12-15-2008

An Agent-Based Model of Viral Marketing: Comparing online-based viral and traditional marketing in a closed system for fashion trends in a competitive setting

Tony S. Wang
University of Pennsylvania, tonywang@wharton.upenn.edu

An Agent-Based Model of Viral Marketing: Comparing online-based viral and traditional marketing in a closed system for fashion trends in a competitive setting

Abstract

Viral marketing, is it hot or just hot air? Is the only thing viral about viral marketing the hype it‟s receiving from the academic community, the corporate world, and the media? This study presents an agent-based model to determine if viral marketing is significantly more effective than traditional marketing in the context of a closed world where two clothing companies compete with one another for consumer loyalty, one using only online viral ads and the only online traditional ads. The study concludes that viral marketing has a small, yet statistically significant advantage over traditional marketing as a whole. The study also finds that viral marketing does significantly better in settings with high populations and high levels of localization.

Keywords

Viral Marketing, Word of Mouth, Agent-Based Modeling, NetLogo, Online Marketing, ABM, Humanties, Philosopy Politcs & Economics, John Gasper, Gasper, John

Disciplines

Advertising and Promotion Management | Marketing | Technology and Innovation

An Agent-Based Model of Viral Marketing:

Comparing online-based viral and traditional marketing in an closed system for fashion trends in a competitive setting

ARTICLE INFORMATION

Tony S. Wang
The Wharton School of the University of Pennsylvania

Keywords:
Viral Marketing
Word of Mouth
Agent-Based Modeling
NetLogo
Online Marketing

Article History:

First submitted on December 15, 2008

ABSTRACT

Viral marketing, is it hot or just hot air? Is the only thing viral about viral marketing the hype it's receiving from the academic community, the corporate world, and the media? This study presents an agent-based model to determine if viral marketing is significantly more effective than traditional marketing in the context of a closed world where two clothing companies compete with one another for consumer loyalty, one using only online viral ads and the only online traditional ads. The study concludes that viral marketing has a small, yet statistically significant advantage over traditional marketing as a whole. The study also finds that viral marketing does significantly better in settings with high populations and high levels of localization.

INTRODUCTION

Viral marketing is a form of marketing in which brand awareness of a product is increased exponentially through self-replicating word of mouth. Much like how a virus spread, viral marketing utilizes advertisements that encourage or incentivize agents to spread the advertisement within their social network (i.e. - friends, families, and peers). In turn, those people will share the advertisement to their social network. In essence, the target audience members also serve as the distributor of the advertisement.

Viral networking is a relatively "new" marketing phenomenon, with companies trying to create viral videos and websites. While the primary dissemination mechanism (word of mouth) of viral ads is by no means new, the concept that consumers can also be salespeople of the product within their social networks by sharing a video or website is. Spending on viral marketing in the past years has drastically increased, showing how well-perceived this new marketing tool is. In fact, 82% of the fastest growing private companies use viral marketing (Ferguson) and US companies spent $4.9 billion on online marketing (both viral and traditional) in the first quarter of 2007, up 26 percent from the same period last year (Allsop).

It is clear that online viral marketing is hot, but is there a significant benefit to viral marketing? It is of importance to explore this phenomenon in more detail. While designing and executing a real-life experiment to study viral marketing is costly and time-consuming, a new computational modeling method called *agent-based modeling* offers a convenient method with which to study viral marketing.

Agent-based modeling (ABM), as a computational modeling tool, is characterized by the use of adaptive agents that are neither fully informed nor fully rational. This underlying basis for agent based modeling makes it an excellent way with which to study society since individuals tend to act in terms of heuristics and societal norms. Furthermore, ABM focuses on a bottom-up approach to modeling society; we code observable micro-motives in the model to predict societal macro-trends.

Of particular interest in constructing this model is to compare viral marketing to traditional marketing. While online viral marketing is certainly popular, is it better than traditional online marketing? Under what conditions does viral marketing perform well in? To answer questions posed in the previous paragraph and explore the topic of viral marketing, I create an ABM to determine if viral marketing offers fashion companies a significant competitive advantage in gaining consume loyalty. The model creates a world in which two clothing products compete for consumer loyalty in the same population. One product will only use online viral marketing while the other product will only use online traditional marketing. The product of interest that this paper considers is clothing.

This paper presents the model and analyzes the results.

ASSUMPTIONS

The model functions on several key assumptions. Understanding these assumptions are key to prevent an over-extrapolation of the results of the model and over-extend the applications of the model. The following is a list of the most important assumptions the model is based on.

Key Assumptions:

1. *Agents live in a closed world*

   I assume that all the agents in the model live in a closed world with only two possible clothing products to choose from. No other clothing trend will enter the system. This assumption is to control against exogenous factors. We only want to consider the direct effect of a viral advertisement that is competing against a traditional advertisement.

2. *Agents are independent but can be swayed by peer pressure*

   This assumption is based on reality, as agents are independent from each other but are influenced by peers and word of mouth. This assumption enables us to study the impact that word of mouth has in the dissemination and conversion of the viral marketing product.

3. *There is only one market segmentation and all agents are target consumers*

   I assume that all the agents in the model have the same tastes and preferences and thus belong to the same market segmentation. I also assume that all the agents are target customers for both clothing products. This assumption is to control for different consumer preferences that would affect the outcome of this model. Having different marketing segmentations would (a) prevent us from studying the direct effect of viral marketing given a competing traditional marketing advertisement as the two products would either target different consumers or would develop a niche market over time. It would also (b) make the model beyond its intended scope.

4. *Agents can only consume one product at a time*

   An agent can only wear one product at a given time. He cannot both be wearing the viral marketing clothing and the traditional marketing clothing. This assumption is to keep the model simple. In real life, an agent can subscribe to more than one fashion trend.

5. *Agents can switch between fashion trends*

   While an agent can only subscribe to one trend at a given time, he may switch to the other trend. The mechanism for switching from one product to another is described in more detail in the methodology section.

6. *Both clothing trends are qualitatively equivalent*

   I assume that both clothing products are the same in terms of price and quality, meaning that the only factor that affects an agent's decision is the nature of the advertisements he sees and the influence of his peers. This assumption controls against any consideration regarding the actual quality of the clothing, which could potentially act as a confounding factor in the results.

7. *Companies will not engage in negative advertisements*

   To control against negative marketing campaigns, I assume that companies will not try to negative influence consumers against the other product. This assumption keeps the model more general as the impact of negative ads changes as the number of products changes as well.

8. *For viral advertisements, there is a difference between the ads ability to interest the turtles and its ability to persuade the turtles*

   While traditional advertisements aim just to persuade a consumer to buy a certain product, viral advertisements can be measured along two characteristics: ability to interest and ability to persuade. A viral ad can be extremely interesting and thus likely to be disseminated over networks; however, as funny or interesting as the advertisement may be, it may not necessarily be persuasive. Thus, I assume that there are two qualities that affect the nature of the viral advertisement- interest and persuasion- while there is only one quality that affects the nature of the traditional advertisement- persuasion.

9. *Agents all have the same influence*

   I assume that all turtles have the same WOM influence. This assumption is to keep the project within its scope as models of influence are another field of inquiry. Furthermore, research has shown that the notion of high network potential individuals does not hold for the most part as social networks overall tend to be homogenous in terms of influence (Smith).

MODEL METHDOLOGY

The model is run on NetLogo, a platform designed to create ABM. The model is primarily run on the interface tab. The information tab gives more information about how the model runs and includes a basic version of the methodology of the model. Lastly, the procedures tab contains the actual coding of the model itself.

This section will go into a detailed overview of how the program runs and assumes knowledge of the NetLogo platform.

The SETUP button creates "x" number of turtles, specified by the POPULATION slider. The turtles are arranged in a circle layout and are connected by undirected links. The links represent social connections between turtles in the world. The formation of the links in the model is based on the Small Worlds Model (Wilensky) found in the NetLogo "Models Library." The small worlds social network configuration utilizes the concept of six degrees of separation to create a world whereby all individuals are connected to a global society. However, within this global society, there are still local communities. The observer can vary the amount of globalization in the model with the GLOBALIZATION slider. The more globalized the world is, the more a turtles connections will come from unique networks that his other peers won't be connected with. The setup button also creates additional connections as specified by the ADDITIONAL-

DENSITY slider. The block of code that pertains to the creation of additional links runs within an "ask links" command. Thus, the maximum percent increase in the density of connections possible for this model is 100%. The reason why I ask the links to create more links is to prevent a world in which everyone is connected to each other, which is highly unrealistic. Finally, the setup button turns most of the turtles own variables false. In other words, the turtles, at this point, haven't seen the ads, aren't persuaded or interested by any of the ads, and aren't wearing either of the clothing yet.

The AWARENESS button first selects which turtles will go online based on the INTERNET-FREQUENCY slider. I assume that only a certain percentage of the agents will be online at any given time. According to a Harris Poll, around 80% of the population goes online (Reuters). Thus, the default value for the slider is 80%; however, the exact percentage of agents that go online is a random figure based on a normal distribution with mean 80% and standard deviation of 5% of the mean. Next, those turtles that are online will either see one, both, or neither of the ads. The percent chance of seeing the ad is based on the assumption that if the turtle is online at this stage, he will be casually surfing the internet and will look at the advertisement if he stumbles across it. Research has shown that around 65% of those using the internet will be using sites where viral advertisements can usually be found, such as an online community, blog, or YouTube (Allsop). Thus, around 65% of the online agents will see the viral advertisement. The exact rate is determined by a random normal distribution with mean of 65% and a standard deviation of 5% of the mean. Similarly, since around 90% of the online community surfs the internet recreationally on websites that traditional advertisements will typically be found on, such as online news portals, magazines, and radios, around 90% of the online community in the model will see the traditional advertisement (Allsop). This percentage also varies in the model by a random normal distribution with mean 90% and standard deviation of 5% of the mean.

The GO buttons run the actual model. There are two GO buttons. The GO (ONCE) button runs the model through one loop of the go procedure. The other button, GO (CONTINUOUS), iterates through the go procedure until all the turtles have subscribed to one of the two fashion trends. The procedures within the go command are as follows.

First, the model generates INTEREST amongst the turtles that have seen the ads. The sliders that affect this block of code are VM-INTEREST-RATE, VM-PERSUASIVENESS and TM-PERSUASIVENESS. The first slider describes the chance that the viral ad will engage the interest of the turtle while the second describes the chance that the viral ad will persuade the turtle to buy the product. The third variable describes the chance that the traditional ad will persuade the turtle to buy the product. For all three variables, the default value is 16% (Allsop). Just like in the AWARENESS command, all percentages are based on a random normal distribution. Thus, the observer can only choose the mean (and standard deviation) of the normal distribution that will be used to generate the actual value for the three variables with the slider but not the actual value used itself.

The DISSEMINATION command asks those turtles that are interested by the viral ad to spread the ad to their peers. The number of peers that an individual who finds the viral ad interesting will spread the ad to is determined by the DISSEMINATION RATE slider, keeping in mind that the actual rate of dissemination is picked randomly from a normal distribution. We assume that the turtle will show the ad to a friend even if the friend has seen it before (as many friends find watching viral videos together a social activity). Also, we assume that turtles that don't go online will be able to see the advertisement. We make this assumption because we assume that if an agent wants to show a connected turtle that doesn't go online the ad, he will show the ad at the library, at his home, or some other location with access to the internet.

Lastly, the FINAL DECISION command runs through a series of codes that enables turtles to make a decision about which product to purchase. There are three total possible sources of influence that affect a turtle's decision: word of mouth, personal experience, and advertisements (Allsop). However, not all three sources of influence will always be used in calculating a turtle's decision.

For turtles that are persuaded by the ad for one product but not the other, there are three blocks of codes that determine which clothing the turtle will purchase.

The first block of code is used if the tick count is equal to zero, meaning that no turtles have bought any clothing before. We do so as personal experience has had no effect in formulating the turtle's and his peers' opinions of the products. When ticks is equal to zero, the two sources of influence that affect the turtle's decision is the advertisement for the company and the word of mouth from the turtle's peers. Research shows that the ratio of influence of word of mouth to advertising is 70% to 30% (Allsop). The influence that word of mouth has on a turtle is calculated by determining the number of peers that also are persuaded by one clothing trend over another. We do not count turtles that are persuaded by both clothing products since they are ambivalent about the situation. These two numbers represents the word of mouth variables that influence the turtle's decision. If we just considered word of mouth (WOM) influence, then the product that more peers support will also be the product that the turtle buys. However, we need to add the effect that ads have on the turtle's decision. Since ads account for 30% of total impact on a turtle's decision, and since the turtle has only been persuaded by one product, we can multiply the number of supportive peers for the product that the turtle is persuaded by by 1.43 or 1 divided by 0.7. This multiplier is derived from the 70:30 ratio previously mentioned. If WOM accounts for 70% of influence and we know the value of the impact WOM has on the turtle's decision, we can algebraically determine the multiplier that will determine the overall impact a product will have on the turtle. The other product is not multiplied by 1.43

because the turtle is not persuaded by the ad for that company. The turtle then chooses the clothing trend with the higher value.

If ticks are greater than zero, then both the turtle and peers might already have chosen one product over another. We need to add the impact that personal experience has on the decision process. If the turtle has not bought the clothing before, then the only impact personal experience has is on the WOM value as it is no longer a simple count of peers persuaded by one ad or the other. We need to consider the role that personal experience has in the formulation of the opinions of the turtle's peers. Research shows that ads have a 20% impact on WOM and personal experience has an 80% impact. Thus, we count the number of peers that support one product and actually wear the product (their WOM consists of their support for the ad and the product itself). Then we multiply the number of peers that don't wear the clothing but are persuaded by the ad by 0.20 (thus they only base their decision on company advertisements). We add the two values up. This method is the new method to calculate the impact of WOM for both products. Then, we multiply the WOM value for the product whose advertisement the turtle has been persuaded by by 1.43. We do not multiply the WOM value for the other product by 1.43. Finally, we compare the two values and the turtle wears the product that has the higher value.

The decision process is slightly different for turtles that are persuaded by one product but not the other and are already wearing the product. We need to add the impact personal experience has in the turtle's decision. Now, there are three components: WOM, the advertisement, and personal experience. The ratio of influence for these three components is 28:13:59 (Allsop). Thus, instead of multiplying the WOM component by 1.43 to obtain the total value, we multiply by 1 divided by 13, or 3.57, to obtain the total value. The rest of the procedure remains the same.

For turtles that are persuaded by both products, there are four blocks of code that determine which clothing the turtle will purchase.

The first two blocks of code are used if the turtle has not worn any of the clothing previously. In this scenario, the turtle will base his decision solely on the WOM of his peers as advertising and personal experience play no role. If ticks are equal to 0, then the WOM is only based on advertisements. If ticks is greater than 0, then the WOM component is based both on personal experience and company advertisements with the appropriate weights attached (turtles whose WOM is only based on advertisements have a 20% influence on WOM).

The other two blocks of code are used if the turtle is wearing one of the two clothing and persuaded by both ads. Now, the turtle's decision is based on personal experience and WOM but not ads, as he is persuaded by both. In this calculation, we need to find out the ratio of influence of WOM to personal experience. That ratio is 33:67 (Allsop). Thus, the WOM multiplier used to obtain the total decision value for the product that he is currently wearing is 1 divided by 0.33, or 3.03. The WOM component for a clothing product is comprised of three sums in this case. First, we count those peers that wear the product but are persuaded by both products. That number if multiplied by 0.80 since their WOM influence only comes from their personal experience. Then, we count those peers that wear the product and are only persuaded by the advertisement for the product they are wearing. That number remains the same as their WOM influence is based on both advertisements for the company and personal experience. Lastly, we take the number of peers that are only convinced by that product's advertisement but aren't wearing the clothing. That number is multiplied by 0.20 as their WOM influence only comes from their support of the company's advertisement. Then, the figures are summed. If the product in question is the product that the turtle is currently wearing, then we multiply that figure by 3.03 to obtain the total decision value that includes both WOM and personal experience. The other value is the sum of the three values. Then the turtle wears the clothing that has the higher decision value.

The final part of the GO command is the UPDATE-PLOTS section, which simply updates the plots on the right hand side. The plots on the right hand side of the model help gauge the spread and effectiveness of the two types of marketing over time. Starting from top to bottom, left to right, the first graph is the ROI graph that shows the number of agents that are wearing the VM clothing and the number of agents that are wearing the TM clothing. Next is the % Wearing VM Clothes graph v. TM Clothes; this graph shows what percentage of the population is wearing VM clothing and what percentage is wearing TM clothing. To check the percentage of people who haven't seen a certain ad, we use the next graph, the % Haven't Seen the Ads graph. Notice that the green line, the line that corresponds to those who haven't seen the TM ad, stays constant. This is on purpose as the model only runs through the traditional marketing advertisement once. However, the red line starts out much higher, which means that the traditional ad has a dissemination advantage in the beginning. The % Undecided Graph shows the % of consumers that have yet to decide which clothing they are going to wear. This may be because they haven't seen the ad or haven't been persuaded. The next graph is the % Seen VM, which shows the percent of agents that have seen the viral ad. Lastly, we have the VM Clothes: TM Clothes ratio. This graph shows the ratio of turtles wearing VM clothes to those turtles wearing TM clothes. It is another measurement of effectiveness in addition to the % Wearing VM Clothes v. TM Clothes graph.

OBSERVATIONS / ANALYSIS OF DATA

To generate data from the model we use the built-in BehaviorSpace software tool. BehaviorSpace is a platform that enables users to perform runs of NetLogo models with varying parameter settings. BehaviorSpace also allows the user to record the data and results of each model run onto an Excel file. I use BehaviorSpace to generate data; I analyze the data on a data analysis software platform called JMP.

In this section, the present paper will go over the analysis of the data and some notable observations of trends and patterns in it.

To address the question of under what conditions does viral marketing do the best in, I vary the two parameter settings that had no default value- GLOBALIZATION and POPULATION. I ran GLOBALIZATION from the values of 0 to 100 in increments of 10 percentage points. I ran POPULATION from the values of 100 to 500 in increments of 100 turtles. The remaining variables in the model stay fixed at their default values. TM-PERSUASIVENESS, VM-PERSUA SIVENESS, and VM-INTEREST-LEVEL are at 16%. ADDITIONAL-DENSITY is at 0%, DISSEMINATION-RATE is at 59%, and INTERNET-FREQUENCY is at 80%. I had BehaviorSpace run each combination of parameter settings 100 times, recoding the results of each step. The resultant data spread contains 92,295 data points, of which 5,500 data points were the final count of turtles wearing viral marketing clothing and turtles wearing traditional marketing clothing. We primarily only consider those 5,500 data points; however, we need to use all the data points when looking at the overall distribution of the data over time.

We are interested in the effectiveness of viral marketing in terms of how many people wear viral marketing clothing at the end of the model. To measure this effectiveness, we will primarily consider two measurements. The first is the ratio of agents that wear the viral marketing clothing to those that wear the traditional marketing clothing (referred to as the VM:TM ratio in the present paper). If the ratio is under 1, then more turtles are wearing the traditional marketing clothing than viral marketing clothing, and the vice versa if the ratio is over 1. The second figure is the percent of agents wearing the viral marketing clothing with regards to all agents wearing clothing (referred to as % VM in the present paper). If this measurement is under .50 (or 50%) then there are more turtles wearing the traditional marketing clothing, and vice versa is this measurement is over .50.

First, let us consider at the overall shape that the measurements follow with regards to time (ticks). When we look at the outputs in NetLogo, both % VM and VM:TM seem to follow a log(x) shape [see Figure 1]. Doing repeated trials on NetLogo confirms that observation [see Appendix B].

Both measurements offer insight into the effectiveness of viral marketing in consumer conversion. However, upon deeper analysis of the data, we will see that it is much better to use % VM as a measurement of effectiveness over VM:TM.

This observation is further substantiated by JMP. When we analyze the data points with a Fit Y by X analysis of both measurements by time (ticks), the best fit the data is still a y to log (x) transformation [See Appendix B]. The R-squared value of the log(x) fit for VM:TM vs. Ticks is 0.52, which is relatively high and compares much better to the R-squared value of 0.49 for the linear fit. The R-squared value for the log(x) fit for % VM is 0.66, which is also high and compares well to the R-squared value of 0.56 for the linear fit. The implications of this observation are elaborated in the "Key Findings" section.

As previously stated, all variables in the model are held constant with the exception of globalization and population. It is important, therefore, to consider the impact that these two specific variables have on the effectiveness of viral marketing. To do so, we need to construct a multiple regression model using these two continuous variables as the explanatory variable and one of the two measurements of effectiveness, VM:TM or % VM, are the response variable.

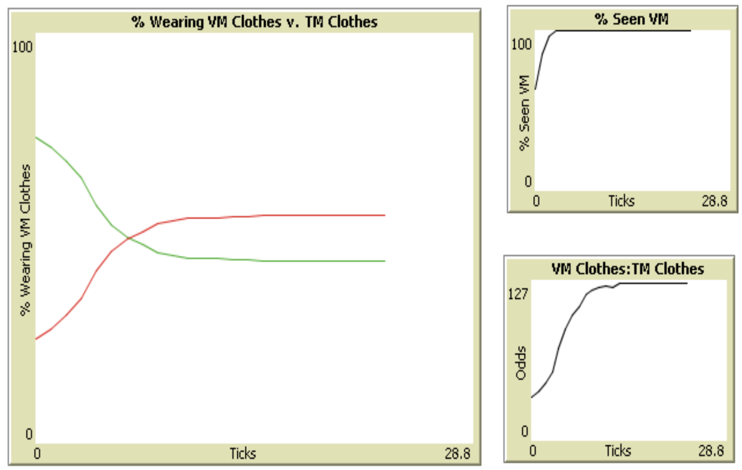


Figure 1: The outputs on the NetLogo interface reveal that both measurements of effectiveness of viral marketing follow a roughly log(x) distribution. The % VM measurement with regards to time is the red line on the lef-hand graph and the VM:TM measurement with regards to time is the bottom right graph.

We will only use the 5,500 data points that record the ending levels of VM:TM and % VM in the multiple regression models. Upon fitting a multiple regression for both measurements I note that there is a serious flaw with using the VM:TM measurement in the multiple regression. As we will see, building a multiple regression model to predict the effectiveness of viral marketing in terms of consumer conversion with VM:TM as our response variable violates one of the three modeling assumptions of a multiple regression. As a result, for multiple regression analysis, we will only use % VM as a measurement of effectiveness and not VM:TM.

The three modeling assumptions that must be met in order to use a multiple regression are that the errors (or residuals) of the regression must be independent, normal, and homoscedastic. As we used BehaviorSpace to create the data both the regression model using VM:TM and the regression model using % VM as the response variable meet this condition. The errors are also relatively homoscedastic (their variance is relatively equal) for both models given the presence of some modest outliers. This is confirmed by looking at the "Residual by Predicted Plot" for both models [See Appendix A]. However, the model using the VM:TM measurement as the y variable violates the assumption of normality [see Figure 2]. The normal quantile plot for the residuals of this model clearly reveal that the residuals for the VM:TM model don't follow a normal pattern.

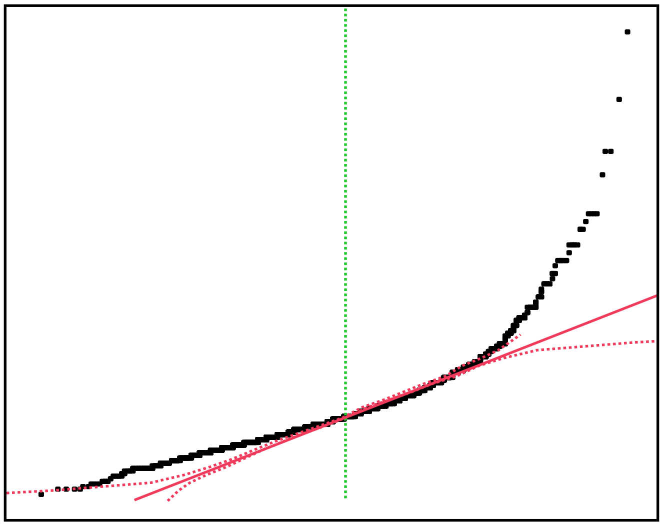


Figure 2: The residuals from the VM:TM regression model clearly don't follow a normal pattern. This is a serious violation of the modeling assumptions of the regression model. We need to find a model that doesn't produce non-normal residuals.

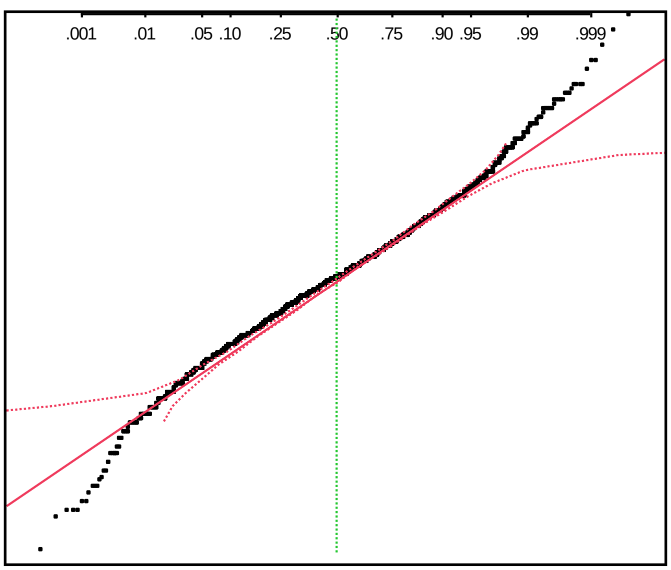


Figure 2: The residuals from the % VM regression model do follow a normal pattern. We use the % VM measurement as a response variable in the mutiple regression model.

This violation of assumed normality in the residuals means that the multiple regression using VM:TM as the response variable is extremely flawed. Luckily, using % VM as a response variable fixes this problem [see Figure 3]. The residuals from this multiple regression are normal, independent, and relatively homoscedastic [for the full JMP output of this regression, see Appendix A].

Using this regression model, we can now quantify the impacts that the variables of globalization and population have on the % VM variable [to see the full output, see Appendix A].

First, we need to determine if the two explanatory variables are significant. We use the standard 0.05 rejection level and find that both variables are significant. Using a t-test, we find that the POPULATION variable has a p-value of 0.0308 and the GLOBALIZATION variable has a p-value of under. 0.0001. Furthermore, an f-test shows that the overall model has a p-value of under 0.0001. Since the model is significant, we can now interpret the slope coefficients of the two variables. POPULATION has a positive slope coefficient, meaning that as population increases, the viral marketing clothing is predicted to have a higher percentage of loyal consumers. GLOBALIZATION has a negative slope coefficient, meaning that as the connections that the turtles have in the model become more globalized (are connected to a larger number of unique networks), the viral marketing clothing is predicted to have a lower percentage of loyal consumers. The best setting for viral marketing is one with high levels of localization (low globalization) and high population.

However, this does not necessary translate into viral marketing being more effective than traditional marketing. We need to see if there is a statistically significant advantage for viral marketing. To do so, we will construct a prediction and confidence interval for the response variable % VM and see if it is significantly higher than 50% at all levels.

In an ideal situation (POPULATION = 500 and GLOBALIZATION = 0), the 95% prediction interval is [0.4229, 0.7001], meaning that there is no significant advantage predicted for viral marketing given such parameters for an individual observation. Similarly, there is no significant advantage predicted for traditional marketing given such parameters for an individual observation. However, the 95% confidence interval is [0.5596, 0.5634], meaning that there is a significant predicted advantage for viral marketing given such parameters for the population as a whole. In a non-ideal situation (POPULATION = 100 and GLOBALIZATION = 100), the 95% prediction interval is. The 95% confidence interval is [0.3676, 0.644798], meaning that there is no significant advantage predicted for viral marketing given such parameters for an individual observation. Similarly, there is no significant advantage predicted for traditional marketing given such parameters for an individual observation. However, the 95% confidence interval is [0.5043, 0.5081], meaning that there is a very slight, yet significant predicted advantage for viral marketing given such parameters for the population as a whole.

Finally, it is important to determine at what mean tick value the percentage of turtles wearing viral marketing clothing is predicted to overtake the percentage of turtles wearing traditional marketing clothing. We do so using the formula derived from the log(x) transformation of the % VM vs. ticks graph using all 92,295data points [see Appendix B]. The formula provided by JMP is %VM = 0.2003676 + 0.1214438*Log (Ticks); from it we determine that the mean tick value at which the % of turtles wearing the viral marketing clothing is predicted to overtake the % of turtles wearing the traditional marketing clothing is 11.79 ticks. If we construct a 95% confidence interval of the mean tick value, the margin of error would be plus or minus 0.000459. The average number of ticks taken before equilibrium is reached (all the turtles have chosen a clothing trend) is 16.64 with a standard deviation of 4.08 [for the full JMP outputs of the fitted y to log(x) model of % VM to ticks using all data points see Appendix B; for the distribution of tick values taken to reach equilibrium see Appendix C].

KEY FINDINGS

Overall, I have found that while traditional marketing is more effective during most of the model, viral advertisement effectively overcome the lack of awareness in the beginning by taking advantage of the ability to quickly generate awareness through word of mouth. In the end, viral marketing is only marginally more effective than traditional marketing, but that advantage is significantly significant. However, the results have such high variance that it is not fruitful to try to predict the effectiveness of an individual viral marketing campaign using the model.

Traditional marketing is more effective during most of the model. The mean tick value at which the % of turtles wearing the viral marketing clothing is predicted to overtake the % of turtles wearing the traditional marketing clothing is 11.79 ticks. As the 95% confidence interval margin of error for the average tick time is so low and as the average ticks before reaching equilibrium is 16.64 with a standard deviation of only 4.08, the traditional marketing clothing is more effective than the viral marketing clothing through most of the model. However, at the end, the viral marketing clothing does significantly better than the traditional marketing clothing.

The reason why viral marketing does much better towards the end of the model is because of the logarithmic shape the spread of its dissemination and conversion take. Thus, at the beginning of the model, viral marketing does poorly. However, as time passes, even if it only catches 16% of the population's interest on the first pass, it can quickly spread around to other turtles. This concave down pattern of dissemination and thus conversion (their interest and persuasion levels are the same) means that in the beginning, growth is extremely fast then levels off towards the end.

As a population parameter, the mean effectiveness of viral marketing is higher than mean effectiveness of traditional marketing once the model reaches equilibrium. We conclude this given the two confidence intervals. As the confidence interval for the effectiveness of viral marketing is significantly higher than the effectiveness of traditional marketing in a non-ideal situation, the confidence interval for the effectiveness of viral marketing in an ideal situation is also significantly higher than the effectiveness of traditional marketing. In both cases, this advantage is small.

However, given the high variance of the results as a whole- the r-squared for the % VM multiple regression model is a mere 0.049537- this model is not appropriate for individual predictions of the estimated effectiveness of a viral marketing campaign. This statement is further corroborated by the prediction intervals for a viral marketing campaign in an ideal and non-ideal situation. In both scenarios, the prediction intervals reveal no significant advantage to using viral marketing or traditional marketing.

As the slope coefficients attached to both POPULATION and GLOBALIZATION in the % VM multiple regression show, viral marketing tends to do better in social networks with high population and high levels of localization.

DISCUSSSION OF RESULTS & AREAS OF FURTHER RESEARCH

This model finds that the "hype" around viral marketing is substantiated to a degree. Viral marketing is slightly more effective than traditional marketing. Furthermore, given the dissemination mechanism of viral marketing, if a viral ad is truly viral and has an extremely high interest rate, then its growth will be explosive. It's no wonder why many companies view viral marketing as a gold nugget of sorts. While it is hard to come across, once found, will generate the company large amounts of awareness and conversion.

While this model helps explain the nature and trajectory of viral marketing and confirm its ability to rapidly generate interest, there is much work remaining is exploring this topic. One such area of research would be studying why high population and levels of localization give viral marketing an advantage in its effectiveness. I hypothesize that high populations give viral marketing an advantage because increasing the number of people simply means that the viral ad will spread with greater numbers. On the other hand, traditional marketing cannot convert the increased population into such an advantage. I hypothesize that high localization gives viral marketing an advantage because it concentrates the people who are wearing the viral marketing clothing. With concentrated numbers of people wearing the viral marketing clothing, their WOM influence and peer pressure on connected groups is much higher, thus having a higher chance of convincing a peer wearing the traditional marketing clothing to switch. The resurgence of the hush puppies in the mid 1990s provides a historical example of how a concentrated group of loyal consumers were able to virally promote a clothing trend through WOM and peer pressure (Gladwell). However, since these hypotheses are just that, hypotheses, a more formal study should go into considering these two variables in the context of viral marketing.

Another area of further research is to produce a model that addresses the weaknesses of the present model. First, the model doesn't take into account the fact that different individuals will have different influences on their peers. While research has shown that people tend to have the similar amounts of influence overall and that the notion of the high potential individual is flawed, it does not refute that fact that certain individuals are held in higher esteem than others. We rely on different people for different types of advice. We may rely on our best friends for fashion advice but the local car mechanic for advice on a new car purchase. As such, not all individuals have the same influence for the same products. This model does not take this consideration into account. Furthermore, we may hold connections on different levels of influence based on how knowledgeable we think they are in the subject matter. Malcolm Gladwell coined a term "mavens" to describe such individuals that are experts in a certain area. Such mavens would not only have more influence and be asked more often for advice, but will also offer advice to more people and share information about their area of expertise more often than non-mavens (i.e.- the dissemination rate of mavens would be higher as well) (Eccleston).

Lastly, given that this ABM only considers online advertisements, it is of potential interest to consider viral marketing in other marketing contexts. Advertisements on other mediums have different default settings in terms of interest and persuasion levels. Furthermore, mediums like billboards and magazines tend to lean heavily towards traditional advertisements as it is hard to make a viral ad on paper- imagine friends passing newspaper clipping to each other as a marketers pass at "viral newspaper marketing." Nevertheless, that does not stop marketers from trying to make viral ads on non-internet mediums. For example, many consider the use of apparel that contains the logo of the company a form of viral marketing, as the consumer both bought the clothing and is marketing it to his friends.

CONCLUSION

We conclude that there is substance to the hype surrounding viral marketing. By constructing an agent-based model, we determined that viral marketing is significantly more effective than traditional marketing in the context of a closed world where two clothing companies competed with one another for consumer loyalty, one using only online viral ads and the only online traditional ads. The study concludes that viral marketing has a small, yet statistically significant advantage over traditional marketing as a whole. The study also finds that viral marketing does significantly better in settings with high populations and settings with high levels of localization.

Viral marketing offers a way for small and large companies alike to overcome large consumer population sizes by quickly and virally penetrating through markets, generating mass awareness and converting large numbers of consumers. While online viral marketing as a whole does provide a significant advantage in terms of effectiveness over online traditional marketing, the results of an individual viral marketing campaign vary so much that it could easily either be total success of failure. Given the high variability of the effectiveness of viral marketing, it is a tool that should not be taken on lightly by a company. Viral marketing, though, is a gold mine waiting to be tapped into by the truly witty, creative, and innovative marketers.

Prominent business researcher Flint McGlaughlin once quipped that "people don't want to be 'marketed to'; they want to be 'communicated with.'" Viral marketing creates a paradigm of consumer involvement that is ideal for many companies. Not only are the consumers the marketers, but they happily fulfill this role by sharing products with their friends and peers. It is perhaps this paradigm that makes viral marketing so successful and so attractive. Nevertheless, much like Walt Whitman noted about great poetry requiring great audiences, at the heart of any great marketing campaign is a great audience.

APPENDIX A
JMP Output for VM:TM and % VM Multiple Regression Model

VM:TM Multiple Regression Model

Summary of Fit

RSquare	0.050518
RSquare Adj	0.050173
Root Mean Square Error	0.38067
Mean of Response	1.20137
Observations (or Sum Wgts)	5500

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	42.38203	21.1910	46.2362
Error	5497	796.56744	0.1449	Prob > F
C. Total	5499	838.94947		<.0001

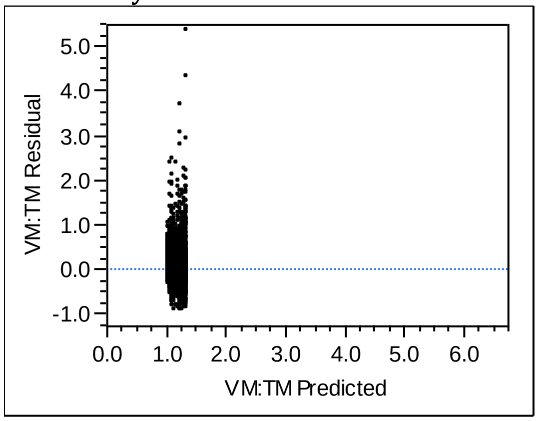
Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	52	37.34730	0.718217	5.1509
Pure Error	5445	759.22014	0.139434	Prob > F
Total Error	5497	796.56744		<.0001
				Max RSq
				0.0950

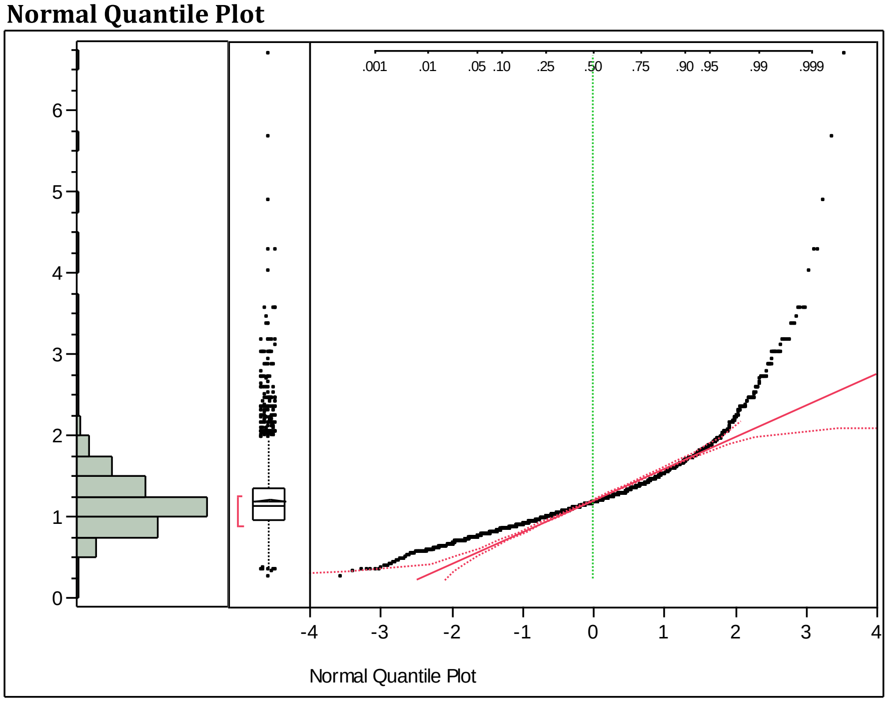
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.3818749	0.014518	95.18	0.0000
Gobalization	-0.00269	0.000162	-16.57	<.0001
Population	-0.000153	3.63e-5	-4.23	<.0001

Residual by Predicted Plot



Normal Quantile Plot



APPENDIX A (cont)
JMP Outputs for VM:TM and % VM Multiple Regression Model

% VM Multiple Regression Model

Summary of Fit

RSquare	0.049537
RSquare Adj	0.049191
Root Mean Square Error	0.069305
Mean of Response	0.533839
Observations (or Sum Wgts)	5500

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	1.376066	0.688033	143.2471
Error	5497	26.402756	0.004803	Prob > F
C. Total	5499	27.778823		<.0001

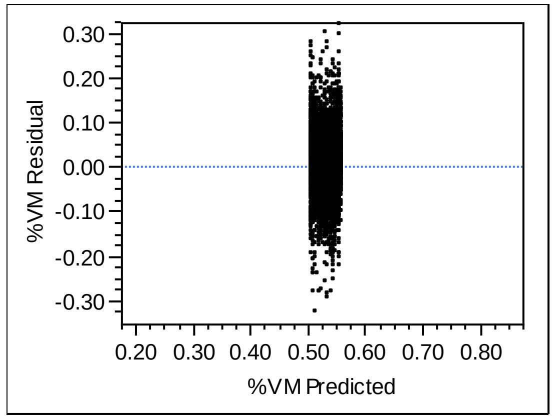
Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	52	1.174842	0.022593	4.8763
Pure Error	5445	25.227914	0.004633	Prob > F
Total Error	5497	26.402756		<.0001
				Max RSq
				0.0918

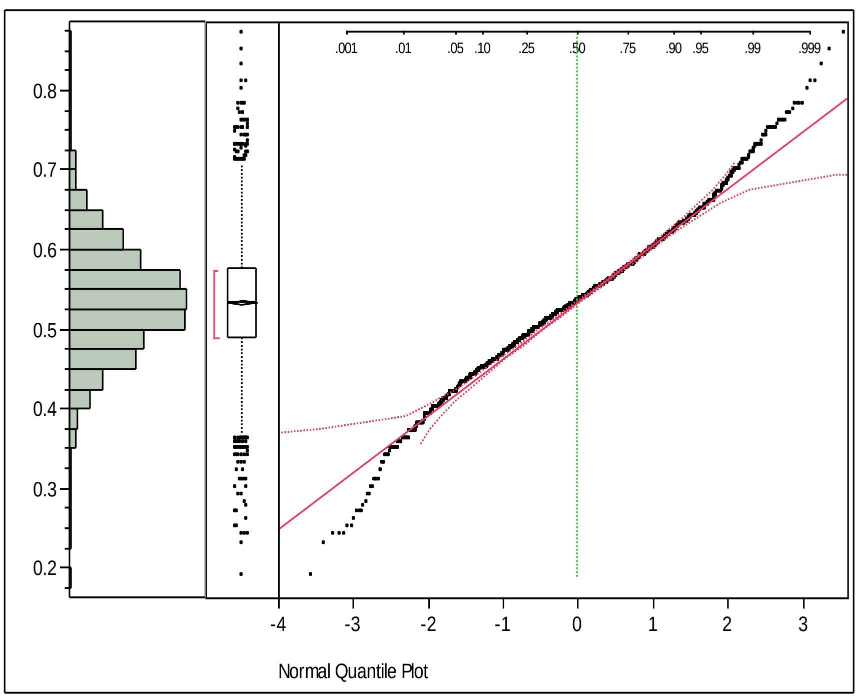
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.5543608	0.002643	209.73	0.0000
Gobalization	-0.000496	2.955e-5	-16.79	<.0001
Population	1.4276e-5	6.608e-6	2.16	0.0308

Residual by Predicted Plot



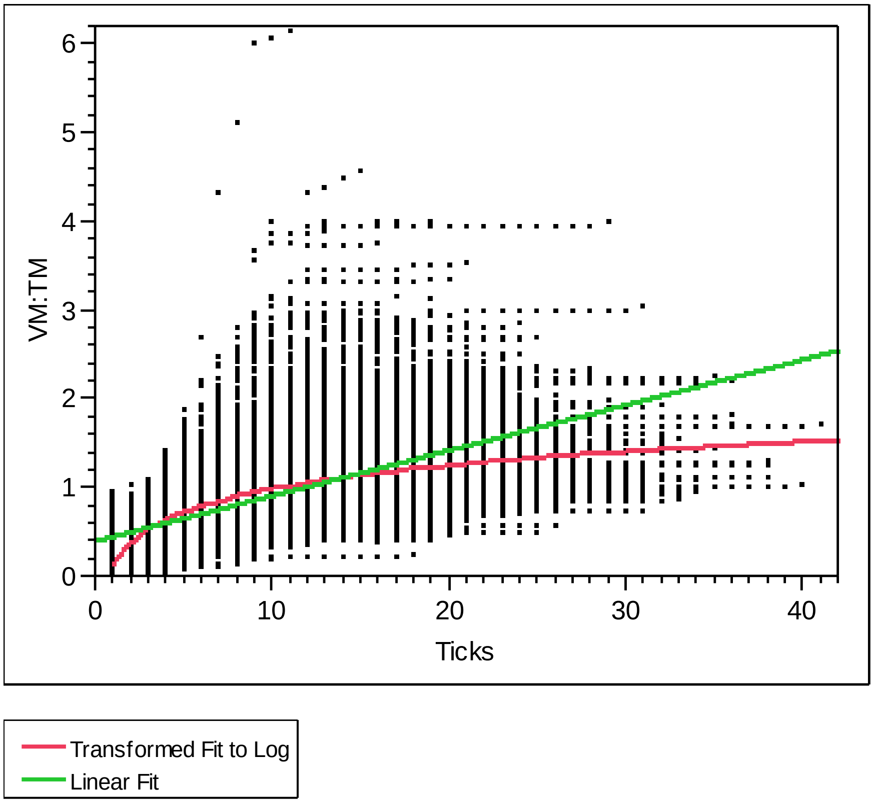
Normal Quantile Plot



APPENDIX B

JMP Outputs for % VM vs. Ticks & VM:TM vs. Ticks Regression Model

% VM vs. Ticks | Y to Log(x) Regression Model and Distribution of Ticks



Transformed Fit to Log
VM:TM = 0.1267348 + 0.3773661*Log(Ticks)

Summary of Fit

RSquare	0.523146
RSquare Adj	0.523141
Root Mean Square Error	0.288499
Mean of Response	0.876112
Observations (or Sum Wgts)	92295

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	8427.410	8427.41	101252.7
Error	92293	7681.680	0.083231	**Prob > F**
C. Total	92294	16109.090		0.0000

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.1267348	0.002539	49.91	0.0000
Log(Ticks)	0.3773661	0.001186	318.20	0.0000

Linear Fit
VM:TM = 0.3953544 + 0.0513261*Ticks

Summary of Fit

RSquare	0.488333
RSquare Adj	0.488328
Root Mean Square Error	0.298844
Mean of Response	0.876112
Observations (or Sum Wgts)	92295

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	1	7866.604	7866.60	88084.16
Error	92293	8242.486	0.089308	**Prob > F**
C. Total	92294	16109.090		0.0000

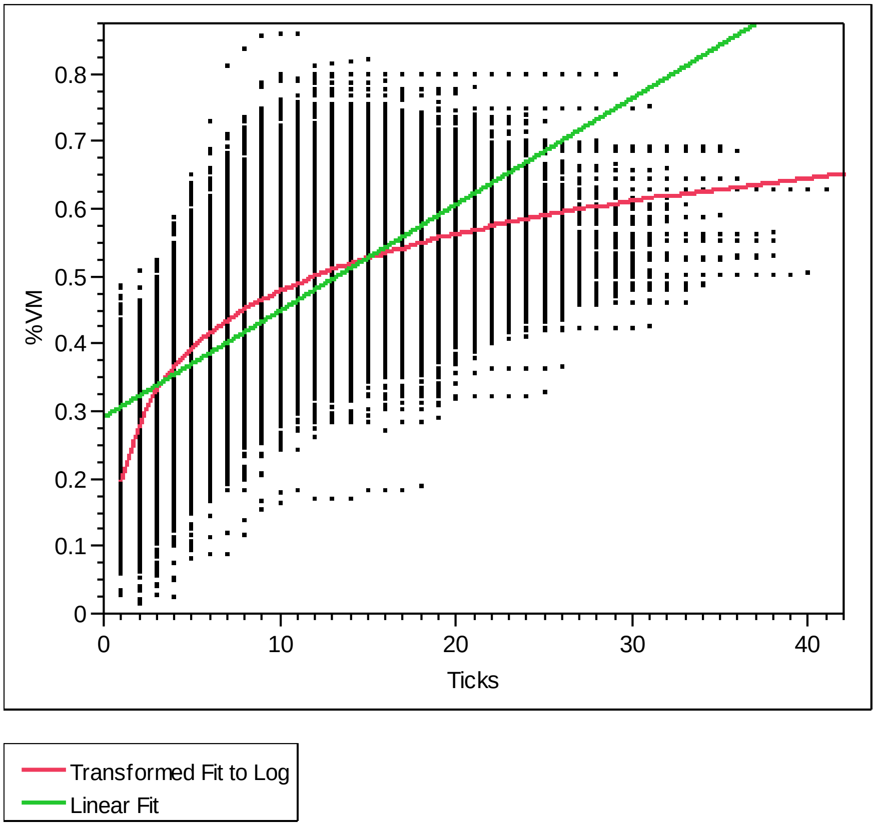
Parameter Estimates

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	0.3953544	0.001895	208.61	0.0000
Ticks	0.0513261	0.000173	296.79	0.0000

APPENDIX B (cont)

JMP Outputs for VM:TM vs. Ticks & % VM vs. Ticks Regression Model

% VM vs. Ticks | Y to Log(x) Regression Model and Distribution of Ticks



Transformed Fit to Log
%VM = 0.2003676 + 0.1214438*Log(Ticks)

Summary of Fit

RSquare	0.660885
RSquare Adj	0.660881
Root Mean Square Error	0.06966
Mean of Response	0.441532
Observations (or Sum Wgts)	92295

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	1	872.8089	872.809	179865.4
Error	92293	447.8581	0.004853	**Prob > F**
C. Total	92294	1320.6670		0.0000

Parameter Estimates

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	0.2003676	0.000613	326.79	0.0000
Log(Ticks)	0.1214438	0.000286	424.11	0.0000

Linear Fit
%VM = 0.2941099 + 0.0157389*Ticks

Summary of Fit

RSquare	0.560101
RSquare Adj	0.560096
Root Mean Square Error	0.079339
Mean of Response	0.441532
Observations (or Sum Wgts)	92295

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	1	739.7067	739.707	117511.9
Error	92293	580.9603	0.006295	**Prob > F**
C. Total	92294	1320.6670		0.0000

Parameter Estimates

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	0.2941099	0.000503	584.55	0.0000
Ticks	0.0157389	0.000046	342.80	0.0000

APPENDIX C
JMP Outputs for Distribution of Ticks Taken to Reach Equilbrium

Distribution of Ticks Taken to Reach Equilibrium

Quantiles

100.0%	maximum	45.000
99.5%		31.000
97.5%		26.000
90.0%		22.000
75.0%	quartile	19.000
50.0%	median	16.000
25.0%	quartile	14.000
10.0%		12.000
2.5%		10.000
0.5%		9.000
0.0%	minimum	7.000

Moments

Mean	16.835091
Std Dev	4.0768983
Std Err Mean	0.0549729
upper 95% Mean	16.94286
lower 95% Mean	16.727322
N	5500

REFERENCES

Bampo, Mauro, et al. "The Effects of the Social Structure of Digital Networks on Viral Marketing Performance." Information Systems Research 19.3 (2008): 273-90.

De Bruyn, Arnaud, and Gary L. Lilien. "A Multi-Stage Model of Word-of-Mouth Influence through Viral Marketing." International Journal of Research in Marketing 25.3 (2008): 151-63.

Duncan J. Watts, Six Degrees: The Science of a Connected Age. (W.W. Norton & Company, New York, 2003), pages 83-100.

Eccleston, Derek, and Luca Griseri. "How does Web 2.0 Stretch Traditional Influencing Patterns?" International Journal of Market Research 50.5 (2008): 591-161.

Ferguson, Rick. "Word of Mouth and Viral Marketing: Taking the Temperature of the Hottest Trends in Marketing." Journal of Consumer Marketing 25.3 (2008): 179-82.

Gladwell, Malcolm. The Tipping Point: How Little Things Can Make a Big Difference. Boston: Little, Brown and Company, 2000.

Masker, Stephen J., and Andrew Somosi. "Marketing to Teens Online." McKinsey Quarterly.4 (2004): 12-.

Reuters. Poll finds nearly 80 percent of U.S. adults go online. 5 November 2007. 2 December 2008 <http://www.reuters.com/article/internetNews/idUSN0559828420071106>.

Smith, Ted, et al. "Reconsidering Models of Influence: The Relationship between Consumer Social Networks and Word-of-Mouth Effectiveness." Journal of Advertising Research 47.4 (2007): 387-97.

Thackeray, Rosemary, et al. "Enhancing Promotional Strategies within Social Marketing Programs: Use of Web 2.0 Social Media." Health Promotion Practice 9.4 (2008): 338-43.

Wilensky, U. (2005). NetLogo Small Worlds Model. <http://ccl.northwestern.edu/netlogo/models/SmallWorlds>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL