

# Optimal Heterophily for Word-of-Mouth Diffusion

**Hikaru Yamamoto**

Seikei University  
3-3-1 Kichijoji-kitamachi,  
Musashino-shi, Tokyo  
yamamoto@econ.seikei.ac.jp

**Naohiro Matsumura**

Osaka University  
1-7 Machikaneyama,  
Toyonaka, Osaka  
matumura@econ.osaka-u.ac.jp

## Abstract

Most past research on word-of-mouth (WOM) communication has focused on the attributes of the WOM sender, such as opinion leadership and innovativeness. However, since WOM communication involves the interaction of the sender and the receiver, it is important to examine the relationship between the two. Past research on personal influence which examines the relational context includes two contradicting arguments: one supporting the power of homophily, and the other supporting the power of heterophily. In this paper, the authors focus on the social relation between WOM senders and receivers and attempt to find the optimal heterophily for WOM diffusion. The results of empirical analysis using both offline survey data and online blogosphere provide evidence of optimal heterophily between WOM sender and receiver. Finally, the patterns of influence are presented to depict the diffusion process.

## Introduction

The emergence of weblogs and social media has made the discovery of ways to build successful WOM marketing programs a central topic in marketing. Most of these programs attempt to find special consumers who are particularly influential in WOM diffusion, and these individuals are given various names such as “influentials”, “e-fluentials”, “influencers”, “evangelists”, “bees”, and “sneezers”.

Part of the reason these programs focus on elites among consumers is that most of the past WOM research has focused on the attributes of the WOM sender, such as opinion leadership and innovativeness.

WOM involves the interaction of sender and receiver, though, so it is important to examine the relationship between the two. The outcome of WOM communication, such as a change in consumer attitude or purchase behavior depends not only on the attributes of the sender, but also on

social relationships. Our work has focused on the social relationship between sender and receiver, especially the homophily. Homophily in the sociology literature is defined as a “principle that contact between similar people occurs at a higher rate than among dissimilar people” (McPherson, Smith-Lovin, and Cook 2001). Both social scientists and computer scientists have done extensive research on this phenomenon (Crandall et al. 2008; Hogg, et al. 2008; Singla and Richardson 2008).

On the other hand, most of the marketing literature uses this term as a synonym for similarity. In this paper, we follow Rogers (2003)’s definition, which defines homophily as “the degree to which two or more individuals who interact are similar in certain attributes”.

The past research on personal influence that has examined the relational context includes two contradicting arguments: one supporting the power of homophily and the other supporting the power of heterophily (Gatignon and Robertson 1985; Rogers and Bhowmik 1970).

When a WOM sender and a receiver are homophilous, the sender is unlikely to know anything more than the receiver. Thus, a WOM receiver is more likely to be exposed to new ideas when interacting with a WOM sender who is dissimilar, but at some point the difference becomes so great that communication suffers. This suggests the presence of optimal heterophily (Alpert and Anderson 1973, Kaigler-Evans, Leavitt and Dickey 1978).

The aim of our research has been to focus on the relational context of WOM and examine the point of optimal heterophily to stimulate WOM communication. This paper is organized as follows. First, the influence diffusion model (IDM), which we use to calculate the influence of each blogger, is presented. After that, we present the result of empirical analysis. The optimal heterophily for the diffusion of WOM is examined in both offline and online environments. Finally, cascade patterns are presented.

## Influence Diffusion Model

In this section, we describe the influence diffusion model (IDM). The IDM was originally an algorithm for measuring values for the influence of messages, senders, and terms from online bulletin boards. Recently, the algorithm was revised and expanded to measure the influence of bloggers (Matsumura, Yamamoto, and Tomozawa 2008).

Throughout the paper, we use the term “blogger” to mean an individual who keeps and updates an online blog. We use the term “postings” to mean an entry created by a blogger on his or her blog.

The IDM recursively calculates the spread of terms in the blogosphere and evaluates the influence of each term, blog entry, and blogger. Figure 1 depicts a simple inter-blog posting relationship.

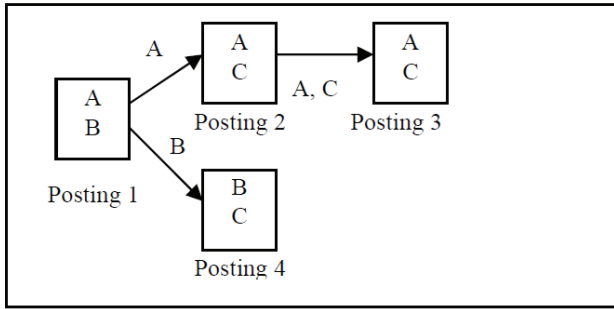


Figure 1. The process of influence diffusion

Posting 1 contains terms A and B, and posting 2, which was posted after Posting 1 and has a link relation (expressed as an “edge”) containing terms A and C. In such a situation, we assume that term A propagated from posting 1 to posting 2. The arrow represents the propagation of the terms. Our assumption is that if the same term appears in all of the postings connecting the two postings with a link or trackback, the blog posted later was “influenced” by the first blog.

Let’s denote the set of terms included in posting 1, 2, 3, and 4 as  $w_1, w_2, w_3, w_4$ . The number of terms propagating from a posting  $x$  on a posting  $y$  ( $x$  precedes  $y$ ) is defined as

$$n_{x \rightarrow y} = |w_x \cap w_y| \quad (1)$$

$|w_x \cap w_y|$  represents the number of terms which appear in all postings between postings  $x$  and posting  $y$ . Next, we define the influence of posting  $x$ , denoted  $i_x$ , as the sum of propagating terms in the blogosphere:

$$i_x = \sum_{y \in \text{all\_postings}} n_{x \rightarrow y} \quad (2)$$

Using the influence of postings, the influence of each blogger can be measured. The influence of blogger  $S_x$ , which is denoted  $I_x$  can be considered as the sum of influence of his or her postings, as shown in Formula 3.

$$I_x = \sum_{y \in \text{all\_postings\_by\_x}} i_y \quad (3)$$

In the next section, the IDM is applied to official data provided by a blog hosting service, and the optimal heterophily between influencers and influencees is examined.

## Empirical Analysis

### Data Description

The data was provided by NIFTY Research Institute, a division of NIFTY Corporation. Nifty Corporation is one of the largest and oldest Japanese Internet service providers, and it provides a blog hosting service and various forms of content. NIFTY Research Institute offers NIFTY Buzz Marketing Solution, a blog mining and consulting service to advertisers.

To determine the optimal heterophily among bloggers, this paper analyzes the diffusion of market-relevant information concerning a new shampoo product from Shiseido. We collected blogs posted from March to August 2006. The blogs containing the name of the shampoo were then screened. A total of 11,001 postings were used for the analysis. These postings were manually investigated by the staff of NIFTY Research Institute to classify them according to the stage of consumer purchase behavior.

Among the 11,001 postings, 3,247 postings (29.5%) were classified as *customer*; i.e., the blogger actually purchased the product. 3,698 postings (33.6%) were tagged as *potential customer*, meaning the postings talked about the product but the blogger did not reach the stage of purchase. 4,056 postings (36.9%) were irrelevant or spam blogs.

As explained in the previous section, we consider a posting to be “influenced” when it is one of two postings sharing a link or trackback relations, and it uses a term appearing in the earlier posting. For this reason, we obtained the link/trackback relationships among the postings: there was a total of 1,333 links among 8,292 blog sites.

### Optimal Heterophily

**Study 1: Optimal Heterophily in the Blogosphere.** In this section, we examine the optimal heterophily in the blogosphere. We used the IDM to calculate the influence of each blogger in the previous section. Here, the influence gap between the sender and the receiver is used as the heterophily measure.

Let’s denote  $I_s$  as the outgoing influence of the sender and  $I_r$  as that of the receiver. The online heterophily  $H_{online}$  is defined as

$$H_{online} = I_r - I_s \quad (4)$$

Figure 2 is a histogram of the influence gap. On the left side of the graph,  $H_{online} < 0$ , meaning that the sender has more influence than the receiver. In this area, the number of WOM pairs increases as heterophily decreases (i.e., as  $H_{online}$  increases).

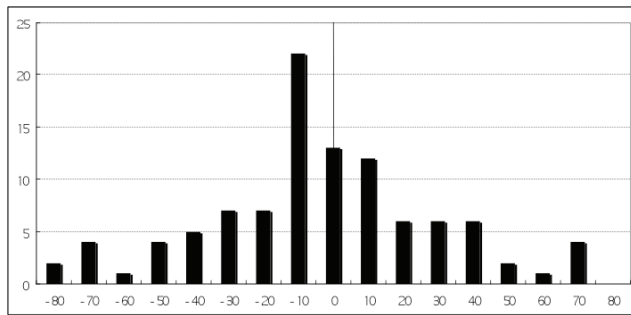


Figure 2. Influence gap in the blogosphere

On the right side of the graph,  $H_{online} > 0$  and the receiver has more influence than the sender. The number of WOM pairs decreases as the gap increases.

The peak of  $H_{online}$  is -10 to -19, suggesting the presence of optimal heterophily between the sender and receiver. Contrary to the commonly accepted notion, the influence does not come from a distant super node, but from ordinary bloggers who are slightly more influential than the receiver.

**Study 2: Optimal Heterophily in the Offline Environment.** To validate our findings in the blogosphere and gain insight into WOM behavior in the offline environment, we analyzed the data on a dyadic study of WOM influence (Yamamoto et al. 2008). The survey was conducted from August 10, 2007 to August 11, 2007, and collected 500 female samples from age 15 to 39. To increase the number of WOM dyads within a limited sample, the topic was broadened from shampoo to cosmetic and beauty products in general.

The respondents were asked to list the initials of five individuals with whom they talk about cosmetics. As a result, a total of 1,684 dyads were collected. The respondents then named the most desirable WOM sender among their list.

From the offline survey, it is difficult to measure the influence of each consumer. For this reason, we measure the category knowledge and define the gap as heterophily. Let's denote  $K_s$  as the category knowledge of the sender and  $K_r$  as that of the receiver. The heterophily is calculated as

$$H_{offline} = K_r - K_s \quad (5)$$

To determine the optimal heterophily, the category knowledge of each respondent (on a 10-point scale) and that of five WOM partners was measured. Since we did not directly measure the category knowledge of the WOM partners, it was measured based on respondents' assumptions.

Figure 3 shows the optimal heterophily in the offline environment. The dots represent actual data. The number of WOM pairs was larger on the left side of the graph, where  $H_{offline} < 0$ . This suggests that the WOM receivers wanted information from someone more knowledgeable regarding the category.

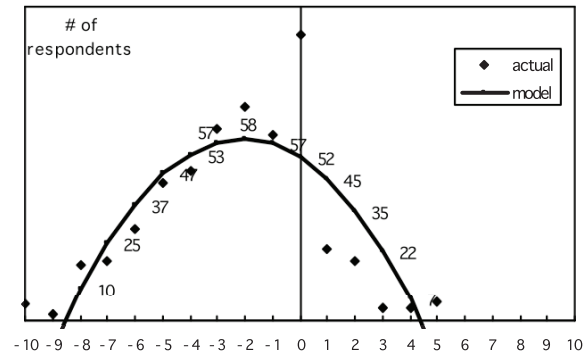


Figure 3. Desirable WOM Sender and Difference in Category Knowledge

Confirming the result from the online environment, the number of WOM dyads increased as the heterophily decreased. The curve is the result of a simple quadratic estimation. The peak of the model is at -2, meaning the WOM receiver wants to receive information from someone slightly more knowledgeable, not from a distant expert.

According to most past research, influence comes from an elite few. Watts and Dodds (2007) examined this "influentials hypothesis" and found that large cascades of influence are driven not by influentials but by a critical mass of easily influenced individuals. The result of our analysis confirms their findings. The influence comes from someone just a little bit knowledgeable, not a distant expert.

### Patterns of Influence in the Blogosphere

Through the analysis of optimal heterophily, we have shown that influence comes from someone just a little bit more knowledgeable, not a distant expert. This suggests that in the diffusion of WOM information, many *grassroots influentials* play the central role instead of a few *super influentials*.

Rogers (2003) states that he became aware of diffusion systems that did not operate at all like a centralized diffusion model. Instead, he points out the importance of decentralized diffusion systems, where new ideas spread horizontally via peer networks.

To understand the patterns of influence from a macro-perspective, here we depict the patterns of influence in the blogosphere. Figure 4 is a directed WOM network where each blogger is represented by a node and an edge is placed between bloggers if there is influence; i.e., propagation of terms. The number located near the edge indicates outgoing influence in particular relationships. The size of the node represents the influence of each blogger. The most influential bloggers are located in the upper part of the figure.

Using the manually tagged stage of consumer behavior, we classified the nodes. A square node represents a *potential customer*, a triangle node represents a *customer*, and a circle node represents a blogger who was a *potential customer* and later became a *customer*. In other words, circle nodes show the network effect where there was an attitude change. Since 46 million dollars was spent to market this shampoo product, this attitude change could be due to the advertising campaign. However, we assume that online WOM communication through the blogosphere also played a part in this change.

In Figure 4, there is no particular special super node. Leskovec, Singh, and Kleinberg (2006) found that the cascade size distribution is approximately heavy-tailed, and any cascade tends to be shallow. Our results are consistent with their findings, as the cascade patterns reveal mostly small sub-graphs.

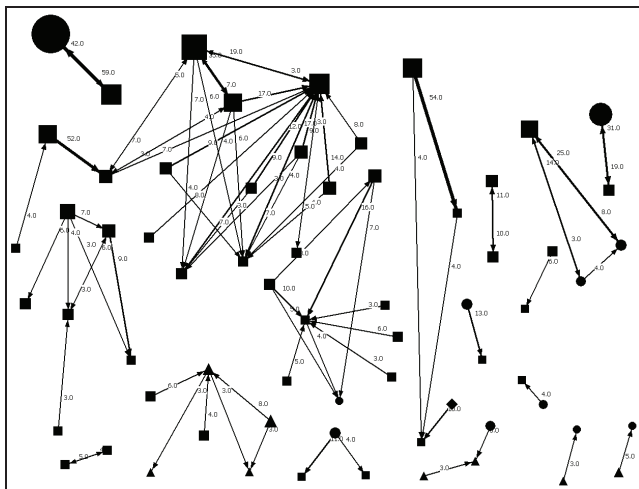


Figure 4. Patterns of Influence in the Blogosphere

## Conclusion

We have shown that there is an optimal heterophily for WOM senders and receivers in both online and offline environments.

Empirical analysis suggests that the people are most influenced by people only slightly more influential than them, not by standout influencers. Our analysis of optimal heterophily and the patterns of influence suggest that in WOM diffusion, many *grassroots influentials* play the central role instead of a few *super influentials*.

Future work will be aimed at further understanding the network effect, such as attitude change by neighbors who are in the range of optimal heterophily, and customization of the model for advertising media evaluation and planning.

## Acknowledgments

We thank Nifty Corporation for sharing its dataset and Dentsu for conducting the joint survey on offline WOM communication.

## References

- Alpert, M. I., and Anderson Jr., W.T. 1973. Optimal Heterophily and Communication Effectiveness: Some Empirical Findings. *Journal of Communication*, 23(3): 328-343.
- Crandall, D.; Cosley, D.; Huttenlocher, D.; Kleinberg, J.; and Suri, S. 2008. Feedback Effects between Similarity and Social Influence in Online Communities, *In Proceedings of KDD 2008*, 160-168.
- Gatignon, H. and Robertson, T. S. 1985. A Propositional Inventory for New Diffusion Research. *Journal of Consumer Research*, 11:849-867.
- Hogg, T.; Wilkinson, D.M.; Szabo, G.; and Brzozowski, M. J. 2008. Multiple Relationship Types in Online Communities and Social Networks. *In Proceedings of the AAAI Symposium on Social Information Processing*.
- Kaigler-Evans, K.; Leavitt, C.; and Dickey, L. 1978. Source Similarity and Fashion Newness as Determinants of Consumer Innovation. *Advances in Consumer Research*, 5:738-742.
- Leskovec, J.; Singh, A.; and Kleinberg, J. 2006. Patterns of Influence in a Recommendation Network, *In Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*.
- Matsumura, N.; Yamamoto, H.; and Tomozawa, D. 2008. Finding Influencers and Consumer Insights in the Blogosphere. *In Proceedings of ICWSM 08*, 76-83.
- McPherson, M.; Smith-Lovin, L.; and Cook, J.M. 2001. BIRDS OF A FEATHER: Homophily in Social Networks. *Annual Review of Sociology*, 27:415-444.
- Rogers, E.M. 2003. *Diffusion of Innovations: Fifth Edition*, NY: The Free Press.
- Rogers, E. M. and Bhowmik, D. K.1970. Homophily-heterophily: Relational Concepts for Communication Research. *Public Opinion Quarterly*, 34(4):523-538.
- Singla, P. and Richardson, M. 2008. Yes, There is a Correlation - From Social Networks to Personal Behavior on the Web, *In Proceedings of WWW 2008*, 655-664.
- Watts, D.J. and Dodds, P.S. 2007. Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*, 34:441-458.
- Yamamoto, H.; Nishida, S.; Morioka, S.; and Yamakawa, S. 2008. Optimal Heterophily and the Impact of Word-of-mouth, *In Proceedings of INFORMS Marketing Science Conference 2008*.