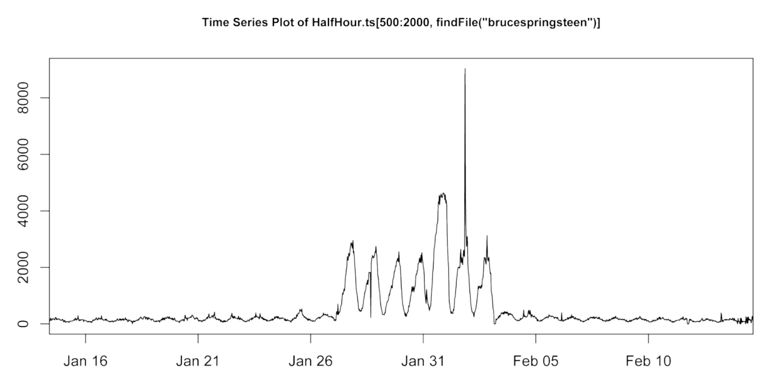
Endogenous and Exogenous Shocks to a Social System: Tracking Artist Page Views and Album Sales

Mark W. Ungerer

Professor Alain Kornhauser
Operations Research and Financial Engineering
Princeton University

May 12, 2008



Bruce Springsteen Myspace profile views experience a large shock following his January 27th, 2009 album release and obtain the most daily views on February 1, 2009 the day of his Superbowl XLIII Half Time Show appearance.

Abstract

This paper studies how people respond to media related events by examining Myspace profile views for 400 music artists. The data was gathered from January 4th, 2009 to April 1st, 2009 at half hour intervals. The patterns that emerge show similar behavior to how people share information in other complex networks such as buying books or watching YouTube videos. I show that Myspace profile views correlate with related Google searches and news articles. This shows Myspace to be a realistic representation of human interest in different artists as events occur. I then examine the list of artists that MTV chooses to promote and conclude that MTV uses no data based or algorithmic method to choose its most popular and new artists, and these promotions have little effect on profile views. Then I examine the correlation between Myspace profile views and album sales for a select 6 artists and find that a positive linear correlation exists. I then use the data to develop a simple model for estimating weekly album sales from profile views for the other 394 artists.

Table of Contents

Acknowledgements

This paper as well as my education here at Princeton represents a culmination of much that has been given to me. I give my greatest thanks to my parents and grandparents for somehow always knowing what is best for me. Without their continued love, support, advice, and guidance I would not be where I am. I would like to thank the teachers and professors I have had who have given me the quantitative and analytical tools that allowed me to perform this research. I would like to thank my friends for their feedback, guidance, and keeping me entertained. I owe my thanks to the Ivy Club and its staff who have kept me well fed throughout my Junior and Senior years. Finally, I would like to extend my sincerest thanks to Professor Alain Kornhauser for his guidance, criticism, and pushing me to do my best.

1

Music in a Social Network

How people discover, access and buy new music is a highly social process. The growth in popularity of an artist is the result of thousands of interactions between individuals connected through an essentially infinitely large social network. Accessing music brings together conscious, sub conscious, cultural, personal, and random factors to result in listening to a limited number of songs in a constrained amount of time. Songs playing on the radio, an ad on a TV show, or a t-shirt someone wears all affect what new music an individual discovers. Understanding the exact factors that lead to each individual in the network consuming the way they do is near impossible. This paper will assume that whatever the result, the reasons that individuals listen to different artists are the result of two types of effects: endogenous and exogenous shocks. Each of these shocks lead to a reordering of an individual's priority queue and affects what song, artist, or Myspace URL he or she visits next.

Personal recommendations can be an important source for finding new music. One recommendation can result in a propagation of new recommendations throughout the network. A recommendation is considered endogenous, meaning "rising from within",[1] and results in a small shock to the overall system. These shocks also take time to propagate and usually occur slowly over time. In the context of music this could be described as "word of mouth" growth in which an artist gains a fan base one fan at a time and the popularity grows in a smoother pattern. Other factors, such as televised concerts, release of a new single, or news of a band breakup can be considered exogenous shocks, meaning "coming from outside". These shocks result in a sudden large-scale jump in artist interest and album sales. In general, these exogenous shocks are short lived. See Table 1.1 for a summary of each shock and examples in the context of music's social network.

The Internet is one of the largest and most complex networks in the world. Through various hardware and software components it brings together millions of people and allows them to search for content, share information and learn about others. By tracking what pages and information people are accessing over time, the Internet provides real time data on what products, people, and topics are important. We now know better than ever before what exactly people are spending their time thinking about, what their eyeballs are watching, and where they are spending their money.

[1] http://dictionary.reference.com/browse/endogenous

Table 1.1 Shocks to a Music Social Network

Type	Description	Example
Endogenous	A small, slower acting shock coming from within the network	A friend recommends a new album
		An individual hears a new song on the radio
		Pandora.com suggests a new artist based upon your past musical interests
Exogenous	A large, sudden shock coming from outside the network, affecting many individuals within	Bruce Springsteen plays at Superbowl XVII
		Kid Rock performs at the Grammy's
		Kelly Clarkson releases a new single in preparation for her upcoming album

And people do more than just surf web pages online. According to recent statistics, people use the Internet for multiple activities, from sending email (90% of Internet users do this), to finding product information (62%), to buying products (26%).[2] People also share their opinions online through blogs, comments, and forums scattered throughout the Internet on thousands of different topics. Increasingly, people are using the Internet as a social network, allowing them to stay in touch with old friends and find new ones. Through these social networks they also learn about what people around them are doing. A social network user can see "status updates" on Facebook.com and read what his or her friends' favorite movies, books, and music are on Myspace.com. They can see someone's career summary on LinkedIn or follow updates from their friends and celebrities at Twitter. The way people use the Internet today, in what has been termed "Web 2.0", is ultimately social. See Table 1.2 for a list of social networks and data on their phenomenal growth and reach.

Table 1.2 Social Networks Worldwide Growth from June 2007 to June 2008

Network	June 2007	June 2008	% Change
Total Internet: Total Audience	778,310	860,514	11%
Social Networking	464,437	580,510	25%
Facebook.com	52,167	132,105	153%
Myspace.com	**114,147**	**117,582**	**3%**
Hi5.com	28,174	56,367	100%
Friendster.com	24,675	37,080	50%
Orkut	24,120	34,028	41%
Bebo.com	18,200	24,017	32%
Skyrock Network	17,638	21,041	19%

Note: Total Worldwide Audience, Age 15+, home and work locations. Source: comScore World Metrix

These social networks are becoming a powerful source of information for millions of individuals. I have chosen to study how people find and access music through social networking site

[2] http://www.stanford.edu/group/siqss/Press_Release/press_detail.html

Myspace.com because of these trends. Social networks are being used increasingly by people of all ages, and music is being accessed increasingly by being downloaded or streamed online.

1.1 MTV, Apple and Other Influential Market Forces

MTV has been seen as a major influence on popular music and popular culture ever since it launched in the early 1980's. It was originally a television station that played music videos from popular artists. Since then, it has evolved into a full programming station with original series geared at a teenage audience. MTV's programming is widely followed by teenagers and therefore capable of influencing what music many teenagers are exposed to, what artists get radio time, and what artists are able to sell large amounts of music. The question still remains as to whether MTV takes artists and makes them popular, or if MTV responds to those artists that are becoming popular through endogenous shocks and makes them "superstars" through promotions. Over the time period January 30, 2009 to April 1, 2009 I have tracked the artists who are listed as "Most Popular", "New Artists", and MTV's own "Artist Picks". Looking at this data along with the artists corresponding trends in album sales and Myspace views will give insight into whether MTV responds to trends or creates them.

Apple, Inc. has become a major player in the music business since it entered the consumer device industry with its iPod device and iTunes platform. Through December 2008, iTunes had sold over 6 billion songs.[3] With such a widely used and distributed platform, iTunes has the ability to influence what music people are aware of and this translates into what music they are paying for. Unlike Myspace, YouTube, or BitTorrent,[4] iTunes has no way to download music for free and therefore iTunes serves as a major source of revenue for many artists. Being promoted by Apple on iTunes home page is quite desirable to many artists and record companies. I have tracked the "iTunes Free Single of the Week" to see what effect being promoted on iTunes has on artist page views. The idea is that people download an artist's free single on iTunes and before buying music, will find more information through the artists Myspace profile. By tracking the growth in Myspace hits, we can get an idea for how much extra interest and sales is driven by Apple.

Finally, dozens of other services such as Pandora.com, which streams free radio based upon your specified musical preferences, affect what music is promoted, played, and purchased. Sites like Playlist.com, where 40 million music fans collaborate to create new playlists for people to enjoy allow music listeners to stumble upon new music related to the music they currently like. And the Facebook application iLike is used by millions of Facebook users to see what songs their

[3] Schiller, Phil. Executive Vice President, Apple Inc. Annual MacWorld Conference, January 5, 2009, San Francisco, Ca.

[4] BitTorrent is a free open source file-sharing application that allows people to transfer media files, including music

friends currently like. These and many other similar services affect what music is listened to, but they act in much more complex, endogenous ways. Thus I have not specifically tracked the promotions or usage statistics for other services. Instead these services such as Pandora, Playlist.com or iLike may result in an artist seeing an increase in traffic, but we expect that increase to be slow. A resulting peak should be endogenous and not attributable to one particular ad, service, or person.

1.2 MYSPACE.COM

Launched in August 2003, Myspace.com has quickly grown to be one of the largest social networking sites in the world. With 100 million unique monthly visitors, Myspace is a site that reaches across the world (Arrington).[5] Furthermore, users are deeply engaged with the site, spending and incredible 266.3 minutes on the site on average per month (Kincaid).

Any individual, band or group can create his or her own Myspace profile. A Myspace artist profile contains information about the artist, free streaming of up to 6 songs, and a list of the artist's friends. See Figure 1.2 for a screenshot of a typical artist profile. Fans can leave comments on the profile, send messages, and listen to music for free. It is the free source of music that has made Myspace artist profiles so widely adopted by fans. In fact, there is no major music artist who does not have a Myspace profile. This property makes Myspace pages a great source for comparing artist popularity, trends over time, and reactions to external events.

Myspace provides information about the total number of profile views each artist has at the current moment. This number is listed on every profile and updated with every new hit. This is therefore easy to track with a script over time. Table 1.3 summarizes the advantages of using Myspace profile views to track interest in music artists.

For almost all independent artists, a Myspace profile is the only online medium for which artists share information with their fans. For almost all major label artists, a Myspace profile is supplemented by a customized website, but the Myspace profile will often have more traffic, more friends, and more comments. For example, multi platinum artist Kelly Clarkson receives about 15,000 visits per month[6] on her website www.kellyofficial.com, and receives around 30,000 hits on her Myspace profile per day.[7] The bottom line is that Myspace artist profiles play a key role to the artist and fan because it is a one stop shop for fans to connect with all their

[5] Traffic measures vary across sources. Quantcast, for example, measures monthly visitors at 70 million. Comscore, is used in this paper as it is a more widely quoted, accepted and used source of internet traffic statistics.
[6] Source: Quantcast (www.quantcast.com)
[7] January 13, 2009. See charts in Appendix.

favorite musicians. By tracking the views over time we can learn about viewer behavior, viral content spread, and response to shocks such as album releases, performances, and promotions.

Figure 1.1 Example of Myspace Artist Profile

Note: Basic information includes profile views, and the streaming music section includes total plays for each song.

Table 1.3 Advantages of using Myspace Profile Views to Gauge artist interest

Reason	Description
Free	Profile views can be accessed by anyone for all artists
Up to date	Profile views are updated with every new hit
Widely Used	Almost every single artist (major label or independent) has a Myspace profile
Real Time	Exogenous shocks occur in real time. For example, when news broke that Rihanna had been abused by her boyfriend Chris Brown (both platinum selling artists), profile views for both artists rose 15x within the half hour.
Protected	Myspace has robots that police the site for foreign scripts and JavaScript, except for Google Analytics. This allows a python script to access it, but not foreign malicious bots.
Accurate	In tests on a secretly created profile, it registers exactly 1 hit per URL load.

2

LITERATURE REVIEW

This section will discuss academic research on Power Law distributions, scale free networks, human dynamics, and endogenous and exogenous shocks. It will give the theoretical basis for my analysis of Myspace profile views and album sales. I will also discuss the details of two papers published by researchers on classifying endogenous and exogenous shocks in Amazon book sales and YouTube video views over time. Both of these papers examine data from a large social network in which people consume and share media. Therefore they provide a good reference point for analysis of Myspace profile views.

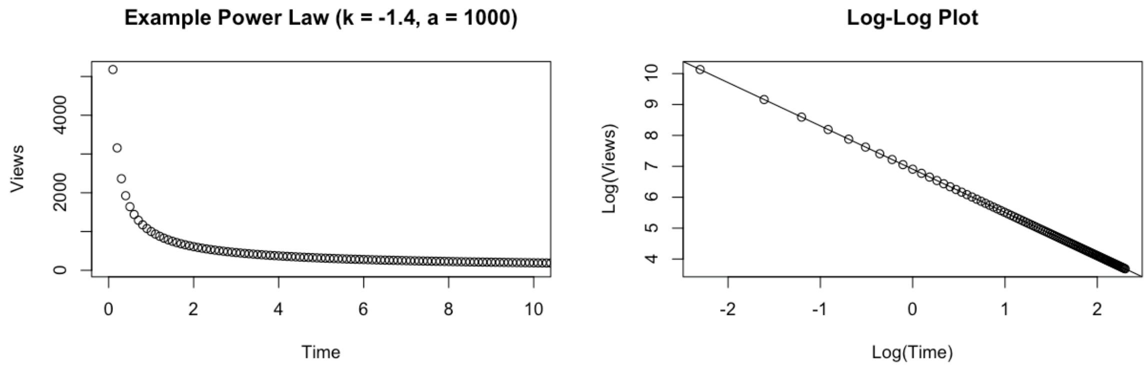
2.1 POWER LAW DISTRIBUTIONS

Power Law distributions describe a mathematical relationship between two quantities in which one variable is the frequency. For example, the size of an earthquake and the frequency with which it occurs behave according to a Power Law distribution. In 1999 Barabasi showed that the World Wide Web follows a Power Law distribution because there are a few highly connected nodes that bridge the network (Barabasi, "Emergence of Scaling"). Barabasi's discovery overturned the previously held thought that the Internet was a randomly connected network and it sparked intense academic interest in finding more Power Law distributions in the natural and manmade worlds. In a Power Law, the fraction of occurrences of some quantity, *x*, is given by eq. 2.1, and Sornette will transform this equation by taking $k = -(1 + \mu)$ (eq. 2.2) to model sales after a peak as a function of time. An example plot is shown below in Figure 2.1.

$$(2.1)\quad f(x) = a\, x^k \qquad\qquad (2.2)\quad P(V) = 1/V^{1+\mu}$$

The most important property of a Power Law distribution is its scale invariance. If we create a new distribution in which $V = 2V$, $P(V)$ only changes by the constant factor 2^k. This property means that if we take the logarithms of both the variables, a linear relationship should emerge with slope = k, no matter the absolute quantity of the variable. By determining the slope empirically, we get the unique parameter for this distribution called its *signature.*

Figure 2.1 Example of Power Law Distribution

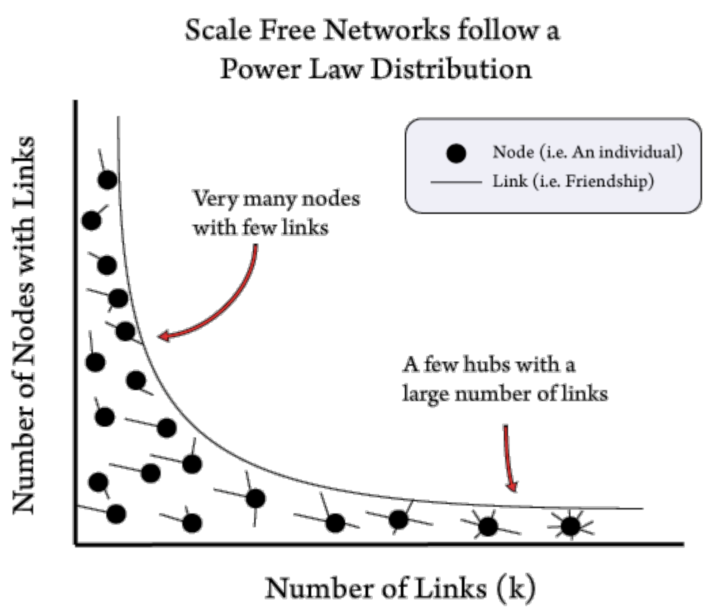


ABOVE: The slope of the line of best fit for the right plot is = k = -1.4.

2.2 SCALE FREE NETWORKS

A scale free network is one whose degree distribution follows a Power Law. The degree of a node is the number of adjacent links. For example, a social network, such as the Myspace community would be scale free if the number of friends of all the users followed a Power Law. Mathematically, this means that the fraction $P(k)$ of individuals in the network with connections to k individuals is given by $P(k) \sim A/k^{\gamma}$ with A as a constant. This formula implies for $\gamma < 2$, the average degree diverges, and for $\gamma < 3$, the standard deviation diverges. Empirical data has shown γ is often in the range [2,3] (Matlis). This means these networks lack a characteristic degree and this is where they derive the name "Scale Free" (Cesar).

Figure 2.2 Property of Scale Free Networks



Occurrences of heavily connected individuals are relatively common (they have a "heavy tail" like the Power Law). These heavily connected nodes (i.e. a very popular person or a website that links to many others, etc) are called "hubs". By definition these networks are not random or uniformly distributed in nodes and edges. Scale free networks have become increasingly important as the World Wide Web, the power grid of the Western United States, and citation networks have been found to exhibit such properties (Matlis).

Scale free networks play an important role in human dynamics because of the implications to how information can be spread throughout the network. It can be shown that as links between nodes are randomly broken in a randomly connected network the ability to transmit information through the network is severely hindered. In a scale free network, however, there is almost no change. The down side is these networks are very vulnerable to targeted hub attacks. Consider the scheduling of plane flights. A targeted attack on hubs such as Chicago, Atlanta, and Dallas would shut down many routes (or at least make it much more costly). The hubs therefore play a key role in keeping the network functioning.

A real world application is fighting the spread of Sexually Transmitted Infections. In this case, if people have sexual partners that are connected through a scale free network, the only way to stop the spread of infection is to target the hub people. Breaking links between less connected individuals will not hinder the diseases ability to spread (Matlis). Thus determining whether or not a network is scale free has important implications on its behavior.

2.3 Human Dynamics Models

In order to predict how information is spread throughout a network, we need a model for how individuals are connected and for the waiting time between cause and action for an individual. If we assume that individual actions are randomly distributed in time (as many models of human dynamics do), it is natural to model behaviors by a Poisson process (Vasquez). A Poisson process is a stochastic process in which events occur independently of each other and continuously. The probability of seeing a number of events (such as number of web site hits) in a given time interval [0,t] is given by Eq. 1, and follows an exponential distribution, with cumulative density function given in Eq. 2:

$$P\left[N(t)=k\right]=\frac{e^{-\lambda t}(\lambda t)^k}{k!} \tag{2.3}$$

$$F(x,\lambda)=1-e^{-\lambda x}, x\geq 0 \tag{2.4}$$

where λ is a parameter that represents the expected amount of events per unit time. From equation 2.4, we can calculate that the mean waiting time is $1/\lambda$. This can be thought of as the expected time for a new page hit. Finally, human dynamics is a single server queuing process in which each individual can serve one task at a time and then move on to the next. In web page context this implies that each individual visits one page at a time.

Research has shown that the best model for a social network is that in which the waiting times between events (communications, sales, views, etc) is modeled with a Power Law and a heavy tail, rather than as exponential through a Poisson process (Barabasi). This result is an important assumption Sornette makes about social networks in building his own mathematical model. Barabasi does address other models for task accomplishment and I list them here along with the resulting distribution of waiting times.

Table 2.1 Models for Task Accomplishment proposed by Barabasi

Model	Description	Resulting Distribution of Waiting Time Until Visitation
First-in-First-out	Artists visited in order they are learned about	**Exponential.** Waiting time until visiting a given artist has exponential tail, meaning artists of similar fame experience same visitation
Random Order	Artists are viewed in random order	**Exponential.** Distribution is also exponential with visitation proportional to fame
Highest Priority Item	Artists are viewed according to personal priority list	**Power Law with exponent = 1**. Some artists will continually be pushed to the bottom resulting in diminished views for some artists and bursts for others

In a paper published along with Vasquez, Barabasi finds that the distribution of response times for email messages follows a Power Law with exponent = 1 (Vasquez, 11), in agreement with his Highest Priority Item model. Sornette also finds that a Power Law fits the acceleration and decay patterns around shocks. This gives strong evidence that assuming the same distribution in Myspace profile views is reasonable.

By analyzing each peak's growth and decay patterns I can determine whether or not the process driving how humans buy books, view YouTube videos, send emails, and accomplish other tasks is the same as how users view Myspace pages and find new music.

2.4 CLASSIFYING ENDOGENOUS AND EXOGENOUS SHOCKS

Many researchers from economics, physics, geology, and finance backgrounds have studied the affects of exogenous and endogenous shocks on complex social systems, networks, and markets. Often these shocks are studied because they are the origins of crises and the ability to predict, plan for, and control such shocks may provide great benefit to the system as a whole. Table 2.1 lists a number of examples.

Didier Sornette, Chair of Entrepreneurial Risks at the Swiss Federal Institute of Technology, has performed much work on socio-economic systems. Two of his recently published papers focus on data sets generated from Internet scraping, are related to social networks, and show similar characteristics to the ways in which people discover music.

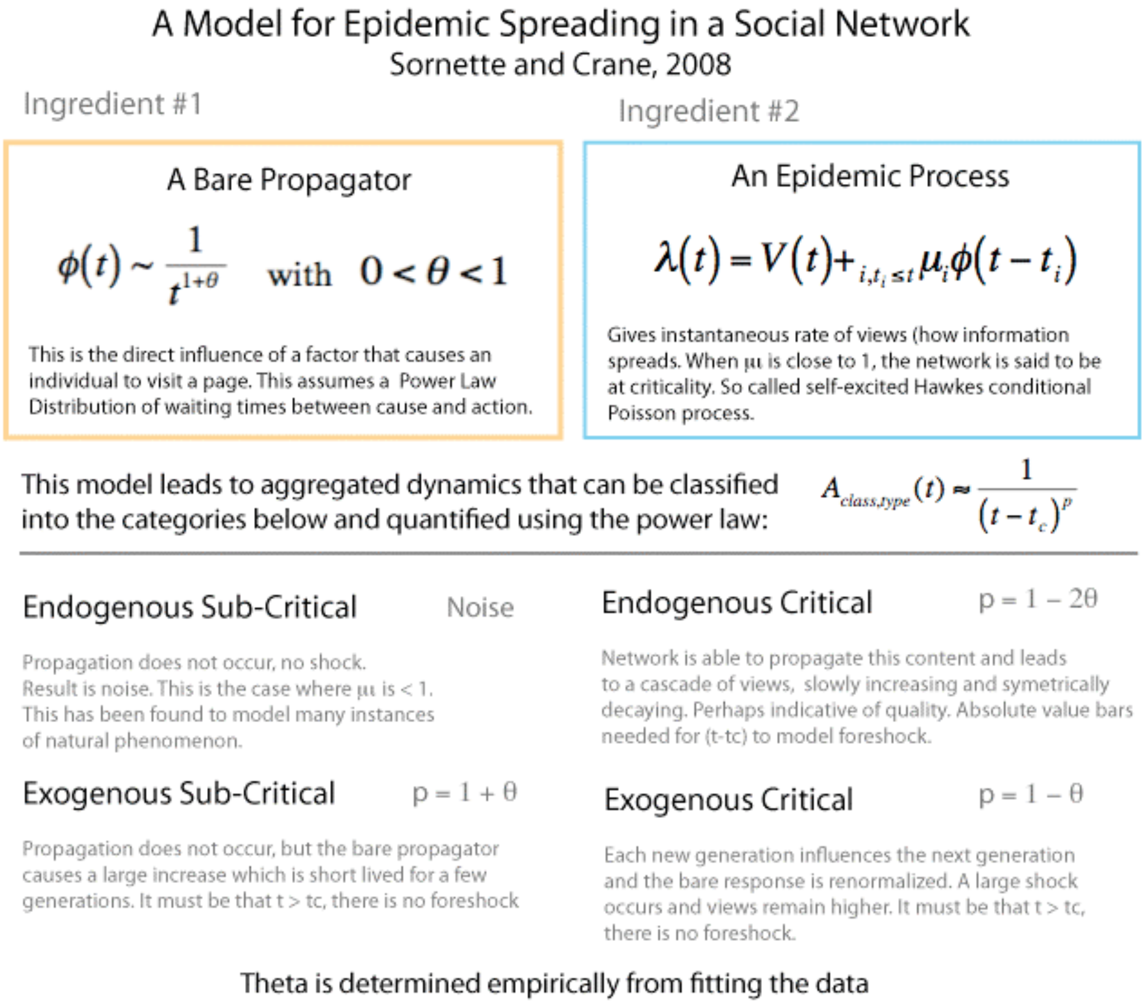
Table 2.2 Real World Examples of Exogenous and Endogenous Shocks (Sornette)

System	Example	Endogenous	Exogenous
Failure of an Engineering Structure	Minnesota Bridge Collapse (2008)	Weight of bridge strained support	Wind or heavy load of cars strained support
Crash in stock market	October 1987	Result of many small sales	Reaction to bad news such as declining GDP
Traffic Gridlock	Los Angeles	People going to work	Dodger's game causes backup
Internet Traffic Spike	Myspace Page views	Artist spreads through word of mouth	Grammy performance causes rush of traffic
Industry Recession	Aviation in 2001	Structural problems	September 11, 2001

The theoretical definitions of endogenous and exogenous shocks are clear and well defined. But how to quantitatively classify endogenous and exogenous shocks based upon data, however, is not. In most cases, we can attribute a large rise in views to a single event such as an album release and call it exogenous. But in other cases we see a peak in profile views, but it is not clear if this is the result of many small shocks accumulating within the Myspace community, or if it is the result of some unseen exogenous shock. Sornette and Crane propose theoretical predictions for what mathematical form these shocks would have based upon two ingredients: a bare propagator and an epidemic spreading process. The resulting classifications are given in Figure 2.3. The key parameter of the model, θ, is what is determined empirically from the data.

The term **criticality** describes how connected a network is. It represents the ability of one user's actions to cascade throughout the network. A network *at criticality* refers to one in which a shock will have a longer effect, penetrate the network more deeply, and result in a slower relaxation following the peak. Mathematically, it means that μ_i in the epidemic equation is close to 1. A *sub-critical* network is one in which shocks are short lived and relaxation occurs very quickly.

Figure 2.3 Model from Sornette and Crane



An **endogenous sub-critical** shock does not form a peak. It is the result of someone in the network performing an action that does not propagate. The result is essentially randomness, similar to the summation of multiple Poisson processes. An **endogenous critical** shock forms a symmetrical peak, with views accelerating up to the peak and then decaying, but in a slower manner than exogenous decay. An **exogenous sub-critical** shock forms an abrupt peak and quickly relaxes back to its former condition. Many users respond to the large external stimuli, and then quickly move on as their actions do not cascade throughout the network. An **exogenous critical** shock shows an abrupt peak that relaxes more slowly as the actions taken in response to the external stimuli propagate throughout the network and other individuals follow the actions of others.

Peak Fraction. Given the equations from Sornette, we see that each type of peak will exhibit a different fraction of total views in the peak. This provides a second mechanism for classifying shocks. The method is as follows: (1) determine peaks as local maximums, (2) calculate fraction of data on peak day, (3) classify according to Figure 2.3. In Chapter 4 I perform this analysis for the complete set of Myspace profile data for all 400 artists.

2.5 APPLICATION: AMAZON BOOK SALES

Fabrice Deschatres and Didier Sornette, looked at the Amazon book sale rankings for thousands of books from Junglescan[8] during 2007-2008. They converted these rankings into sales through an algorithm (Deschatres, 3). Using a model similar to that described in Figure 2.1, Sornette attempts to find and classify peaks according to the acceleration pattern of sales and to the decay pattern.

Finding peaks. Deschatres and Sornette qualify a peak as a local maximum over a 3 month time window which is at least k = 2.5 times larger than the average of the time series over the same time window, and they then fit the sales dynamics by a Power Law:

$$(2.5) \quad S(t) \sim \frac{A}{(t - t_c)^p}$$

Where t_c represents the critical time, taken to be very close to the peak. They then look at a time interval of 1 week around the critical time to perform the fit (Deschatres, 14). Notice that Sornette puts the exponent in the denominator, so his results will find a positive value for the exponent.

Results. Deschatres and Sornette find 1,013 peaks out of 14,000 books, and find that the exponents calculated from the Power Law fit fall into two clusters. One cluster corresponds to the endogenous equations with p = (1- 2θ) = 0.4, and one to the exogenous equation with p = (1-θ) = 0.7. Here 0.4 and 0.7 were determined empirically from the data. Solving for θ gives **θ = 0.3** in both cases.

Conclusions. Deschatres concludes that sales shocks can be classified into two categories based on acceleration pattern of sales or based on the exponent of relaxation, although they do admit that there may be a "continuum between" the two types. They also claim that social networks have converged close to criticality (μ = 1) based upon the empirical data. This implies that one user's actions inspire other actions endlessly throughout the network. Finally, they state "studies of critical phenomena show that very different systems can exhibit fundamental similarities… generally referred [to] as universality" (Deschatres, 25). An important part of my analysis will

[8] Junglescan maintains a MySQL database of Amazon product rankings scraped directly from the Amazon.com site

be determining whether or not this universality holds true in the network of Myspace music listeners. That is, do critical points also exhibit similar acceleration and decay patterns? Is the Myspace network close to criticality as well?

I will adopt the following methods of analysis from Deschatres and Sornette and apply it to peaks found in Myspace profile views:

- Determine peaks based upon a parameter, k, representing multiple of peak value over average value of time series. I will determine the optimal k empirically
- Assume sales dynamics behave according to a Power Law around peaks
- Fit data dynamically to the time frame around the peak that gives the best R-squared
- Attempt to classify peaks into two categories: exogenous and endogenous

2.6 APPLICATION: YOUTUBE VIRAL VIDEO SPREAD

To take our analysis of peaks one step further, we could strive to make qualitative statements about the underlying objects based upon the shape of the peak. That is, rather than simply classifying shocks as exogenous or endogenous we should use these patterns to explain how the community views this product, video, or artist. We can answer questions such as, is this product likely to remain popular, and for how long? How deeply engaged with this product is the community? If information about content was embedded in the dynamics, that would make studying the dynamics even more valuable.

Richard Crane and Didier Sornette give such qualitative labels with YouTube videos by tracking the views of 5 million YouTube videos over time. By studying the patterns that emerge, they classify the videos as "viral", "quality", or "junk". First, they perform similar analysis as Deschatres and Sornette performed on Amazon book sales over time. They find similar results: peaks can be classified as the result of two types of shocks, however, they find a third cluster of peaks whose decay parameters correspond to exogenous sub-critical shocks. This class was not found in Amazon book sales, indicating parts of the YouTube network are not at criticality.

They then discuss a prediction of the model called *peak fraction*. The peak fraction, F, is the fraction of views observed on the peak day relative to total cumulative views. The three different types of peaks each have a different F resulting from how quickly views accelerated and how quickly they decayed afterwards. Table 2.3 summarizes the ranges of peak fractions Crane and Sornette used for YouTube video classification.

Table 2.3 Qualitative Labels and Peak Fraction

Qualitative Label	Type of Shock	Peak Fraction	Description of Peak
Viral	Endogenous Critical	0% < F < 20%	Precursory growth through word of mouth leading to peak and symmetric relaxation as content spreads throughout network
Quality	Exogenous Critical	20% < F < 80%	A burst of activity which causes a cascade of views throughout the network
Junk	Exogenous Sub-critical	80% < F < 100%	A burst of activity that does not spread or have a lasting impact

Source: Crane, 3-4

Viral videos are of high quality because the community's behavior indicates it cares about the content and each user spends time to share the content with others, resulting in the cascade of views and slow relaxation. In fact, this group of videos could be considered the highest quality or long lasting of all categories. "Quality" videos obtain this label because after a large shock (which in YouTube's case is often because the video was placed on the home page) views decay more slowly, indicating that the community continued to view and share this video beyond the immediate attention from the shock. Presumably it was the quality of the video that led to such a slow relaxation. "Junk" videos get their label from the steep relaxation that occurs after an exogenous shock. After a spike in views due to promotion, chance, or other external factors, the community quickly forgets about the video, does not share it virally, and moves on.

To summarize, Sornette et al take a simple epidemic model based upon a propagator and a branching process to derive Power Law equations describing sales or views dynamics over time. They analyze real world data in Amazon book sales and YouTube video views and find that the patterns that emerge fit the model well with a parameter of $\theta = 0.3$.

3

DATA OVERVIEW

This paper utilizes three data sources, which were generated specifically for this paper. Every artist studied was selected for a specific reason, Myspace profile views and MTV promotions were tracked for these select artists, and album sales data was gathered to complement this particular group of artists as well. In this respect, the data is highly customized and makes the most out of just a 3-month time frame.

3.1 ARTIST SELECTION PROCESS

Appendix A2 lists all 400 artists studied. Each artist was selected for a specific reason. Ninety-one of the artists had albums being released between January 6, 2009 and April 1, 2009. I expect artists with upcoming album releases to show variation in the amount of interest and page views, in particular we might expect to find a peak somewhere around the time of the album release. Promotional campaigns, single releases, performances, and other exogenous factors make this group of artists particularly prime for studying shocks. I expected endogenous shocks to play an important role over time as word of mouth propagates news of the upcoming album throughout the network. I also expected media attention to result in several sharp exogenous shocks before and after the albums are released.

3.2 MYSPACE PROFILE VIEWS

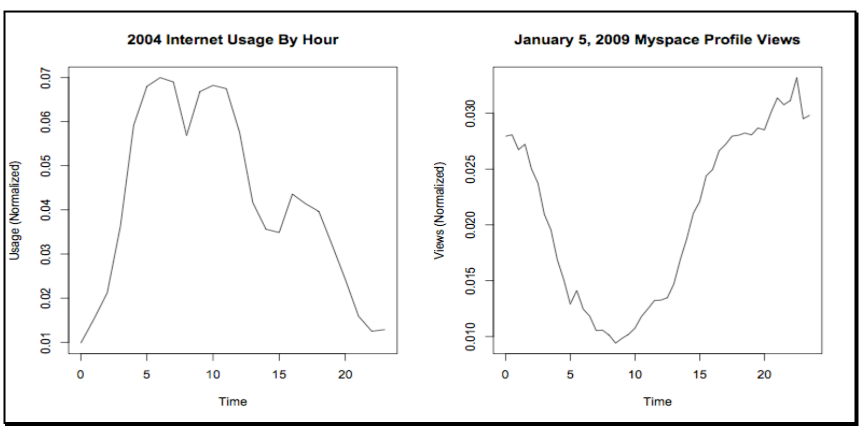
I gathered the profile views for 179 artists at half hour intervals from January 4th, 2009 to April 1, 2009, and at daily intervals for 222 artists from March 1, 2009 to April 1st, 2009. I decided to sample the data at half hour intervals because it allows a day's worth of data to appear smooth. This time frame tells us whether people respond to an event within a 30-minute interval. For all practical purposes, if views spike within a 30-minute interval of a news release, this can be considered as a reaction in real time. Thus we can track behavior as events such as performances unfold in real time. We can answer questions such as are viewers more interested in the artist before the performance, after the performance, or during it? From examining the data, it appears that the clear answer is viewers are most interested during an event and interest falls off quickly afterwards. There will be more on this in Chapter 4 and in the conclusion, Chapter 7.

Conclusions based on empirical observations are only as good as the underlying data. For this reason, I took extra measures to ensure quality in the data and that I knew exactly how the profile views I tracked were recorded. I contacted Myspace engineers to help in this effort. Furthermore, I periodically examined the data set to confirm the correct number of points had been sampled for each artist, and I frequently looked at the data to see if it made intuitive sense.

Alexandra Galasso, an Account Manager in Myspace's New York City office, contacted the Myspace engineering team in New York regarding how the profile views that shows up on each artist's page is generated. She confirmed that Javascript was not allowed on the site and that they have robots that police for foreign scripts which helps ensure quality in the views data.

Examining the data periodically gave me the most confidence in its accuracy. The data was periodic by day, showed responsiveness to important events (such as the Grammy's or Superbowl), and appeared consistent across artists. Most artists reached maximum views rates from 8PM EST to 11PM EST. For a group of American artists and a site that targets young people using it for leisure, this is what we would expect. We expect the most hits to come when people are home from work and school and 8PM EST to 11PM EST captures these hours. In opposition, general Internet traffic peaks during the day as businesses and people use the Internet to email and perform work related functions. These observations are shown in Figure 4.1, and indicate the Myspace user population accesses Myspace during leisure time from home. Note that the plot on the left is from 2004 and for adults (aged 18-54).

Figure 3.1 Myspace Hits For January 5, 2009, EST versus 2004 Adult Daily Usage



Source: Media in Mind, for Adults 18-54 Source: Myspace Profile Data

The python script takes 3 minutes on average to execute. Successively it (1) reads in the webpage www.myspace.com/artistname, (2) finds the string "Profile Views: * #######", (3) parses that string for the profile views, and (4) writes this number into a text file. See Appendix A1.1 for the complete script.

There were some problems in gathering the data. Occasionally, the script could not open the webpage, or could not find the profile views. When I encountered a problem, I had the script write a "0" into the text file. I then went back and wrote over the 0's with the average of the previous time period's views and the next time period's views. It is extremely rare (occurs less than 0.01% of the time) that there are two successive intervals for the same file in which profile views could not be obtained. This method of averaging gives a very good approximate data point and I had no problems after fixing the data in this way.

Caveat. I want to include one important caveat about the data and what can be concluded from it. The profile views that are generated purely reflect the interest of the Myspace community, and not necessarily that of larger society. I expect there would be a high correlation between Myspace and all consumers, because Myspace has hundreds of millions of users, however, they are not the same thing. Thus if Myspace profile views spike 10x their normal level, that does not guarantee that interest in this artist is now 10x greater everywhere. It simply means the Myspace community is giving that artist more attention. To address this issue, Section 4.3 compares volume of Google Searches for various artists with Myspace profile views. The result is that Myspace appears to correlate quite highly with general interest.

Who is the Myspace community? Using data from Quantcast, we get a slice of what the 100 million unique monthly visitors look like. Table 3.1 below summarizes their properties. The data shows that Myspace users are mostly young people. Myspace has almost 2.5x more teen users than the Internet as a whole. This fact is likely why Myspace music pages have become so ubiquitous in the music industry. It is also good for my purposes because it is the same demographic that MTV is trying to reach with their promotions, programming, and content. Furthermore, a larger proportion of music consumption occurs by young people.[9]

Table 3.1 Myspace user demographic

Demographic	Percent	Compared to Internet Average (Index=100)	Demographic	Percent	Compared to Internet Average (Index =100)
Male	42%	85	Other	2%	105
Female	58%	114	No kids	52%	79
Ages 3-11	3%	34	Has kids	48%	138
Ages 12-17	29%	245	Income $0-30k	21%	114
Ages 18-34	45%	156	Income $30-60k	27%	103
Ages 35-49	15%	57	Income $60k-100k	27%	97
Ages 50+	8%	34	Income $100k+	25%	90
Caucasian	69%	86	No college	53%	117

[9] Boorstein, Eric. Senior Thesis.

African America	12%	146	College	38%	92
Asian	4%	81	Grad School	10%	66
Hispanic	14%	206			

Source: Quantcast, as of February 27, 2009.

3.3 MTV Promoted Artists, Featured Artists, and New Artists

I gathered the list of promoted, featured, and new artists listed on MTV.com through a python script and the use of regular expressions. I wrote this data to a CSV file. The purpose here is to see how MTV changes those artists they promote and feature over time. I was curious to see if they moved artists from one list to another (that is do new artists become featured, and then become most popular?).

I had no problems gathering this data, although it ended up being very static. Almost every day the listing was the same. This means the list of artists, unlike the Billboard Hot 100, is done by a group of people choosing artists and listing them manually, rather than by some algorithm that takes into account what people are really listening to, buying, and going to see in concert.

I also tracked which artists were played in MTV promo spots on television. These are brief 10-15 second slots in which MTV shows an artist's music video at the end of a regularly scheduled program. As an artist was played on MTV, I marked it down in my artist file and checked to see how Myspace views responded. I noticed two things. The artists promoted on television were the same as the website (this is reassuring that the data set gathered online actually represents those artists MTV wants to promote), and the promotion did not result in an exogenous shock to profile views for the artist. This means that MTV does not have the sole power to produce exogenous shocks on a daily basis.

3.4 Album Sales Data

With $850 in funding I was able to purchase album sale data for 6 major label artists from Neilsen Soundscan at weekly intervals over the time frame January 4, 2009 – April 1, 2009.[10] I chose artists who had experienced album releases, large exogenous shocks, MTV promotions, or trends over the time period. This varied selection allows us to determine how responsive album sales are to performances and promotions. Given more funding, I would recommend buying all artists album sale data and performing similar regression analysis against each artists profile views.

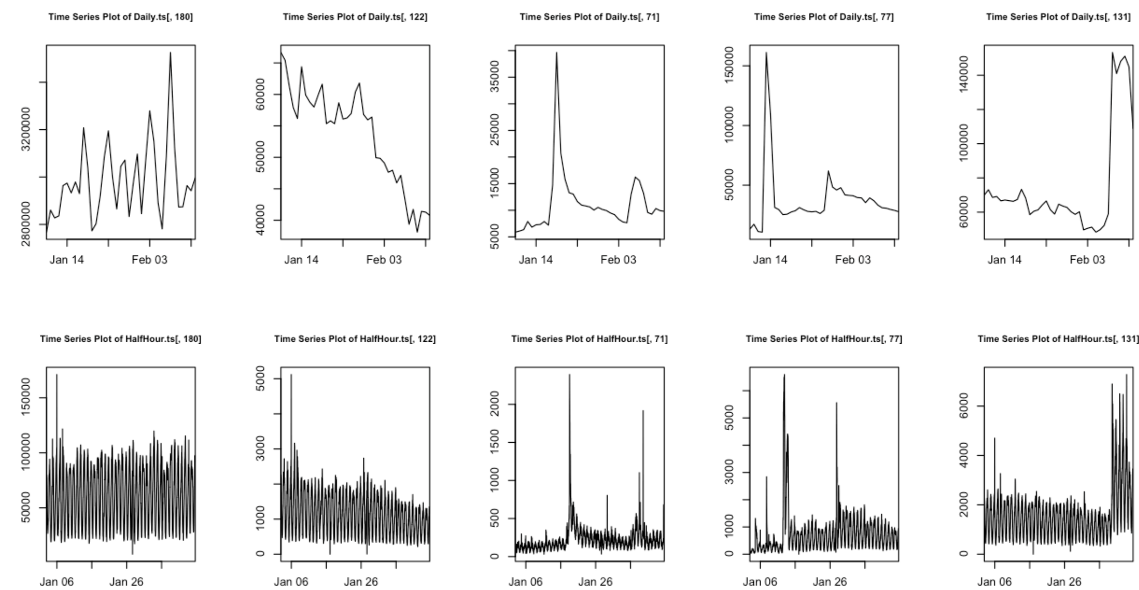
[10] Neilsen Soundscan is an information system that tracks music and music video products throughout North America.

Table 3.2 List of artists whose album sale data was purchased

Artist	Reason
Bruce Springsteen	Album release on 1/27/09, Superbowl Appearance on 2/1/09
Miley Cyrus	Experienced sudden rise in views after release of a new single, off an album released almost a year prior
Lady Gaga	One of the few prolonged upward trends in Myspace views. Also an MTV Most Popular, New Artist, and Artist Pick. Only artist to be on all three lists
U2	Album “New Line on the Horizon” released on 3/3/09. Highly publicized performance at Barack Obama’s inauguration
Carolina Liar	MTV New Artist, iTunes Free Single of the Week on 3/6/09, album released 5/19/2008
Rihanna	Was abused by boyfriend and multi platinum artist Chris Brown on 2/9/09, causing her to miss scheduled Grammy performance and gain large amounts of media attention

4

MYSPACE PROFILE VIEW ANALYSIS



ABOVE: Plots of daily (top) and half hour (bottom) profile views on the same time scale. Left to right: all artists total, Fall Out Boy, U2, Kelly Clarkson, and Rihanna. Large spikes, trends, and relaxation are all quite obvious to the naked eye. This section discusses methods for categorizing, modeling, and interpreting these shocks.

This chapter analyzes Myspace profile views for 400 major label music artists by examining the foreshock and aftershock signatures for peaks in daily views. I compare my empirical results to peaks in Amazon book sales and YouTube video views, and find that Myspace has similar viewer behavior. Therefore, I conclude that Sornette's proposed model of epidemic buyers connected in a network applies to how people find and access music in a social network, and that the Myspace network is close to criticality (meaning user's actions can propagate throughout the network causing a cascade of activity).

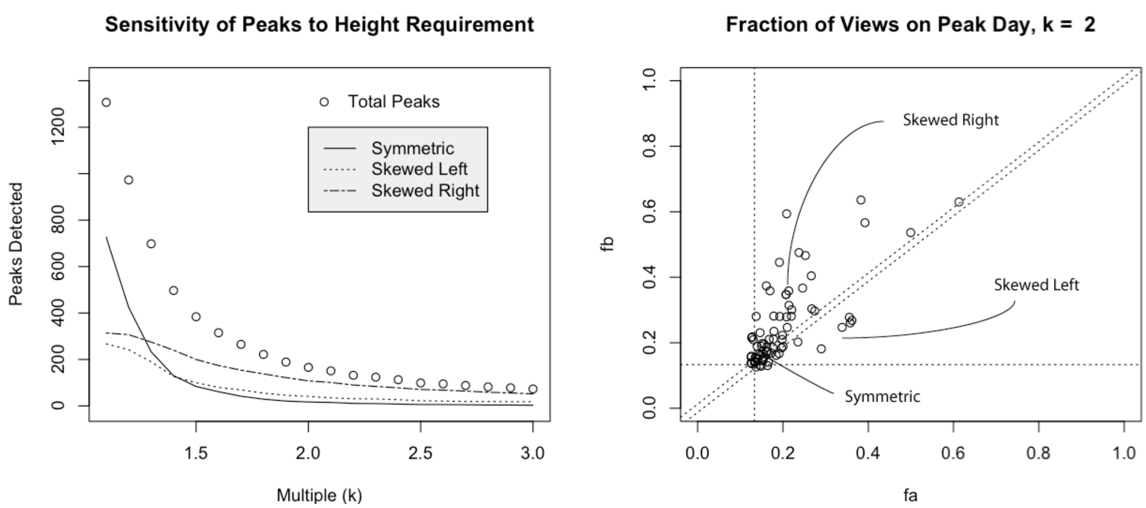
This chapter also compares bursts of activity in Myspace views to the corresponding volume of searches on Google following a large shock. This gives a measure of how closely the Myspace demographic (a younger, slightly female audience) is aligned with general human interest. It also provides insight into what events the Myspace demographic finds more or less interesting than the general public. Finally, I examine qualitatively what type of events cause an artist's profile to receive the most attention and for how long this attention lasts. I find evidence to suggest that television appearances and scandals do the most to increase views. I also find that any shock only has an impact for a week or less. This indicates that news travels very quickly throughout the network, is consumed and then forgotten.

4.1 Finding Peaks: Criteria and Classification

I define a peak as any data point that is a local maximum with 2 weeks of data surrounding it. After testing different time frames, two weeks resulted in the most peaks with high log-log regression R-squared values. Furthermore, for a peak to be kept, there must be no other data point in the two week time frame that is a local maximum that is k = 2.0 times greater than its surrounding data points. This ensures that analyzing our peaks will not be contaminated by other shocks. In sum, two weeks appears to be the timeframe in which Myspace profile views respond according to an organized dynamic. After that noise, other shocks, or a flatlining dominates and no Power Law decay pattern is observed.

A peak has a daily views value that is *k* times greater than the average of the week before or the week after. This ensures the peak is sufficiently big enough. In order to determine the most appropriate value for *k*, I plotted the number of peaks found against values of *k* = {1.1, 1.2, …, 2.9, 3}. From examining the resulting plot (Figure 4.1a), I determined that a parameter of 2.0 is most appropriate for this data set. This is the point at which the number of peaks found begins to flatten out, indicating that we are no longer capturing peaks that are simply the result of noise or random fluctuations.

Figure 4.1 (a) Peak Detection as a Function of k, and (b) Peak Shapes for k=2.0



(a) Number of peaks detected as a function of k along with the shapes of those peaks. (b) Each point is (f_a, f_b) for one peak, calculated as [Views(peak day) / sum(Views on seven days before or after peak)].

Peak Fraction. Given *k* = 2.0 I found 73 peaks (see Figure 4.5 for a sample of these peaks). The first thing we are interested in is exactly how large are these peaks and what are their shapes? We can answer this question by calculating the ratio of views on the peak day to sum of all views for the week before and the week after the peak.

Fraction Views Before = f_b = Views(Peak Day) / Σ (Views Week Before + Peak Day)
Fraction Views After = f_a = Views(Peak Day) / Σ (Views Week After + Peak Day)

We can compare the fraction of views before to the fraction after to determine the peak's skew. A large f_b or f_a means the peak rises high above the level of views before or after the peak respectively. We might expect news events to be unpredictable and spontaneously generate interest. This would cause $f_b > f_a$. This is the essence of a shock and a relaxation. Figure 4.1(b) confirms this prediction as most points lie above the line y=x.

Roughly 6% of the peaks are symmetric (defined by $|f_a - f_b| < 0.025$), 21% are skewed-left and 71% are skewed-right. Although we expected and found significantly more peaks skewed-right, existing models do not predict or explain a peak that is skewed-left. I propose that this is because the music community (and film community) is highly susceptible to "hype". This term describes increasing interest due to anticipation of a future event and it is often the aim of music, film and television marketing campaigns. But once the event occurs, the hype is gone and people lose interest more quickly than it built up. YouTube videos, however, do not have "hype" because there is no event in the future for which the community would be anticipating. To model this dynamic mathematically, I propose that we expect to see an endogenous Power Law growth as hype and marketing efforts spread throughout the network and then an abrupt drop. Although finding peaks skewed-left may be unexpected, upon considering the hype marketing that drives the music industry, this result is not surprising at all.

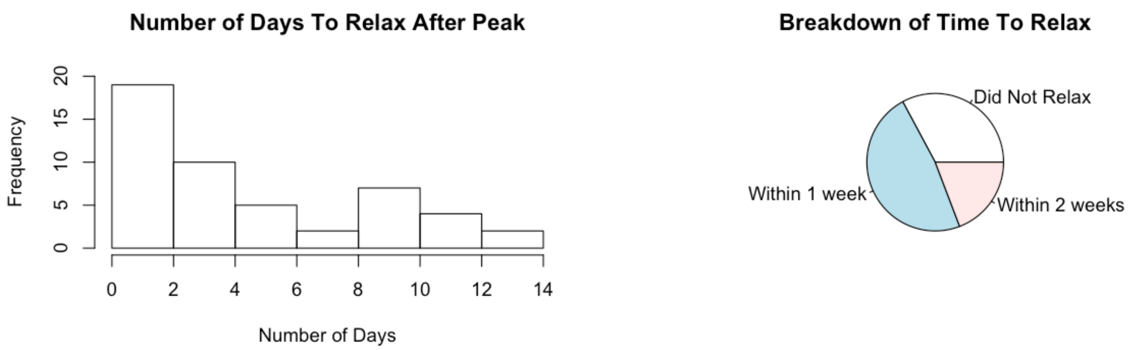
To the largest peaks have values of 0.6. Letting X = views on peak day and Y = average daily views the week before peak, we can solve X / (X + 7Y) = 0.6 to find that X is roughly 10 times greater than Y. This implies our peaks are a rise of two to tens times the previous week's average value. To summarize: most peaks are the result of a large shock and relaxation, few are truly symmetric, and peak sizes range from 2 to 10 times their normal values.

Time to Relax. A simple method to determine the length of each shock is to calculate the number of days until the daily views is back to within 10% of the average daily views two weeks before the peak. This assumes that the views two weeks before are in an average state and have not begun significant acceleration. A glance at the peaks found confirms this is true.

An alternative method would be to calculate when the curve begins to flatline, indicating it is no longer relaxing. However, I believe that if views flatline at a level higher than before the shock, this shock has permanently increased views and should be given credit for this. Thus we consider views relaxed only when it has return to normal levels over the whole period of time.

After performing the calculations in the first method (days until views return to previous average) I find that 24 peaks out of 73 did not relax within 2-weeks after the peak day. Thus 1/3 of shocks could be classified as "long-term". These shocks have an effect on the viewing population that is sustained and significant. Most peaks, however, were very short lived, returning to their previous levels almost immediately. In the left plot in Figure 4.2 we see that it was most common for peaks to relax within 2 days following the peak! That means that many of these shocks were a one-day event and normal levels returned the next day. I propose that either the short attention span of viewers or a very efficient information sharing system must exist to explain this behavior. We recall that a peak is at least 200% larger than the views before it so these are not small shocks by any means, and yet we see a return to previous levels very quickly.

Figure 4.2 Time For Views To Relax Following a Peak



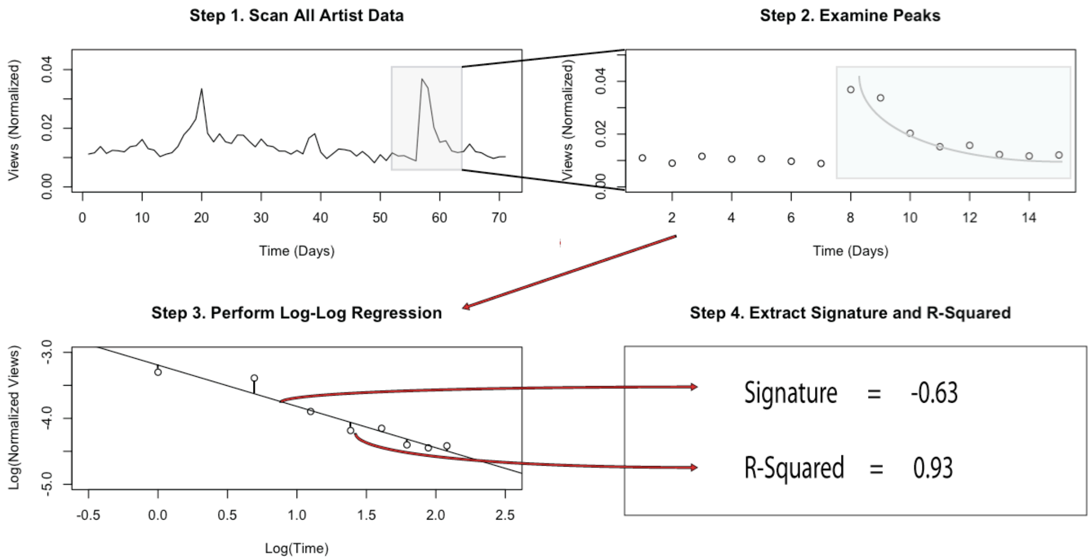
Roughly 33% of peaks did not return to the average level of two weeks before the peak within 2 weeks following the peak (long-term). Roughly 48% of peaks relaxed within the first week (short-term), and roughly 19% of peaks relaxed within the second week following the peak (medium-term).

Power Law fit. Now we have our 73 peaks and understand their shapes. Next we fit the foreshocks and aftershocks of each peak to a Power Law distribution and extract each shock's signature, α, given in the following form:

$$A(t) = \frac{A}{(t - t_c)^{\alpha}} \tag{5}$$

The basic method for finding and analyzing peaks is diagramed in Figure 4.3. The slope of the resulting regression is referred to as the Power Law's signature. The signature is equal to the exponent, α, in equation 5. The log-log signatures found for the foreshocks and aftershocks are plotted in Figure 4.4 (top) along with the R-squared fit values.

Figure 4.3 Method for Analyzing Peak Signatures

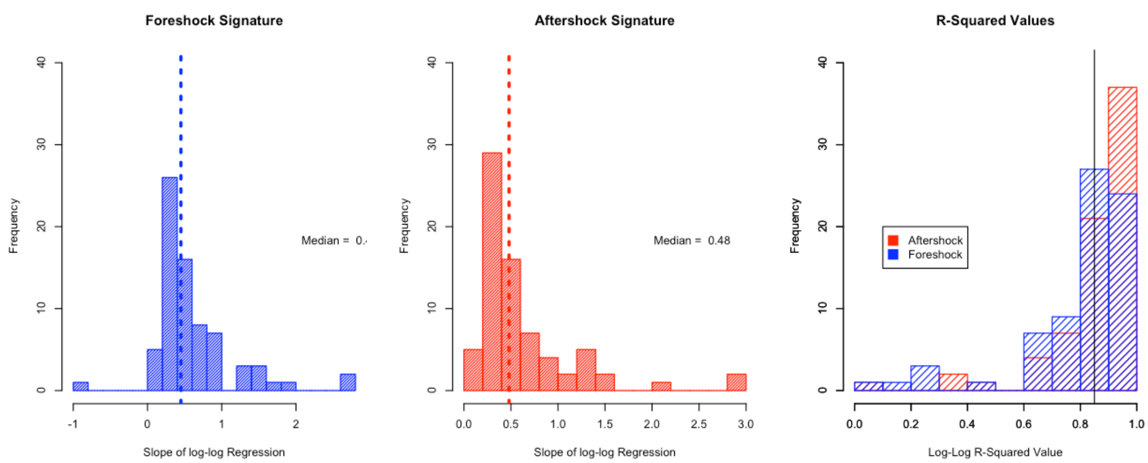


Determining tc. We do not know beforehand what critical time (t_c) and what range to use for the fit. The critical time represents the time before the peak at which the Power Law relationship begins. It is taken to be near the peak, but its exact value is not known. To find it, I tried 50 values of tc = 0.1, 0.2, …, 5, as suggested by Crane for YouTube video views. For each value of t_c, I record the resulting squared deviations from a least squares fit on the logarithm of the data. I then take t_c which minimizes these deviations.

Determining time frame. We do not know over what time frame after the peak the views relax according to the Power Law given above, and this time frame could be different for different peaks. We assume that we need at least 1 week of data to get an accurate fit and so we set minimum days = 7. We let maximum days run up to 14 days. We perform 8 new fits for each of the t_c values (for a total of 8x50 = 400 regressions for each peak). Again, we take the time frame that minimizes the squared deviations. In this way we have a regression that best fits the data through a proper t_c value and through a time frame that captures the Power Law regression and not noise or other trends.

The resulting log-log signatures (Figure 4.4) cluster around a value of 0.35 for both the foreshock and aftershock signatures. A second, smaller cluster appears around 1.35. Importantly, these values do not change if we filter the signatures to only include those that came from regressions with high R-squared values or if we change the value of *k* we used to qualify peak height, or if we change the window we allow our fit to examine. We also recall that Power Laws are scale invariant and therefore the size of an artists views should not matter in determining its signature (so a widely known artist with tens of thousands of page views can be compared directly to an artist with hundreds). This implies that our estimate of clusters at 0.35 and 1.35 are robust.

Figure 4.4 Results from Log-Log Regressions on 73 peaks



The Power Law formula given before (equation 5) is the same as Sornette's equation for an exogenous shock to a critical network, derived from his model of a bare propagator and an epidemic branching process with exponent = 1 - θ. I now determine the key parameter, θ, from the empirically determined exogenous signature. Since we only found 2 clusters,

For exogenous critical:

$$\begin{aligned} 1-\theta &= p1 \\ 1-\theta &= 0.35 \\ \boldsymbol{\theta} &= \mathbf{0.65} \end{aligned}$$

For endogenous critical:

$$\begin{aligned} 1-2\theta &= p2 \\ 1-2\theta &= 0.35 \\ \boldsymbol{\theta} &= \mathbf{0.325} \end{aligned}$$

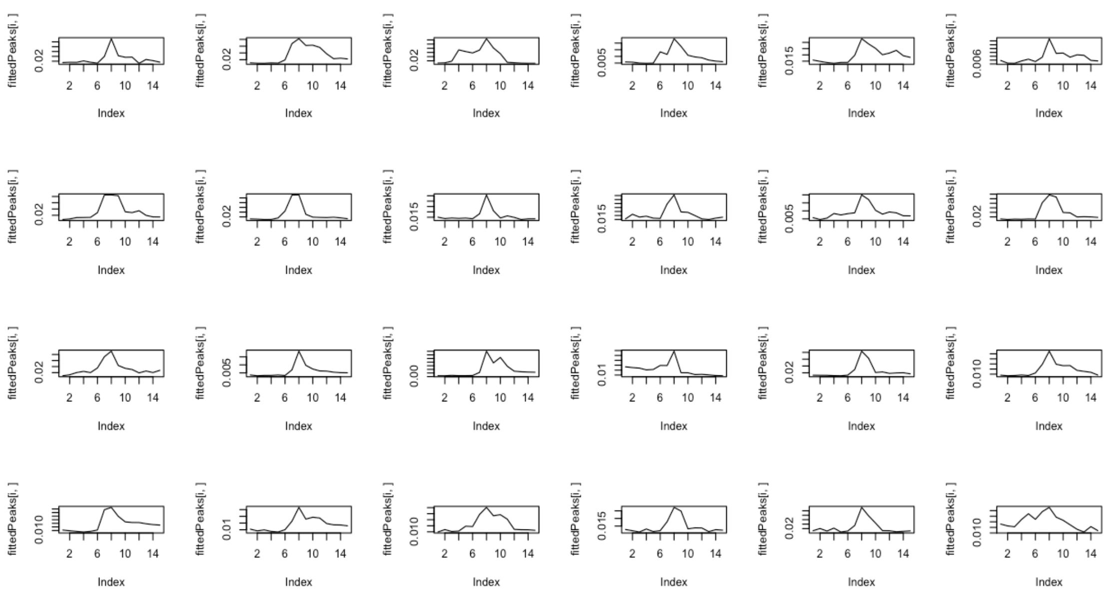
For exogenous sub-critical:

$$\begin{aligned} 1+\theta &= p3 \\ 1+\theta &= 1.35 \\ \boldsymbol{\theta} &= \mathbf{0.35} \end{aligned}$$

Sornette found θ = 0.3 from examining Amazon book sale signatures and YouTube video views. Unfortunately, we have a problem because if most of our peaks (the large cluster) are of the exogenous critical type, the estimate of θ is 0.65 which is not consistent with exogenous sub-critical or with Sornette's findings. If however, the large cluster is endogenous critical, then the estimates of 0.325 and 0.35 match well with each other and with Sornette's findings.

I had expected to find exogenous critical cluster at 0.7, and although some peaks do exist with this signature, a clear cluster does not exist at this value. This implies that either the model is not a fit or that most peaks in Myspace profile views are the result of much shorter-lived endogenous growths and relaxations. One piece of evidence confirming this conclusion is that the cluster exists at 0.35 in both the foreshock and afterschock signatures, which must be the case for endogenous critical because it is a symmetric peak. If this is true, Sornette's model of epidemic buyers connected in a network describes how people find and access music in a social network.

Figure 4.5 24 Out of the 38 Fitted Peak Decays (R-squared > 0.8)



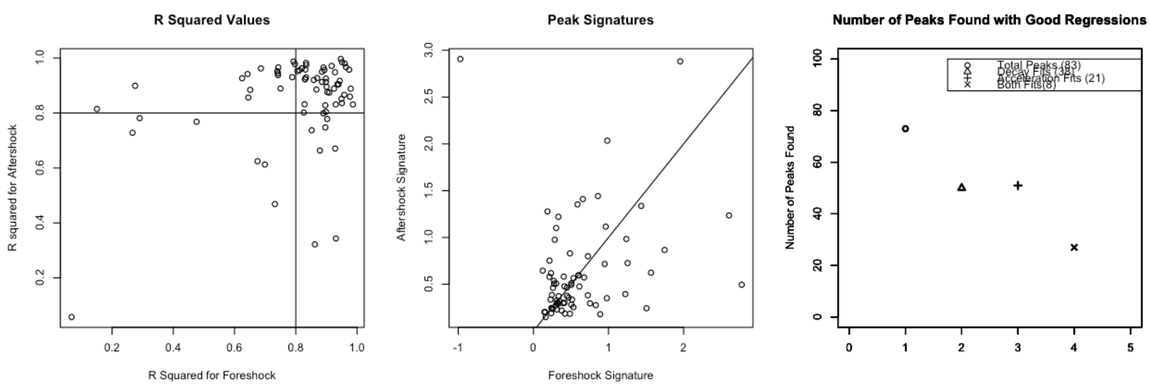
X-axis: 15 day span centered on peak. **Y-axis**: Normalized profile views (Ni = Views on day i / sum[all views]) . The shape, skew, and size of the peaks varies, although most peaks appear quite steep, sudden, and quick to return to previous view levels.

Goodness of Fit. After extracting the key model parameter, θ, I then examine the R-squared values for each of the fits (Figure 4.4, right plot). Overall the fits are very good. This is of course the result of the careful tuning by my algorithm in the initial steps. By making sure the right critical time and time interval we have increased our chances of fitting the Power law. One thing is apparent from this histogram: aftershocks fit a Power Law better than the foreshocks. 50 of 73 aftershocks have an R-squared over 0.8, whereas 40 foreshocks attain such a high fit and the gap is even higher when we look at R-squared over 0.9. This is evidence that the relaxation aftershock obeys a Power Law better than the foreshock. Therefore many of these peaks may be the result of abrupt exogenous shock related to news or media event. Examining a plot of 24 of the fitted peaks confirms their tendency to skew right (Figure 4.5).

Figure 4.6a plots the paired R-squared values for each peak. This chart tells a similar story. For those peaks with a good aftershock fit, they are no more likely to have a foreshock that is a good fit. Phrased another way, the aftershock usually obeys a Power Law, even if the foreshock does not. However, if a foreshock has a good fit it is much more likely to have an aftershock that is also a good fit, indicating there is some cascade effect creating the accelerating views which then relaxes after the peak. This indicates most peaks are abrupt and relax according to a Power Law. And for those peaks that experience a Power Law acceleration, the resulting decay is also according to a Power Law.

Figure 4.5b shows that there is one main cluster of signatures and it centers around (0.4, 0.4). It also shows that the foreshock signatures are slightly more negative, likely the result of forcing a Power Law fit to an abrupt peak, resulting in a larger exponent in absolute value. This cluster corresponds to exogenous critical growths and decays. Finally, Figure 4.5c shows us that of all 83 peaks, 38 of them have a good fit for the aftershock relaxation, 17 for the foreshock acceleration, and only 8 have a good fit for both. All of these plots reinforce that peaks relax according to a Power Law and accelerate abruptly.

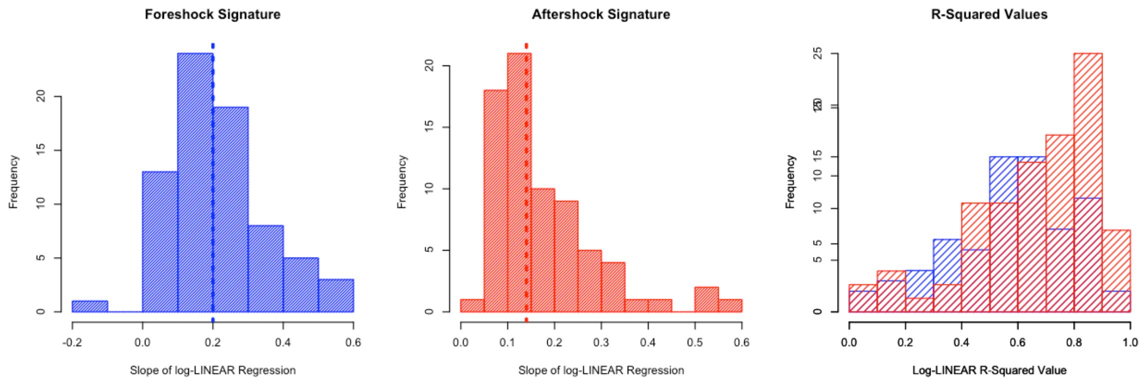
Figure 4.6 (a) R-Squared Values, (b) Signatures, and (c) Quality of Regressions



Log-linear Regression. At first glance, many people might assume that these peaks are exponentially distributed. So for completeness, we assume that there exists a linear relationship between time and the log of the views leading up to the peak and following the peak. We do this with the assumption that the model was wrong and a Power Law is not appropriate (although there is much evidence to indicate a Power Law is appropriate). We take the views to relax according to $\mathbf{V(t) = e^{pt}}$, where p is again the signature we are extracting. Performing the analysis in R (see Appendix A1.3) we obtain the histograms in Figure 4.7. The plots have the same axis as the log-log plots above them. The results here are clear: the R-squared values are much lower. With a log-linear regression, 16 peaks have R-squared values greater than 0.8 in the aftershock, and only 2 peaks have R-squared greater 0.8 in foreshock as well this is more than a 50%

reduction in fitted peaks. Thus we can conclude a linear-log relationship does not exist between time and views following a shock and looking further into the decay parameter is not meaningful.

Figure 4.7 Results From Log-Linear Regressions on 73 Peaks

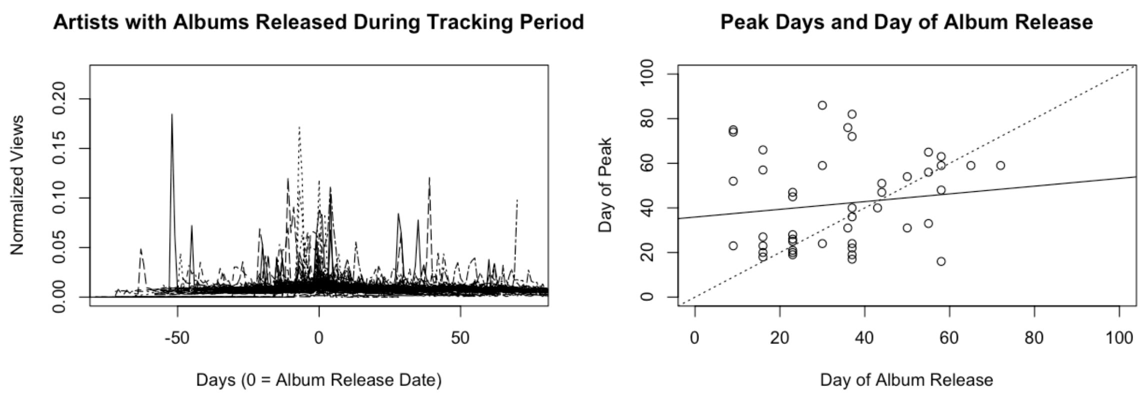


Linear Regression. For completeness, I also performed linear regressions on the peaks and the resulting signatures were widely varied (no clustering) and the R-squared values were very low.

4.2 Effect of Album Releases On Views

We would expect views to be higher around the time of an album release, and the data confirms this. Artists who have albums being released experience a much higher volume of shocks and those shocks are much more pronounced, however, they are not consistently centered on the release date. We see that it is more likely that interest peaks before the release date. This is because artists often post new lyrics, singles, and messages to their fans which drives a high volume of traffic and fans continually check back. It thus appears that the optimal window for marketing efforts is from 10 days before album release to 5 days after the release. More will be said about views and album releases in Chapter 6.

Figure 4.8 Album Releases and Peak Days



Above left: daily-normalized profile views for all 90 artists with albums released during the period January 1, 2009 to April 1, 2009. All plots were re-centered to have the day of the album release on day 0. We see there is most activity 1-2 weeks before the album release, although significant and somewhat sporadic activity exists throughout the whole time frame. Above right: a plot showing the day of album release and the peak days for each artist. For example, a few artists had albums released on day 8, and after analyzing these artists profile views, peaks were found on day 20, 52, and 77. If peak days truly occurred on the day of the album release, we would see a linear correlation on the 45-degree line, shown dashed here. However, the correlation is much weaker than might have been expected.

4.3 GOOGLE SEARCHES AND MYSPACE VIEWS

Google search volume is often used as a proxy for general human interest. By comparing changes in Myspace profiles views to changes in Google searches we can determine if Myspace is a reasonable proxy for general human interest, and whether Myspace represents a "normal" slice of the population or if it is composed of a different demographic. Google Trends gives us the tool to do this comparison (www.google.com/trends).

I look at three shocks to compare Myspace users and the world's response to various news events. The first two are Rihanna and Chris Brown who both received huge media attention when news broke on February 9th, 2009 that Chris Brown had physically abused Rihanna, causing them both to miss their scheduled Grammy appearances. The third example is for Carolina Liar, a band who appeared on the TV show "90210" on February 4th, 2009 and was featured as the iTunes free single of the week on March 6, 2009.

In Figure 4.9 we see the number of searches for the artists Rihanna and for Chris Brown on the bottom left. It shows that as soon as news broke searches spiked to an unprecedented level for both artists. Similarly, their Myspace profile views went from a laggardly downward trend to a huge spike in views. Perhaps some of the views on Myspace were the result of Google searches turning up the links to the profiles (Rihanna's Myspace profile is hit #4 on Googling "Rihanna" and Chris Brown's profile is hit #3). But more likely, this is strong evidence that the Myspace community reacts in the same way as the larger community.

Table 4.1 attempts to answer whether or not the scale of the reaction is the same in general interest (Google searches) and the Myspace community (profile views). It appears that the reaction is smaller from the Myspace community. This is likely because news involving domestic abuse made headlines in major newspapers and television programs. Thus it reached a demographic of older people who may have had no interest or background with Rihanna or Chris Brown beforehand. This led to them turning to Google to find more information about these two artists.

Figure 4.9 Google Searches and Myspace Profile Views

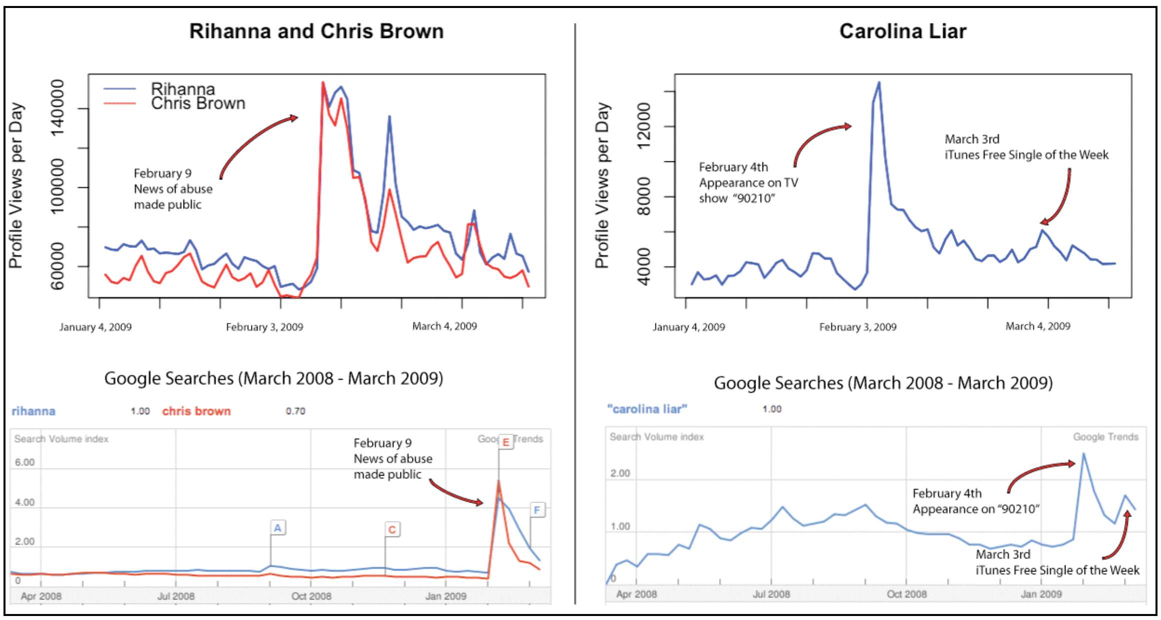


Table 4.1 Google and Myspace Respond to Exogenous Shocks

Data Source	Week Before Shock	Week Following Shock	Increase
	Rihanna		
Google (search index)	.7	4.5	6.4x
Myspace (avg daily views)	51,526	148,382	2.9x
	Chris Brown		
Google	.4	5.35	13.3x
Myspace	50,067	141,768	2.8x
	Carolina Liar		
Google	.86	2.50	2.9x
Myspace	3,404	9,556	2.8x

Google values represent a relative search index adjusted to so that 1.0 is the average value during the interval March 2008 – March 2009. Myspace values represent the average daily profile views. Rihanna and Chris Brown shocks occur on February 9th, 2009 as news broke that he had abused her. Carolina Liar shock occurs on February 4th, 2009 when the band appeared on the TV show "90210".

One of the best case studies for determining the affects that iTunes, exogenous shocks, and word of mouth growth has on sales is from the band Carolina Liar. They released an album called "Coming to Terms" on May 19, 2008. So unlike many of the other major label music artists I have studied, Carolina Liar did not have large attention and promotional campaigns around the time of their album release. Instead they have continued to tour and promote their music through smaller venues and channels. During the period I tracked their album sales and profile views they made an appearance on a popular TV show, "90210" and even had one of their songs featured as the iTunes free single of the week. They were also the only band to be listed on MTV's "New Artist" list for the entire duration of January – April 2009. All of this occurred a full 10 months after the release of their album! Figure 4.9 shows Carolina Liar's profile views and Google Searches.

Clearly the views and searches appear to correlate over the 3-month period of January 2009 – March 2009. Table 4.1 shows that the correlation is almost spot on. After their 90210 appearance views and Google searches for the band increased 2.8x and 2.9x respectively. This is an amazingly close relationship. Unlike Rihanna and Chris Brown, the "90210" appearance was known only to those who watch the show or have friends who do. The viewers who watch "90210" are of a younger, teenage audience, likely similar to the demographic that makes up Myspace. Unfortunately, viewer demographics for "90210" are not available, but we can surmise that this similar demographic is what resulted in the almost exact jump in Google and Myspace.

This analysis confirms that Myspace reacts to news with the same speed and direction as Google searches and the general public; however, the intensity depends on the type of news. News that is widely broadcasted (such as Rihanna and Chris Brown abuse scandal, Bruce Springsteen at the Superbowl, U2 playing at Barack Obama's inauguration concert) shows stronger reactions from the general public. News that is focused to teens appears to correlate with speed, direction and size to the reaction in the Myspace community. Chapter 6 will examine how strongly album sales correlate with Myspace views and Google searches. We will then be able to answer whether it is worthwhile for a record company to get their artists at an event like Barack Obama's inauguration concert, as opposed to a smaller venue like a teen facing television program.

5

MTV Promoted Artists

I tracked MTV's list of "Most Popular", "Artist Picks", and "New Artists" artists from January 10th, 2009 to April 1st, 2009. These lists are displayed on MTV.com and these artists receive promotions in the form of TV short music video segments at the end of regularly scheduled programming. The purpose of tracking these lists was to discover what methods MTV uses to decide which artists they promote and ultimately try to answer the question, does MTV drive popular music or does MTV respond to those artists that are becoming popular on their own? Does MTV have the ability to make someone a superstar?

Artists on these lists gain large exposure from the daily traffic to MTV.com. According to Viacom's 2008 10-K filing, MTV.com reaches 8 million monthly viewers with 67 million video streams. Viacom writes "Our on-air programming drives traffic to our digital properties and vice versa, allowing convergent, or cross-platform, advertising sales" (Viacom, 3). Viacom's *Media Network* segment (including MTV and BET and related properties) received a combined 87 million unique monthly visitors during the year 2008. Exposure to this audience represents a significant marketing opportunity for the artists fortunate enough to make these three lists.

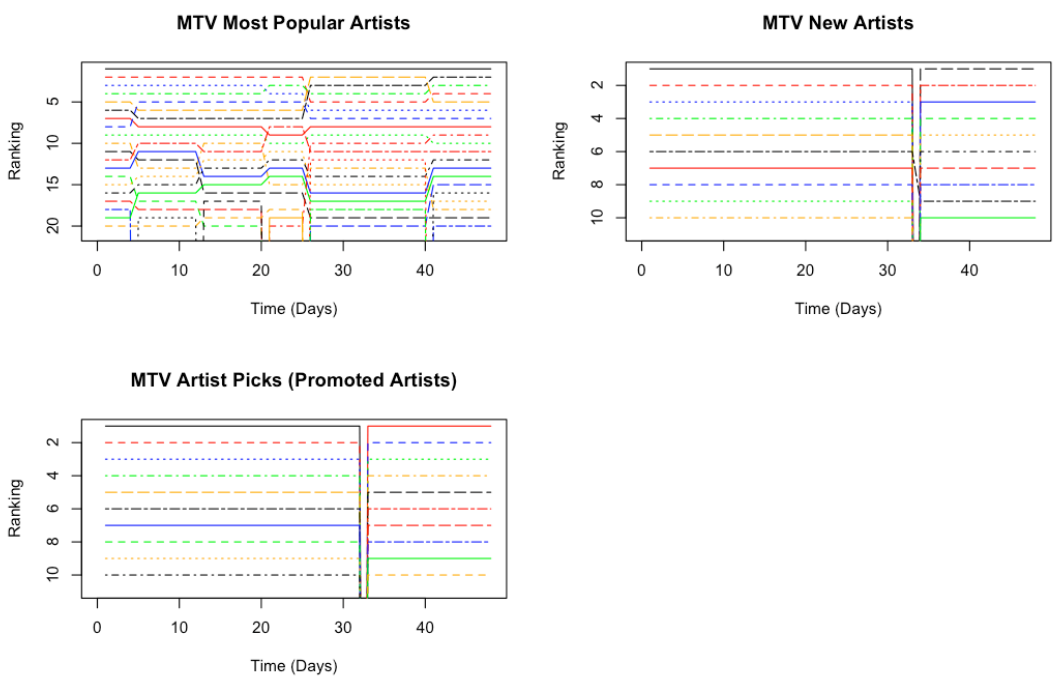
5.1 WHO, WHEN AND WHY

In short, all three lists are incredibly static. Unlike the Billboard lists of top 100 tracks by weekly sales volume, MTV's lists stayed constant for months at a time. The Billboard list top 10 changes, often significantly from week to week. It seems unlikely that a company as large and media focused as MTV would not be updating its list due to laziness or lack of attention. We can assume that by promoting artists for an extended period of time they are more likely to gain traction, increase interest in this artist and be able to sell more content, videos, and music from that artist. But as we will see, being listed and promoted on these does not increase Myspace profile views (which we use as a proxy for popularity) with statistical significance. In fact, in the case of Most Popular artists, there is a negative correlation! Those artists experience lower profile views while listed on the Most Popular artist list than when they are not listed. Thus we conclude that MTV's method of promotion is out of touch and ineffective.

To understand these lists visually we plot each artist's trajectory on the list over time. These graphs are shown in Figure 5.1 Each line in each graph shows the progress of one artist over time.

For example, in "Most Popular" the solid red line that begins at Ranking = 7 represents Lady Gaga. She moved down one spot on day 3, down one more spot on day 20, and back up one spot on day 25. One pattern common to all lists is that MTV changes its lists all at one time. Thus we can be reasonably certain that adjustments to who MTV is promoting at any given time is not the result of algorithms based upon sales, requests, or page views. New Artists and Artist Picks were only changed once in a 2-month period. Clearly MTV is hoping that by listing and promoting these artists for an extended period of time will expose their audience to new music that they might consume, watch videos from, and spend more time on MTV.com and watching MTV programming.

Figure 5.1 MTV's list of "Most Popular", "New Artists", and "Artist Picks"



Most Popular. Throughout the whole 2-month period, Lil' Wayne was listed as the Most Popular artist (solid black line at the top), however during the last weeks he was not even in the top 20 of the Billboard Hot 100.[11] Clearly, this list is not based on sales. It is difficult to know without speaking with MTV management directly whether they also monitor requests or page views to decide who is the "most popular". Other familiar names such as Britney Spears, Rihanna, and Chris Brown made the most popular list as well. Further evidence of non algorithmic decision making is shown by the frequency with which MTV changes the most

[11] The BillBoard Hot 100 is a weekly published list of the top 100 selling singles in the country available at www.billboard.com

popular list. It is not weekly, nor is it regularly spaced. Furthermore changes are made in groups suggesting a team of people may be planning the updates and then rolling them out all at once.

New Artists. MTV updated the New Artist list once in the 2-month period, and the only artist to survive that change was Carolina Liar, the band who appeared on the show 90210 and was featured as iTunes Free Single of the Week. Carolina Liar was the only band on this list before the change to gain success. It appears that MTV will stick with artists who begin to gain success, but will filter out those who have not gained traction. This behavior makes sense and indicates that MTV's strategy to is to spend time promoting a set of new "underground" artists and only keeping those artists who actually become popular.

Artist Picks. This list is composed partly of the same artists as the Most Popular list, but like the New Artist list, it is filtered periodically. This list was changed only once and all artists were replaced. These artists receive heavy promotion on MTV's television programs and appear to be artists that MTV think hold particular promise and interest for its audience. The fact that the list is the biggest name artists, it is filtered infrequently, and changed all at once seems to suggest that MTV may be receiving payment for promoting these artists.

5.2 Correlation with Myspace Views

If MTV has the power to influence people the artists on these lists would experience a boost while listed. As more people saw this artist listed and shared them with friends, an endogenous acceleration would occur and the artists page views would increase over time. In particular, the New Artist and Artist Pick lists should certainly benefit from in terms of Myspace Profile hits and album sales as they benefit from the extra marketing. For "Most Popular" artists, however, we might expect views to be constant. After all, these most popular artists are likely listed because of previous success, promotional campaigns, and high album sales. At best, MTV promotions would keep them at a high-sustained amount of views and sales over time.

Unfortunately, these intuitive predictions are not played out in the data. It appears that profile views over time do experience some long-term trends, but it is largely related to album releases, and not to what MTV is doing to promote them. In order to turn plots into quantitative results, I performed a dependent means T-test on the daily views while artists were listed on each of the lists and while they were not. To be clear: I had two data points for each artist. The first was that artists average daily profile views while listed. The second was average daily views while not listed. Thus any artist who was not added to or removed from the list during the time frame (such as Lil' Wayne on the Most Popular list) was not considered because there was no reference against which to judge.

The results are unexpected. At the significance level $\alpha = 0.05$ none of the tests are significant. So it appears that MTV does not pick or promote artists who are in a significantly more popular phase of their career. This also rules out any possibility that MTV is having a significant impact on artist profile views. This seems to answer the question, "Does MTV drive popular music or does popular music drive MTV?" - MTV does not drive popular music. It reacts to popular music and lists popular artists, but it does not have the ability to make an artist a star.

Table 5.1 T-Test Results

List	N	Avg % Change	P-Value
Most Popular	7	- 2.26%	0.5897
New Artists	9	+10.76%	0.0722
Artist Picks	18	+0.75%	0.4059

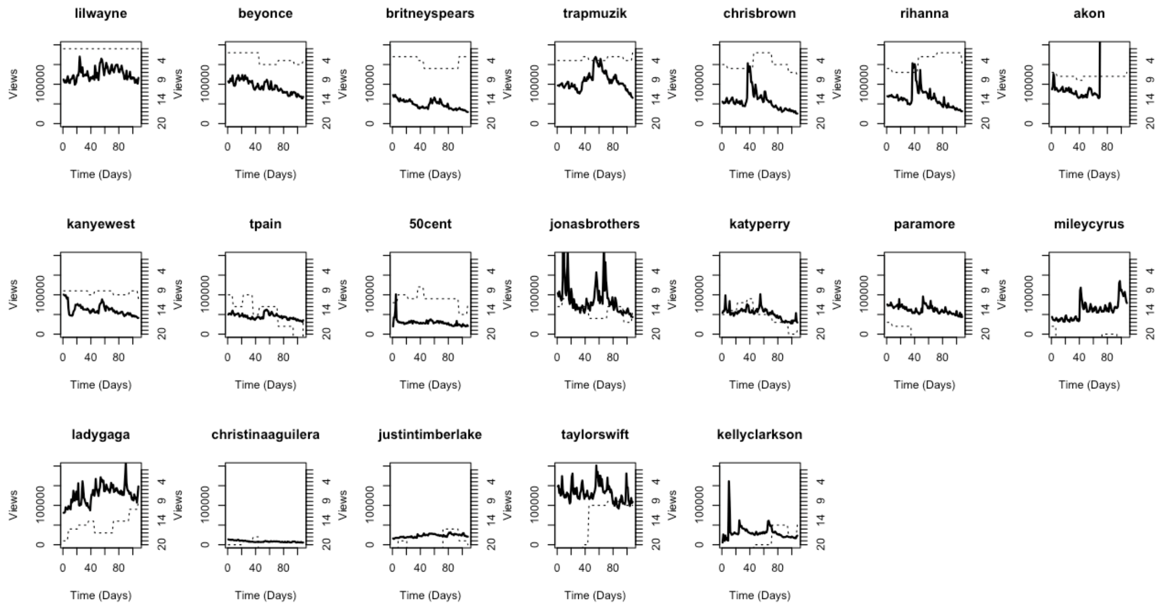
Results of dependent means T-test on average daily profile views while listed and average daily views while not listed for various artists. Not every artist listed could be tested because (a) I did not have the Myspace profile view data, or (b) They were on the list the whole time so there was no reference point.

For the Most Popular list, not only is there not a significant increase in views for the artist while listed but there is actually a negative relationship! This means on average profile views for each artist increase after being removed from the list, or they drop once added! It appears that MTV is either not paying attention to data or is using another method to determine who is listed on the Most Popular artist list. What that alternative method is I leave to the decision of my reader, but I would suggest beginning examination with the relationship between MTV and the four major record labels Universal, Sony BMG, EMI, and Warner. These four labels account for 70% of record sales and represent every artist on MTV's Most Popular list (Lamb).

The New Artists list represents those artist that are up and comers – new to the scene and going perhaps expected to be the big stars one day. For this reason, it is not surprising that we see the highest p-value here. There is a positive correlation between views and being listed, but it is not quite statistically significant. All artists except one received more views while MTV promoted them, but they did not receive much increase (10% on average). Finally, the Artist Pick list appears to have a slight increase in views but nothing worthy of consideration.

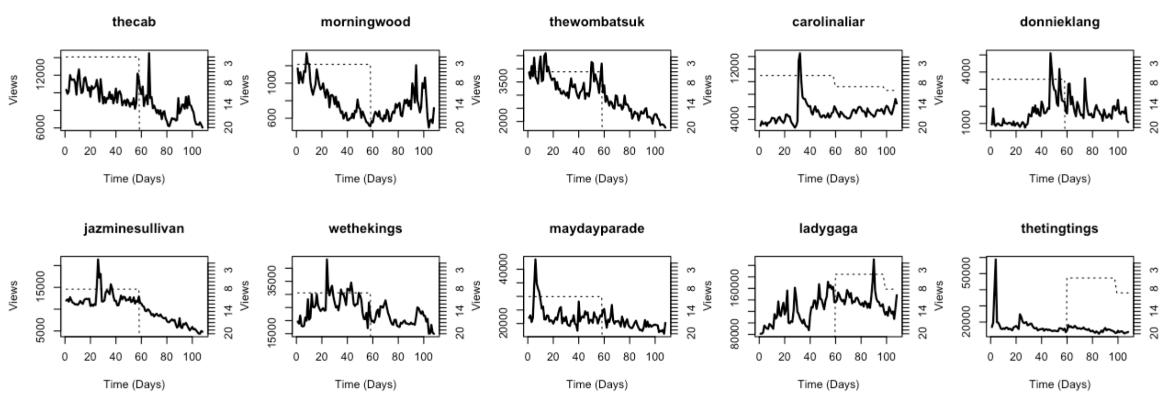
Figures 5.2 - 5.4 show the artist rankings superimposed on artist profile views. This is the data from which the T-tests were performed on. See the notes below each graph for details and discussions of each individually.

Figure 5.2 Profile Views and Ranking on MTV's Most Popular Artist List



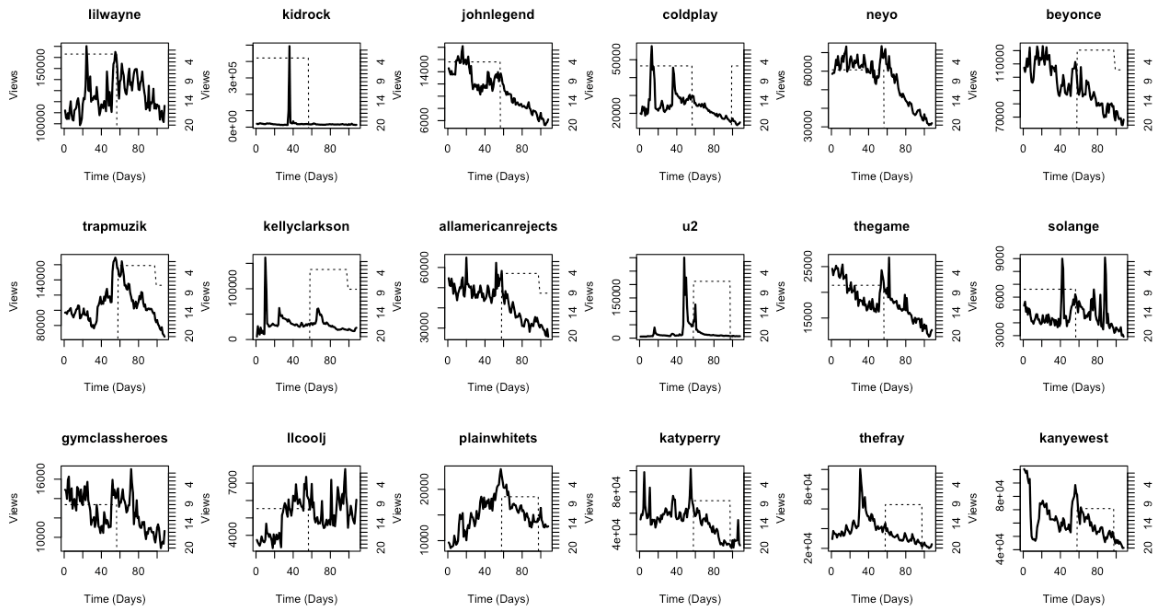
The plots above give artist profile views (dark black line) and ranking on MTV's top 20 Most Popular Artist list (dashed line). We would generally expect a positive correlation. As profile views increase or decrease we would expect the placement on the Most Popular list to increase of decrease as well. However, as shown Table 5.1, not only does a positive relationship not exist but the average views while not on the list are greater than the average views while on the list.

Figure 5.3 Profile Views and Placement on MTV's New Artist List



The plots above show artist profile views (dark black line) and listing on MTV's New Artist List. We might expect profile views to be larger while listed than while not. Being on this list exposes the artists to MTV.com's daily traffic (Alexa.com rank of 417 worldwide) as well as on air promotions including short video segments at the end of MTV regular TV programming. As seen in Table 5.1, a positive relationship does exist but it is not significant at the 0.05 level (but would be significant at the 0.10 level). We must note that this positive correlation does not imply that MTV promotions are causing the increase, although it is certainly a likely candidate.

Figure 5.4 Profile Views and Placement on MTV's Artist Pick List



The plots above show artist profile views (dark black lines) and listing on MTV's Artist Pick list. One noticeable trait for these artists is that it appears they are listed shortly after a large spike or increase in views and while the artist is listed views seem to linearly decrease. Thus MTV's artist picks are nothing more than what people liked a few weeks before and MTV's decision to pick them has no ability to sustain that popularity. In many cases (Kelly Clarkson, U2, The Fray) the large spike was due to a new album release. It is surprising MTV did not pick these artists in anticipation of or at the time of the album release. They are consistently behind.

5.3 MTV'S ADVERTISING REVENUE

Given that none of the T-tests were significant and no noticeable pattern exists between the Most Popular list and the top selling artists of the week (given by the BillBoard Hot 100), it is natural to inquire what methods MTV is using to generate these lists. We may speculate that MTV uses a combination of selective discretion and payments from record companies to develop these lists. After all, MTV's greatest asset is its place in pop culture and the ability to reach millions of viewers. MTV was named the "Best Global Pure Media Brand" for the ninth year in a row by a 2008 Business Week Study (Viacom). This is of great value to record companies who have products aimed directly at MTV's target demographic.

MTV, as well as VH1, Nickelodeon, BET, and other stations, is owned by Viacom, Inc. In its 2008 10-K filing, Viacom reports that its *Media Networks* segment generates 54% of its revenue through advertising. This includes the sale of ads both on television and on digital properties (such as MTV.com). Media Networks has formed "strategic relationships" to distribute content with AOL, Myspace, Hulu, Bebo, and more. So clearly MTV is attempting to direct and gain traffic from the artists Myspace pages. But the filing is not clear about what the terms of the deals are with each of these companies, and there is no mention of relationships with any record labels. However, if these record companies are paying advertising clients (not affiliate partners), they certainly would not be listed here. Across all of Viacom's properties Media Networks segment accounts for 100% of advertising revenue. This figure stood at $4.7 billion, or 32% of Viacom's $14.6 in annual 2008 revenue. This is clearly a very large and important part of their business, but it is not clear from the 10-k filing exactly what clients, industries or platforms account for the total.

Without direct comment from MTV and their executives, we are left to speculation. But certainly with the data given, it appears that the selected artists, time frames, and decisions to alter these lists is not algorithmic and likely relies on a combination of selection by MTV and payments from artists record labels.

6

ALBUM SALES AND PROFILE VIEWS CORRELATION

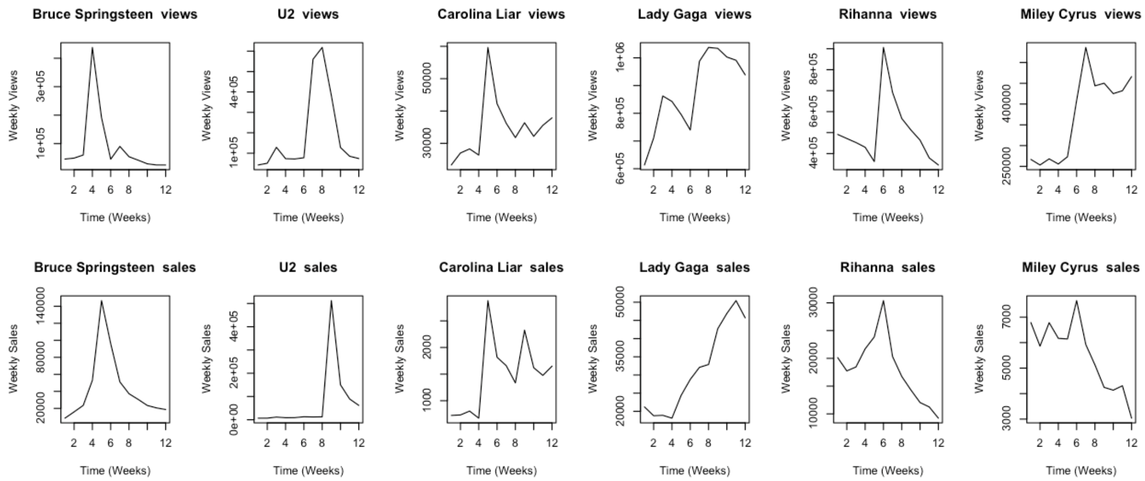


Figure 6.1 Profiles Views (top) and Album Sales (bottom) for 6 artists

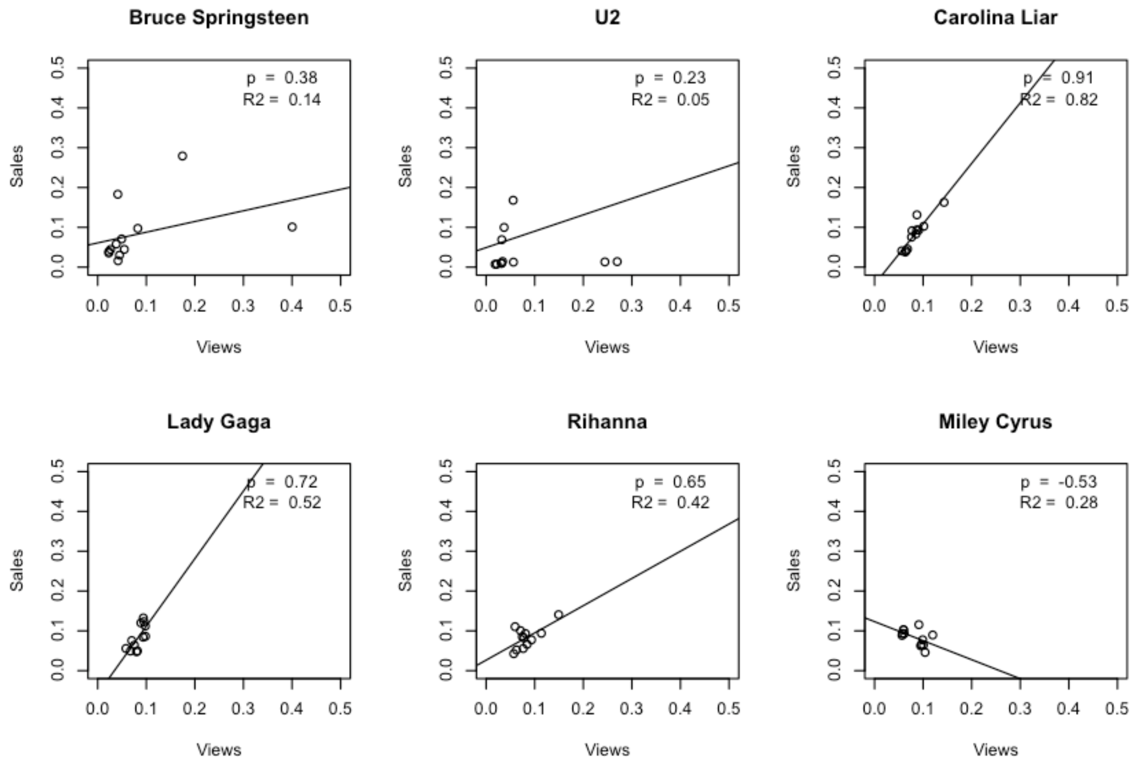
What record companies, marketing teams, and managers want to know is how does the interest in my artist's Myspace page affect revenue generation? That is, if people travel to an artists Myspace site, are they likely to also go out and buy the album? It seems that all record companies have decided that having a Myspace profile will increase the chances of making money through album sales, appearances, tours, and sponsorship because all major label artists have their own Myspace page and all of them offer free streaming of up to 6 songs in full length. Therefore they are willing to sacrifice free plays from the Myspace profile in hopes of further revenues from album sales, merchandise, and touring.

6.1 LINEAR REGRESSION ON NORMALIZED DATA

Figure 6.1 above shows the profile views and accompanying album sales for 6 artists at weekly intervals. The album sales were purchased from Nielsen Soundscan. The first pattern observable to the naked eye is a clear correlation between the spikes, trends, and dips for most of these artists. We also notice that it appears the profile views may spike slightly before albums sales. This result is likely the result of the hype phenomenon discussed in Chapter 4.

In order to best understand the relationship quantitatively, I regressed normalized profile views against normalized album sales with the expectation that increases in profile views would lead to increases in sales and a positive linear correlation would exist. Almost all Myspace artist profiles have prominent links to “Buy on iTunes” or Amazon, or a number of other online music retailers, so these profiles serve not just as a place to share information with fans, but to direct them to revenue generating sites. For the most part, this positive correlation is found in the data.

Figure 6.2 Correlation Between Views and Sales



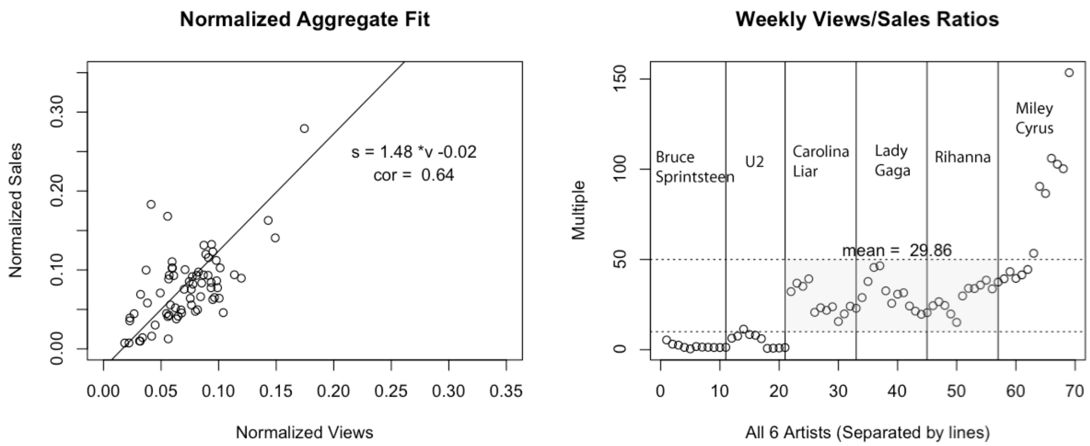
A plot of normalized weekly views (x-axis) against normalized weekly album sales (y-axis) for the 12-week period from January 11th, 2009 to April 4th, 2009. There exists a positive linear correlation for all artists, except Miley Cyrus. The R-squared values are low, indicating we should not expect a perfect linear fit, and the shape of the fit implies that we should not expect a 1:1 ratio between an increase in views and sales.

Bruce Springsteen and U2 have a few outliers. These points are weeks of high profile views (x value) and correspondingly lower sales (y value). These are points that correspond to hype in anticipation of an upcoming album release. One may be tempted to shift the values over a week, but the problem is that very quickly after release (within 1 week), the correlation between views and sales returns to a true week-to-week relationship. Bruce Springsteen and U2 are also the only 2 artists that had an album released during this time frame. This suggests that in the absence of hype, profile views and album sales correlate well. During times of excitement, however, the timing between fluxes of views and sales are off by a about a week.

To fix this issue, I removed these “hype” data points and calculated the correlation and line of best fit for all of the data. In this way we use these 6 artists as a training set and arrive at an equation we could use to estimate sales from views for our other 396 artists. The equation derived from a least squares regression is:

$$\text{Normalized Weekly Sales} = 1.48 * \text{Normalized Weekly Views} - 0.02$$

Figure 6.3 Total Fit for Views and Sales Excluding Outliers



Above left: is all 6 artists aggregated together, removing the 3 outlier points which were the result of hype (causing views to greatly exceed sales because the album was not yet released). We then take these normalized points and fit a linear regression and get the correlation. **Above right:** is the multiple of raw weekly views to raw weekly sales. We can see that it differs for each artist. For Bruce Springsteen and U2, who had much anticipated albums released the album sales is very high relative to views because it is a period of intense sales and because these artists have distinctly older audiences than the others. For artists who have had an album out for a while (such as Miley Cyrus, the last segment on the right) the continued interest dwarfs current sales. For the other 3 artists, they are in a middle ground where sales and views appear somewhat constant. Thus I use this data to estimate the multiple a typical teen focused artist might expect for his or her album sales to profile views. This multiple is around 30. For an artist with an older population (such as U2 and Bruce Springsteen), this multiple is around 3.

From this regression, we see that the correlation is positive and of medium strength (Pearson correlation of 0.64). Certainly from just examining the graph we can tell that any estimate is not truly to be expected, but on the average may give a good measure of an artist's normalized sales for that week.

6.2 ESTIMATING SALES

We run into a problem if we try to use this to predict real sales. We can get an estimate of the weekly-normalized sales from profile views, but how do we translate to actual sales numbers and revenue? That is, is there a consistent relationship between sales volumes and total profile views? Does an artist with twice as much traffic on their Myspace page sell twice as many records? One way to answer this would be to get one sales data point, which may be easier and less expensive to obtain the whole sales data set. From that first data point we could back out what the other sales must be.

Another option is to examine multiples of raw sales/views for the data we have and use those multiples to estimate the sales from views for other artists. These multiples are shown in the right plot in Figure 6.3. In this chart each vertical band separates one artist from another. We notice that each artist seems to have a characteristic range of multiples. This makes sense because each artist differs in how involved their fans are with the Myspace community. For an artist with a fan base that uses Myspace heavily, views may greatly exceed sales. For an artist whose fans do not use Myspace, sales may be much larger relative to daily views. But as we have seen in the correlations plot, the relative changes seem to stay consistent no matter what.

To a large extent, these characteristic ranges have to do with age of the fan base. For U2 and Bruce Springsteen, both artists who gained fame over 20 years ago in the 1980's, we do not expect Myspace profile views to be proportionately as big as their fan base. Myspace was launched in 2003 and attracts a largely teenage audience. For simplicity, we can consider two types of artists: "teen" and "general". The first refers to artists whose marketing is based towards a teen audience of similar composition to the Myspace community. A General artist refers to an artist whose appeal is to older or to a more alternative demographic, not captured by Myspace. In reality there are not just 2 types of artists and every artist probably lies somewhere in between these extremes so we can create a variable to represent the demographic with a value of 1 being a purely teen audience and a value of 0 being a purely older audience.

Table 6.1 Estimating Sales from Profile Views

Artist Type	Description	Sales Estimate
Teen	Fan base is young and typical of the Myspace community	$S = V / 30$
General	Fan base is older or more alternative, not typical of the Myspace community	$S = V / 3$
Mix	Fan base is a mixed demographic. Estimate parameter, i, between 0 (General) and 1 (Teen)	$S = V / (3 + 27i)$

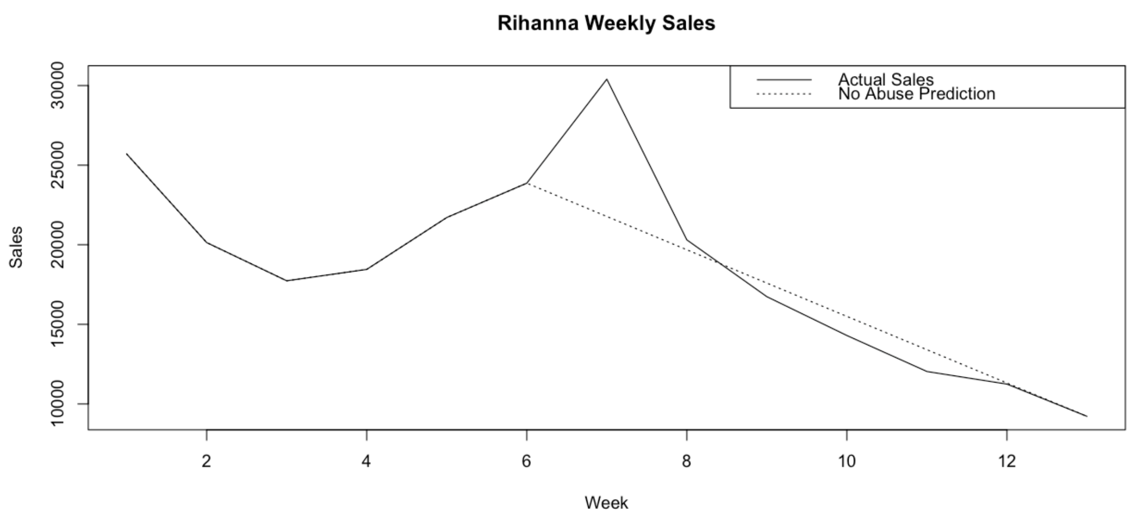
Using these equations, I can estimate sales for any of the 400 artists. As a few test examples, Table 6.2 lists sales estimates over a 10-week period from the week ending January 4th, 2009 to March 8th, 2009. I selected the artists Chris Brown (relation to Rihanna), Britney Spears (wildly popular) and The Fray (widely publicized album release).

Table 6.2 Weekly Sales Estimates for Various Artists (in 000's)

Artist	i	1	2	3	4	5	6	7	8	9	10	Total
Britney Spears	1.0	16.1	13.9	13.4	13.5	11.3	9.5	9.6	11.5	13.7	13.0	125
Chris Brown	1.0	13.0	13.4	13.1	12.6	11.2	28.9	20.2	15.8	15.7	13.4	157
The Fray	0.9	8.7	8.8	9.8	11.1	18.1	14.2	12.0	10.3	10.0	8.4	111

Given the sales data we can estimate how much extra revenue was generated as a result of Rihanna being abused by boyfriend, Chris Brown. We look at the increase in sales and the relaxation, integrating the area above what the estimated trend line would have been for each week following the abuse scandal.

Figure 6.4 Rihanna Weekly Sales



A plot showing the total weekly sales for Rihanna, along with a linear estimate of would be sales in the absence of the large shock caused by news of the abuse she received from boyfriend Chris Brown.

The total difference between actual sales and predicted sales is only 5,757 albums. Album pricing averages $15 across most retailers, thus resulting in $86,300 in additional revenue. Consider Rihanna's previous weekly sales, this is not as significant as we might have expected. It is only 24% increase on one week of sales! The lesson here is that even news of a negative nature will increase sales, but it will not do so as much as profile views (which spiked over 8 times). That is people consume the news but do not reach into their pockets to purchase the music right away. Perhaps Rihanna will see longer term benefits from this media exposure when she releases her next album and has a name that is more familiar, more talked about, and more hyped up given her past history.

From examining the profile views and sales for the six artists I was able to purchase data for, we can conclude a correlation exists. We can conclude that it is in general positive and linear, with a slope near +1. We can use this data set as a training data set to make estimates for all the other artists weekly sales. I have done so, but without being able to purchase further data, it is impossible to tell whether these estimates are accurate or not. Given more time and money, this is the first analysis I would perform. I would use the new data to further refine the linear relationship and try to further refine the “multiples method” of estimating sales.

7

CONCLUSIONS

This paper has attempted to take data that is the result of millions of people interacting to make decisions in a constrained environment and make sense of it. We have looked at how profile views for 400 artists change over time and attempted to model and understand the patterns that emerge. When someone visits an artists profile they are making a conscious decision to get information about that artist. Usually it is because the user wants to listen to a song, but it can also be to read the artist's blog, see that artist's other friends, or leave a comment for the artist. In any case, this interaction is intimate and shows that the user has been told about this artist and cared enough to spend his or her time with that artist's page. In turn the user may tell other friends to visit this page and so the network grows.

After a big piece of news comes out, such as an album release, new single, performance, or television appearance, more users flock to that artist. Over time the interest decays and does so in a predictable way. This paper has shown that the best fit for the acceleration and decay patterns around peaks is a Power Law. This is interesting because a Power Law is also being found in many other networks such as citations, the Internet itself, and the waiting time between executions of tasks. Sornette and Crane have found this to hold true in the online network, YouTube, and in online commerce site Amazon.com by examining the sales of books over time. They propose an epidemic model of buyers connected through a network and determine empirically the model parameter. I found a model parameter of 0.35, similar to the value of 0.3 they had found in YouTube and Amazon book sales.

Classifying exogenous and endogenous shocks summary:

- The community is susceptible to "hype" - the building up of views that relax very quickly after the peak. This phenomenon is not found in most other networks.
- Most peaks are short lived and relax within 1 week (~50%). Roughly 33%, however, do not relax within 2 weeks after a peak, indicating a shock that has a lasting impact on artist interest
- Model parameter (see Figure 2.3 for the model) ~ 0.35 determined empirically.
- No significant exogenous critical cluster found.
- Parameter is similar to that found in other networks.
- The 0.35 cluster leads us to believe many parts of the Myspace network are at criticality (news propagates easily from person to person).

- The existence of a cluster near 1.35 indicates some artists do not have a fan base whose network is at criticality.
- There is not a high level of correlation between the day of an album release and the peak days. Although we may expect maximum views to be found around an album release, this does not appear to be the case.

Next I examined Google Trends data to understand whether a flux of people visiting a Myspace profile was limited to this community or if it represented an increase in interest worldwide. The clear answer is that Myspace views and Google searches correlate highly, although in some cases the multiple of increase in Google searches was a stronger reaction than the Myspace community. I speculate this is because news events large enough to reach major outlets such as Newspapers, CNN, and other blogs reach a new audience that is not fully captured by Myspace. Nevertheless, it is reassuring to know that information spread in the Myspace network behaves similarly to that in the general public. It even suggests that studying and modeling the data from Google Trends directly (on a variety of different subjects and topics) may provide for another example to show people sharing information in a way that is best modeled by a Power Law.

Google Search and Myspace Profile View Summary:

- A shock in Myspace views is also a shock to related Google Search.
- Multiple of increase is not always the same in Google as Myspace.
- Generally, Google reacts more strongly to large news.
- Hypothesis that Google searches also obey similar Power Law dynamic.
- Hypothesis that Sornette's model would apply equally well.

As well as tracking Myspace profile views, I have tracked the artists listed by MTV as "Most Popular", "New Artists", and "Artist Picks". I examined the average profile views for these artists while they were listed by MTV and the average views while they were not listed. The expectation is that MTV chooses artists to list who are in a popular phase of their career, and perhaps even that MTV promotions through MTV.com and short televised video play would directly increase Myspace views. It turns out this is not the case with statistical significance. In the case of Most Popular artists profile views were actually higher on average while not listed. Further examination of these lists (Figure 5.1) shows that the are not changed with regularity in time intervals, indicating the choices are not driven by data (such as sales, weekly requests, or video views). I speculate that MTV changes artists based on who its clients such as the big four record companies wish to promote, although I have no definitive evidence indicating payment is necessary to be listed.

MTV Promotional List Summary:

- MTV's methods of promotion are hard to understand.
- They are not algorithmic or data driven.
- The time intervals are infrequent.
- The artists being promoted do not experience statistically significantly higher profile views during periods of promotion.
- The "Most Popular" artists, in fact, experience fewer views on average.
- MTV does not list any major record company as a partner, but has \$4.7B out of \$15B total revenue from advertising. This could be where some of the promotional decisions are derived from.

.

Finally, I examined the correlation between weekly profile views and weekly album sales over a 13-week period for 6 select artists. Two of these artists had albums released during the 13-week period and the other 4 had significant shocks, trends or activity during the time frame. This made them prime for studying whether media exposure resulted in proportionate increases in sales. Figure 6.1 shows the plots for all the artists, and it is clear that a positive linear correlation exists in all but one case. I will note that I do not have access to the conversion data to understand why or how people who bought the album decided to do so. That is, did the increase in Myspace views lead to people clicking through to iTunes or deciding to go visit the record store? Or did purchasing the album cause them to become curious and read more about the artist at their Myspace page? Likely the answer is a bit of both and given more time and access to the conversion data, I would examine this relationship. This would give an even better idea of how important an artists Myspace profile is to generating sales and getting fans to become paying customers.

<u>Album Sales and Profile Views Summary:</u>

- 6 Artists were examined and a positive linear correlation existed for 5 out of the 6.
- The R-squared values for normalized weekly sales against normalized weekly views were not very high, with an aggregate R2 of 0.64.
- Analyzing the ratio of views to sales shows us that each artist has a characteristic range
- These ranges are likely proportionate to how much of the artist's fan base uses Myspace
- I estimate the multiple range to go from 3 for an artist with limited Myspace fan base to 30 for an artist with a fan base who uses Myspace frequently.
- Using this data, we can create a model to estimate sales (S) from views (V).
- Model: $S = V / (3 + 27 * i)$.
- In this model, *i* is a somewhat subjective parameter ranging from 0 to 1. 0 indicates very little to no Myspace usage in the fan base, and 1 indicates a fan base whose demographics align perfectly with Myspace.

BIBLIOGRAPHY

Arrington, Michael. *Facebook Now Nearly Twice the Size of Myspace WorldWide.* 22 Jan. 2009. http://www.techcrunch.com/2009/01/22/facebook-now-nearly-twice-the-size-of-myspace-worldwide/

Barabasi, Albert-Laszlo. "The origin of bursts and heavy tails in human dynamics." *Nature*, May 12, 2005.Volume 435. http://arxiv.org/pdf/cond-mat/0505371.

Barabási, A-L. and Albert, R. "Emergence of Scaling in Random Networks." *Science* **286**, 509-512, 1999.

Barabási, A-L.; Albert, R.; and Jeong, H. "Mean-Field Theory for Scale-Free Random Networks." Physic *A* **272**, 173-187, 1999.

Cesar A. Hidalgo R. and Albert-Laszlo Barabasi (2008) "Scale-free networks." Scholarpedia, 3(1):1716

Crane, R. Sornette, D. "Robust Dynamic Classes Revealed by Measuring the Response Function of a Social System." *Proceedings of the National Academy of Sciences,* 2008; 105 (41): 15649 DOI: 10.1073/pnas.0803685105

Crane, R. Sornette, D. "Viral, Quality, and Junk Videos on YouTube: Separating Content From Noise in an Information Rich Environment." http://drrileycrane.googlepages.com/VQJ-YouTube.pdf

Deschatres, F. Sornette, D. "The Dynamics of Book Sales: Endogenous versus Exogenous Shocks in Complex Networks." *Physical Review E.* Vol. 72. Issue 1. Id 016112. http://arxiv.org/abs/physics/0412171

Eubank, S. et al. "Structure of Social Contact Networks and Their Impact on Epidemics." Chapter from AMS-DIMACS Special Issue on Epidemiology. 2006. http://gladiator.ncsa.uiuc.edu/PDFs/networks/eubank2006.pdf

Galasso, Alexandra. <agalasso@myspace.com> "RE: Myspace Profile Views". 20 Feb. 2009. Personal Email (20 Feb. 2009).

Kincaid, Jason. *Rules of Engagement: Users are Spending More Time on Myspace.* TechCrunch.12 Feb. 2009. http://www.techcrunch.com/2009/02/12/rules-of-engagement-users-are-spending-more-time-on-myspace/

Lamb, Bill. *Top 4 Major Pop Record Labels.* About.com. 1 May 2009. http://top40.about.com/od/popmusic101/tp/majorlabels.htm

Matlis, Jan. "Quick Study: Scale Free Networks". 4 Nov. 2002. Article on Computer World. Accessed March 17, 2009. http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=75539

Perelló, Josep, Masoliver, Jaume, Kasprzak, Andrzej and Kutner, Ryszard. *A Model for Interevent Times with Long Tails and Multifractality in Human Communications: An Application to Financial Trading*. 18 July. 2008. http://ssrn.com/abstract=1162706

Quantast. May 1, 2009. http://www.quantcast.com/docs/display/info/FAQ

Small, Michael. "Scale-Free Network." From *MathWorld*--A Wolfram Web Resource, created by Eric W. Weisstein. http://mathworld.wolfram.com/Scale-FreeNetwork.html

Sornette, D. *Endogenous versus Exogenous Origin of Crises*. 2 Feb. 2008. http://arxiv.org/pdf/physics/0412026v1

Toto, Serkan. *Researcher Claims "Attention Spirals" Hold Key to Predicting Success of YouTube Videos*. TechCrunch. 18 Nov. 2008. http://www.techcrunch.com/2008/11/18/researcher-claims-attention-spirals-hold-key-to-predicting-success-of-youtube-videos/

Viacom, Inc. (2008). 10-k Annual Report 2008. Retrieved May 8, 2008 from Viacom website at http://www.viacom.com/news/News_Docs/78157ACL.PDF.

Vazquez, Alexei, et al. *Modeling bursts and heavy tails in human dynamic*. October 12, 2005. http://arxiv.org/abs/physics/0510117v1

Appendix A1

I include the code for the scripts that gathered the data and performed the data analysis that resulted in the charts found throughout this paper. I have done this to provide as much transparency as possible. Please note that some scripts (such as the MATLAB script) have been shortened to make it more readable. I encourage readers to reuse any of these scripts as necessary and I can provide them in full file format. Please contact me at mungerer@princeton.edu if you are interested.

Table A1.1 Listing of Scripts

File Name	Language	Description
Daily.py	Python	Scrapes the profile views from 400 artists. Intended to be executed via cron on a daily basis
removeZeros.m	MATLAB	Takes raw profile views from text files and writes views to new files with any zeros removed by averaging
ParseData.r	R	Reads in, analyzes and performs regressions on profile views
Plats	R	Produces all plots found in Chapters 1-4
MTV.r	R	Reads in artist lists, plots, and analyzes the lists with respect to profile views
albumSales.r	R	Reads in album sales data from Nielsen Soundscan and produces correlations between sales and views

A1.1 Python Script

```
#!/usr/bin/python
####################################################################
# Mark W. Ungerer
# Princeton University
# Operations Research and Financial Engineering
# mungerer@princeton.edu
#
# - Filename: daily.py
# - Description: scrapes data from social networking site Myspace
# - Details: uses urllib and regular expressions to find the profile
# views for different artists and writes that data point into a
# file.
# - It is intended for this file to be executed once every day
# - Execution: >>> daily.py
####################################################################

import urllib
import re

######   ARTISTS TO TRACK #######
w = list()
w.append("glasvegas")
w.append("thegourds")
w.append("erinmccarley")
…
< all other artists here>
…
w.append("eltonjohn")
w.append("tilatequila")
w.append("hoobastank")
################################

# Regular Expressions
pattern = re.compile('Profile\sViews:\D*\d*')
digits  = re.compile('\d{1,13}')

# Get Data for each artist
for i in range(1, len(w)+1):
        # Open file
        filename = 'Thesis/Daily/' + str(i) + '.txt'
        f = open(filename, 'a')

        # Read in webpage (check for connection error)
        try:
                # Get webpage
                webpage = urllib.urlopen('http://www.myspace.com/'+w[i-1]).read()

                # Extract profile views (check for match not found)
                re_profile = re.search(pattern, webpage)
                if re_profile != None:
                        profile = re_profile.group()
                        views   = re.search(digits, profile).group()
                        f.write(views + '\n')
                        print views + ' for ' + filename
                else:
                        print 'ERROR: Profile views not found in webpage, writing 0 into file ' +
str(i) + '.txt'
                        f.write('0' + '\n')

        # There was a problem connecting and opening URL
        except:
                print "ERROR: Problem opening web page " + w[i-1] + ", writing 0 into file " +
filename
                f.write('0' + '\n')

        f.close() # Close file
```

A1.2 MATLAB Data Manipulation (Raw Views to Views Per Interval)

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Mark W. Ungerer
% Princeton University
% Operations Research and Financial Engineering
% mungerer@princeton.edu
%
% Filename: removeZeros.m
%
% Gets rid of zeros with averages of previous data points
%
% Must specify source of data:
%   Can be "Data" or "Daily" ... capitalization must be
%    correct!
% OUTPUT: writes new files into a folder "NewData"
% that must be in the same directory and be EMPTY!
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function removeZeros()

    % Define constants
    maxViewsPerInterval = 12000;
    maxViewsPerDay = 600000;

    % Fix data for both daily and half hour views
    %update = 'Removing zeros from Data/'
    %removeZerosFrom('Data/', maxViewsPerInterval, sizeData());

    update = 'Removing zeros from Daily/'
    removeZerosFrom('Daily/', maxViewsPerDay, sizeDaily());


    function removeZerosFrom(dataSource, maxViews, M)

        % Get Parameters
        N     = numArtists();
        data  = zeros(N,M);
        views = zeros(N, M-1);

        % READ IN DATA FOR THIS FILE
        for i = 1:N
            f         = strcat(dataSource, int2str(i), '.txt');
            data(i,:) = textread(f, '%u', 'delimiter', '\n');
        end

        update = '...data read in, now removing zeros and converting to views...'

        % REPLACE ZEROS WITH AVERAGES
        for i = 1:N
            for j = 1:M-1
                if data(i, j+1) == 0

                    % Check for multiple 0's in a row
                    start = j;
                    final = j+2;

                    % Find end of string of 0's
                    while data(i, final) == 0
                        if final >= M
                            final = final + 1;
                        else
                            break;
                        end
                    end

                    % Calculate increment and write new values
                    increment = round((data(i, final) - data(i, start)) / (final-start));
                    for k = (j+1):(final-1)
                        data(i, k) = data(i, k-1) + increment;
                    end
                end
            end
        end

        % Convert this raw data to views per interval
        for i = 1:N
            for j = 1:M-1
                tempViews = data(i, j+1) - data(i,j);
                if tempViews < 0
                    views(i,j) = 0;
                else
                    if tempViews > maxViews
                        views(i,j) = views(i,j-1);
```

```
                else
                    views(i,j) = tempViews;
                end
            end
        end
    end

    update = '...Normalizing data...'
    %%%%%%%%%%%%%%%%%%%% NORMALIZE %%%%%%%%%%%%%%%%%%%%%
    norms = zeros(N, M-1);      % normalized views per interval
    totalViews = zeros(M-1,1);  % views per interval all artists
    normTotalViews = zeros(M-1,1); % normalized views all artists
    sums = zeros(N,1);

    for i = 1:N
        sums(i) = sum(views(i,:));
        for j = 1:M-1
            norms(i,j) = views(i,j)/sums(i);
        end
    end

    % Normalize total views vector
    for i = 1:M-1
        totalViews(i) = sum(views(:,i));
    end

    sumTotalViews = sum(totalViews);
    for i = 1:M-1
        normTotalViews(i) = totalViews(i)/sumTotalViews;
    end


    update = '...writing to new files...'

    % WRITE TO NEW FILES
    for i = 1:N

        % Open new file, destroying old contents
        newFolder = strcat('New', dataSource);

        f   = strcat(newFolder, int2str(i), '.txt');
        fid = fopen(f, 'w');

        % Write every data point
        for j = 1:M-1
            viewsLine = strcat(int2str(views(i,j)), '\n');
            fprintf(fid, viewsLine);
        end

        fclose(fid);
    end

    % Open total.txt
    newFolder = strcat('New',dataSource);
    f = strcat(newFolder, 'total.txt');

    fid = fopen(f, 'wt');

    % Write data
    for i = 1:M-1
        totals = sum(views(:,i));
        viewsLine = strcat(int2str(totals), '\n');
        fprintf(fid, viewsLine);
    end

    % Close file
    fclose(fid);

  end
end
```

A1.3 R Scripts (Data handling, plotting, time series)

ParseData.r reads in data, turns it into a time series and creates necessary functions.

```
###################################################
# Mark Ungerer
# Princeton University
# Operations Research and Financial Engineering
# mungerer@princeton.edu
#
# Filename: ParseData
#
# Description:
# - Reads in profile view data and turns it into time series objects
# - Creates functions findArtist and findFile to help find which
#   artist corresponds to which data column
# - Finds Peaks
# - Analyzes shapes of peaks
# - Fits peaks to log-log regression
# - Calculates R Squared Values
#
###############################################################
######## LIST OF MAJOR VARIABLES CREATED BY THIS SCRIPT #######
# DATA:
# allDaily, normDaily, allData = main data sets
# Daily.ts, NormDaily.ts, HalfHour.ts = main time series
# allNorm, normDaily, normMajor = normalized values
#          normDaily is for smaller data set (179 artists)
#               but has an extra 57 days of data
#          normMajor is for 403 major label artists
#          allNorm combines them with 57 days of 0's to make them
#               equal in length
#
# PEAKS:
# peaks     = (numPeaks x 2) with each row corresponding to
#             (artist number, peak point)
# peaksData = (numPeaks x 15) with each row corresponding
#             to one peak, with critical point at [i,8]
# numPeaksArray = (30 x 1) as numPeaks as a function of k
#             with k = 1.1, 1.2, ..., 3.0
# shapes    = (20 x 3) corresponding to k = 1.1, ..., 3
#             and each row to (numSymm, numLeft, numRight)
# classifyPeaks = 3x1 shapes for k = optimal (2.0)
# peakFractions = (numPeaks x 2) to fraction views in peak
#                 relative to (week before, week after)
#
# REGRESSION:
# signatures  = (numPeaks x 2) with each row = power law signature
#               for (foreshock, aftershock)
# rsquare     = (numPeaks x 2) with each row = r squared value for
#               log-log linear fit
# numFits     = num peaks with fore + decay r squared > 0.8
# numDecays   = num peaks with decay r squared > 0.8
# numLinLogFits = num peaks w/ fore + decay r squared > 0.8
# numLinLogDecays = num peaks w/ decay r squared > 0.8
# fittedPeaks = (numFits x daysForPeak) with each row corresponding to
#               peak who had a good enough regression on both
# fittedDecays= (numDecays x daysForPeak) for peaks with good decay fit
# goodFits    = (numFits x 1) corresponding to artist the good
#               fit came from
# goodDecays  = (numDecays x 1) corresponding to artists the good
#               decay fits came from. Good for finding those
#               "example" peaks
# alphas      = (numPeaks x 2) with each row corresponding to
#               alphas from plfit (see documentation in plfit.r)
#               for foreshock, aftershock
#
##############################################################
```

```
# CLEAR WORKSPACE
rm(list = ls(all = TRUE))

# IMPORT RENE CARMONA'S LIBRARY
library(Rsafd)

# GET DATA
setwd("/Users/mungerer/Dropbox/Thesis/Web Scraping/NewData/")
allData <- read.table("allData.txt", header=TRUE)
setwd("/Users/mungerer/Dropbox/Thesis/Web Scraping/NewDaily/")
allDaily <- read.table("allDaily.txt", header=TRUE)
setwd("/Users/mungerer/Dropbox/Thesis/Web Scraping/NormDaily/")
normDaily <- read.table("normDaily.txt", header=TRUE)
setwd("/Users/mungerer/Dropbox/Thesis/Web Scraping/NewMajor/")
allMajor <- read.table("allMajor.txt", header=TRUE)
setwd("/Users/mungerer/Dropbox/Thesis/Web Scraping/NormMajor/")
normMajor <- read.table("normMajor.txt", header=TRUE)
allNorm = do.call('cbind', c(normDaily, normMajor))

# SIZE OF DATA SETS
nArtists = 180+223 # 403 artists total
M_data   = length(allData[,1])
M_daily  = length(allDaily[,1])

# MAKE DAILY TIME SERIES
Daily.ts     <- timeSeries(positions = timeSequence(from = timeDate("2000-01-05"), length=M_daily,
by="day"), allDaily)
normDaily.ts <- timeSeries(positions = timeSequence(from = timeDate("2000-01-05"), length=M_daily,
by="day"), allDaily)

# MAKE HALF HOUR TIME SERIES
N   <- round((M_data) /2)
POS <- sort(c(timeSequence(from = ISOdate(2009, 1, 4, hour = 0, min = 0, sec = 0, tz = "EST"),
length=N, by="hour", allData), timeSequence(from = ISOdate(2009, 1, 4, hour = 0, min = 30, sec = 0, tz
= "EST"), length=N-1, by="hour", allData)))
HalfHour.ts <- timeSeries(positions = POS, length=M_data, allData)

# MAKE DATA WEEKLY
weeks        = trunc(M_daily/7)
dataByWeeks = matrix(ncol=180, nrow = weeks)
dataPointsFilledIn = 0
sevens       = seq(1,weeks*7,7)
for (i in 1:180)
{
	dataPointsFilledIn = 0;
	for (j in sevens)
	{
		dataPointsFilledIn = dataPointsFilledIn + 1;
		dataByWeeks[dataPointsFilledIn,i] = sum(allDaily[j:(j+6),i])
	}
}

# Extract seasonal, trend and components
#daily.stl <- sstl(Daily.ts[,180]) # doesn't work, not seasonal
hh.stl <- sstl(HalfHour.ts[,180])

# CREATE FUNCTION TO GET ARTIST NAME FROM NUMBER
artists = c('glasvegas', 'thegourds', 'erinmccarley', 'officialtsool', 'wholewheatbread',
'jessiekilguss', 'animalcollectivetheband', 'mydeardisco', 'toopuretodie', 'varsityfanclub',
'heatherheadley', 'ohnonotstereo', 'illinois', 'mattyork', 'orthewhale', 'combichrist',
'antonyandthejohnsons', 'lisahannigan', 'reelbigfish', 'johnfrusciantemusic', 'umphreysmcgee',
'modernskirts', 'andrewbird', 'loneydear', 'richiebookermarley', 'ciara', 'lisalopesthegreatest',
'franzferdinand', 'duncansheik', 'brucespringsteen', 'badplus', 'dentmay', 'twotonguesrock',
'melindadoolittle', 'thefray', 'bowwow', 'chriscornell', 'vonbondies', 'lilymusic', 'mosdef',
'courtneylove', 'benlee', 'taylorhicks', 'indiaarie', 'mirandaleerichards', 'bustarhymes',
'missyelliott', 'warreng', 'dexterromweberduo', 'morrissey', 'theappleseedcast', 'plushgun',
'trailofdead', 'robynhitchcock', 'tommykeeneband', 'anaegge', 'brokenspindles', 'wheelsonfire',
'faunts', 'kovakuk', 'sevencaves', 'taxidoll', 'leftsidebrain', 'hatebreed', 'robiemusic',
'karlwilson', 'lorien', 'zacknichols2', 'nekocase', 'justintownesearle', 'u2', 'bellx1',
'shortwavefade', 'officialmadeleinepeyroux', 'handsomefurs', 'mstrkrft', 'kellyclarkson', 'sncmusic',
'nestysangrenueva', 'margotandthenuclearsoandsos', 'deptofeagles', 'lauraizibor', 'thisisbrutha',
```

```
'rascalflatts', 'sugarland', 'zacbrownband', 'montgomerygentry', 'bradpaisley', 'alanjackson',
'billycurrinton', 'dierksbentley', 'jameyjohnson', 'blakeshelton', 'keithurban', 'taylorswift',
'jimjones', 'kingsofleon', 'anberlin', 'riseagainst', 'incubus', 'offspring', 'shinedown',
'enticetheband', 'themterribles', 'teganandsara', 'carolinaliar', 'morningwood', 'wethekings',
'maydayparade', 'donnieklang', 'jazminesullivan', 'eminem', 'thecab', 'thewombatsuk', 'coldwarkids',
'mastodon', 'mikamyspace', 'klaxons', '3oh3', 'secondhandserenade', 'mudvayne', 'britneyspears',
'mariahcarey', 'mileycyrus', 'christinaaguilera', 'allamericanrejects', 'themaine', 'paramore',
'theacademyis', 'beyonce', 'rihanna', 'akon', 'foreverbrandy', 'kanyewest', 'ludacris', 'trapmuzik',
'50cent', 'kardinaloffishall', 'falloutboy', 'kevinrudolfmusic', 'coldplay', 'jacksmannequin',
'justintimberlake', 'jonasbrothers', 'johnlegend', 'pinkspage', 'solange', 'lilwayne', 'neyo',
'llcoolj', 'kidrock', 'metrostation', 'ladygaga', 'katyperry', 'thegame', 'hamiltonsmusic', 'tpain',
'gymclassheroes', 'soad', 'plainwhitets', 'ofarevolution', 'amywinehouse', 'vampireweekend',
'theveronicas', 'thedodos', 'titusandronicus', 'semipreciousweapons', 'redjumpsuit', 'thetingtings',
'jordinsparks', 'leonalewis', 'aliciakeys', 'chrisbrown', 'threesixmafia', 'celinedion' ,'metallica',
'billy_joel', 'eltonjohn', 'tilatequila', 'total',
'gabriellacilmi','chriscornell','lionelrichie','jadakiss','thedreamteam','methodman','sum41','ronbrowz
','razorlight','whitelies','thisisutada','thelovewillows','kerihilson','chrisettemichele','thirtysecon
dstomars','aaliyah','acdc','aerosmith','againstme','alanismorissette','alkalinetrio','amberpacific','a
mygrant','amypearsonmusic','anastaciaofficial','antiflag','ash','asherrothmusic','ashleytisdale','atmo
sphere','atreyurock','avengedsevenfold','avrillavigne','backstreetboys','badcompany','beastieboys','bj
ork','blackeyedpeas','blackrebelmotorcycleclub','blacksabbath','blindside','blink182','blondie','blurt
heband','bonjovi','boyslikegirls','bulletformyvalentine','camilatodocambio','carlysimon','carrieunderw
ood','celinedion','officialchakakhan','chamillionaire','cherconquerstheworld','chevelle','daughtry','c
ky','coheedandcambria','collectivesoul','coolio','craigdavid','daftpunk','damienrice','danielpowter','
danielaonline','danitykane','davematthewsband','thedavestewart','davidarchuleta','davidbowie','officia
ldavidcook','deathcabforcutie','destinyschild','diddy','dido','disturbed','dixiechicks','dmx','drdre',
'dragonforce','duranduran','elviscostello','elvispresley','enriqueiglesias','ericclapton','evanescence
','everclear','faithhill','fionaapple','flaminglips','flyleaf','foofighters','fortminor','funeralforaf
riend','genesismusic','georgeharrison','georgemichael','glassjaw','gloriaestefan','gnarlsbarkley','goo
goodolls','goodcharlotte','gorillaz','greenday','gunsnroses','gwenstefani','hardfi','hoobastank','inte
rpol','inxs','jamesblunt','jamesingrammusic','janesaddictionmyspace','janetjackson','jenniferlopez','j
essicasimpson','jet','jewel','johnmayer','jordyntaylor','journey','kellyrowland','kylieminogue','leona
lewis','lifehouse','lillix','lindaronstadt','linkinpark','lostprophets','lupefiasco','madonna','marily
nmanson','marionraven','maroon5','maryjblige','mastodon','matchboxtwenty','megadeth','metallica','mich
aelbuble','michaeljackson','michellebranch','michellewilliams','mikejones','moby','morrissey','motleyc
rue','muse','mychemicalromance','nas','nataliecole','nataliemerchant','natashabedingfield','nellyfurta
do','newkidsontheblock','nickelback','nirvana','norahjones','notoriousbig','oasis','oldirtybastard','o
zzyosbourne','payableondeath','panicatthedisco','paparoach','paulmccartney','paulwall','officialpaulac
ole','tenclub','phish','pixies','queen','rkelly','rem','radiohead','ratm','redhotchilipeppers','rickym
artin','rilokiley','robthomas','robbiewilliamsofficial','rodstewart','therollingstones','ryancabrera',
'samanthajade','seanpaul','shakira','simpleplan','slipknot1','staind','storyoftheyear','sugarray','tak
ingbacksunday','talibkweli','thebeatles','thecure','thedarkness','thedecemberists','eaglesmusic','thek
illers','smashingpumpkins','thestrokes','theverve','thevines','thewhitestripes','theymightbegiants','t
hirdeyeblind','threedaysgrace','tinaturner','tompetty','tonibraxton','toriamos','tracychapman','trapt'
,'officalunclek','usher','vanhalen','velvetrevolver','whitneyhouston','willsmithproduction','wutang','
yellowcard','total2');

# RETURNS THE ARTIST URL (STRING) FROM A FILE NUMBERS
findArtist <- function(i) {
      if (i > 403 || i < 1) return(0);
      return(artists[i]);
}
# RETURNS THE FILE NUMBER (INTEGER) FROM THE ARTIST URL
findFile  <- function(sArtist) {
      for (i in 1:403){ if (sArtist == artists[i]) { return(i); } }
      return('NOTFOUND');
}

#################################################################
# ALBUM RELEASE DATE... SET UP DATA
# Entry (i,j) means artist i, release date j
daysPast = c(2, 2, 2, 2, 2, 5, 9, 9, 9, 9, 9, 9, 9, 16, 16, 16, 16, 16, 16, 16, 16, 16, 23, 23, 23,
23, 23 ,23, 23, 23, 30, 30, 30, 30, 30, 30, 30, 30, 36, 36, 36, 37, 37, 37, 37, 37, 37, 37, 37, 43,
44, 44, 44, 44, 44, 44, 44, 44, 44, 44, 50, 50, 50, 51, 51, 55, 55, 55, 58, 58, 58, 58, 58, 65, 65,
72, 72)
daysPastMajor = c(72, 65, 65, 65, 65, 72, 72, 72, 65, 72, 79, 79, 79, 87)
numAlbumReleases = 77 + 14;
critical = matrix(ncol=2, nrow=numAlbumReleases)
for (i in 1:77){ critical[i,] = c(i,daysPast[i]); }
for (i in 78:(78+13)) {critical[i,] = c((i+103), daysPastMajor[i-77])}
```

```
##################################################################
# GET NUMPEAKS, AND SHAPE FOR DIFFERENT VALUES OF k
findShape <- function(k)
{
	numPeaks = 0;
	# IMPORTANT: nrow must be >= numpeaks found
	peaks = matrix(nrow = 10000, ncol = 2);

	# CHOOSE DATA SET
	v <- allNorm;

	# PEAK IS LOCAL MAX AND k GREATER THAN SURROUNDING
	for (i in 1:nArtists) {
		for (j in 8:(M_daily-7)) {
			if((   v[j,i] > k*mean(v[(j-7):(j-1),i])
				||  v[j,i] > k*mean(v[(j+1):(j+7),i]))
				&&  v[j,i] > max(v[(j+1):(j+7),i])
				&&  v[j,i] > max(v[(j-7):(j-1),i])
				&&  min(v[(j-7):(j-1),i]) > 0
			  )
			{
				numPeaks = numPeaks + 1;
				peaks[numPeaks, 1] = i;
				peaks[numPeaks, 2] = j;
			}
		}
	}
	numPeaksArray[l] = numPeaks;
	# Write data into array
	daysForPeak = 15
	peaksData = matrix(ncol=daysForPeak, nrow=numPeaks);
	for (i in 1:numPeaks)
	{
		start  = peaks[i,2]-7;
		if (start < 1) start = 1;
		end    = start + daysForPeak - 1;
		artist = peaks[i,1];
		y      = v[start:end,artist];
		peaksData[i,] = y;
	}

	# FRACTIONS OF VIEWS IN PEAK
	peakFractions = matrix(nrow=numPeaks, ncol=2)
	for (i in 1:numPeaks)
	{
		peakFractions[i,1] = max(peaksData[i,]) / (sum(peaksData[i,1:7]) + max(peaksData[i,]));
		peakFractions[i,2] = max(peaksData[i,8]) / (sum(peaksData[i,9:15]) + max(peaksData[i,]));
	}

	# Determine shape of peaks
	# classifyPeaks = (# symmetric, # skewed right, # skewed left)
	classifyPeaks = c(0,0,0)
	dif = seq(0, numPeaks)
	dif = peakFractions[,2] - peakFractions[,1];
	symCutoff = 0.01
	for (i in 1:numPeaks)
	{
		t <- abs(dif[i])
		if (t < symCutoff)
		{ classifyPeaks[1] = classifyPeaks[1] + 1 }
	}
	for (i in 1:numPeaks)
	{
		if (dif[i] > 0.01)
		{ classifyPeaks[2] = classifyPeaks[2] + 1 }
	}
	for (i in 1:numPeaks)
	{
		if (dif[i] < -0.01)
		{ classifyPeaks[3] = classifyPeaks[3] + 1 }
	}
```

```
	return(classifyPeaks)

} # End function findShape()

############## Get shape vs. k #############
# CHOOSE DATA SET
v <- allNorm
#v <- normDaily;

# LOOP THROUGH ALL k
shapes = matrix(nrow=20, ncol=3)
numPeaksArray = matrix(nrow=20, ncol=1);
for (l in 1:20)
{
	k = 1.0 + (l/10);
	shapes[l,] = findShape(k);
	numPeaksArray[l] = sum(shapes[l,]);
}

######################################################
##### Find Peaks and Regress for Specific k ##########
# PARAMETERS
k              = 2.0
minDays        = 7      # Need at least 1 week to get fit
daysAfterPeak  = 14     # Search up to two weeks
t_critical_max = 3
Rthreshold     = 0.85
######################################################

# NECESSARY VARIABLES
numPeaks       = 0;
peaks          = matrix(nrow = numPeaksArray[10], ncol = 2);
daysBeforePeak = daysAfterPeak
daysForPeak    = daysAfterPeak*2 + 1
criticalDay    = daysBeforePeak+1
t_critical     = seq(.1,t_critical_max,.1)
maxDays        = daysAfterPeak
foreThreshold  = Rthreshold
afterThreshold = Rthreshold

# FUNCTION: findPeak. SEARCHES FOR A PEAK, USED TO FIND OTHER SHOCKS
findPeak <- function(data, kp) {
	for (i in 5:(length(data)-1)) {
		if (    data[i] > kp*mean(data)
		     && data[i] > max(data[i-4:i-1])
		     && data[i] > data[i+1]
		  ) { return(TRUE); } }
	return (FALSE);
}

# INDENTIFY PEAKS THAT MEET CRITERIA:
# 1. Must be a local maximum
# 2. Must be greater than k times avergae of data before or after
# 3. Must not have another shock following
numPeaks = 0;
peaks = matrix(nrow=1000, ncol=2)
for (i in 1:nArtists) {
	for (j in criticalDay:(M_daily-daysBeforePeak)) {
		if((   v[j,i] > k*mean(v[(j-daysBeforePeak):(j-1),i])
			&&  v[j,i] > k*mean(v[(j+1):(j+daysBeforePeak),i]))
			&&  v[j,i] > max(v[(j+1):(j+daysBeforePeak),i])
			&&  v[j,i] > max(v[(j-daysBeforePeak):(j-1),i])
			&&  min(v[(j-daysBeforePeak):(j-1),i]) > 0
			#&&  !findPeak(v[j:(j+daysAfterPeak),i],2)
		  )
		{
			numPeaks = numPeaks + 1;
			peaks[numPeaks, 1] = i;
			peaks[numPeaks, 2] = j;
		}
	}
```

```
}
# SHORTEN ARRAY TO ONLY LENGTH
peaks = peaks[1:numPeaks,]

# WRITE DATA FOR THESE PEAKS INTO AN ARRAY
peaksData = matrix(ncol=daysForPeak, nrow=numPeaks);
for (i in 1:numPeaks)
{
	start         = peaks[i,2]-daysBeforePeak;
	end           = start + daysForPeak - 1;
	peaksData[i,] = v[start:end,peaks[i,1]];
}

# GET FRACTIONS OF VIEWS IN PEAKS
peakFractions = matrix(nrow=numPeaks, ncol=2)
for (i in 1:numPeaks)
{
	peakFractions[i,1] = max(peaksData[i,]) / (sum(peaksData[i,1:criticalDay]));
	peakFractions[i,2] = max(peaksData[i,]) / (sum(peaksData[i,criticalDay:(daysForPeak)]));
}

# GET AFTERSHOCK AND FORESHOCK AND THEIR LOGS
afterShock = peaksData[,criticalDay:daysForPeak]
foreShock  = matrix(ncol=criticalDay, nrow=numPeaks)
for (i in 1:numPeaks)
{ for (j in 1:criticalDay){	foreShock[i,j] = peaksData[i, (1+criticalDay)-j]; } }
logAfter	= log(afterShock)
logFore		= log(foreShock)

# SIGNATURES AND R SQUARED VARIABLES!
signatures	= array(0, dim=c(numPeaks, 2))
rsquare		= array(0, dim=c(numPeaks, 2))

# LOOP THROUGH ALL PEAKS
for (i in 1:numPeaks) {
	bestDaysA = 0;
	bestDaysF = 0;

	# TRY DIFFERENT t_critical VALUES
	for (t in 1:length(t_critical))
	{
		# GET TIME VECTOR
		logT = log(seq(t_critical[t], (t_critical[t] + maxDays),1))

		# TRY EACH LENGTH TO GET BEST FIT
		for (numDays in minDays:maxDays) {

			# GET FIT
			foreFit  = lsfit(x = logT[1:numDays], y=logFore[i, 1:numDays]);
			afterFit = lsfit(x = logT[1:numDays], y=logAfter[i,1:numDays]);

			# GET SUM SQUARES TOTAL
			SStotFore  = sum((logFore[i, 1:numDays] - mean(logFore[i, 1:numDays]))^2)
			SStotAfter = sum((logAfter[i,1:numDays] - mean(logAfter[i,1:numDays]))^2)

			# GET TOTAL ERROR TERMS
			SSerrFore  = sum(foreFit$resid^2)
			SSerrAfter = sum(afterFit$resid^2)

			# R Squared = 1 - (SSE / SST)
			tempRF = 1 - (SSerrFore/SStotFore);
			tempRA = 1 - (SSerrAfter/SStotAfter);

			# IS FORESHOCK FOR THIS TIME INTERVAL BETTER?
			if (tempRF > rsquare[i,1]) {
				rsquare[i,1]    = tempRF;
				signatures[i,1] = foreFit$coef[2];
			}
			# IS AFTERSHOCK FOR THIS TIME INTERVAL BETTER?
			if (tempRA > rsquare[i,2]){
				rsquare[i,2]    = tempRA;
```

```
                        signatures[i,2] = afterFit$coef[2];
                  }
            }
      }
}
# FLIP SIGNS TO REPRESENT AS EXPONENT IN DENOMINATOR
signatures = -signatures

# NOW GET THOSE THAT FIT WELL (above threshold)
goodSignatures = matrix(nrow=numPeaks, ncol=2)
fittedDecays   = matrix(ncol=daysForPeak, nrow=numPeaks)
fittedFore     = matrix(ncol=daysForPeak, nrow=numPeaks)
fittedPeaks    = matrix(ncol=daysForPeak, nrow=numPeaks)
# REMEMBERS ARTIST AND PEAK LOCATION
goodDecays     = matrix(nrow=numPeaks, ncol=2)
goodFore       = matrix(nrow=numPeaks, ncol=2)
goodFits       = matrix(nrow=numPeaks, ncol=2)
# KEEP TRACK OF HOW MANY GOOD FITS THERE ARE
numDecays      = 0;
numFore        = 0;
numFits        = 0;

for (i in 1:numPeaks)
{
      if (rsquare[i,2] > afterThreshold)
      {
            # Decay fit is good enough
            numDecays                   = numDecays + 1;
            fittedDecays[numDecays,]    = peaksData[i,];
            goodDecays[numDecays,]      = peaks[i,];
            goodSignatures[numDecays,1] = signatures[i,1];
            goodSignatures[numDecays,2] = signatures[i,2];
      }
      if (rsquare[i,1] > foreThreshold)
      {
            # Acceleration fit is good enough
            numFore              = numFore + 1;
            fittedFore[numFits,] = peaksData[i,];
            goodFore[numFits,]   = peaks[i,];
      }
      if (rsquare[i,1] > foreThreshold && rsquare[i,2] > afterThreshold)
      {
            # Both fits are good enough
            numFits               = numFits + 1;
            fittedPeaks[numFits,] = peaksData[i,];
            goodFits[numFits,]    = peaks[i,];
      }
}
# SHORTEN DOWN
goodDecays = goodDecays[1:numDecays,]
goodFore   = goodFore[1:numFore,]
goodFits   = goodFits[1:numFits,]

##########################################################
######### PERFORM LOG-LINEAR FIT AND REGRESSION ##########
# Code here
# Removed for brevity. Similar to above with log taken of views only

##########################################################
# FIGURE OUT HOW LONG THE SHOCKS LAST

# GO THROUGH EACH PEAK
daysToRelax = matrix(0, ncol=73);
for (i in 1:numPeaks)
{
      # GET BASELINE FOR THIS DATA SET AS AVERAGE OF 2 WEEKS BEFORE
      baseLine = mean(peaksData[i,1:7]);
      gotDays = FALSE;

      # GO THROUGH EACH DATA POINT AFTER PEAK
```

```
        for (j in (criticalDay+1):length(peaksData[1,]))
        {
                # IF WITHIN 10% OF BASELINE, CONSIDER IT 'RELAXED'
                if (peaksData[i,j] < 1.1*baseLine && !gotDays) {
                        daysToRelax[i] = j - criticalDay;
                        gotDays = TRUE;
                }
        }
}
numPeaksDidNotRelax = 0;
numPeaksWithinWeek1 = 0;
numPeaksWithinWeek2 = 0;
relaxationProps = matrix(0, nrow=3);
for (i in 1:numPeaks) {
        if      (daysToRelax[i] == 0) { numPeaksDidNotRelax = numPeaksDidNotRelax + 1; }
        else if (daysToRelax[i] < 8)  { numPeaksWithinWeek1 = numPeaksWithinWeek1 + 1; }
        else                          { numPeaksWithinWeek2 = numPeaksWithinWeek2 + 1; }
}
relaxationProps[1] = numPeaksDidNotRelax/73;
relaxationProps[2] = numPeaksWithinWeek1/73;
relaxationProps[3] = numPeaksWithinWeek2/73;
```

Plots.r creates all the plots found in this paper. After ParseData.r has been run, all the variables necessary for each plot have been created.

```
##############################################################
# Mark Ungerer                                               #
# Princeton University                                       #
# Operations Research and Financial Engineering              #
# mungerer@princeton.edu                                     #
#                                                            #
# Filename: Plots                                            #
# Description: Plots data in order that plots appear in paper#
#                                                            #
# Warning: must have run ParseData                           #
##############################################################

# Plots all graphs 32 at a time
par(mfrow=c(4,8))
for (i in 1:32) { plotTitle = findArtist(i); plot(allDaily[,i], type="l", main=plotTitle) }
par(mfrow=c(4,8))
for (i in 33:64) { plotTitle = findArtist(i); plot(allDaily[,i], type="l", main=plotTitle) }
par(mfrow=c(4,8))
for (i in 65:96) { plotTitle = findArtist(i); plot(allDaily[,i], type="l", main=plotTitle) }
par(mfrow=c(4,8))
for (i in 97:128) { plotTitle = findArtist(i); plot(allDaily[,i], type="l", main=plotTitle) }
par(mfrow=c(4,8))
for (i in 129:150) { plotTitle = findArtist(i); plot(allDaily[,i], type="l", main=plotTitle) }
par(mfrow=c(4,8))
for (i in 151:180) { plotTitle = findArtist(i); plot(allDaily[,i], type="l", main=plotTitle) }

# Plot weekly graphs
par(mfrow=c(4,8))
for (i in 1:32) { plotTitle = findArtist(i); plot(dataByWeeks[,i], type="l", main=plotTitle) }
par(mfrow=c(4,8))
for (i in 33:64) { plotTitle = findArtist(i); plot(dataByWeeks[,i], type="l", main=plotTitle) }
par(mfrow=c(4,8))
for (i in 65:96) { plotTitle = findArtist(i); plot(dataByWeeks[,i], type="l", main=plotTitle) }
par(mfrow=c(4,8))
for (i in 97:128) { plotTitle = findArtist(i); plot(dataByWeeks[,i], type="l", main=plotTitle) }
par(mfrow=c(4,8))
for (i in 129:150) { plotTitle = findArtist(i); plot(dataByWeeks[,i], type="l", main=plotTitle) }
par(mfrow=c(4,8))
for (i in 151:180) { plotTitle = findArtist(i); plot(dataByWeeks[,i], type="l", main=plotTitle) }

###############################################################
# FIGURE 1.1
# Daily Internet Usage Statistics
x <- seq(0,23)
y <- c(1.295, 2.02, 2.8, 4.8, 7.81, 8.97, 9.23, 9.1, 7.5, 8.81, 9, 8.9, 7.6, 5.5, 4.7, 4.6, 5.75,
5.46, 5.23, 4.23, 3.2, 2.1, 1.65, 1.7)
y_norm <- y / sum(y)
day <- allData[1:48,180]
x2 <- seq(0,23.5, by=0.5)
day_norm <- day / sum(day)
# Plot
par(mfrow=c(1,2))
plot(x, y_norm, xlab="Time", ylab="Usage (Normalized)", main="2004 Internet Usage By Hour", type="l")
plot(x2, day_norm, xlab="Time", ylab="Views (Normalized)", main="January 5, 2009 Myspace Profile
Views", type="l")
par(mfrow=c(1,1))

###############################################################
#                       CHAPTER 2                             #
###############################################################
# Power Law
x = seq(.5,10,.1)
y = 1000*x^(-0.4)
theLogX = log(x)
theLogY = log(y)
par(mfrow=c(1,2))
```

```
plot(x,y, main="Power Law Between X and Y (exp = -0.4)")
plot(theLogX, theLogY, main="Log-Log Relationship between X and Y")
abline(lsfit(theLogX, theLogY))
par(mfrow=c(1,1))


###############################################################
#                       CHAPTER 4                              #
###############################################################
# FIGURE 4.0 (unlabeled, 2x5)
par(mfrow=c(2,5))
plot(Daily.ts[,findFile('brucespringsteen')])
plot(Daily.ts[,findFile('u2')])
plot(Daily.ts[,findFile('ladygaga')])
plot(Daily.ts[,findFile('rihanna')])
plot(Daily.ts[,findFile('carolinaliar')])
plot(Daily.ts[,findFile('mileycyrus')])
plot(Daily.ts[,findFile('britneyspears')])
plot(Daily.ts[,findFile('trapmuzik')])
plot(Daily.ts[,findFile('falloutboy')])
plot(Daily.ts[,findFile('kellyclarkson')])
par(mfrow=c(1,1))

# Half Hour Time Series
plot(HalfHour.ts[,180])    # total
plot(HalfHour.ts[,122])    # britney spears
plot(HalfHour.ts[,71])     # u2
plot(HalfHour.ts[,77])     # kelly clarkson
plot(HalfHour.ts[,131])    # rihanna

##############################################################
# FIGURE 4.1 (1x2)
# Peak detection(k), shape, and shape for k = 2.0
par(mfrow=c(1,2))
# K AND TOTAL PEAKS FOUND
kArray <- seq(from=1.1, to=3.0, by = 0.1)
plot(x=kArray, y=numPeaksArray, ylab="Peaks Detected", xlab="Multiple (k)", main="Sensitivity of Peaks
to Height Requirement",ylim=c(-10,1400))
points(x = kArray, y = shapes[,1], type="l", lty=1)
points(x = kArray, y = shapes[,2], type="l", lty=3)
points(x = kArray, y = shapes[,3], type="l", lty=6)
legend(1.6,670, legend=c("Total Peaks"), pch=1, box.lty=0)
legend(1.6,600, legend=c("Symmetric", "Skewed Left", "Skewed Right"), lty=c(1,3,6), bg="#EEEEEE")

# PEAK FRACTIONS FOR k = 2.0
plot(x = peakFractions[1:numPeaks,2], y = peakFractions[1:numPeaks,1], ylab="fb", xlab="fa",
xlim=c(0,1), ylim=c(0,1))
title(main=paste('Fraction of Views on Peak Day, k = ',k))
cutoffLine = k / criticalDay;
abline(a =  0.0125, b = 1, lty=3)
abline(a = -0.0125, b = 1, lty=3)
abline(h = cutoffLine, lty=3)
abline(v = cutoffLine, lty=3)
par(mfrow=c(1,1))

##############################################################
# Figure 4.2 Time to Relax
par(mfrow=c(1,2))
hist(daysToRelax[daysToRelax>0], main="Number of Days To Relax After Peak", xlab="Number of Days",
ylim=c(0,20))
thenames = c("Did Not Relax", "Within 1 week", "Within 2 weeks")
pie(relaxationProps, labels=thenames, main="Breakdown of Time To Relax")

##############################################################
# Figure 4.3 The process
# Should be edited in abode illustrator
par(mfrow=c(2,2))
# 1. Scan all artist data
plot(normDaily[,22], type="l", main="Step 1. Scan All Artist Data", xlab="Time (Days)", ylab="Views
(Normalized)", ylim=c(0,.045))
# 2. Identify Peaks
```

```
plot(peaksData[12,], main="Step 2. Examine Peaks", xlab="Time (Days)", ylab="Views (Normalized)")
# 3. Look at loglog regression of decay
plot(x=logT, y=logAfter[12,], main="Step 3. Perform Log-Log Regression", xlab="Log(Time)",
ylab="Log(Normalized Views)")
abline(lsfit(logT, logAfter[12,]))
# 4. Determine signature (-0.6265548) and rsquare (0.9308366)
plot(0,0, main="Step 4. Extract Signature and R-Squared")
par(mfrow=c(1,1,))

##############################################################
# FIGURE 4.4 (1x3 histograms)
# Foreshock and Aftershock Signatures and R Squared Values
plotLogLogResults <- function()
{
	histYHeight = 25
	par(mfrow=c(1,3))

	# HISTOGRAM 1 - Foreshock Signatures
	hist(signatures[,1], main="Foreshock Signature", xlab="Slope of log-log Regression", col="blue",
density=50, ylim=c(0,histYHeight), breaks=10)
	tempMedian = round(median(signatures[,1]), 2)
	text = paste("Median = ", tempMedian)
	abline(v=tempMedian, lty=3, col="blue", lwd=3)
	legend(1.8, 20, box.lty=0, legend=c(text))

	# HISTORGRAM 2 - Aftershock Signatures
	hist(signatures[,2], main="Aftershock Signature", xlab="Slope of log-log Regression", col="red",
density=50, ylim=c(0,histYHeight), breaks=10)
	tempMedian = round(median(signatures[,2]), 2)
	text = paste("Median = ", tempMedian)
	abline(v=tempMedian, lty=3, col="red", lwd=3)
	legend(1.8, 20, box.lty=0, legend=c(text))

	# HISTOGRAM 3 - Double layered histogram of R Squared Values
	hist(rsquare[,2], main="", xlab="", col="red", density=20, xlim=c(0,1), ylim=c(0,histYHeight),
breaks=10)
	par(new=T)
	hist(rsquare[,1], main="R-Squared Values", xlab="Log-Log R-Squared Value", col="blue",
density=20,  xlim=c(0,1), ylim=c(0,histYHeight), breaks=10)
	legend(0.1, 20, legend = c("Aftershock", "Foreshock"), col = c("red", "blue"), pch=c(15,15),
pt.cex=2, lty = c(0, 0))
	abline(v=.85, lty=1, lwd=1)
	par(mfrow=c(1,1))
}
plotLogLogResults()


##############################################################
# FIGURE 4.5 (4x8)
# Example of peaks found, plot first 32
par(mfrow=c(4,8))
for (i in 1:32)
{
	start  = peaks[i,2]-7;
	end    = start + daysForPeak - 1;
	artist = findArtist(peaks[i,1]);
	x      = seq(from=start, to=end, by=1);
	plot(x, peaksData[i,],type="l", xlab="Days", ylab="Views", main=artist)
}

##############################################################
# FIGURE 4.6 (1x2)
# Plot of r squared pairings and signature pairings
plotRSquardAndSignaturePairings <- function()
{
	par(mfrow=c(1,3))
	plot(rsquare[,1], rsquare[,2], main="R Squared Values", xlab="R Squared for Foreshock", ylab="R
squared for Aftershock")
	abline(v=.8)
	abline(h=.8)
```

```
	plot(signatures[,1], signatures[,2], main="Peak Signatures", xlab="Foreshock Signature", 
ylab="Aftershock Signature")
	abline(a=0, b=1)
	# Number of peaks found and how many had good regressions
	plot(x=1, pch=1, lwd=2, numPeaks, main="Number of Peaks Found with Good Regressions", xlab="", 
ylab="Number of Peaks Found", xlim=c(0,5), ylim=c(0,100))
	par(new=T)
	plot(x=2, pch=2, lwd=2, numDecays, main="Number of Peaks Found with Good Regressions", xlab="", 
ylab="Number of Peaks Found", xlim=c(0,5), ylim=c(0,100))
	par(new=T)
	plot(x=3, pch=3, lwd=2, y=length(rsquare[rsquare[,1]>.8,1]), main="Number of Peaks Found with 
Good Regressions", xlab="", ylab="Number of Peaks Found", xlim=c(0,5), ylim=c(0,100))
	par(new=T)
	plot(x=4, pch=4, lwd=2, numFits, main="", xlab="", ylab="", xlim=c(0,5), ylim=c(0,100))
	legend(1.75, 100, legend=c("Total Peaks (83)", "Decay Fits (38)", "Acceleration Fits (21)", 
"Both Fits(8)"), pch=c(1,2,3,4))
	par(mfrow=c(1,1))
}
plotRSquardAndSignaturePairings()

#################################################################
# FIGURE 4.7 (1x3) Histograms
# Now for the log-linear Regressions
logLinearResultsPlot <- function()
{
	par(mfrow=c(1,3))
	histYHeight = 25
	# HISTOGRAM 1 - Foreshock Signatures
	hist(llSignatures[,1], main="Foreshock Signature", xlab="Slope of log-LINEAR Regression", 
col="blue",  density=50)
	tempMedian = round(median(llSignatures[,1]), 2)
	text = paste("Median = ", tempMedian)
	abline(v=tempMedian, lty=3, col="blue", lwd=3)
	legend(-1.8, 29, box.lty=0, legend=c(text))

	# HISTOGRAM 2 - Aftershock Signatures
	hist(llSignatures[,2], main="Aftershock Signature", xlab="Slope of log-LINEAR Regression", 
col="red",   density=50)
	tempMedian = round(median(llSignatures[,2]), 2)
	text = paste("Median = ", tempMedian)
	abline(v=tempMedian, lty=3, col="red", lwd=3)
	legend(-1.8, 29, box.lty=0, legend=c(text))

	# HISTOGRAM 3 - Double layered histogram of R-squared values
	hist(llRsquare[,1], main="R-Squared Values", xlab="Log-LINEAR R Squared Value", col="blue", 
density=20, xlim=c(0,1), ylim=c(0,histYHeight))
	par(new=T)
	hist(llRsquare[,2], main="", xlab="Log-LINEAR R-Squared Value", col="red", density=20, 
xlim=c(0,1))
	legend(0.1, 28, legend = c("Aftershock", "Foreshock"), col = c("red", "blue"), pch=c(15,15), 
pt.cex=2, lty = c(0, 0))
	par(mfrow=c(1,1))
}
logLinearResultsPlot()

##############################################################
# FIGURE 4.8 (one big plot)
# Plot all artists with Release Date Aligned

par(mfrow=c(1,2))
n = length(critical[,1]);
scaleFactor = 90/(90-57);

plot(x=seq(-critical[1,2],(M_daily-critical[1,2]-1)),y=(allNorm[,critical[i,1]]/scaleFactor), 
ylim=c(0,.225), xlim=c(-75,75), xlab="Days (0 = Album Release Date)", ylab="Normalized Views", 
main="Artists with Albums Released During Tracking Period", type="l")

for (i in 1:n)
{
	start = -1*critical[i,2];
	end = M_daily + start - 1;
```

```
        x = seq(start, end);
        if (i > 77){ y = allNorm[,critical[i,1]]/scaleFactor;       }
        else{ y = allNorm[,critical[i,1]]; }

        points(x,y, type="l", lty=i)
}

# Plot to see if peaks correlate with album releases
x <- matrix(ncol = numPeaks, nrow=1)
for (i in 1:numPeaks) { x[i] = daysPast[peaks[i,1]]; }
plot(x, y = peaks[1:numPeaks,2], xlab = "Day of Album Release", ylab = "Day of Peak", main = "Peak
Days and Day of Album Release", xlim=c(0,100), ylim=c(0,100))
abline(a=0, b=1, lty=3)
lsReleasePeaks <- lsfit(x = peaks[1:45,1], y=peaks[1:45,2])
abline(lsReleasePeaks)
par(mfrow=c(1,1))

##############################################################
# FIGURE 4.9
# Rihanna and Chris Brown Plot (along with Google Trends)
# Edited to form 1 graph in photoshop
par(mfrow=c(1,2))
plot(allDaily$rihanna, type="l", main="Rihanna and Chris Brown", xlab="Days", ylab="Profile Views per
Day", ylim=c(0,150000), col="blue", lwd=2)
points(allDaily$chrisbrown, type="l", col="red", lwd=2)
legend("bottomleft", legend=c("Rihanna", "Chris Brown"), box.lty=0, col=c("blue", "red"), lty=1)
plot(allDaily$carolinaliar, main="Carolina Liar", type="l", col="blue", lwd=2, ylim=c(0,15000))
par(mfrow=c(1,1))


############################################################
# Good signatures histogram
par(mfrow=c(1,2))
hist(goodSignatures[,2], main="Aftershock Signature from Good Fits", breaks=9)
abline(v=round(median(goodSignatures[1:numDecays,2]),2))
hist(goodSignatures[,1], main="Foreshock Signature from Good Fits", breaks=9)
abline(v=round(median(goodSignatures[1:numDecays,1]),2))
par(mfrow=c(1,1))

##############################################################
# Plot the peaks with a good power law fit
par(mfrow=c(4,8))
for (i in 1:32) {
        artistString = findArtist(goodDecays[i])
        plot(fittedDecays[i,], type="l", xlab="Days", ylab="Views", main = artistString)
}
par(mfrow=c(1,1))

##############################################################
# FIGURE 4.9
# Time series Analysis for Half Hour Data
# Plot seasonality (commented out because no plotting here)
par(mfrow=c(2,2))
plot(HalfHour.ts[,180])
legend("topright", legend=c("Original Data"), lty=1, box.lty=0)
plot(hh.stl$sea)
legend("topright", legend=c("Seasonal Component"), lty=1, box.lty=0)
plot(hh.stl$trend)
legend("bottomright", legend=c("Trend Component"), lty=1, box.lty=0)
plot(hh.stl$rem)
legend("topright", legend=c("Remainder Component"), lty=1, box.lty=0)
par(mfrow=c(1,1))
```

MTV.r reads in the artist lists, plots them, and analyzes them for trends. It cross references the data with profile views so ParseData.r must have been executed first.

```
###########################################################
# Mark Ungerer
# Filename: MTV.r
# Plots MTV promoted artists, new artists, and most popular and compares them to their Myspace profile
# views over the same time period.
############################################################

# GET RAW DATA SCRAPED FROM MTV.COM
setwd("/Users/mungerer/Dropbox/Thesis/Data and Papers/MTV")
mostPopular <- read.table("mostpopular.txt", sep = ",", header=TRUE, stringsAsFactors = FALSE,
strip.white = TRUE)
newArtists <- read.table("newArtists.txt", sep = ",", header=TRUE, stringsAsFactors = FALSE,
strip.white = TRUE)
artistPicks <- read.table("artistpicks.txt", sep = ",", header=TRUE, stringsAsFactors = FALSE,
strip.white = TRUE)
M_MTV = length(mostPopular[,1])
M_MTV = length(newArtists[,1])
M_MTV = length(artistPicks[,1])

# ADD LIST OF ARTISTS TO CHECK THROUGH (must match spelling in MTV files)
# MOST POPULAR (MP)
MPList = c('lilwayne', 'beyonce', 'britneyspears', 'trapmuzik', 'chrisbrown', 'rihanna', 'akon',
'pussycatdolls', 'kanyewest', 'tpain', 'pink', '50cent', 'jonasbrothers', 'notoriousbig', 'katyperry',
'eminem', 'paramore', 'mileycyrus', 'ladygaga', 'christinaaguilera', 'pornoforpyros', 'blink182',
'souljaboytellem', 'justintimberlake', 'taylorswift', 'kellyclarkson', 'MIA', 'michaeljackson',
'ryanadams')
MPMyspace = c('lilwayne', 'beyonce', 'britneyspears', 'trapmuzik', 'chrisbrown', 'rihanna', 'akon',
'kanyewest', 'tpain', '50cent', 'jonasbrothers', 'katyperry', 'paramore','mileycyrus', 'ladygaga',
'christinaaguilera', 'justintimberlake', 'taylorswift', 'kellyclarkson')
MPCorr      = c(1,2,3,4,5,6,7,9,10,12,13,15,17,18,19,20,21,22,23)

# NEW ARTIST DATA (NA)
NAList      = c('thecab', 'santogold', 'morningwood', 'shwayze', 'thewombatsuk', 'carolinaliar',
'donnieklang', 'jazminesullivan', 'wethekings', 'maydayparade', 'asherroth', 'chesterfrench',
'kidcudi', 'ladygaga', 'thetingtings', 'kuroma', 'ronbrowz', 'gsboyz', 'colinmunroe',
'whitetieaffair', 'electrikred', 'ojdajuiceman', 'djeightyeightkeys')
NAMyspace   = c('thecab', 'morningwood', 'thewombatsuk', 'carolinaliar', 'donnieklang',
'jazminesullivan', 'wethekings', 'maydayparade', 'ladygaga', 'thetingtings')
NACorr      = c(1,3,5,6,7,8,9,10,14,15)

# MTV's OWN ARTIST PICKS (AP)
APList      = c('pink', 'lilwayne', 'kidrock', 'johnlegend', 'coldplay', 'neyo', 'beyonce',
'trapmuzik', 'kellyclarkson', 'allamericanrejects', 'thekillers', 'u2', 'thegame', 'solange',
'gymclassheroes', 'llcoolj', 'plainwhitets', 'katyperry', 'thefray', 'kanyewest', 'ciara',
'britneyspears', 'mileycyrus', 'jamiefoxx')
APMyspace   = c('lilwayne', 'kidrock', 'johnlegend', 'coldplay', 'neyo', 'beyonce', 'trapmuzik',
'kellyclarkson', 'allamericanrejects', 'u2', 'thegame', 'solange', 'gymclassheroes', 'llcoolj',
'plainwhitets', 'katyperry', 'thefray', 'kanyewest')
# Index of artists on APlist that I am tracking views
APCorr      = c(2,3,4,5,6,7,8,9,10,12,13,14,15,16,17,18,19,20,22,23)

####################################################################
# FUNCTION getRankings
# Creates a matrix with each column corresponding to one artist
# and each row to the place they were on that day
# i.e. a value of 5 in row 2 means they were 5th most popular on day 2)
getRankings <- function(list, rawRankings)
{
	N        = length(list)
	M_MTV    = length(rawRankings[,1])
	M_List   = length(rawRankings[1,])
	rankings = matrix(ncol = N, nrow = M_MTV)
	for (i in 1:N){
		currentArtist = list[i];
		for (j in 1:M_MTV) {
			rankings[j,i] = 100;
```

```
                    for (k in 1:M_List) {
                          if (currentArtist == rawRankings[j,k]) {
                                rankings[j,i] = k;
                          }
                    }
              }
       }
       return(rankings)
}

#####################################################################
# FUNCTION: getDifferences. TAKES LIST OF ARTIST AND THEIR RANKINGS
# AND CALCULATES AVERAGE VIEWS WHILE ON LIST AND WHILE NOT
getDifferences <- function(XXMyspace, XXCorr, XXRankings) {
       # CREATE NECESSARY VARIABLES
       numXXMyspace = length(XXMyspace);
       XXAvg        = matrix(0, nrow = numXXMyspace, ncol=3)

       # LOOP THROUGH ALL ARTISTS
       for (i in 1:numXXMyspace)
       {
              # GET DATA
              views   = allDaily[23:M_daily,findFile(XXMyspace[i])]; # 23 less MTV data points
              rank    = XXRankings[,XXCorr[i]];
              iRank   = 0; iNoRank   = 0;
              sumRank = 0; sumNoRank = 0;

              # LOOP THROUGH EACH DAY
              for (j in 1:M_MTV)
              {
                    # RANK BELOW 20 MEANS NOT ON LIST
                    if (rank[j] > 20) {
                          iNoRank   = iNoRank + 1;
                          sumNoRank = sumNoRank + views[j];
                    } else {
                          iRank   = iRank + 1;
                          sumRank = sumRank + views[j];
                    }
              }
              # MAKE SURE NOT TO DIVIDE BY 0!
              if (iRank   != 0) { XXAvg[i,1] = (sumRank / iRank) ;      } # Rank
              if (iNoRank != 0) { XXAvg[i,2] = (sumNoRank / iNoRank) ;  } # No Rank
              if (iRank   != 0 && iNoRank != 0)
                               { XXAvg[i,3] = XXAvg[i,1] - XXAvg[i,2]; } # Difference
       }
       # GET RID OF ZERO ENTRIES
       temp  = matrix(ncol=3, nrow=numXXMyspace); index = 0;
       for (i in 1:numXXMyspace) {
              if (XXAvg[i,3] != 0){ index = index + 1; temp[index,] = XXAvg[i,]; }
       }
       # RETURN SHORTENED VECTOR
       return(temp[1:index,]);
}

###############################################################
#                  CALL THE FUNCTIONS                         #
###############################################################
# Get Rankings in usable form
MPRankings = getRankings(MPList, mostPopular);
NARankings = getRankings(NAList, newArtists);
APRankings = getRankings(APList, artistPicks);

# Get difference between views while on list and while not
MPAvg = getDifferences(MPMyspace, MPCorr, MPRankings);
NAAvg = getDifferences(NAMyspace, NACorr, NARankings);
APAvg = getDifferences(APMyspace, APCorr, APRankings);

# Perform t-test to see if
# Is [(views while listed) - (views while not listed)] > 0?
tTestMP = t.test(MPAvg[,3], conf.level=.95, alternative="g");
tTestNA = t.test(NAAvg[,3], conf.level=.95, alternative="g");
```

```
tTestAP = t.test(APAvg[,3], conf.level=.95, alternative="g");
tTestResults = cbind(c(mean(MPAvg[,3]), mean(NAAvg[,3]), mean(APAvg[,3])), c(tTestMP$p.value,
tTestNA$p.value, tTestAP$p.value));
colnames(tTestResults) = c("Mu", "P-Value");

################################################################
#                           PLOTS                              #
################################################################
# Figure 5.1 (2x2)
# Plot of MTV's artist lists over time
# Needs photoshop to complete bottom right corner with information
par(mfrow=c(2,2))
plot(x = seq(1,M_MTV), y=MPRankings[,1], ylim = c(21,1), xlim=c(0, M_MTV), xlab="Time (Days)",
ylab="Ranking", main="MTV Most Popular Artists", type="l", lty=1, col='red')
for (i in 2:length(MPRankings[1,])) {
       if (i %% 5 == 1) { pColor = 'red'; }
       if (i %% 5 == 2) { pColor = 'black'; }
       if (i %% 5 == 3) { pColor = 'blue'; }
       if (i %% 5 == 4) { pColor = 'green'; }
       if (i %% 5 == 0) { pColor = 'orange'; }
       points(x=seq(1,M_MTV), y=MPRankings[,i], type="l", col=pColor, lty=i)
}
plot(x = seq(1,M_MTV), y=NARankings[,1], ylim = c(11,1), xlim=c(0, M_MTV), xlab="Time (Days)",
ylab="Placement", main="MTV New Artists", type="l", lty=1)
for (i in 2:length(NARankings[1,])) {
       if (i %% 5 == 1) { pColor = 'black'; }
       if (i %% 5 == 2) { pColor = 'red'; }
       if (i %% 5 == 3) { pColor = 'blue'; }
       if (i %% 5 == 4) { pColor = 'green'; }
       if (i %% 5 == 0) { pColor = 'orange'; }
       points(x=seq(1,M_MTV), y=NARankings[,i], type="l", col=pColor, lty=i)
}
plot(x = seq(1,M_MTV), y=APRankings[,1], ylim = c(11,1), xlim=c(0, M_MTV), xlab="Time (Days)",
ylab="Placement", main="MTV Artist Picks (Promoted Artists)", type="l", lty=1)
for (i in 2:length(APRankings[1,])) {
       if (i %% 5 == 1) { pColor = 'black'; }
       if (i %% 5 == 2) { pColor = 'red'; }
       if (i %% 5 == 3) { pColor = 'blue'; }
       if (i %% 5 == 4) { pColor = 'green'; }
       if (i %% 5 == 0) { pColor = 'orange'; }
       points(x=seq(1,M_MTV), y=APRankings[,i], type="l", col=pColor, lty=i)
}
par(mfrow=c(1,1))

###############################################################
# FIGURE 5.2
# PLOT MYSPACE VIEWS AND RANKINGS FOR EACH ARTIST ON SAME PLOT
# FUNCTION: plotMyspaceAndRankings
plotMyspaceAndRankings <- function(rows, cols, artistList, rankings, corr, yHeightLimit)
{
       par(mfrow=c(rows,cols));
       for (i in 1:length(artistList))
       {
              if (yHeightLimit != 0) {
                     plot(allDaily[,findFile(artistList[i])], ylab="Views", main=artistList[i],
                     xlab="Time (Days)", type="l", ylim=c(0,yHeightLimit), lwd=2)
           } else {
              plot(allDaily[,findFile(artistList[i])], ylab="Views", main=artistList[i],
                     xlab="Time (Days)", type="l", lwd=2)
           }
              par(new=TRUE)
              plot(rankings[,corr[i]], ylab="", xlab="", main="", type="l", lty=3, axes=FALSE,
ylim=c(20,0))
              axis(side=4,20:1, ylab="Ranking")
       }
}
# MOST POPULAR ARTISTS
plotMyspaceAndRankings(3,7, MPMyspace, MPRankings, MPCorr, 200000) # MOST POPULAR
plotMyspaceAndRankings(2,5, NAMyspace, NARankings, NACorr, 0) # NEW ARTISTS
plotMyspaceAndRankings(3,6, APMyspace, APRankings, APCorr, 0) # MTVS ARTIST PICK
```

albumSales.r reads in and analyzes the sales data purchased from Nielsen Soundscan. It then uses the weekly profile views data generated in ParseData.r to calculate the correlations and estimate multiples between sales and views.

```
###################################################
# Mark Ungerer
# Princeton University
# Operations Research and Financial Engineering
# mungerer@princeton.edu
#
# Filename: albumSales.r
#
# Description:
# - Reads in album sale data for 6 artists
# - Plots and regresses on data
#
###############################################################

##########################################################
# IF HAVEN'T RUN ParseData.r IMPORT CARMONA'S LIBRARY
#library(Rsafd)

# GET DATA
setwd("/Users/mungerer/Dropbox/Thesis/Data and Papers/Data Final")
albumSales <- read.table("albumSales.txt", header=FALSE, sep=",", strip.white=TRUE)
M_sales <- length(albumSales[1,])
artistNames <- c('Bruce Springsteen', 'U2', 'Carolina Liar', 'Lady Gaga', 'Rihanna', 'Miley Cyrus')
albumArtists <- c('brucespringsteen', 'u2', 'carolinaliar', 'ladygaga', 'rihanna', 'mileycyrus')

##########################################################
# CALCULATE R SQAURED VALUE ON A SET OF DATA
Rsquared <- function(X,Y) {
	fit    = lsfit(X, Y);
	SStot  = sum((Y - mean(Y))^2);
	SSerr  = sum(fit$resid^2)
	R2 = 1 - (SSerr/SStot);
	return(R2);
}

##########################################################
# GET DATA NORMALIZED AND IN USEFUL FORMAT
asaViews   = matrix(ncol=6, nrow=12);
asaSales   = matrix(ncol=6, nrow=12);
nasaViews = matrix(ncol=6, nrow=12);
nasaSales = matrix(ncol=6, nrow=12);
cors       = matrix(nrow=6, ncol=1);
Rs         = matrix(nrow=6, ncol=1);
vsSlopes   = matrix(nrow=6, ncol=1);
vsIntercepts = matrix(nrow=6, ncol=1);
for (i in 1:6)
{
	# GET DATA IN NICE VECTOR
	asaViews[,i]  = dataByWeeks[1:12,findFile(albumArtists[i])];
	asaSales[,i]  = t(albumSales[i,2:13]);

	# NORMALIZE
	nasaViews[,i] = asaViews[,i] / sum(asaViews[,i]);
	nasaSales[,i] = asaSales[,i] / sum(asaSales[,i]);

	# GET CORRELATION AND R-SQUARED
	X = nasaViews[,i]; Y = nasaSales[,i];
	cors[i] = cor(X, Y);
	Rs[i] = Rsquared(X,Y);
	vsSlopes[i] = lsfit(X, Y)$coef[2];
	vsIntercepts[i] = lsfit(X,Y)$coef[1];
}

###############################################
#    GET VIEWS AND SALES WITHOUT OUTLIERS     #
###############################################
```

```
viewsAndSales = matrix(0, nrow=(12*6-3), ncol=2);
nviewsAndSales = matrix(0, nrow=(12*6-3), ncol=2);

# LOOP THROUGH
index = 1;
skipped = '';
for (i in 1:6)
{
	for (j in 1:12)
	{
		# SKIP DATA POINT 4 WHEN i = 1
		if ((i == 1 && j == 4) || (i == 2 && (j==7 || j==8))){
			skipped = paste(skipped, '(', i, ',', j, ') ');
		}
		else {
			nviewsAndSales[index,1] = nasaViews[j,i];
			nviewsAndSales[index,2] = nasaSales[j,i];
			viewsAndSales[index,1]  = asaViews[j,i];
			viewsAndSales[index,2]  = asaSales[j,i];
			index = index + 1;
		}
	}
}

# GET CORRELATION AND EQUATION
nviewsAndSalesCor = cor(nviewsAndSales[,1], nviewsAndSales[,2]);
nviewsAndSalesFit = lsfit(nviewsAndSales[,1], nviewsAndSales[,2]);
a = nviewsAndSalesFit$coef[1];
b = nviewsAndSalesFit$coef[2];

# GET THE MULTIPLES BETWEEN SALES AND VIEWS RAW
nviewsAndSalesMult = nviewsAndSales[,1] / nviewsAndSales[,2];
viewsAndSalesMult = viewsAndSales[,1] / viewsAndSales[,2];
navgMult = mean(nviewsAndSalesMult);
t = viewsAndSalesMult[viewsAndSalesMult>10]
avgMult  = mean(t[t<50]);
newAvgMult = mean(viewsAndSalesMult[1:21])

###################################################
# ESTIMATE SALES FOR A FEW ARTISTS
salesEstimates <- matrix(0, nrow=3, ncol=11);
viewsData <- matrix(0, nrow=1, ncol=10);
for (i in 1:3)
{
	index = 0;
	if (i == 1) { index = 1.0; viewsData = dataByWeeks[,findFile('britneyspears')]; }
	if (i == 2) { index = 1.0; viewsData = dataByWeeks[,findFile('chrisbrown')]; }
	if (i == 3) { index = 0.9; viewsData = dataByWeeks[,findFile('thefray')]; }
	for (j in 1:10)
	{
		salesEstimates[i,j] = viewsData[j] / (3 + 27*index);
	}
}
for (i in 1:3) salesEstimates[i,11] = sum(salesEstimates[i,1:10]);

###################################################
# FIGURE OUT HOW MUCH EXTRA RIHANNA MADE AFTER ABUSE
rihannaSales <- t(albumSales[5,])
rihannaSalesNoAbuse <- t(albumSales[5,])

# CALCULATE HOW MUCH SHE LOST BY LINEAR COMPARISON
slope = (rihannaSales[6] - rihannaSales[13]) / 7;
for (i in 1:6) { rihannaSalesNoAbuse[6+i] = rihannaSales[6] - (i*slope); }
diffInSales <- rihannaSales - rihannaSalesNoAbuse;
totalDiff <- sum(diffInSales)

# PLOT
plot(rihannaSales, type = 'l', main='Rihanna Weekly Sales', xlab="Week", ylab="Sales")
points(rihannaSalesNoAbuse, type='l', lty=3)
legend('topright', legend=c('Actual Sales', 'No Abuse Prediction'), lty=c(1,3))
```

```
####################################################
#                   PLOTS                          #
####################################################

# PLOT THE TOTAL FIT
par(mfrow=c(1,2))
plot(nviewsAndSales[,1], y=nviewsAndSales[,2], xlab="Normalized Views", ylab="Normalized Sales",
xlim=c(0,.35), ylim=c(0,0.35), main="Normalized Aggregate Fit");
text(0.27, 0.25, paste('s =', round(b, digits=2), '*v',round(a, digits=2)));
text(0.27, 0.22, paste('cor = ', round(nviewsAndSalesCor, digits=2)));
abline(nviewsAndSalesFit);

# PLOT THE MULTIPLE
plot(viewsAndSalesMult, xlab="All 6 Artists (Separated by lines)", ylab="Multiple", main="Weekly
Views/Sales Ratios");
text(40,55, paste('mean = ', round(avgMult, digits=2)))
abline(h=10, lty=3); abline(h=50, lty=3);
abline(v=11); abline(v=21); abline(v=33); abline(v=45); abline(v=57)
#############################################################
# PLOT VIEWS AND SALES VERSUS TIME ON 2x6 CHART
for (i in 1:6) {
	plot(x = seq(1, 12), y = asaViews[,i], ylab="Weekly Views", xlab="Time (Weeks)",
main=paste(artistNames[i], ' views'), type="l");
}
for (i in 1:6) {
	plot(x = seq(1, M_sales-1), y = albumSales[i,2:13], xlab="Time (Weeks)", ylab = "Weekly Sales",
main=paste(artistNames[i], ' sales'), type="l");
}

############################################################
# SCATTERPLOT OF VIEWS AND SALES FOR EACH WEEK ON 2x3 CHART
par(mfrow=c(2,3));
for (i in 1:6)
{
	# PLOT AND ADD TRENDLINE
	plot(nasaViews[,i], nasaSales[,i], xlab="Views", ylab="Sales", main=artistNames[i],
xlim=c(0,0.5), ylim=c(0,0.5));
	text(0.38, 0.48, paste('p  = ', round(cors[i], digits=2)));
	text(0.38, 0.44, paste('R2 = ', round(Rs[i],   digits=2)));
	#text(0.38, 0.40, paste('s = ', round(vsSlopes[i], digits=2),
	#                      'v + ', round(vsIntercepts[i], digits=2)));
	abline(lsfit(x=nasaViews[,i], y=nasaSales[,i]));
}
par(mfrow=c(1,1));
```

Appendix A2

This section lists all artists that were tracked for this project, their record label, and the reason they were interesting to study during the January 2009 – April 2009 time period.

A2.1 Full Artist List

The following table lists all the artists that were tracked, their unique Myspace URL, relevant notes on why they were tracked, and their record label. There are 400 total artists. From "Glasvegas" to "Tila Tequila" (Artists 1-180) I tracked daily and half hour views from January 4th to April 1st, 2009. For the rest of the artists, I tracked daily views from March 2 to April 1, 2009.

Table A1.1 Artist List (400 Total)

Artist	Myspace URL	Notes	Label
Glasvegas	glasvegas	Album: 1/06/09	Sony BMG
The Gourds	thegourds	Album: 1/06/09	Unsigned
Erin McCarley	erinmccarley	Album: 1/06/09	Universal
Soundtrack of our Lives	officialtsool	Album: 1/06/09	Warner
Whole Wheat Bread	wholewheatbread	Album: 1/06/09	Fighting Records
Jessie Kilguss	jessiekilguss	Album: 1/09/09	Exotic Bird Recordings
Animal Collective	animalcollectivetheband	Album: 1/13/09	Animal
My Dear Disco	mydeardisco	Album: 1/13/09	Dancethink
Too Pure to Die	toopuretodie	Album: 1/13/09	TrustKill Records
Varsity Fanclub	varsityfanclub	Album: 1/13/09	EMI
Heather Headley	heatherheadley	Album: 1/13/09	EMI
Oh No Not Stereo	ohnonotstereo	Album: 1/13/09	G'fuck Y'self
Illinois	illinois	Album: 1/13/09	+1 Records
Matt York	mattyork	Album: 1/20/09	Rock Ridge Music
Or, The Whale	orthewhale	Album: 1/20/09	Seany Records
Combichrist	combichrist	Album: 1/20/09	Out of Line Records
Antony And the Johnsons	antonyandthejohnsons	Album: 1/20/09	Secretly Canadian
Lisa Hannigan	lisahannigan	Album: 1/20/09	ATO Records
Reel Big Fish	reelbigfish	Album: 1/20/09	Independent
John Frusciante	johnfrusciantemusic	Album: 1/20/09	Record Collection Music
Umphrey's McGee	umphreysmcgee	Album: 1/20/09	SCI Fidelity
Modern Skirts	modernskirts	Album: 1/20/09	Modern Skirts Recording
Andrew Bird	andrewbird	Album: 1/27/09	Fat Possum
Loney Dear	loneydear	Album: 1/27/09	Polyvinyl Records
Richie Booker	richiebookermarley	Album: 1/27/09	Unsigned
Ciara	ciara	Album: 1/27/09	Sony BMG
Lisa "Left Eye" Lopes	lisalopesthegreatest	Album: 1/27/09	Sony BMG
Franz Ferdinand	franzferdinand	Album: 1/27/09	Domino
Duncan Sheik	duncansheik	Album: 1/27/09	Atlantic
Bruce Stringsteen	brucespringsteen	Album: 1/27/09	Sony BMG
Bad Plus	badplus	Album: 2/03/09	Universal
Dent May & His Magnificent Ukulele	dentmay	Album: 2/03/09	Paw Tracks
Two Tongues	twotonguesrock	Album: 2/03/09	Vagrant Records
Melinda Doolittle	melindadoolittle	Album: 2/03/09	Hi Fi Recordngs
The Fray	thefray	Album: 2/03/09	Sony BMG
Bow Wow	bowwow	Album: 2/03/09	Sony BMG
Chris Cornell	chriscornell	Album: 2/03/09	Universal
Von Bondies	vonbondies	Album: 2/03/09	Majordomo Records
Lily Allen	lilymusic	Album: 2/09/09	EMI
Mos Def	mosdef	Album: 2/09/09	Warner
Courtney Love	courtneylove	Album: 2/09/09	Unknown
Ben Lee	benlee	Album: 2/10/09	New West Records
Taylor Hicks	taylorhicks	Album: 2/10/09	Vanguard Records
India Arie	indiaarie	Album: 2/10/09	Universal
Miranda Lee Richards	mirandaleerichards	Album: 2/10/09	Nettwerk Music Group
Busta Rhymes	bustarhymes	Album: 2/10/09	Universal
Missy Elliott	missyelliott	Album: 2/10/09	Warner
Warren G	warreng	Album: 2/10/09	Death Row, Others
Dex Romweber Duo	dexterromweberduo	Album: 2/10/09	Bloodshot
Morrissey	morrissey	Album: 2/16/09	Decca
The Appleseed cast	theappleseedcast	Album: 2/17/09	The Militia Group

Plushgun	plushgun	Album: 2/17/09	Tommy Boy
..And You Will Know Us By…	trailofdead	Album: 2/17/09	Merge Records
Robyn Hitchcock	robynhitchcock	Album: 2/17/09	Yep Roc Records
Tommy Keene	tommykeeneband	Album: 2/17/09	Matador Records
Ana Egge	anaegge	Album: 2/17/09	Grace/Parkinsong
Broken Spindles	brokenspindles	Album: 2/17/09	Blank.wav
Wheels on Fire	wheelsonfire	Album: 2/17/09	Big Legal Mess
Faunts	faunts	Album: 2/17/09	Friendly Fire Recordings
Kovak	kovakuk	Album: 2/17/09	Kovak Records
Seven Caves	sevencaves	Album: 2/23/09	Diffusion Records
Taxi Doll	taxidoll	Album: 2/23/09	Antidote Media
Left Side Brain	leftsidebrain	Album: 2/23/09	Sugar Shack Records
Hatebreed	hatebreed	Album: 2/24/09	Unsigned
Robie	robiemusic	Album: 2/24/09	RobieVille Inc Records
Karl Wilson	karlwilson	Album: 2/28/09	None
Lorien	lorien	Album: 2/28/09	Instant Karma/Virgin
Zack Nichols	zacknichols2	Album: 2/28/09	Unsigned
Neko Case	nekocase	Album: 3/03/09	Anti
Justin Townes Earle	justintownesearle	Album: 3/03/09	Bloodshot
U2	u2	Album: 3/03/09	Universal
Bell X1	bellx1	Album: 3/03/09	BellyUp/YepRoc/Warner
Shortwave Fade	shortwavefade	Album: 3/03/09	Unsigned
Madeleine Peyroux	officialmadeleinepeyroux	Album: 3/10/09	Rounder Records
Handsome Furs	handsomefurs	Album: 3/10/09	Sub Pop Records
MSTRKRFT	mstrkrft	Album: 3/17/09	Dim Mak/ Downtown
Kelly Clarkson	kellyclarkson	Album: 3/17/09	Sony
Straight No Chaser	sncmusic	BB Breaking In	
DJ Nesty	nestysangrenueva	BB Breaking In	
Margot & the Nuclear So So's	margotandthenuclearsoandsos	BB Breaking In	
Department of Eagles	deptofeagles	BB Breaking In	
Laura Izibor	lauraizibor	BB Breaking In	
Brutha	thisisbrutha	BB Breaking In	
Rascal Flatts	rascalflatts	BB Country Hot 10	
Sugarland	sugarland	BB Country Hot 11	
Zac Brown Band	zacbrownband	BB Country Hot 12	
Montgomery Gentry	montgomerygentry	BB Country Hot 13	
Brad Paisley	bradpaisley	BB Country Hot 14	
Alan Jackson	alanjackson	BB Country Hot 15	
Billy Currington	billycurrinton	BB Country Hot 16	
Dierks Bentley	dierksbentley	BB Country Hot 17	
Jamey Johnson	jameyjohnson	BB Country Hot 18	
Blake Shelton	blakeshelton	BB Country Hot 19	
Keith Urban	keithurban	BB Country Hot 20	
Taylor Swift	taylorswift	BB Country Hot 21	
Jim Jones	jimjones	BB Rap Hot 10	
Kings of Leon	kingsofleon	BB Rock Hot 10	
Anberlin	anberlin	BB Rock Hot 11	
Rise Against	riseagainst	BB Rock Hot 12	
Incubus	incubus	BB Rock Hot 13	
Offspring	offspring	BB Rock Hot 14	
Shinedown	shinedown	BB Rock Hot 15	
Entice	enticetheband	Friends	
Them Terribles	themterribles	Friends	
Tegan and Sara	teganandsara	MTV featured	
Carolina Liar	carolinaliar	MTV New Artist	iTunes free single, 3/6
Morningwood	morningwood	MTV New Artist	
We the Kings	wethekings	MTV New Artist	
Mayday Parade	maydayparade	MTV New Artist	
Donnie Klang	donnieklang	MTV New Artist	
Jazmine Sullivan	jazminesullivan	MTV New Artist	
Santogold	santogold	MTV New Artist	
The Cab	thecab	MTV New Artist	
The Wombats	thewombatsuk	MTV New Artist	
Cold War Kids	coldwarkids	MTV.com featured	
Mastodon	mastodon	MTV.com featured	
Mika	mikamyspace	MTV.com featured	
Klaxons	klaxons	MTV.com featured	
3oh3	3oh3	MTV.com featured	
Secondhand Serenade	secondhandserenade	MTV.com Video	
Mudvayne	mudvayne	MTV.com Video	
Britney Spears	britneyspears	MTV.com Video	
Mariah Carey	mariahcarey	MTV.com Video	

Miley Cyrus	mileycyrus	MTV.com Video	
Christina Aguilera	christinaaguilera	MTV.com Video	
All American Rejects	allamericanrejects	MTV.com Video	
The Maine	themaine	MTV.com Video	
Paramore	paramore	MTV.com Video	
The Academy Is	theacademyis	MTV.com Video	
Beyonce	beyonce	MTV.com Video	
Rihanna	rihanna	MTV.com Video	
Akon	akon	MTV.com Video	
Brandy	foreverbrandy	MTV.com Video	
Kanye West	kanyewest	MTV.com Video	
Ludacris	ludacris	MTV.com Video	
TI	trapmuzik	MTV.com Video	
50 Cent	50cent	MTV.com Video	
Kardinal Offishall	kardinaloffishall	MTV.com Video	
Fall Out Boy	falloutboy	MTV.com Video	
Kevin Rudolf	kevinrudolfmusic	MTV.com Video	
Coldplay	coldplay	MTV.com Video	
Jack's Mannequin	jacksmannequin	MTV.com Video	
Justin Timerblake	justintimberlake	MTV.com Linked	
Jonas Brothers	jonasbrothers	MTV.com Linked	
John Legend	johnlegend	MTV.com Pick	
Pink	pink	MTV.com Pick	
Solange	solange	MTV.com Pick	
Lil Wayne	lilwayne	MTV.com Pick	
Ne-Yo	neyo	MTV.com Pick	
LL Cool J	llcoolj	MTV.com Pick	
Kid Rock	kidrock	MTV.com Pick	
Metro Station	metrostation	MTV.com Popular	
Lady Gaga	ladygaga	MTV.com Popular	
Katy Perry	katyperry	MTV.com Popular	
The Game	thegame	MTV.com Popular	
Charles Hamilton	hamiltonsmusic	MTV.com Popular	
T-Pain	tpain	MTV.com Popular	
Gym Class Heroes	gymclassheroes	MTV.com Popular	
System of A Down	soad	Myspace popular	
Plain White Ts	plainwhitets	Myspace popular	
OAR	ofarevolution	Myspace popular	
Amy Winehouse	amywinehouse	Rolling Stone	
Vampire Weekend	vampireweekend	Rolling Stone	
The Veronicas	theveronicas	Rolling Stone	
The Dodos	thedodos	Rolling Stone	
Titus Andronicus	titusandronicus	Rolling Stone	
Semi Precious Weapons	semipreciousweapons	Rolling Stone	
The Red Jumpsuit Apparatus	redjumpsuit	Rolling Stone	
The Ting Tings	thetingtings	Top Ringtones	
Jordin Sparks	jordinsparks	Top Ringtones	
Leona Lewis	leonalewis	Top Ringtones	
Alicia Keys	aliciakeys	Top Ringtones	
Chris Brown	chrisbrown	Top Ringtones	
Three 6 Mafia	threesixmafia	Top Ringtones	
Celine Dion	celinedion	Touring	
Metallica	metallica	Touring	
Billy Joel	billy_joel	Touring	
Elton John	eltonjohn	Touring	
Tila Tequila	tilatequila	Famous	
Gabriella Cilmi	gabriellacilmi	Album: 3/17/09	Universal
Chris Cornell	chriscornell	Album: 3/10/09	Universal
Lionel Richie	lionelrichie	Album: 3/10/09	Universal
Jadakiss	jadakiss	Album: 3/10/09	Universal
The Dream	thedreamteam	Album: 3/10/09	Universal
Method Man	methodman	Album: 3/17/09	Universal
Sum 41	sum41	Album: 3/17/09	Universal
Ron Browz	ronbrowz	Album: 3/17/09	Universal
Razorlight	razorlight	Album: 3/10/09	Universal
White lies	whitelies	Album: 3/17/09	Universal
Utada	thisisutada	Album: 3/24/09	Universal
The Love Willows	thelovewillows	Album: 3/24/09	Universal
Keri Hilson	kerihilson	Album: 3/24/09	Universal
Chrisette Michele	chrisettemichele	Album: 3/31/09	Universal
30 Seconds to Mars	thirtysecondstomars	Major Label Artist	EMI
Aaliyah	aaliyah	Major Label Artist	Universal

AC/DC	acdc	Major Label Artist	Sony BMG
Aerosmith	aerosmith	Major Label Artist	Universal
Against Me!	againstme	Major Label Artist	Warner
Alanis Morissette	alanismorissette	Major Label Artist	Warner
Alkaline Trio	alkalinetrio	Major Label Artist	Sony BMG
Amber Pacific	amberpacific	Major Label Artist	Warner
Amy Grant	amygrant	Major Label Artist	Warner
Amy Pearson	amypearsonmusic	Major Label Artist	Sony BMG
Anastacia	anastaciaofficial	Major Label Artist	Sony BMG
Anti-Flag	antiflag	Major Label Artist	Sony BMG
Ash	ash	Major Label Artist	Warner
Asher Roth	asherrothmusic	Major Label Artist	Universal
Ashley Tisdale	ashleytisdale	Major Label Artist	Warner
Atmosphere	atmosphere	Major Label Artist	Warner
Atreyu	atreyurock	Major Label Artist	Warner
Avenged Sevenfold	avengedsevenfold	Major Label Artist	Warner
Avril Lavigne	avrillavigne	Major Label Artist	Warner
Backstreet Boys	backstreetboys	Major Label Artist	Sony BMG
Bad Company	badcompany	Major Label Artist	Warner
Beastie Boys	beastieboys	Major Label Artist	EMI
Björk	bjork	Major Label Artist	Warner
Black Eyed Peas	blackeyedpeas	Major Label Artist	Universal
Black Rebel Motorcycle Club	blackrebelmotorcycleclub	Major Label Artist	EMI
Black Sabbath	blacksabbath	Major Label Artist	Warner
Blindside	blindside	Major Label Artist	Warner
Blink 182	blink182	Major Label Artist	Universal
Blondie	blondie	Major Label Artist	Warner
Blur	blurtheband	Major Label Artist	EMI
Bon Jovi	bonjovi	Major Label Artist	Universal
Boys Like Girls	boyslikegirls	Major Label Artist	Sony BMG
Bullet For My Valentine	bulletformyvalentine	Major Label Artist	Sony BMG
Camila	camilatodocambio	Major Label Artist	Sony BMG
Carly Simon	carlysimon	Major Label Artist	Warner
Carrie Underwood	carrieunderwood	Major Label Artist	Sony BMG
Celine Dion	celinedion	Major Label Artist	Sony BMG
Chaka Khan	officialchakakhan	Major Label Artist	Warner
Chamillionaire	chamillionaire	Major Label Artist	Universal
Cher	cherconquerstheworld	Major Label Artist	Warner
Chevelle	chevelle	Major Label Artist	Sony BMG
Chris Daughtry	daughtry	Major Label Artist	Sony BMG
CKY	cky	Major Label Artist	Warner
Coheed and Cambria	coheedandcambria	Major Label Artist	Sony BMG
Collective Soul	collectivesoul	Major Label Artist	Warner
Coolio	coolio	Major Label Artist	Warner
Craig David	craigdavid	Major Label Artist	Warner
Daft Punk	daftpunk	Major Label Artist	EMI
Damien Rice	damienrice	Major Label Artist	Warner
Daniel Powter	danielpowter	Major Label Artist	Warner
Daniela Castillo	danielaonline	Major Label Artist	Warner
Danity Kane (Bad Boy/Atlantic)	danitykane	Major Label Artist	Warner
Dave Matthews Band	davematthewsband	Major Label Artist	Sony BMG
Dave Stewart	thedavestewart	Major Label Artist	Warner
David Archuleta	davidarchuleta	Major Label Artist	Sony BMG
David Bowie	davidbowie	Major Label Artist	EMI
David Cook	officialdavidcook	Major Label Artist	Sony BMG
Death Cab for Cutie	deathcabforcutie	Major Label Artist	Warner
Destiny's Child	destinyschild	Major Label Artist	Sony BMG
Diddy	diddy	Major Label Artist	Warner
Dido	dido	Major Label Artist	Sony BMG
Disturbed	disturbed	Major Label Artist	Warner
Dixie Chicks	dixiechicks	Major Label Artist	Sony BMG
DMX	dmx	Major Label Artist	Sony BMG
Dr. Dre	drdre	Major Label Artist	Universal
DragonForce	dragonforce	Major Label Artist	Warner
Duran Duran	duranduran	Major Label Artist	Sony BMG
Elvis Costello	elviscostello	Major Label Artist	EMI
Elvis Presley	elvispresley	Major Label Artist	Sony BMG
Enrique Iglesias	enriqueiglesias	Major Label Artist	Universal
Eric Clapton	ericclapton	Major Label Artist	Warner
Evanescence	evanescence	Major Label Artist	Warner
Everclear	everclear	Major Label Artist	EMI
Faith Hill	faithhill	Major Label Artist	Warner

Fiona Apple	fionaapple	Major Label Artist	Sony BMG
Flaming Lips	flaminglips	Major Label Artist	Warner
Flyleaf	flyleaf	Major Label Artist	Warner
Foo Fighters	foofighters	Major Label Artist	EMI
Fort Minor	fortminor	Major Label Artist	Warner
Funeral for a Friend	funeralforafriend	Major Label Artist	Warner
Genesis	genesismusic	Major Label Artist	Sony BMG
George Harrison	georgeharrison	Major Label Artist	EMI
George Michael	georgemichael	Major Label Artist	Sony BMG
Glassjaw	glassjaw	Major Label Artist	Warner
Gloria Estefan	gloriaestefan	Major Label Artist	Sony BMG
Gnarls Barkley	gnarlsbarkley	Major Label Artist	Warner
Goo Goo Dolls	googoodolls	Major Label Artist	Warner
Good Charlotte	goodcharlotte	Major Label Artist	Sony BMG
Gorillaz	gorillaz	Major Label Artist	EMI
Green Day	greenday	Major Label Artist	Warner
Guns N' Roses	gunsnroses	Major Label Artist	Universal
Gwen Stefani	gwenstefani	Major Label Artist	Universal
Hard-Fi	hardfi	Major Label Artist	Warner
Hoobastank	hoobastank	Major Label Artist	Warner
Interpol	interpol	Major Label Artist	EMI
INXS	inxs	Major Label Artist	Sony BMG
James Blunt	jamesblunt	Major Label Artist	Warner
James Ingram	jamesingrammusic	Major Label Artist	Warner
Jane's Addiction	janesaddictionmyspace	Major Label Artist	Warner
Janet Jackson	janetjackson	Major Label Artist	Universal
Jennifer Lopez	jenniferlopez	Major Label Artist	Sony BMG
Jessica Simpson	jessicasimpson	Major Label Artist	Sony BMG
Jet	jet	Major Label Artist	Warner
Jewel	jewel	Major Label Artist	Warner
John Mayer	johnmayer	Major Label Artist	Sony BMG
Jordyn Taylor	jordyntaylor	Major Label Artist	
Journey	journey	Major Label Artist	Sony BMG
Kelly Rowland	kellyrowland	Major Label Artist	Sony BMG
Kylie Minogue	kylieminogue	Major Label Artist	Warner
Leona Lewis	leonalewis	Major Label Artist	Sony BMG
Lifehouse	lifehouse	Major Label Artist	Warner
Lillix	lillix	Major Label Artist	Warner
Linda Ronstadt	lindaronstadt	Major Label Artist	Warner
Linkin Park	linkinpark	Major Label Artist	Warner
Lost Prophets	lostprophets	Major Label Artist	Warner
Lupe Fiasco	lupefiasco	Major Label Artist	Warner
Madonna	madonna	Major Label Artist	Warner
Marilyn Manson	marilynmanson	Major Label Artist	Universal
Marion Raven	marionraven	Major Label Artist	Universal
Maroon 5	maroon5	Major Label Artist	Universal
Mary J. Blige	maryjblige	Major Label Artist	Universal
Mastodon	mastodon	Major Label Artist	Warner
Matchbox Twenty	matchboxtwenty	Major Label Artist	Warner
Megadeth	megadeth	Major Label Artist	Warner
Michael Bublé	michaelbuble	Major Label Artist	Warner
Michael Jackson	michaeljackson	Major Label Artist	Sony BMG
Michelle Branch	michellebranch	Major Label Artist	Warner
Michelle Williams	michellewilliams	Major Label Artist	Sony BMG
Mike Jones	mikejones	Major Label Artist	Warner
Moby	moby	Major Label Artist	Warner
Morrissey	morrissey	Major Label Artist	Warner
Mötley Crüe	motleycrue	Major Label Artist	Warner
Muse	muse	Major Label Artist	Warner
My Chemical Romance	mychemicalromance	Major Label Artist	Warner
Nas	nas	Major Label Artist	Universal
Natalie Cole	nataliecole	Major Label Artist	Warner
Natalie Merchant	nataliemerchant	Major Label Artist	Warner
Natasha Bedingfield	natashabedingfield	Major Label Artist	Sony BMG
Nelly Furtado	nellyfurtado	Major Label Artist	Universal
New Kids on the Block	newkidsontheblock	Major Label Artist	Universal
Nickelback	nickelback	Major Label Artist	Warner
Nirvana	nirvana	Major Label Artist	Universal
Norah Jones	norahjones	Major Label Artist	EMI
Notorious B.I.G.	notoriousbig	Major Label Artist	Warner
Oasis	oasis	Major Label Artist	Sony BMG
Ol' Dirty Bastard	oldirtybastard	Major Label Artist	Warner

Ozzy Osbourne	ozzyosbourne	Major Label Artist	Sony BMG
P.O.D.	payableondeath	Major Label Artist	Warner
Panic at the Disco	panicatthedisco	Major Label Artist	Warner
Papa Roach	paparoach	Major Label Artist	Warner
Paul McCartney	paulmccartney	Major Label Artist	EMI
Paul Wall	paulwall	Major Label Artist	Warner
Paula Cole	officialpaulacole	Major Label Artist	Warner
Pearl Jam	tenclub	Major Label Artist	Sony BMG
Phish	phish	Major Label Artist	Warner
Pixies	pixies	Major Label Artist	Warner
Queen	queen	Major Label Artist	EMI
R. Kelly	rkelly	Major Label Artist	Sony BMG
R.E.M.	rem	Major Label Artist	Warner
Radiohead	radiohead	Major Label Artist	EMI
Rage Against The Machine	ratm	Major Label Artist	Sony BMG
Red Hot Chili Peppers	redhotchilipeppers	Major Label Artist	EMI
Ricky Martin (Latin America)	rickymartin	Major Label Artist	Sony BMG
Rilo Kiley	rilokiley	Major Label Artist	Warner
Rob Thomas	robthomas	Major Label Artist	Warner
Robbie Williams	robbiewilliamsofficial	Major Label Artist	EMI
Rod Stewart	rodstewart	Major Label Artist	Warner
Rolling Stones	therollingstones	Major Label Artist	EMI
Ryan Cabrera	ryancabrera	Major Label Artist	Warner
Samantha Jade	samanthajade	Major Label Artist	Sony BMG
Sean Paul	seanpaul	Major Label Artist	Warner
Shakira	shakira	Major Label Artist	Sony BMG
Simple Plan	simpleplan	Major Label Artist	Warner
Slipknot	slipknot1	Major Label Artist	Warner
Staind	staind	Major Label Artist	Warner
Story of the Year	storyoftheyear	Major Label Artist	Warner
Sugar Ray	sugarray	Major Label Artist	Warner
Taking Back Sunday	takingbacksunday	Major Label Artist	Warner
Talib Kweli	talibkweli	Major Label Artist	Warner
The Beatles	thebeatles	Major Label Artist	EMI
The Cure	thecure	Major Label Artist	Warner
The Darkness	thedarkness	Major Label Artist	Warner
The Decemberists	thedecemberists	Major Label Artist	EMI
The Eagles	eaglesmusic	Major Label Artist	
The Killers	thekillers	Major Label Artist	Universal
The Smashing Pumpkins	smashingpumpkins	Major Label Artist	EMI
The Strokes	thestrokes	Major Label Artist	Sony BMG
The Verve	theverve	Major Label Artist	EMI
The Vines	thevines	Major Label Artist	EMI
The White Stripes	thewhitestripes	Major Label Artist	EMI
They Might Be Giants	theymightbegiants	Major Label Artist	Warner
Third Eye Blind	thirdeyeblind	Major Label Artist	Warner
Three Days Grace	threedaysgrace	Major Label Artist	Sony BMG
Tina Turner	tinaturner	Major Label Artist	EMI
Tom Petty & The Heartbreakers	tompetty	Major Label Artist	Warner
Toni Braxton	tonibraxton	Major Label Artist	Sony BMG
Tori Amos	toriamos	Major Label Artist	Sony BMG
Tracy Chapman	tracychapman	Major Label Artist	Warner
Trapt	trapt	Major Label Artist	Warner
Uncle Kracker	officalunclek	Major Label Artist	Warner
Usher	usher	Major Label Artist	Sony BMG
Van Halen	vanhalen	Major Label Artist	Warner
Velvet Revolver	velvetrevolver	Major Label Artist	Sony BMG
Whitney Houston	whitneyhouston	Major Label Artist	Sony BMG
Will Smith	willsmithproduction	Major Label Artist	Sony BMG
Wu-Tang Clan	wutang	Major Label Artist	Universal
Yellowcard	yellowcard	Major Label Artist	EMI

Pledge Statement

I pledge my honor that this paper represents my own work and was completed in accordance to the Honor Code and University regulations.

Signed,

Mark W. Ungerer