MULTI-STEP “WORD-OF-MOUTH” COMMUNICATION INFLUENCE IN MARKETING: A MATHEMATICAL MODEL

ELY MERZBACH AND JACQUES PICARD

Abstract. The number of mathematical models measuring “Word-of-Mouth” influence in marketing has been quite limited so far. The authors propose a new model based on the theory of branching processes that provides a framework for the computation of the effect of specific marketing communication. It also takes into account “spin-off” referrals. This model measures Word-of-Mouth impact on marketing and also permits to compute the probability of exhaustion of new customers at a specific given time.

Key words and phrases. Word-of-Mouth, advertising, mathematical model, branching process, communication, marketing.
1. Introduction

The power of “Word-of-Mouth” advertising, or the influence of non-commercial and non-public sources of information on purchase activities, has been recognized for a long time (Katz and Lazarfeld (1955), Brooks (1957)). Since then many researches have been done on the subject.


The impact of redress and incentives by marketers on Word-of-Mouth has been examined by Blodgett et al. (1993), Maxham (2001), Swanson et al. (2001), Wirtz et al. (2002), Moore et al. (2004), Hocutt et al. (2005).


The interaction between use of internet and Word-of-Mouth was the subject of other studies: Kieker et al. 2001), Boush et al. (2001), Phelps et al. (2004), Datta et al. (2005), Gruen et al. (2006), Thorson (2006), Vilpponen et al. (2006), Alon et al. (2006), Wojnicki (2006).


The question surrounding the management of this form of communication has drawn the attention of: Wirtz et al. (1977), Bayus (1985), Haywood (1989), File (1992), Ennew et al. (2000), Wirtz et al. (2002), Gremler et al. (2001).

The interface between national culture and Word-of-Mouth advertising has also been investigated: Gothard (2001), Bruce (2004), Lam et al. (2005), Fong et al. (2006), Podoshen (2006).

However, not much research has been done in finding mathematical models which measure quantitatively “Word-of-Mouth” influence in marketing. Among the few works in this area are the following:
Monahan (1984) presents a model of the diffusion of a new product in which the rate of sales depends upon the advertising and Word-of-Mouth communication between customers and noncustomers. However, the required conditions are very stringent and the main result has no direct practical application, as far the estimation of the effect of work is concerned.

In Harrison-Walker (2001), a scale to measure Word-of-Mouth communication is developed. This scale is based on statistical hypothesis testing: Different hypotheses concerning the relations between commitment service quality and Word-of-Mouth activity or Word-of-Mouth praise are treated. The empirical results are very partial. Moreover, no effort is made to measure the Word-of-Mouth itself and its influence along time.

Helm (2003) attempts to discuss different approaches to the calculation of positive Word-of-Mouth, leading to a monetary referral value of a company’s customers. She underlines the difficulty “to develop a quantitative model to capture a marketing phenomenon that most researchers and managers shrink back from addressing at all”. Later she writes: “Measuring a phenomenon does not necessarily mean that it has to be monetized”, concluding that the “Word-of-Mouth” is “outside the domain of pure metrics”.

Hogan et al. (2004) show that it is possible to quantify the way in which Word-of-Mouth often complements and extends the effects of advertising. Individual customer behavior is directed onto a measure of long-term firm value, obtaining “Customer Life-time Value”. It is calculated as the expected net profit that will be received from a customer over a specific time horizon. However, it does not take into consideration many different cases. For example, only the effect of positive Word-of-Mouth is examined, but not the effect of negative Word-of-Mouth. Moreover, only the case where the propensity to spread Word-of-Mouth is uniform is studied.

Lee et al. (2006) use some company data and simulation in order to empirically investigate the effect of Word-of-Mouth in estimating CLV (Customer Life-time Value). Here, too, negative Word-of-Mouth is not considered. Moreover, the model described is very specific and does not cover the general case of Word-of-Mouth.

Finally, in Dwyer (2007), a metric is defined for measuring the value a community assigns each Word-of-Mouth instance and the value the community assigns to the members that create them. However, only electronic Word-of-Mouth is discussed and some very specific assumptions are necessary for defining his metric.

In contrast, in our mathematical model presented here, instead of trying to measure the Word-of-Mouth value of a single customer, we have attempted to be more comprehensive and have computed the total value of Word-of-Mouth for a marketing communication.
2. The Mathematical Model

The mathematical model presented here is random and based on the classical theory of branching processes.\footnote{For a basic understanding of the theory of branching processes, see Karlin-Taylor (1975).}

Consider a product which must be marketed in a large population.

Each person who hears about this product can be either “non-influenced” or “influenced but not-buyer”, or a “buyer”. In other words, each person belongs to one and only one of the three states \{0, 1, 2\}:

\begin{align*}
0 & = \text{non-influenced} \\
1 & = \text{influenced, but not-buyer (thereafter called “influenced”)} \\
2 & = \text{buyer}.
\end{align*}

The “Word-of-Mouth” influence works in the following manner:

Each influenced person, and each buyer, every day meets different people with whom he or she might talk about the product and so exerts some influence on them.

The model given here is aimed to specify the expected number of influenced persons and the expected number of buyers in the population. Also, we will present some tools permitting to compute the distribution, the mean and the variance of the “power of influence” of the individuals of the population.

Moreover, we will present a tool which will enable the computation of the probability of “extinction”: No further people are influenced or become buyers.

The mathematical model constructed here is a kind of “two-type branching process”, defined first by T. Harris (1963) for biology models and developed later using vector-valued martingale theory. In Gonzales-Martinez-Mota (2006), this model is generalized in the case the population is “dependent of springs”.

We use essentially the same notation as in Karlin-Taylor (1975) as well as other classical mathematical results from branching theory. Let \( I_n \) and \( B_n \) be the number of individuals who are influenced persons and buyers, respectively, in the \( n \)th generation. We may write

\begin{align*}
I_{n+1} &= \sum_{j=1}^{U_n} \xi_j^{(1)} + \sum_{j=1}^{V_n} \xi_j^{(2)}, \\
B_{n+1} &= \sum_{j=1}^{U_n} \zeta_j^{(1)} + \sum_{j=1}^{V_n} \zeta_j^{(2)},
\end{align*}

where \( \xi, \zeta \) denote the influence coefficients, and \( \xi_j^{(1)}, \xi_j^{(2)}, \zeta_j^{(1)}, \zeta_j^{(2)} \) are the contributions of each influenced person and buyer, respectively, in the \( n \)th generation.
where \((\xi_j^{(i)}, \zeta_j^{(i)})\) are independent, identically distributed, random vectors with distribution
\[
\Pr\{\xi_j^{(i)} = k, \zeta_j^{(i)} = l\} = p_i(k, l), \quad k, l = 0, 1, 2, \ldots,
\]
for \(j = 1, 2, \ldots\) and \(i = 1, 2\).

Here \(p_i(k, l) \geq 0\) and \(\sum_{k,l=0}^{\infty} p_i(k, l) = 1\) for \(i = 1, 2\).

In other words, \(p_1(k, l)\) and \(p_2(k, l)\) are the probabilities that an influenced person and a buyer, respectively, produces \(k + l\) direct “descendants” of which \(k\) are influenced persons and \(l\) are buyers.

The time unit is generally a day, but in some circumstances, it may be something else (e.g., a week).

We assume the process begins with a single individual, i.e., we assume either
\[
I_0 = 1 \quad \text{and} \quad B_0 = 0 \quad (1)
\]
or
\[
I_0 = 0 \quad \text{and} \quad B_0 = 1. \quad (2)
\]

We introduce the pair of two-dimensional probability generating functions
\[
\varphi^{(i)}(s, t) = \sum_{k,l=0}^{\infty} p_i(k, l) s^k \cdot t^l, \quad i = 1, 2,
\]
that is,
\[
\varphi^{(1)}_n(s, t) = \sum_{k,l=0}^{\infty} \Pr\{I_n = k, B_n = l \mid I_0 = 1, B_0 = 0\} s^k \cdot t^l,
\]
\[
\varphi^{(2)}_n(s, t) = \sum_{k,l=0}^{\infty} \Pr\{I_n = k, B_n = l \mid I_0 = 0, B_0 = 1\} s^k \cdot t^l.
\]

The generating function of (1) is
\[
\varphi^{(1)}_0(s, t) \equiv s,
\]
and that of (2) is
\[
\varphi^{(2)}_0(s, t) \equiv t.
\]

In addition,
\[
\varphi^{(i)}_1(s, t) = \varphi^{(i)}(s, t) \quad \text{for} \quad i = 1, 2.
\]

It is not difficult to prove the following important formula:
\[
\varphi^{(i)}_{n+m}(s, t) = \varphi^{(i)}_m(\varphi^{(1)}_n(s, t), \varphi^{(2)}_n(s, t)) \quad (3)
\]
for \(i = 1, 2\) and \(n, m = 0, 1, 2, \ldots\).
Computation of the expectations (the number of influenced people and the number of buyers on each day) can be done in two different ways, either using generating functions and formula (3), or using the theory of vector-valued martingales.

Let us introduce the following notation. Let $X_n = (I_n, B_n)$ be the two-dimensional vector with components $I_n$ and $B_n$. Let

- $m_{11} = E\{I_1 | I_0 = 1, B_0 = 0\} = E\xi^{(1)}$,
- $m_{12} = E\{B_1 | I_0 = 1, B_0 = 0\} = E\zeta^{(1)}$,
- $m_{21} = E\{I_1 | I_0 = 0, B_0 = 1\} = E\xi^{(2)}$,
- $m_{22} = E\{B_1 | I_0 = 0, B_0 = 1\} = E\zeta^{(2)}$,

and introduce the matrix of expectations

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}.$$  

Thus $m_{11}$ and $m_{12}$ are the expected numbers of influenced people or buyers, respectively, produced by an influenced person, and $m_{21}$ and $m_{22}$ are the expected numbers of influenced people or buyers, respectively, produced by a buyer.

**Theorem 1.**

$$E[X_{n+r} | X_n] = X_n M^r \quad \text{for} \quad r, n = 0, 1, 2, \ldots \quad (4)$$

**Remark:** This matrix identity means that if $M$ is invertible, then $\{X_nM^{-n}, \ n = 0, 1, \ldots \}$ is a vector-valued martingale. Therefore, since it is non-negative, this vector-valued martingale converges almost surely. Then, either $X_n \to (0, 0)$, or $X_n \to (0, \infty)$, or $X_n \to (\infty, 0)$, or $X_n \to \infty$ at an exponential rate.

**Proof of the Theorem.** (See Karlin-Taylor (1975).) The proof for $r = 1$ proceeds directly. Thus,

$$E[X_{n+1} | X_n] =$$

$$\left( E \left[ \sum_{j=1}^{U_n} \xi_j^{(1)} + \sum_{j=1}^{V_n} \xi_j^{(2)} \ | \ (I_n, B_n) \right], E \left[ \sum_{j=1}^{U_n} \xi_j^{(1)} + \sum_{j=1}^{V_n} \xi_j^{(2)} \ | \ (I_n, B_n) \right] \right)$$

$$= (m_{11} I_n + m_{21} B_n, m_{12} I_n + m_{22} B_n)$$

$$= (I_n, B_n) \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

$$= X_n \cdot M.$$
We now assume that relation (4) holds for \( r \) and prove it for \( r + 1 \). By the Markov property for \( \{X_n\} \), we have

\[
E[X_{n+r+1} \mid X_n] = E\{E[X_{n+r+1} \mid X_{n+r}, \ldots, X_n] \mid X_n\} = E\{E[X_{n+r} \mid X_n] \} = E\{X_{n+r} \mid X_n\} \cdot M \quad \text{(using the induction hypothesis)}
\]

\[
= X_n M^{r+1}.
\]

\( \square \)

Notice and denote the following:

\[
\frac{\partial \varphi^{(1)}_n(s, t)}{\partial s} \bigg|_{s=t=1} = E[I_n \mid I_0 = 1, B_0 = 0] = m_{11}^{(n)},
\]

\[
\frac{\partial \varphi^{(1)}_n(s, t)}{\partial t} \bigg|_{s=t=1} = E[B_n \mid I_0 = 1, B_0 = 0] = m_{12}^{(n)},
\]

\[
\frac{\partial \varphi^{(2)}_n(s, t)}{\partial s} \bigg|_{s=t=1} = E[I_n \mid I_0 = 0, B_0 = 1] = m_{21}^{(n)},
\]

\[
\frac{\partial \varphi^{(2)}_n(s, t)}{\partial t} \bigg|_{s=t=1} = E[B_n \mid I_0 = 0, B_0 = 1] = m_{22}^{(n)},
\]

and

\[
M^{(n)} = \begin{bmatrix}
    m_{11}^{(n)} & m_{12}^{(n)} \\
    m_{21}^{(n)} & m_{22}^{(n)}
\end{bmatrix}.
\]

**Corollary.**

\[
M^{(n)} = M^n \quad \text{and} \quad E[X_n] = E(I_n, B_n) = (EI_n, EB_n) = E[X_0]M^n,
\]

\( n = 1, 2, \ldots \).

The proof is immediate from (4).

We can proceed in the same manner for the variances:

\[
v_{11}^{(n)} = \text{Var}[I_n \mid I_0 = 1, B_0 = 0]
\]

\[
v_{12}^{(n)} = \text{Var}[B_n \mid I_0 = 1, B_0 = 0]
\]

\[
v_{21}^{(n)} = \text{Var}[I_n \mid I_0 = 0, B_0 = 1]
\]

\[
v_{22}^{(n)} = \text{Var}[B_n \mid I_0 = 0, B_0 = 1],
\]

and

\[
V^{(n)} = \begin{bmatrix}
    v_{11}^{(n)} & v_{12}^{(n)} \\
    v_{21}^{(n)} & v_{22}^{(n)}
\end{bmatrix}.
\]

There is no direct formula for computing the variance \( \text{Var}(X_n) \) since it depends also on the correlation between the two processes \( \{I_n\} \) and \( \{B_n\} \). For a one-type branching process
\{I_n\}, supposing \(I_0 = 1\), we have

\[
\text{Var}(I_n) = \begin{cases} 
vm^{n-1}m^{-1} & \text{if } m \neq 1 \\
vm & \text{if } m = 1
\end{cases}
\]

where \(m = m^{(1)}_{11} = E(I_1)\) and \(v = v^{(1)}_{11} = \text{Var}(I_1)\).

A similar formula holds for the one-type branching process \(\{B_n\}\).

More generally, we can compute the \(4 \times 4\) covariance matrix.

The formulas are very cumbersome and therefore they will not be given here. Notice only that the matrix variance \(V^{(n)}\) is linear with \(n\) if and only if the determinant of \(M\) is one.

For this two-dimensional branching process, we may introduce the “extinction” probabilities

\[
\begin{align*}
\pi^{(1)} &= \Pr\{I_n = B_n = 0 \text{ for some } n \mid I_0 = 1, B_0 = 0\}, \\
\pi^{(2)} &= \Pr\{I_n = B_n = 0 \text{ for some } n \mid I_0 = 0, B_0 = 1\}.
\end{align*}
\]

It is well-known (Frobenius theorem) that if \(M\) is a matrix with positive elements (symbolically written here as \(M \gg 0\)), then the eigenvalue of largest magnitude is positive and real. This eigenvalue is designated as \(\rho(M) = \rho\).

It is convenient to introduce the vector notations

\[
\begin{align*}
\phi(u) &= (\varphi^{(1)}(s,t), \varphi^{(2)}(s,t)), \\
\phi_n(u) &= (\varphi^{(1)}_n(s,t), \varphi^{(2)}_n(s,t)), \\
\pi &= (\pi^{(1)}, \pi^{(2)}), \\
1 &= (1, 1).
\end{align*}
\]

Then we may state a general result on the extinction probability \(\pi\) and a method of computation for this probability:

**Theorem 2.** Assume that the components of \(\phi(u)\) are not linear functions of \(s\) and \(t\) and that \(M \gg 0\) (every element of \(M\) is positive). Then \(\pi = 1\) if the largest eigenvalue \(\rho\) of \(M\) does not exceed one and \(\pi \ll 1\) if \(\rho > 1\). (The notation \(u \ll v\) signifies that \(v - u\) has positive components.) In the case \(\rho > 1\), \(\pi\) is the smallest nonnegative solution of

\[
u = \phi(u), \quad u \ll 1.
\]

Moreover, if \(q\) is any vector in the unit square other than \(1\), then \(\lim_{n \to \infty} \phi_n(q) = \pi\), and the only nonnegative solutions of (5) are 1 and \(\pi\).
The proof of the theorem can be found in Karlin-Taylor (1975).

The fact that $\phi_n(q)$ converges to $\pi$ gives a practical (but quite cumbersome) way for computation of the probability of extinction.

In practical situations, the process is stopped at a certain point of time (i.e., after a certain number of days) $D$. This stopping time may be constant or random. It may be random if the number of days is not known in advance and may depend on many other parameters. By Doob’s optional stopping martingale theorem (see Feller (1958)), the process $\{X_{n\wedge D}M^{-(n\wedge D)}, n = 0, 1, \ldots\}$ is a vector-valued martingale, and therefore $E[X_{n\wedge D}] = E[X_0]E[M^{n\wedge D}]$. We also can compute the expectations of the total number of influenced people and buyers until time $D$ by the following new formula:

$$E\left[\sum_{n=0}^{D} X_n\right] = \sum_{\alpha} E[X_0] \sum_{n=0}^{D} M^\alpha P\{D = \alpha\}.$$

Therefore, if the initial values are known, and the matrix of expectations $M$ is known, this formula gives the mean value of influenced persons and the mean value of buyers.

General variance and covariance formulas, as well as a Functional Central Limit Theorem, will be discussed in a forthcoming paper.

### 3. Discussion and Conclusion

For the first time a model of branching process is being used to measure Word-of-Mouth impact in marketing. It is a model that fits well with the nature and structure of the Word-of-Mouth process, and as such provides a framework (that may be adjusted for different product-market situations), for the calculation of the effect of a specific marketing communication. It also takes into consideration "spin-off" referrals (referrals by non-buyers, enthusiastic about a new offering). The lack of consideration for that effect had been a major drawback in previous models.

One of the main advantages of this model is its simplicity. If one knows that a person has bought a product, or was favorably impressed by it, one should be able to forecast the number of buyers that this initial phenomenon will bring about, after a certain amount of time. The only inputs that have to be known for this purpose are the elements of Matrix $M$. In other words, we only need to compute the different expectations: $m_{11}$, $m_{12}$, $m_{21}$, $m_{22}$. These values can be obtained quite easily, on the basis of historical or sample data.

One could also predict the monetary value of a marketing communication (for example by a salesperson) in a virgin market. This can be done by evaluating the number of people...
having bought or having been influenced by that communication, and then forecasting through our model the number of sales that would result after a certain period of time.
REFERENCES


Department of Mathematics, Bar-Ilan University, 52900, Ramat-Gan, Israel
E-mail address: merzbach@macs.biu.ac.il

Netanya Academic College, Netanya, Israel
E-mail address: picard.jacques@uquam.ca