tend to exhibit more Slide and Slide & Diffuse patterns, while more popular products tend to have a higher share of Diffuse patterns.*

The early period implications

Given the above, it is interesting to see to what extent knowledge of the early period immediately after launch can help us explain the eventual popularity of a particular freeware. In Figure 3 we present a (log-scaled) scatter plot of the eventual popularity vs. the 1st month popularity of all individual products, noting the various archetypes. The overall correlation between the sum of downloads and the 1st month downloads is moderately high (ρ = 0.54 and ρ = 0.73 for log-scaled popularity). It is clearly shown how the archetypes classification we created – which was based on percentiles and did not take absolute popularity into account – is clearly related to popularity, both in the 1st month and overall. Indeed, the difference between the archetypes in these two factors was significant using a two-sample Hotelling's t^2 test.

Figure 3 provides some interesting insights. Firstly, although we saw that Diffuse patterns are associated more with popular products, ***given eventual popularity level***, the highest 1st month downloads are those of Slides. This may not be very surprising when considering that the first month is the highest overall for a Slide pattern. If a product reached a certain level of eventual popularity, it had to be higher early on to be considered a Slide.

However this does not mean that that greater first month downloads are associated with Slides. For a given first month popularity, we see that it is the Diffuse patterns that are responsible for the higher eventual popularity cases, and that Slides are underrepresented in the higher levels of eventual popularity[2].

Result 4a: *Patterns that start popular tend to stay popular.*

Result 4b: *Controlling for first month downloads, Slide & Diffuse patterns will get fewer downloads than Diffuse patterns; and Slide patterns will get the least downloads.*

Predicting popularity

Recapitulating the results above, we see that many patterns begin with a drop; that the patterns can be grouped into three separate archetypes; and that these archetypes, as well as the pattern's first-month popularity, are highly associated with overall popularity.

One may ask if we can use this information to predict the ultimate popularity of a certain product based on the first few months of its activity (using real, non-smoothed data). One insightful prediction option is to see if we can predict the overall decile of popularity (e.g., Top 10%; 11-20%) for these patterns using these early covariates. We use cluster analysis to classify the patterns into the ten popularity deciles[3].

For the classification process, for each pattern the following covariates were used:

[2] We also ran an OLS regression of the pattern on total downloads that replicated these findings, with the first month as a control. For the SourceForge dataset, Slides are associated with 90,000 downloads fewer than Diffuse patterns (for a given first-month popularity level); and S&D patterns with 60,000 downloads fewer than Diffuse.

[3] We used the K-Nearest Neighbor (KNN) algorithm (Punj and Stewart 1983). We used 70% of the data for training and k-fold cross-validation of the KNN model, and tested the predictions on the remaining 30%. Classification using a multinomial logistic regression yielded similar results.

- The number of downloads in the first month (log transformed)
- A drop dummy (1 if the downloads in the second month are lower than the first month, and 0 otherwise)
- The parent category (out of 16) of the pattern
- The intended audience (out of 29 options) of the pattern. “Intended audience” reflects a self-classification by creators of who will use the product (for example “developers”, “end users”, or “security”).
- The launch month of the pattern

The drop dummy and first-month popularity were the focal covariates (we did not include the pattern in the classification, as the pattern is not known in the first few months), and the other covariates are used as control.

The naïve estimator for success is 10% (if we randomly pair each pattern and popularity decile). Using only the data on the number of downloads in the first month, we were able to reach 22.6% success in classification (2.26 times better than the naïve estimator). While the drop dummy alone did not add much (10.7% classification success), together with the first drop, the success in prediction grows to 24.1% (6.6% improvement over the first month alone). When adding the control variables, they do not improve the prediction’s success.

Since the graphs examined are the unfiltered graphs, they might be noisy at the start, so that a longer period of examination may be needed to consider if there is a drop or not. Looking more than one month ahead, and setting the drop dummy when there is still a drop in the third month (rather than the second month), increases prediction success to 25.3%.

We also examined the ability to predict popularity if we split the data into three segments (top, mid, and bottom popularity); the naïve predictor should achieve 33.3% success in such a scenario. Using KNN, the prediction success was 61.5% (1.85 times better than the naïve predictor), and 62.7% when using three months for the drop dummy.

Another prediction option is to compare several linear regression models using combinations of the covariates described above (e.g., first month downloads, drop dummy), and examine the models using the single holdout method (Foutz and Jank 2010). In this approach, we hold out the i^{th} pattern out of the sample, estimate the model based on the n-1 remaining patterns, and then compute the MSE and MAPE for the overall n predicted values. We found that covariates such as the features of the software (i.e., category, intended audience) are less helpful in predicting the success of the pattern than are these covariates (see Appendix 2 for the full analysis).

Robustness Checks

We conducted a number of robustness checks to see to what extent the results are driven by the data and methods used. In particular we looked at three types of checks: A different product (smartphone apps), versions, and product features.

Other products: the case of smartphone apps

A question can be asked as to what extent findings from PC-based, open-source freeware can be generalized to other freeware environments. In particular, smartphones have become a prominent freeware distribution outlet, to the extent that the vast majority of smartphone apps are freeware (Olson 2013). Data on large-scale smartphone app adoption over time is not readily available for researchers (Garg and Telang 2013). We were yet able to obtain the cooperation of a global firm, which we will call "Mobility" so as not to reveal its identity, that is a player in a market of helping businesses to create free smartphone apps that can be used as part of their business. Under this business model, Mobility creates the app and helps manage it for the client for a monthly fee. Mobility clients are varied and include service providers such as restaurants, artists who distribute their music, educational institutions, and non -profits. These clients offer free apps created on the Mobility platform to their own end users, who are typically individual customers or prospects. Mobility can track these apps' downloads by end users over time.

The Mobility dataset is more limited than that of SourceForge in some aspects. The Mobility apps are specific to certain service providers, so are naturally relevant to much smaller market segments. In addition, unlike the case of open-source software, there is an entity (Mobility clients) that may make dedicated effort to push the freeware via external effects, which we do not observe. The time span we have for Mobility downloads is also more limited: a one-year window of weekly data for downloaded apps (between February 2011 and November 2013).

Due to the smaller magnitude of adoption, we used data on apps that had at least 50 downloads, taking a minimum of 52 weeks, and truncated at 52 weeks. We thus had weekly adoption data for 7,479 smartphone apps. We repeated the Functional Data Analysis (FDA) as in the first data set, this time resulting in 14 clusters, resulting in similar shapes to those we found in the SourceForge dataset (Figure 4).

We found largely similar results to the case of SourceForge products. The patterns that emerged were grouped again in the same order of size as in SourceForge into the archetypes of Diffuse (55.2%), Slide (25.7%), and Slide & Diffuse (19.1%). We can see that while the share of Diffuse is close to that of SourceForge, the share of Slides is higher at the expense of S&D. Here too, we observed the ubiquity of drops at the beginning of the pattern, this time in 64.1% of the apps (vs. 61.3% in the SourceForge dataset), with an average drop length of 12.8 weeks (vs. 9 months in the freeware dataset). Table 2 further describes the similarities between these datasets.

Figure 2b shows the relationship between archetype share and popularity decile for Mobility. For Slide and for Diffuse, the patterns are similar to that of SourceForge freeware. S&D, which has the smallest share of the three for Mobility, is relatively stable in popularity. Looking at smaller popularity cases here, we see that the share of Diffuse pattern drops beneath that of Slide patterns for the last three deciles. Therefore, the regularity that created the "funnel" in Figure 2a continues, and for lower download numbers turns into an X shape. As with the case of SourceForge, we saw a correlation (ρ = 0.92 vs 0.54 in SourceForge) between the first-week popularity and overall popularity.

Versions

One might wonder to what extent the introduction of versions (software updates) affects popularity and the archetypes. Most of the products in our data set (63%) have released several versions over the examined life cycle. The number of versions or their release patterns revealed neither as a driving force in popularity nor in archetype. When examining patterns with only one version (no newer versions were released over the life cycle of the product), we also witness the same funnel of popularity as in Figure 2a. Also, in the case of Mobility apps, no new versions were released, and the correlation between popularity and archetype were even stronger. Finally, we examined the case of the S&D patterns. We wondered if the difference between popular and less popular S&D patterns could be explained by the difference in number of versions over the Slide period of the pattern. Interestingly, we found that in the case of S&D, unpopular (rather than popular) products released more versions (and versions per month) over their Slide periods.

Product features

Another issue is to what extent product features can drive the types of patterns we see. To examine this, we considered two types of features from SourceForge. One is product category: SourceForge data are divided into 16 parent categories, 56 sub-categories, and over 300 sub-sub categories (patterns may belong to one or more of these categories). The other feature is "intended audience" as suggested by the product's creators (e.g., developers, end users, security). 29 types of intended audiences were identified in SourceForge.

We did not find any correlation between our fundamental findings and the product types we could identify. Consider for example Figure 5, which is equivalent to Figure 2a - the relationship between popularity and archetype ubiquity - for the 16 parent categories offered by SourceForge. Moreover, in Table 3 we present the distribution of patterns within each parent category, excepting one category (Formats and Protocols, which is the smallest category with only a few hundred products). The funnel-like relationship of the figures shown is consistent across parent categories.

Discussion

Our large-scale analysis demonstrates the non-traditional growth pattern for freeware. We saw that most freeware growth patterns begin with some drop in sales, identified three archetypes of growth patterns – Diffuse, Slide & Diffuse, and Slide – and found that their share in the growth pattern of freeware depends heavily on the freeware's popularity, with the share of Diffuse growing monotonically with product popularity. We see this pattern for various kinds of freeware and product categories.

An interesting implication of these findings is that the tendency of new product growth research to focus on highly successful products may lead us to miss the growth dynamics of less popular products, which in the case of freeware can be a significant part of the market. It is a good question to what extent this phenomenon is freeware specific. Given that what we know about new product growth in general largely comes from successful products, it is not clear whether the shape of growth for less successful products in conventionally analyzed categories will follow the same trajectory as the successful ones. It is clearly a promising avenue for future research.

An essential implication of these findings is that the product lifecycle theory and generalizations, which were built using data on categories such as durables, services, and pharmaceuticals, should be re-considered vis-à-vis the emerging markets for freeware. The common generalization of an s-shaped curve that begins with growth, with turning points at takeoff, slowdown ("saddle"), and peak (Golder and Tellis 2004) may be more complicated for zero-price markets. Interestingly, behavioral research has suggested that consumer behavior toward products changes greatly when the price drops to zero (Shampanier et al. 2007). We may need to take a different look at the parts of the market that are characterized by zero-price offers.

Next, we want to consider the implications of these results on our understanding of the evolution of markets for free products. Of course, in the absence of individual-level behavioral data, our ability to explain adoption behavior is limited. Yet aggregate demand patterns have played a major role in our understanding of new product adoption theory, and the unique nature of our data presents a good opportunity to re-consider the dynamics in freeware markets.

A notable result we see is the early drop in demand. In some cases, this drop continues for the entire lifecycle (Slides). In others, it changes after a substantial decline to a growth process (S&D). In fact, even among the classic Diffuse growth process we identified, about a third start with a slight drop early on. Building on previous research, one could argue that this pattern is the symptom of a period with an absence of social influence among customers, (Fourt and Woodlock 1960; Gatignon and Robertson 1985). Thus failures, which are characterized by little positive word of mouth, should take on a Slide-like pattern (Garber et al. 2004). Our findings that Slides are disproportionately represented among less popular products are consistent with this explanation.

In the case of S&D, this explanation may also be consistent with the two-segment view of new product growth, where new products diffuse initially to an early market, and only later to a mainstream market (Goldenberg et al. 2002; Van den Bulte and Joshi 2007). When the communication within segments is based on both internal and external influence, we may see a temporary decline in growth or a “saddle” (Goldenberg et al. 2002). If one assumes that the first segment is affected by external influence only, it is straightforward to see that an S&D pattern as observed here will emerge.

Yet, there are reasons to question the no-social-influence explanation for the drop. Firstly, the rationale of the r-shaped pattern for low-involvement supermarket goods is that coordinated advertising and promotions can drive sales over time. However, unlike supermarket brands, freeware advertising budgets may be limited, if available at all (Comino and Manenti 2005; Foresman 2012). In addition, what we do know about the adoption processes of products such as smartphone apps is the opposite of this assumption: Social influence plays a major role in individuals’ download decisions (Aharony et al. 2011). The business literature also highlights the significant role of social influence in this respect, especially given the very large quantity of freeware available (Kolirin 2012). This is clearly true for the many kinds of freeware with network externalities and in particular games, which are a very popular freeware category that often exhibits network effects.

Indeed, market research surveys show that for smartphone apps, in addition to browsing the app store, word-of-mouth communications from friends and family play a central role in download decision (Haselmayr 2013; MTV 2011). In fact, browsing for freeware by itself may be highly affected by internal influence. Given the many choices consumers face, demand can depend largely on making it onto the bestseller charts (ADA 2014; Carare 2012). Since these charts are affected by the level of previous adoption, the number of previous users plays a central role in the chart-based choice.

The initial shock and freeware adoption

We suggest that the patterns witnessed here can be explained by an expected disproportional external influence on freeware adoption early on, yet this does not imply the absence of social influence at that time. These dynamics resemble the market for movies, where producers invest much of their media campaign budget, as well as product offerings in term of screens, at (or before) launch, which helps to explain the r-shaped nature of demand for successful movies (Sawhney and Eliashberg 1996). Yet, at the same time, movies enjoy word of mouth, which is particularly strong early on (Liu 2006). The relationship to popularity, however, is inverse for movies compared to what we see here for freeware: Where smaller-scale “sleeper” movies grow in an s-shaped manner, blockbusters will often exhibit a large-scale Slide (Ainslie et al. 2005); while for freeware, the Diffuse is the most common shape.

Why might early-period external effect be high for freeware? Firstly, note that given the large number of choices in the market, a new freeware has to fight for attention, and for the chance that consumers will discover it. This chance is largely affected by mentions from external sources. These can be “New” or “Staff-Picked” charts on influential download sites or reviews by bloggers and on relevant social media sites. A social media recommendation by one of a few individuals would be considered “external” rather than “internal” influence, as it does not directly affect the number of earlier adopters.

Due to the constant flow of freeware into the market, there is a good chance that these external effects will be particularly strong early on in the life of a new freeware, when attention is given to newcomers. It can also follow actions by freeware developers who want to push their products. Indeed, it is argued in the practitioners literature that given the noise of many other products, freeware producers have a very short period of time in which to generate the groundswell that can lead to attention by sources such as the charts in the app stores, and thus they must act early on (Rice 2013). Consequently, those freeware developers who invest in marketing may often do it in “burst campaigns” that are meant to get them on consumers’ radar early in the game (ADA 2014).

Such dynamics will result in an adoption process that may include a disproportionate effect of early external influence, which decays after a while, yet alongside the presence of internal influence from previous users. Can this process explain the patterns we observe? To further explore this issue, we created – in the spirit of the Generalized Bass model (Bass et al. 1994) – a Flexible Communication Model (FCM) extension to the Bass diffusion model (see Appendix 3 for details). In the Bass model new product growth is driven by two kinds of influence: Internal Influence comes from earlier adopters. Internal Influence can represent edffects such as word of mouth and imitation. External influence, on the other hand, comes from marketing efforts- Advertising, sales people and promotions. n the FCM model, both External Influence and Internal Influence can change (exponentially) over time. Using Non Linear Least Squares, we fitted this model, as well as the basic Bass model as a comparison, to the 18 clusters of Figure 1. A number of results are apparent:

- The Changing Influence Model fit well the diverse patterns of Figure 1, with an average Adjusted R-Square of over 96% (Appendix 3). It is interesting to compare it to the basic Bass model, which as expected, has a problem fitting the Slide & Diffuse pattern, or a Diffuse that begins with a slight drop (the average R-square of the Bass model is 76%).

- Looking at the parameter value of the FCM (Table 5), we see that a Slide does not imply the absence of internal influence. Where the Bass model will set internal influence to zero to capture a Slide pattern, under FCM, Slides are characterized by both internal and external influence.

- The fit results suggest that both external and internal influence are largely declining with time. This effect is consistent with individual-level findings on the diffusion of technological products (Risselada et al. 2014), and in particular with our assumption on the role of external influence. The only instance of an increase over time is that of internal influence for Slide & Diffuse, which indeed can explain the phase-transition-like shape. While internal influence is low initially, if it increases over time, later the pattern can change to a diffuse-like growth.

The overall picture for freeware growth

Given the above, we can try to paint an overall picture freeware developers can expect to see when they introduce a new product to the market.

Freeware do not come in one package. There are a number of growth archetype patterns. A changing role of external and internal influence and in particular an ***initial*** strong external effect can drive the type of pattern that will emerge.

Expect an initial drop. Most freeware entry to the market is followed by some drop in demand. Some will continue as Slide, and S&D. Some will be smaller drops that will quickly turn into a Diffuse pattern.

Popularity matters for shape. The ubiquity of growth patterns changes much with products' overall popularity. As products become popular, there is a higher prevalence of a Diffuse pattern, with a very high share of Diffuse among the very popular. Among the Diffuse patterns, the share of early period drops is lower for the more popular products. The pattern/popularity change is continuous across popularity deciles, with no "phase transition".

Early popularity still predicts best. Given the complexity of the growth patterns, it is hard to predict eventual popularity, or even growth pattern outcome, by early popularity numbers. What we see is that, ceteris paribus, the extent of early popularity is the best predictor of eventual popularity: Highly popular freeware early on tend to be ultimately popular.

Concluding comments

This study is just a step on the road to understanding growth markets for freeware, and much work remains for future research. For example, the exact effect of rating and rankings by websites and other consumers may have a strong effect on freeware pattern growth. Previous research suggests that these effects are not straightforward, and may vary in various cases, for example based on the product's popularity (Duan et al. 2009). A closer look may enable us a better understanding of the growth of freeware.

It is also a good question to what extent low-priced, yet still not free, software and applications grow in comparison to freeware. Would the zero price effect (Shampanier et al. 2007) make such products similar to higher priced offers, or do they still resemble the freeware patterns we identify? In a more general sense, some of the dynamics we discussed here such as the higher early external effect may be relevant to software growth in general. The new product growth literature (Peres et al. 2010) may need to take a close look at online growth as a particular case of diffusion. The richness of online data is a great opportunity in this regard. One of the promising avenues is by obtaining individual-level adoption data (Katona et al. 2011). This will enable better examination of the aggregate patterns observed here, and further explore the behavioral background to the market-level pattern. The research on freeware growth is just in its infancy.

Appendix 1: Clustering the growth curves using Functional Data Analysis

Functional Data Analysis (FDA) is a collection of techniques in statistics for the analysis of curves or functions. Going into the details of the techniques is beyond the scope of this paper, yet interested readers are welcomed to explore the specific methods we used using the references below.

At first stage, we smoothed the graphs using Hodrick-Prescott (HP) filter to separate any short-term noises, retaining only the long-term trend (Chandrasekaran and Tellis 2011; Hodrick and Prescott 1997). We also converted the data to percentages, to remove the effect of popularity on our shape analysis. For the remainder of the paper, unless clearly stated otherwise, we will consider only the filtered, five-year patterns.

The second step of the FDA approach is Principal Component Analysis (PCA), reducing the number of dimensions in the data to five principal components. We used the broken stick selection criterion to select the number of relevant components, adding information above ordered random proportions (Jackson 1993). Next, we grouped the patterns using k-means clustering on the five selected principal components. We used the "Jump" approach (Sugar and James 2003) to find the optimal number of clusters (k) to be 18. Figure 1 shows the mean patterns for these clusters (Figures follow references throughout).

Appendix 2: **Comparing Prediction Models in the SourceForge Dataset**

In the SourceForge dataset, we demonstrated that many patterns begin with a drop, that the patterns can be grouped into three separate archetypes, and that these archetypes as well as the pattern's first-month popularity are highly associated with the pattern's overall popularity (results 1-4). Above, we use this information to better predict the ultimate popularity of a certain pattern from the first two months of its activity (using real, non-smoothed data), leading to predictions two times better and more than the naïve prediction. Another analysis is to compare several regression models using a combination of these covariates, using accuracy measures such as mean absolute percentage error (MAPE) or mean squared error (MSE). We use the same covariates as in the classification model above, this time using the popularity (measured by the log-transformed total number of downloads) as the dependent variable. For each pattern, the following covariates were used (from the non-filtered data):

1. The number of downloads in the first month (log transformed)
2. A drop dummy (1 if the downloads in the second month are lower than the first month and 0 otherwise)
3. The parent category (out of 16) of the pattern
4. The intended audience (out of 29) of the pattern
5. The launch month of the pattern

Table 4 describes the selected variables for each of the compared models and their accuracy and fit measures. We use the single holdout method (Foutz and Jank 2010), where we hold out the i^{th} pattern out of the sample, estimate the model based on the n-1 remaining patterns, and then compute the MSE and MAPE for the overall n predicted values. We can see that while using all covariates leads to the best fitted model, the features coefficients add very little to the fit, and the first-month downloads followed by the drop dummy lead to the largest increase in the model's accuracy, similar to what we found in the classification analysis above.

Appendix 3: **A Flexible Communication Model (FCG) and the Fit to the 18 Clusters**

To examine the effect of differential communication intensity over time in the context of the freeware patterns we observe, we created a flexible communications growth model (FCG). We begin with the Bass function and let both the external as well as the internal parameters be a varying function of time:

(1) $$\frac{dx}{dt} = \left(pe^{-\delta_1 t} + qe^{-\delta_2 t} \cdot \frac{x(t)}{M} \right) \left(M - x(t) \right)$$

If decay parameters δ_1 and δ_2 are positive, social influence intensity decays with time. Note that we have translated the data into percentages, and standardized it so that $F(60) = 1$, where $f(t)$ and $F(t)$ are the percentage and cumulative percentage of adopters, respectively. Thus, the equation has to be translated into percentages in the standard manner by dividing both sides by the market potential. However, in our data, we have right censoring, and so sales at Period 60 are not zero. Therefore, when $F = 1$, sales are still positive, which is captured by the M parameter where M might be larger than 1. The final model is given by the following:

(2) $$f(t) = \left(pe^{-\delta_1 t} + qe^{-\delta_2 t} \cdot \frac{F(t)}{M} \right) \left(M - F(t) \right)$$

Indeed this describes the shapes of the 18 clusters well with an average R-Square of 96.5%. An example is given in Figure 6, a Slide & Diffuse pattern, where the external coefficient begins large but declines rather rapidly, while the word-of-mouth parameter starts small yet increases over time. In Table 5 we average the resultant coefficients by category to which the cluster belongs: Slide, Slide & Diffuse, and Diffuse.

References

ADA (2014); "Discoverability: How to Get Noticed in a Marketplace Overflowing with Apps," Working paper, Application Developers Alliance, Washington, DC.

Aharony, Nadav, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland (2011), "Social fMRI: Investigating and shaping social mechanisms in the real world," *Pervasive and Mobile Computing* 7 (6), 643-659.

Ainslie, Andrew, Xavier Drèze, and Fred Zufryden (2005), "Modeling movie life cycles and market share," *Marketing Science* 24 (3), 508-517.

Anderson, Chris (2009), *Free: The Future of a Radical Price*, Hyperion, New York.

Bass, Frank M. (1969), "A New Product Growth for Model Consumer Durables," *Management Science* 15 (5), 215-227.

Bass, Frank M., Trichy V. Krishnan, and Dipak C. Jain (1994), "Why the Bass model fits without decision variables," *Marketing science* 13 (3), 203-223.

Bejan, Adrian, and Sylvie Lorente (2012), "The S-Curves Are Everywhere," *Mechanical Engineering* 134 (5), 44-47

Carare, Octavian (2012), "The impact of bestseller rank on demand: evidence from the app market," *International Economic Review* 53 (3), 717-742.

Chandrasekaran, Deepa, and Gerard J. Tellis (2007), "A Critical Review of Marketing Research on Diffusion of New Products," in Naresh K. Malhotra, Review of Marketing Research, 3, Armonk: M.E. Sharpe, 39-80.

Chandrasekaran, Deepa, and Gerard J. Tellis (2011), "Getting a Grip on the Saddle: Chasms or Cycles?," *Journal of Marketing* 75 (4), 21-34.

Chandrasekaran, Deepa, Joep Arts, Gerard J. Tellis and Ruud Frambach (2013), "Pricing in the International Takeoff of New Products," *International Journal of Research in Marketing* 30 (3), 249-264.

Chandrashekaran, Murali, Raj Mehta, Rajesh Chandrashekaran and Rajdeep Grewal (1999), "Market motives, distinctive capabilities, and domestic inertia: A hybrid model of innovation generation," *Journal of Marketing Research* 36 (1), 95-112.

Cheng, Hsing Kenneth, and Yipeng Liu (2012), "Optimal software free trial strategy: The impact of network externalities and consumer uncertainty," *Information Systems Research* 23 (2), 488-504.

Comino, Stefano, and Fabio M. Manenti (2005), "Government policies supporting open source software for the mass market," *Review of Industrial Organization* 26 (2), 217-240.

Daniel, Sherae, Ritu Agarwal, and Katherine J. Stewart (2013), "The Effects of Diversity in Global, Distributed Collectives: A Study of Open Source Project Success," *Information Systems Research* 24 (2), 312-333.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2009), "Informational cascades and software adoption on the internet: an empirical investigation," *Management Information Systems Quarterly* 33 (1), 23-48.

Dunn, John E. (2011), "Free Antivirus Grabs More Market Share, Claims Opswat Survey," *Techworld*, June 8, http://news.techworld.com/security/3284838/free-antivirus-grabs-more-market-share-claims-opswat-survey.

Evanschitzky, Heiner, Martin Eisend, Roger J. Calantone, and Yuanyuan Jiang (2012), "Success Factors of Product Innovation: An Updated Meta-Analysis," *Journal of Product Innovation Management* 29 (S1), 21-37.

Foresman, Chris (2012), "iOS app success is a "lottery": 60% (or more) of developers don't break even," *arstechnica*, May 4, http://arstechnica.com/apple/2012/05/ios-app-success-is-a-lottery-and-60-of-developers-dont-break-even.

Fourt, Louis A., and Joseph W. Woodlock (1960), "Early Prediction of Market Success for New Grocery Products," *Journal of Marketing* 25 (2), 31-38.

Foutz, Natasha Zhang, and Wolfgang Jank (2010), "Research Note—Prerelease Demand Forecasting for Motion Pictures Using Functional Shape Analysis of Virtual Stock Markets," *Marketing Science* 29(3), 568–79.

Garber, Tal, Jacob Goldenberg, Barak Libai, and Eitan Muller (2004), "From density to destiny: Using spatial dimension of sales data for early prediction of new product success," *Marketing Science* 23 (3), 419-428.

Garg, Rajiv, and Rahul Telang (2013), "Inferring app demand from publicly available data," *MIS Quarterly* 37 (4), 1253-1264.

Gatignon, Hubert, and Thomas S. Robertson (1985), "A Propositional Inventory for New Diffusion Research," *Journal of Consumer Research* 11(4), 849-867.

Ghose, Anindya, and Sang Pil Han (2014), "Estimating Demand for Mobile Applications in the New Economy," *Management Science,* 60 (6), 1470-1488.

Goldenberg, Jacob, Barak Libai, and Eitan Muller (2002), "Riding the Saddle: How Cross-Market Communication Can Create a Major Slump in Sales," *Journal of Marketing* 66 (2), 1 - 16.

Golder, Peter N., and Gerard J. Tellis (1997), "Will it ever fly? Modeling the takeoff of really new consumer durables," *Marketing Science* 16 (3), 256-270.

Golder, Peter N., and Gerard J. Tellis (2004), "Growing, Growing, Gone: Cascades, Diffusion, and Turning Points in The Product Life Cycle," *Marketing Science* 23 (2), 207-218.

Greve, Henrich R. (2011), "Fast And Expensive: The Diffusion of a Disappointing Innovation," *Strategic Management Journal* 32 (9), 949-968.

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

Griffin, Abbie, and Albert L. Page (1996), "PDMA success measurement project: recommended measures for product development success and failure," *Journal of Product Innovation Management* 13 (6), 478-496

Gupta, Sunil (2013), "For Mobile Devices, Think Apps, Not Ads," *Harvard Business Review*, 91 (3), 70-75.

Haruvy, Ernan, and Ashutosh Prasad (2005), "Freeware as a Competitive Deterrent," *Information Economics & Policy* 17 (4), 513-534.

Haselmayr, Melanie (2013), "App Discovery: Why Can't Anyone Figure This Out Yet?," *Forbes*, August 15, http://www.forbes.com/sites/allbusiness/2013/08/15/app-discovery-why-cant-anyone-figure-this-out-yet.

Hodrick, Robert J., and Edward C. Prescott (1997), "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking* 29 (1), 1–16.

Jackson, Donald A. (1993), "Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches," *Ecology* 74 (8), 2204–14.

Jiang, Zhengrui, and Sumit Sarkar (2010), "Speed Matters: The Role of Free Software Offer in Software Diffusion," *Journal of Management Information Systems* 26 (3) 207-240.

Jiang, Zhengrui, Frank M. Bass, and Portia Isaacson Bass (2006), "Virtual Bass model and the left-hand data-truncation bias in diffusion of innovation studies," *International Journal of Research in Marketing* 23 (1), 93-106.

Jung, Euy-Young, Chulwoo Baek and Jeong-Dong Lee (2012), "Product survival analysis for the app store," *Marketing Letters* 23 (4), 929-941.

Katona, Zsolt, Peter Pal Zubcsek, and Miklos Sarvary (2011), "Network effects and personal influences: The diffusion of an online social network," *Journal of Marketing Research* 48 (3), 425-443.

Kolirin, Assaf (2012), "Amid Too Many Choices, App Consumers Rely on Word-of-Mouth Marketing," *Online Marketing Strategies and Tactics*, May 30, http://www.emarketingandcommerce.com/article/amid-too-many-choices-app-consumers-rely-word-of-mouth-marketing/1.

Kumar, Vineet (2014), "Making "Freemium" Work," *Harvard Business Review* 92 (5), 27-29.

Kumar, Vineet, Brett R. Gordon, and Kannan Srinivasan (2011), "Competitive strategy for open source software," *Marketing Science* 30 (6), 1066-1078.

Lakka, Spyridoula, Christos Michalakelis, Dimitris Varoutas, and Draculis Martakos (2012), "Exploring the determinants of the OSS market potential: The case of the Apache web server," *Telecommunications Policy* 36 (1), 51-68

Lee, Young-Jin, and Yong Tan (2013), "Effects of Different Types of Free Trials and Ratings in Sampling of Consumer Software: An Empirical Study," *Journal of Management Information Systems* 30 (3), 213-246.

Liu, Yong (2006), “Word of mouth for movies: Its dynamics and impact on box office revenue,” *Journal of marketing* 70 (3), 74-89.

Madey, Greg, ed., (2013), “The SourceForge Research Data Archive (SRDA),” *University of Notre Dame*, February 14, http://srda.cse.nd.edu.

Mäkinen, Saku J., Juho Kanniainen, and Ilkka Peltola (2014), “Investigating Adoption of Free Beta Applications in a Platform-Based Business Ecosystem,” *Journal of Product Innovation Management* 31 (3), 451-465

Mallapragada, Girish, Rajdeep Grewal, and Gary Lilien (2012), “User-Generated Open Source Products: Founder's Social Capital and Time to Product Release,” *Marketing Science* 31 (3), 474-492.

Meade, Nigel, and Towhidul Islam (2006), “Modelling and forecasting the diffusion of innovation–A 25-year review,” *International Journal of Forecasting* 22 (3), 519-545.

Moe, Wendy W., and Peter S. Fader (2002), “Using Advance Purchase Orders to Forecast New Product Sales,” *Marketing Science* 21 (3), 347-364.

MTV (2011), “MTV Networks' Mobile Apps Study Reveals the Life Cycle of an App: From Discovered to Discarded,” *PRNewswire*, June 7, http://www.prnewswire.com/news-releases/mtv-networks-mobile-apps-study-reveals-the-life-cycle-of-an-app-from-discovered-to-discarded-123348433.html.

Niculescu, Marius F., Dong Jun Wu (2014), “Economics of free under perpetual licensing: Implications for the software industry,” *Information Systems Research* 25 (1), 173-199.

Olson, Parmy. 2013, “The Win For Games: They Grab Two-Thirds Of App Store Sales, “ *Forbes*, September 19, http://www.forbes.com/sites/parmyolson/2013/09/19/the-win-for-games-they-grab-two-thirds-of-app-store-sales.

Peres, Renana, Eitan Muller, and Vijay Mahajan (2010), “Innovation diffusion and new product growth models: A critical review and research directions,” *International Journal of Research in Marketing* 27 (2), 91-106.

Punj, Girish, and David W. Stewart (1983), “Cluster analysis in marketing research: review and suggestions for application,” *Journal of Marketing Research* 20 (2), 134–48.

Rangaswamy, Arvind, and Sunil Gupta (2000), “Innovation Adoption and Diffusion in the Digital Environment: Some Research Opportunities,” In *New-Product Diffusion Models,* eds. V. Mahajan, E. Muller, Y. Wind, 75-96. *New* York: Springer.

Rice, Kelly (2013), “Why pre-launch hype is the key to app success,” *Kinvey*, May 2, http://www.kinvey.com/blog/2545/why-prelaunch-hype-is-the-key-to-app-success.

Risselada, Hans, Peter C. Verhoef, and Tammo HA Bijmolt (2014), “Dynamic effects of social influence and direct marketing on the adoption of high-technology products,” *Journal of Marketing* 78 (2), 52-68.

Rogers, Everett M. (2003), *Diffusion of Innovations*. New York, N.Y.: The Free Press.

Rubin, Ben Fox (2013), “The Dirty Secret of Apps: Many Go Bust,” *Wall Street Journal*, March 7, http://online.wsj.com/news/articles/SB10001424127887324582804578346221047028366.

Sawhney, Mohanbir S., and Jehoshua Eliashberg (1996), “A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures,” *Marketing Science* 15 (2), 113-131.

Shampanier, Kristina, Nina Mazar, and Dan Ariely (2007), “Zero as a special price: The true value of free products,” *Marketing Science* 26 (6), 742-757.

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.

Sugar, Catherine A., and Gareth M. James (2003), “Finding the number of clusters in a dataset,” *Journal of the American Statistical Association* 98 (463) 750-763.

Van den Bulte, Christophe, and Stefan Stremersch (2004), “Social contagion and income heterogeneity in new product diffusion: A meta-analytic test,” *Marketing Science* 23 (4), 530-544.

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

Washington, Kelli, and Richard K. Miller (2010), "FREE," In *Consumer Marketing 2010*, 77-81. Atlanta, GA.: Richard K. Miller & Associates.

Whitmore, Andrew, Namjoo Choi, Anna. Arzrumtsyan (2009), “Open-source software: The role of marketing in the diffusion of innovation,” *Information Technology and Control* 38 (2), 91-101.

Woods, Dan (2013), “The Battle of the Freemium and Enterprise Business Models in the Task Management Market,” *Forbes*, March 18, http://www.forbes.com/sites/danwoods/2013/03/18/freemium-enterprise-business-models-task-management-attask-asana.

Yogev, Guy (2012), *The Diffusion of Free Products: How Freemium Revenue Model Changes the Strategy and Growth Of New Digital Products*, Saarbrücken, Germany: Lap Lambert Academic Publishing.

Young, H. Peyton (2009), “Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning,” *The American economic review* 99 (5), 1899-1924.

Tables and Figures

Table 1: Downloads Statistics

Decile	Mean	Median	St. Dev.
Top 10%	547,617	63,335	6,647,566*
11% - 20%	13,851	12,827	4,271
21% - 30%	5,823	5,655	1,184
31% - 40%	3,114	3,059	502
41% - 50%	1,868	1,851	255
51% - 60%	1,192	1,183	146
61% - 70%	793	787	90
71% - 80%	539	536	58
81% - 90%	370	368	41
Bottom 10%	248	247	30

* At the top 10% are 329 projects with more than 1 million downloads, while the median is 63,335; this drives up the standard deviation.

Table 2: Comparing SourceForge and Mobility datasets

Descriptive	**SourceForge**	**Mobility**
Projects in the dataset	59,343	7,479
Clusters found in the dataset	18	14
Time frame	60 months	52 weeks
Share of patterns with drop in the start of the life cycle	61.3%	64.1%
Average drop length	9 months	12.8 weeks
Mean drop by pattern: Diffuse	3.5 months	4 weeks
Mean drop by pattern: Slide	11.7 months	15.1 weeks
Mean drop by pattern: Slide & Diffuse	7.5 months	11 weeks
Archetype Distribution: Diffuse	58.5%	55.2%
Archetype Distribution: Slide	15.6%	25.7%
Archetype Distribution: Slide & Diffuse	25.9%	19.1%
Correlation between 1st month downloads and overall downloads	$\rho = 0.54$	$\rho = 0.92$

Table 3: Parent Category Statistics

Parent Category	No. of products in category	Diffuse	Slide	S&D
Development	16,539	59.1%	13.9%	27.0%
Internet	11,151	56.4%	18.4%	25.2%
System Administration	8,870	60.0%	13.8%	26.2%
Communications	7,229	55.9%	18.4%	25.7%
Science & Engineering	6,325	66.6%	7.7%	25.7%
Games	6,178	52.0%	19.9%	28.0%
Security & Utilities	4,234	58.6%	13.6%	27.8%
Audio & Video	3,887	61.4%	13.7%	24.9%
Business & Enterprise	3,767	58.0%	14.7%	27.2%
Home & Education	2,770	62.3%	9.4%	28.3%
Graphics	2,560	63.2%	13.1%	23.7%
Desktop Environment	2,077	55.1%	17.7%	27.2%
Other/Non-listed Topic	1,267	59.8%	14.5%	25.7%
Multimedia	917	61.9%	12.0%	26.1%
Mobile	439	58.8%	14.4%	26.9%
Formats and Protocols	289	57.8%	13.8%	28.4%
Average	4,906	59.2%	14.3%	26.5%

Table 4: Model comparison

Model	Description	Covariates	MAPE (%)	MSE
1	Product features	3,4	18.03	3.043
2	First month downloads	1	13	1.68
3	Data available on first month	1,3,4,5	12.86	1.641
4	Drop in second month	2	18.28	3.112
5	Downloads and second month drop	1,2	12.31	1.523
6	All covariates	1,2,3,4,5	12.19	1.496

Table 5: Average coefficients for each category of the 18 clusters

	Slide	**Slide & Diffuse**	**Diffuse**
External Coefficient – *p*	0.09	0.03	0.01
External Coefficient Decay Parameter - δ_1	0.34	0.21	0.10
Internal Coefficient – *q*	0.24	0.06	0.18
Internal Coefficient Decay Parameter - δ_2	0.15	-0.01	0.05
Market Potential – *M*	1.68	1.54	2.89
Adjusted R-Square	99%	94%	96%

Figure 1: The mean patterns for the 18 SourceForge clusters (plotted as popularity over time)

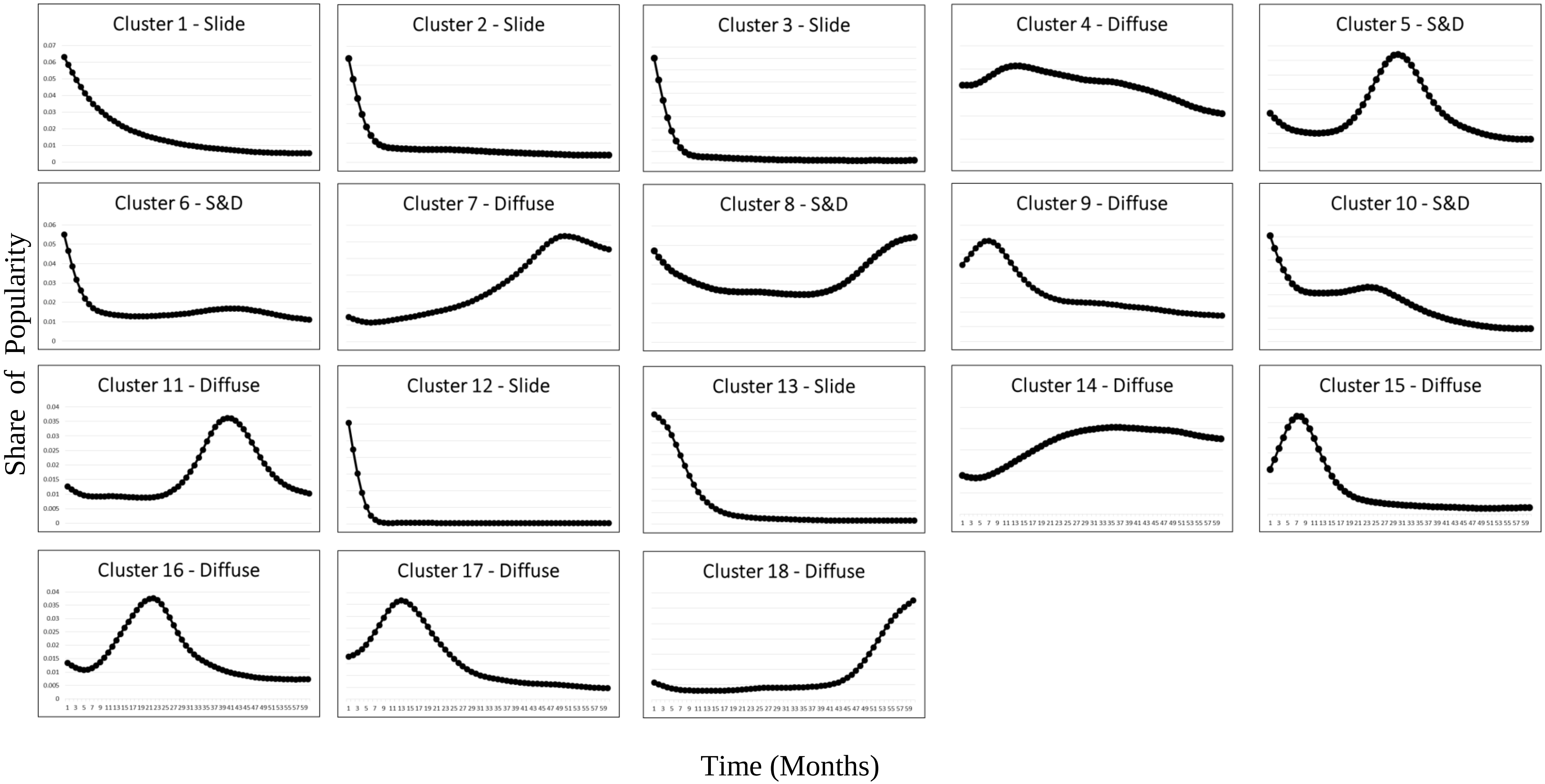

Figure 2: Examining pattern distribution by downloads in SourceForge and Mobility datasets

Figure 2a: Pattern distribution by downloads: SourceForge Freeware

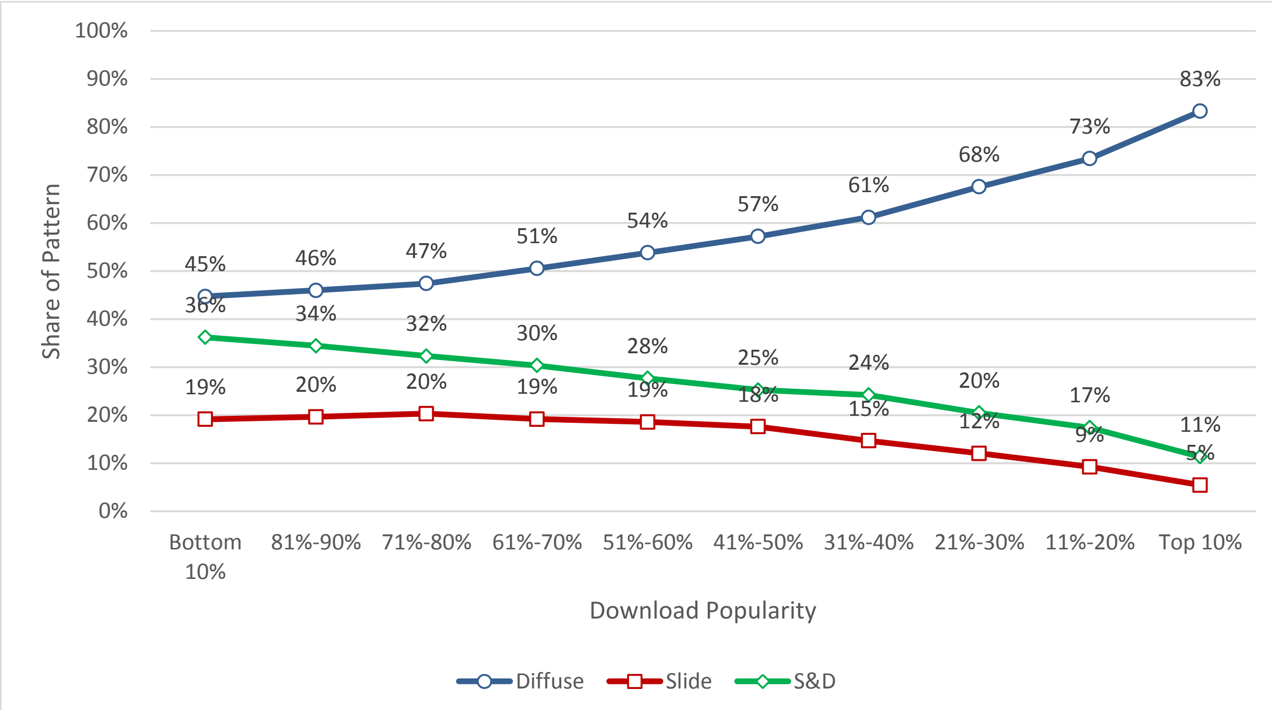

Figure 2b: Pattern distribution by downloads: Mobile Apps

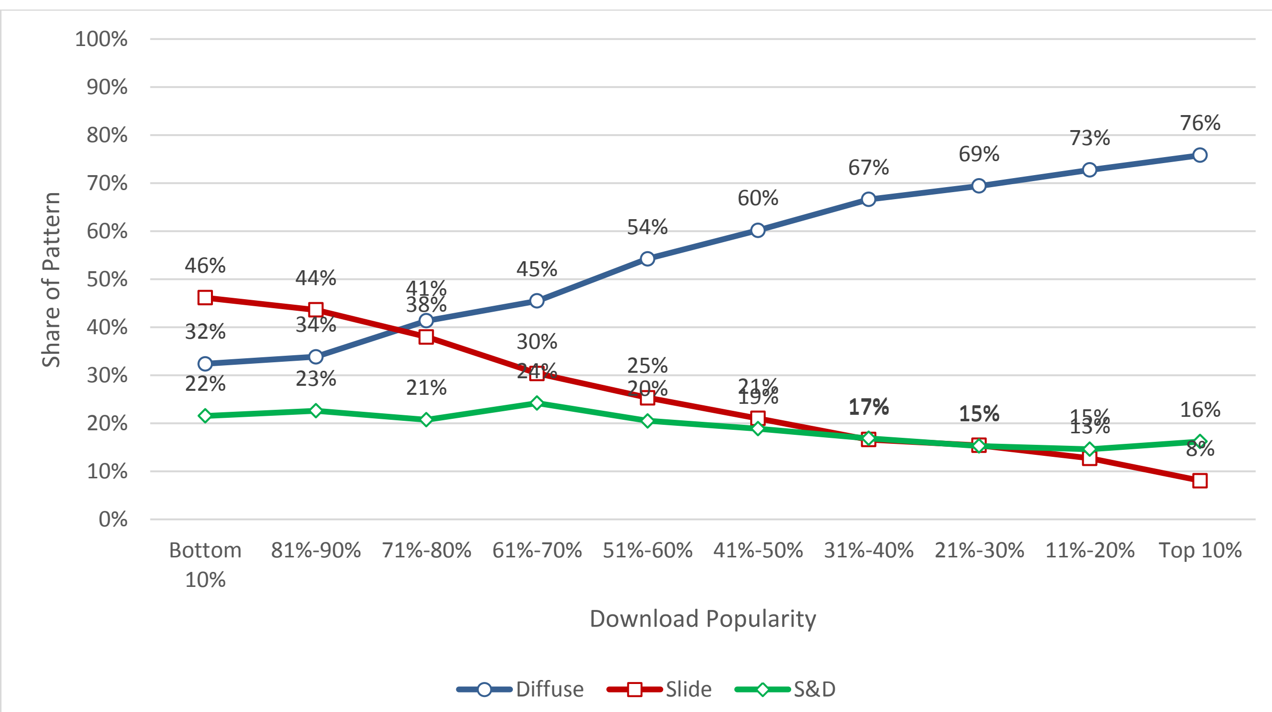

Figure 3: Pattern popularity vs. 1st month popularity (log scaled)

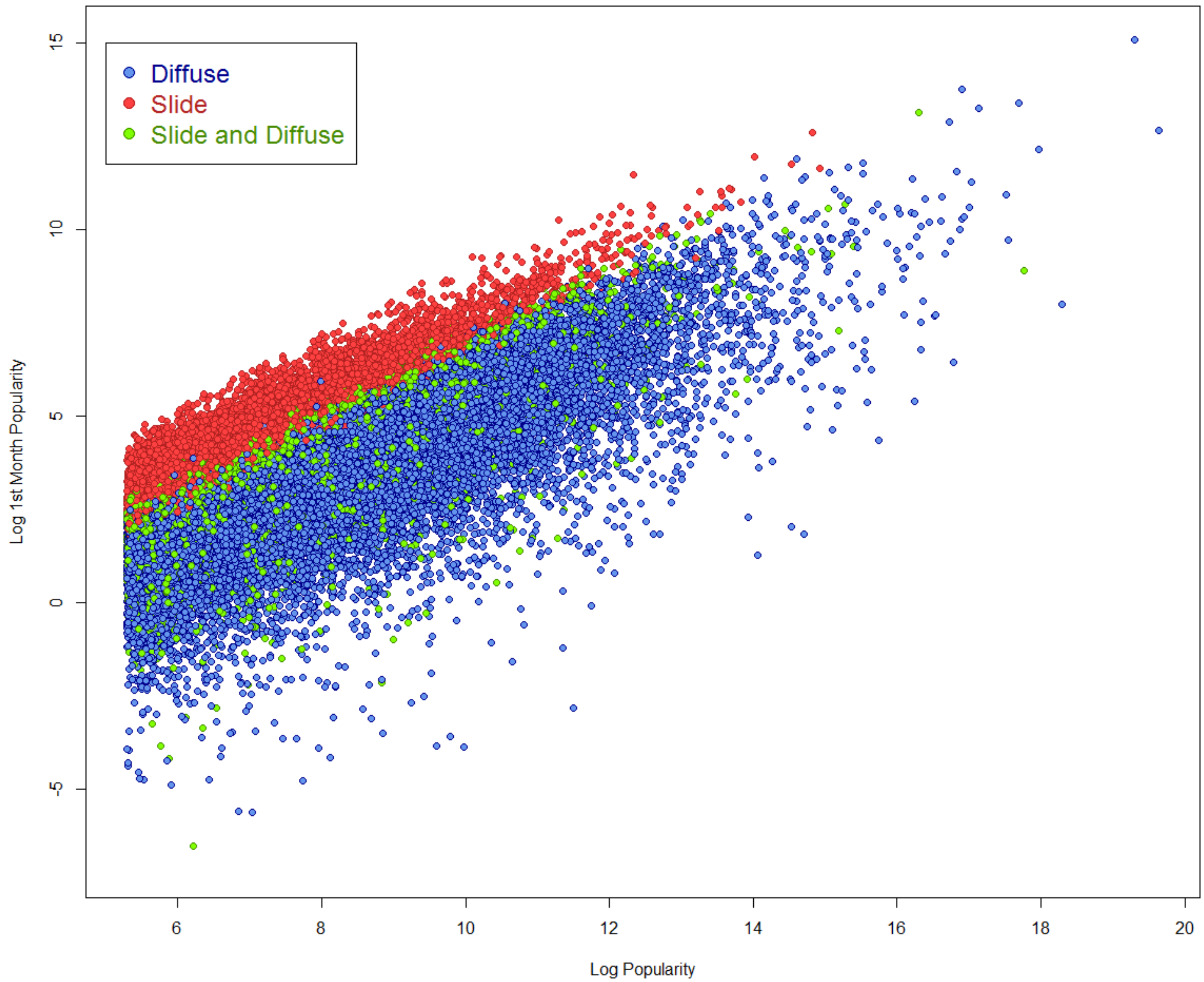

Figure 4: The mean patterns for the 14 Mobility clusters (plotted as popularity over time)

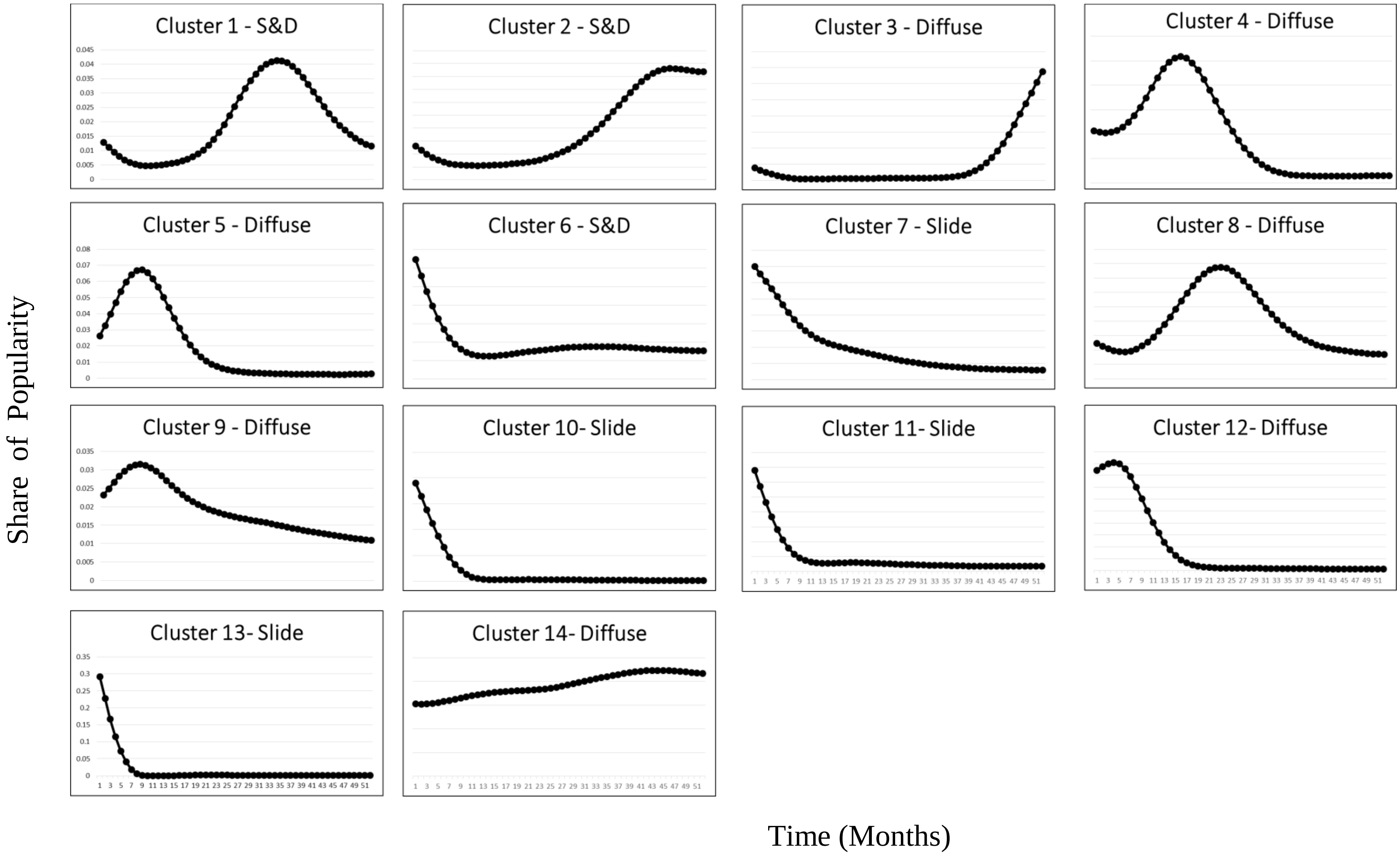

Figure 5: Pattern distribution by popularity within SourceForge's parent categories

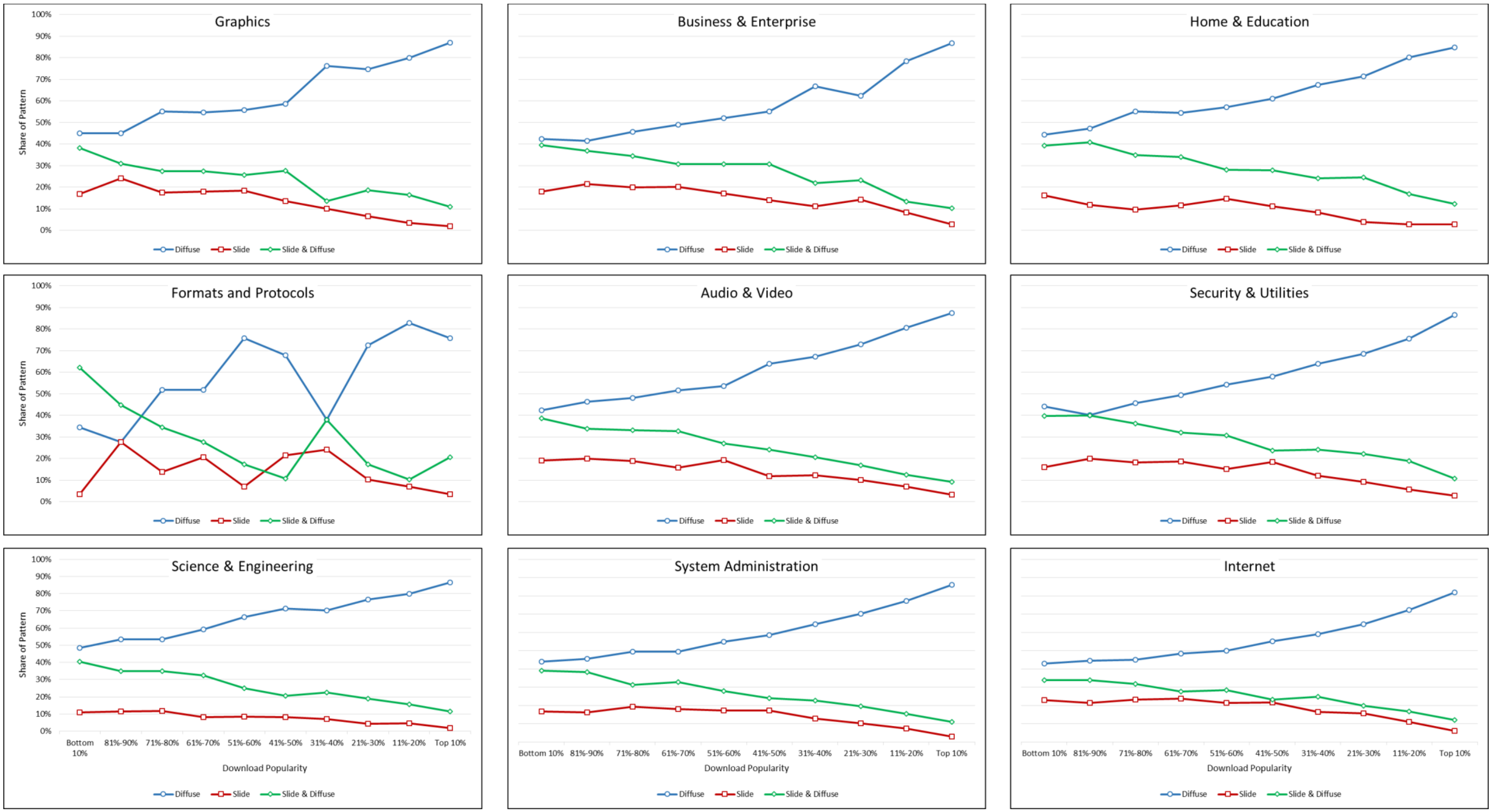

Figure 5: Pattern distribution by popularity within SourceForge's parent categories (continued)

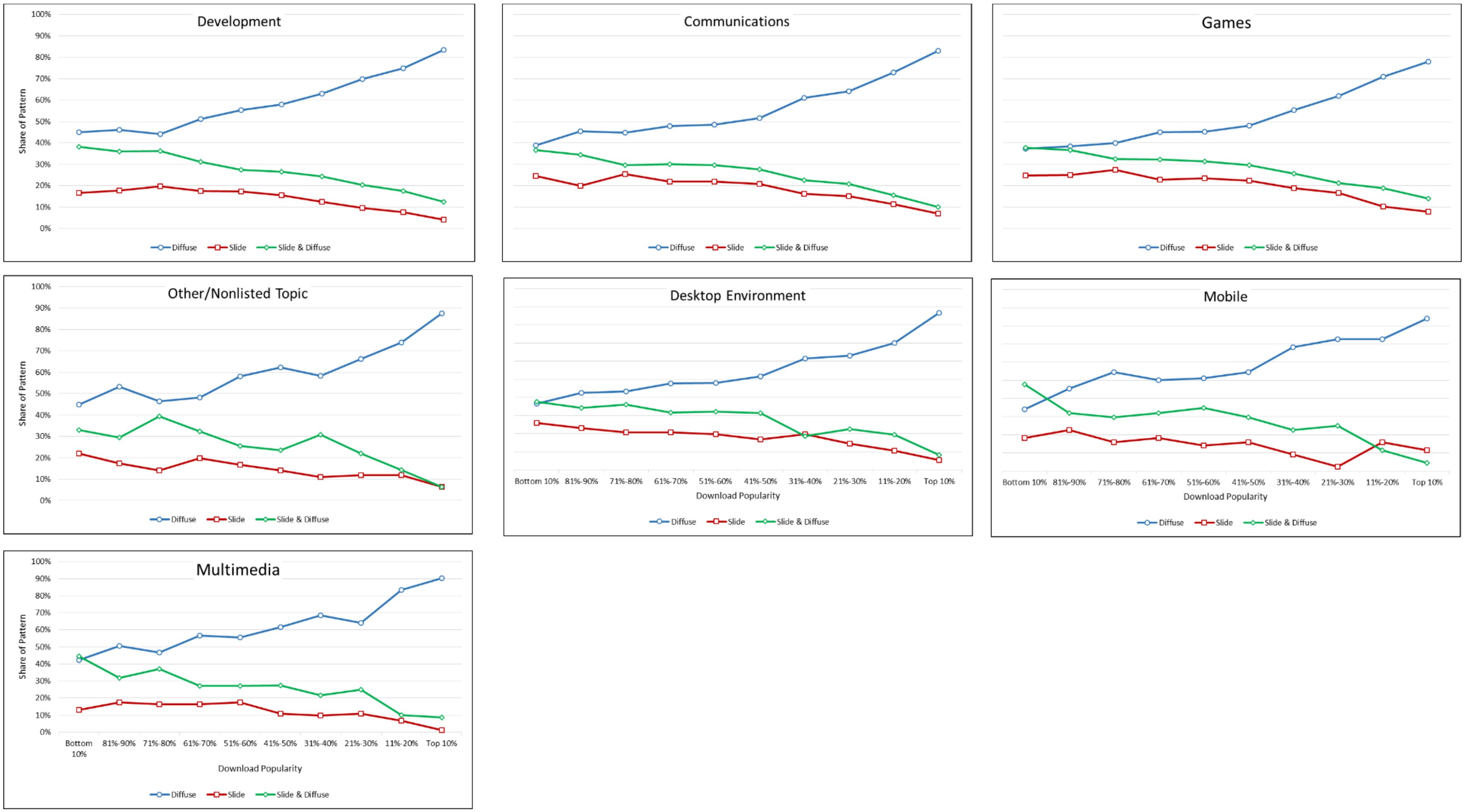

Figure 6: Cluster 6 with a rapid declining external shock, and constant small word of mouth