# Unveiling the secret of information rediffusion process on social media from information coupling perspective: a hybrid approach of machine learning and regression model[‡]

Zhen Yan[1,*], Rong Du[2] and Hua Wang[1,*]

[1] Xi'an Jiaotong University, Shaanxi 710049 Xi'an, China
[2] Xidian University, Shaanxi 710126 Xi'an, China

**Abstract**

Given the popularity and prevalence of communication through social media platforms, it is critical to determine the mechanisms that diffuse and rediffuse information. Prior studies have examined the impacts of a range of news item characteristics on the spread of information. However, little research has yet explored the influence that information coupling might have on the commenting and reposting behavior of users. Using the Sina Microblog site, we modeled three information couplings – emotional coupling, semantic coupling, and cognitive coupling – to determine whether they have any influence on the spread of information. We also examined whether opinion leaders wield a moderating influence in these relationships. Building on the cardinal literature and theories, we find that emotional and semantic coupling contributes more to commenting, whereas cognitive and emotional coupling both influence reposting more. Both these findings are supported by construal-level theory. Opinion leaders have a positive correlation with reposting, which is also supported by two-step flow theory. Overall, this research deepens our present understanding of information rediffusion at the comment and reposting levels. Our findings highlight the importance of considering information coupling from a linguistic point of view and of considering the influence of opinion leaders. This research also opens up interesting opportunities for further study on the role that information coupling might play given a comprehensive view of user-generated content (UGC). The outcomes of this study should help social media platforms and their users better understand how information spreads on social media.

**Keywords**

information coupling, two-fixed model, construal-level theory, two-step flow theory, information rediffusion

## 1. Introduction

In the post-internet era, communicating through social media has become a ubiquitous part of daily life. This not only gives rise to massive amounts of information more sensitive to public health information, they have also become more likely to get information about public health emergencies from social media (Becker & Gijsenberg 2022). This is because they believe that information sharing and communicating with others will provide them with more up-to-date and transparent information more quickly (Wang et al., 2022). The Sina Microblog, one of the world's biggest social media platforms, was an important and popular form of human-media interaction during the pandemic and has continued to be so ever since. There is no doubt that social technologies and constantly evolving internet technologies are transforming information diffusion, rediffusion, and the way people acquire information and knowledge. It is therefore paramount to explore the factors that influence these rediffusion processes and

Joint Workshop of the 5th Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE2024)and the 4th AI+ Informetrics (ALL2024),April 23-24,2024, Changchun, Jilin, China and Online

* Corresponding author.
jessieyan92@163.com (Z. Yan); durong@mail.xidian.edu.cn (R. Du); seablue@xjtu.edu.cn (H. Wang)

the mechanisms by which the coupling of information content and context influence the process.

Some scholars have studied information diffusion processes from the perspective of user behavior, such as information sharing (Fu & Shen, 2014), reactions to information (Kim et al., 2023), and interactions with information (Jensen et al., 2013), while others have studied the content of information, including the emotions conveyed (Naskar et al., 2020) and the topics discussed (Chen et al., 2020;Kim et al., 2023). According to Chen et al. (2020), two main online behaviors influence information diffusion through social networks: commenting and reposting. Commenting provides platforms and sources of information rediffusion while reposting facilitates information rediffusion because of the structure of the Internet.

Information coupling, as an association of topically related documents for managing and manipulating coupled information extracted from the database (Bhowmick et al., 1998), refers to the degree of difference between information source and the User-generated-content (UGC), the content that is created by members of the general public and distributed over the internet (Daugherty et al. 2008, Krumm et al. 2008), in the present study. Information coupling also has been studied from content-congruence and topic consistency aspects, respectively (Peng et al., 2020; Kim et al., 2023). However, we have very little knowledge on how information coupling influences information rediffusion process, which arouses and promotes information rediffusion extremely, is neglected. To fill this research gap, this study concentrates on the factors that influence the information rediffusion process from the perspective of information coupling, i.e. the difference between the information source (hereto as the news) and the UGC. There are three main research questions we seek to answer:

Research Questions 1: How does information coupling influence information rediffusion in terms of commenting?

Research Questions 2: How does information coupling influence information rediffusion in terms of reposting?

Research Questions 3: How do opinion leaders affect information rediffusion?

To answer these research questions, we designed a moderated nonlinear model as a way of exploring which factors influence the information rediffusion process and how. The empirical setting for this study is news of public health emergencies and the UGC associated with this news, crawled from the Sina Microblog. These difference between the two types of information – news and UGC – form the information coupling. Our research exerts efforts on the information coupling from sematic, typology, and cognition perspectives, employs a two-way fixed moderated nonlinear model (i.e., comment-fixed effect model and repost-fixed effect model).

# 2. Theoretical background and Conceptual model

## 2.1 Summarization of theoretical background

Overall, prior studies have extensively studied the paradigm of networks and the motivations behind UGC and user behavior in the information diffusion process. Some scholars have developed algorithms based on information propagation theory, such as the SIR model (Xu et al., 2020; Harrigan et al., 2021), while others have used technical means to reveal any emotional influences at play (Singh et al., 2020; Chen et al., 2020; Diwali et al., 2023). However, information couplings comprising the origin of information with UGC has received less attention as has the contribution such couplings make to the information diffusion process. Our review indicates that specific user activities along with the content of the information to be spread have the greatest influence over whether the informationwill be disseminated.

## 2.2 Conceptual model of the present work

Drawing insights from the previous literature, the impact of information rediffusion is reflected in the total sum of comments and reposts. Given the structure of social networks, more comments should attract greater user attention, while more reposts should expand the sphere of exposure. In other words, reposts spread attention wider and further while comments increase the level of scrutiny given to some news (Shiau et al., 2017).

In addition, the information rediffusion mechanism is also stimulated by information coupling. Emotions and topics, the most significant aspects of information content, reveal personal attitudes (Qiao et al., 2022; Yin et al., 2023). As mentioned, emotional couplings refer to the similarity of the feelings in an information source and its associated UGC. Here, extreme UGC is usually associated with intense emotions, and therefore may contain incoherent arguments (Yin et al., 2023). Indeed, to express strong case for or against an information source, an incentivized user needs to deliver a particularly coherent argument that covers many details, thus giving rise to semantic meaning. For this reason, we therefore assume that both emotional and semantic coupling influence information rediffusion. Further, due to individual differences in cognition, the cognitive influence of some news also plays an important role in delivering information. Metaphor, as the surface expression of cognition, is regarded as cognitive coupling, which is also one of the independent variables in this study.

However, the structure of social networks means that information diffusion will also depend on the relationships between users. These relationships directly influence information diffusion but opinion leaders, who have large numbers of followers, also indirectly influence the information rediffusion process. Therefore, opinion leaders, as one important facet of social networks, is a moderating variable in this study. The control variables include gender, whether the user is verified, the number of posts the user has made on the platform, and the number of users a user is following (László et al., 2023; Lin et al., 2022; Liu et al., 2023). For the whole view of the conceptual model for this study, we illustrate it on Fig.1 on Appendix 1.

# 3. Methodology

## 3.1 Overview of the research framework

Our dataset, which comprises 4,017 pieces of news and 416,358 pieces of UGC was crawled from Sina Microblog. The period of study is 1 Dec 2021 to 1 Jun 2022, All of the news relates to public health emergencies because this type of news is particularly interesting to the public (Li et al., 2020).Then we removed the several words UGC and resaved 415,473 pieces of UGC (i.e. remove repeated data and symbol-only data and Jieba word split).

As discussed in the literature review, we drew the factors for study from the literature. We modelled emotional coupling, semantic coupling, and cognitive coupling using a machine learning approach and negative binominal regression models to measure the influence of these factors on information rediffusion. The influence of opinion leaders was modelled as a moderating effect (Wang et al., 2022). Finally, we conclude the working mechanism of information rediffusion and apply them on management practice. Details follow in Figure 2 on Appendix 1.

## 3.2 Variables description and measurement

We took comments and reposts as our dependent variables, while the independent variables are emotional coupling, semantic coupling, and cognitive coupling. The influence of opinion leaders was modelled as a moderating variable. Opinion leaders were defined as those with more than 10,000 followers and Big V badge on the Sina Microblog. Table 1 in Appendix 1 shows the definitions, formulas and measurement metrics for each variable.

We devised two fixed models to estimate the two different dependent variables, i.e., a commenting model and a reposting model. All of the dependent variables were measured in terms of frequency.

All the measurements of variables are illustrated on Appeendix 1.

# 4. Results and Findings

## 4.1 Comment model

Table 2 presents the results of the main regressions used to test the effects of the three types of couplings on information rediffusion. Note that we standardized all continuous independent variables to leverage the comparison of effect sizes. We first entered the control variables in Model 1 and then added the three coupling variables and the moderate variable to Models 2-5 in a stepwise fashion. We then compared the R2 of Models 2-5 with Model 1, which was taken as the baseline model, and found that adding the three coupling variables along with the moderating variable significantly improved the model's fit ($p<0.001$).

Model 2, which includes all the control variables, tests the influence of emotional coupling (M=1.084,SD=0.557). The correlation shows that emotional coupling attracts more comments ($\beta 1=$ 1.007**), which induces that when the difference between UGC and the news on emotional intensity increase at 1, the one comment of the UGC is added. Thus, emotional intensity has a positive effect on information rediffusion at the comment level. Model 3, which tests semantic coupling, shows that this type of coupling is also positively related to information rediffusion at the comment level ($\beta 2=$ 0.667***, $p < 0.001$). This result indicates that a great similarity between the news and the UGC on semantic level will significantly increase the number of comments made against the item. Model 4, which tests the influence of cognitive coupling on comments, also indicates a positive correlation. Thus, the more cognitively similar the news and the UGC, the more comments the item will attract ($\beta 3=$ 0.637*** ,$p < 0.001$). Opinion leaders, as a moderating variable, also have a positive effect on comments ($\beta 4=$ 0.227* ,$p < 0.05$).

## 4.2 Repost model

The results of the negative binominal model tests to assess how the variables influence reposting behavior are shown in Table 3. Model 6 contains the control variables and is regarded as the baseline of the reposting model. Compared to Model 1 in Table 2, Model 6 demonstrates that gender and whether the user is verified contributes more significantly to reposting than to comments ($\beta 5=$ 0.857** ,$p < 0.01$).

Models 7-10 portray the stepwise regressions for the independent and moderating variable. In Model 7, emotional coupling is shown to have a positive influence on reposting ($\beta 1=$ 946**,$p < 0.01$), indicating that differences in emotional coupling attract more frequent reposts. Semantic coupling also significantly affects reposting, as indicated by Model 8 ($\beta 2=$ 0.417*** ,$p < 0.001$), while cognitive coupling also significantly influences reposting behavior as demonstrated by the results from Model 9 ($\beta3= 0.668^{***}, p < 0.001$). The moderating variable, opinion leaders, has a greater positive influence on reposting than it does on commenting ($\beta4= 3.388^{**}, p < 0.01$), as shown by Model 10 (Table 3) when compared to Model 5 (Table 2). This phenomenon explicitly displays the "nudge" effect of opinion leaders in social network as two-step flow theory posits.

### 4.3 Moderating factors

In terms of the moderating effect of opinion leaders between information coupling and rediffusion, the data indicate that the interactions of opinion leaders with emotional coupling, semantic coupling, and cognitive coupling are significantly correlated with each other (see Model 11 of Table 4 and Model 12 of Table 4).

Models 11 and 12 also demonstrate that opinion leaders exert a different influence over commenting behavior to reposting. Opinion leaders will attract a greater number of comments through emotional intensity ($\beta1= 2.317^{***}$, $p < 0.001$) and relying on cognitive expressions ($\beta3= 2.304^{***}$, $p < 0.001$). However, to attract more reposts, opinion leaders need to motivate users through semantic content ($\beta2= 2.359^{***}$, $p < 0.001$) and, again, cognitive expressions ($\beta3= 2.707^{***}$, $p < 0.001$). Overall, similarity in metaphorical expression is the most important factor in an opinion leader receiving comments and reposts on social media.

## 5 Conclusion and implication

The overarching conclusions from this research are that emotional and semantic coupling prompt information rediffusion through comments, while reposting typically depends on emotional and cognitive coupling. Further, opinion leaders contribute more to reposting behavior than to commenting. Compared to previous studies, the specific contributions of this study can be summarized as follows.

Although previous studies on the diffusion of information report that content needs to be written in a certain way or placed in a certain context in order to be perceived easily by others, emerging evidence from B2C platforms suggests that the concreteness of lexical cues can influence the beliefs and mindsets of users as they read and make sense of UGC (Peng et al. 2020; Jörg et al., 2023). However, few of these studies have examined the cognitive cues underlying content at the lexical level. Building on and going beyond recent studies, we applied metaphorical expressions, the linguistic surface of cognition, to determine the effect of cognitive coupling.

In theory, Figure 3 in appendix shows that the difference in emotional intensity between a piece of news and some UGC is a highly significant factor as shown by the green curve in Fig. 3 , which fluctuates dramatically. This is consistent with previous findings (Yin et al., 2023) and is supported by cognitive dissonance theory (Festinger, 1962). Cognitive dissonance refers to the psychological state of discomfort or stress triggered by factors such as contradictory information in the environment, or the inconsistency of one's beliefs with their actions or new information. Individuals realize that it's difficult to process self-contradictory information (Alter & Oppenheimer, 2009) which is always presented as less attention paid. Fig.3 portrays the sentiment polarity of the news (the blue color curve), UGC (the red color curve), and their difference (the green color curve). It shows that when the difference of news and UGC in emotion intensity fluctuates largely, the emotional intensity of UGC changes largely as well. The sentiment polarity of the different shows that contradictory directly contribute to the increase of cognitive dissonance in the evaluation of the same attributes among different information content. At the same time, the polarity of emotional intensity always accompanied with less frequency of comments or repost. Therefore, our results suggest that as the difference in emotional intensity becomes larger, as supported by cognitive dissonance theory, it negatively influences how UGC is perceived as manifest by lower numbers of comments and reposts.

The interaction effects of opinion leaders with three types of information coupling also represent a prominent cue that opinion leaders positively influence the number of comments mainly through expressing intense emotions, which can shape others' thinking and mindsets. However, using different metaphorical expressions, especially converse metaphors helps opinion leaders to attract more reposts. More specifically, spatial metaphor, such as *up, increase, support*, is always bound to *down, doubt of the facts, bottom* in UGC of opinion leaders which receives more repost. For examples, the number of patients always described as extremely higher with less treatment, which portrays an opposite picture in public health emergencies and reaches more comments and reposts. Besides, the structural metaphor *"the pandemic is a war"* is used to map the public health emergencies to war, thus many expressions on war is used to described the emergencies. The *doctors* and *nurses* are described as *soldiers* and *heroes,* which provides a more specific picture of the fierce situation in public health emergencies. This type of metaphors used by opinion leaders is attracted more comments or reposts as well.

**References:**


[1]. Alter AL, Oppenheimer DM (2009). Uniting the tribes of fluency to form a metacognitive nation. Personality Soc. Psych. Rev. 13(3), 219–235.

[2]. Becker, M., & Gijsenberg, M.J. (2022). Consistency and Commonality in Advertising Content: Helping or Hurting? International Journal of Research in Marketing.

[3]. Bhowmick Sourav ; Ng, Wee-Keong & Lim Ee Peng. Information coupling in web databases. (1998). Conceptual Modeling - ER ’98: 17th International Conference on Conceptual Modelling, Singapore, 1619 November 1998: Proceedings, 1507, 92-106.

[4]. Chen Yanzhen, Rui, Huaxia & Whinston Andrew B. (2021). Tweet to the Top? Social Media Personal Branding and Career Outcomes. MIS Quarterly,45(2):499-534.

[5]. Diwali, K. Saeedi, K. Dashtipour, M. Gogate, E. Cambria & A. Hussain. Sentiment Analysis Meets Explainable Artificial Intelligence: A Survey on Explainable Sentiment Analysis. in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2023.3296373

[6]. Festinger L.. (1962). A Theory of Cognitive Dissonance (Stanford University Press, Stanford, California).

[7]. Fu, Xin, & Shen, Yun (2014). Study of collective user behaviour in Twitter: A fuzzy approach. Neural Computing and Applications, 25(7), 1603–1614.

[8]. Ge M.S., Mao R. & Cambria E. Explainable Metaphor Identification Inspired by Conceptual Metaphor Theory. 36th AAAI Conference on Artificial Intelligence. FEB 22-MAR 01, 2022,10681-10689.

[9]. Kim, E., Ding, M. (Annie), Wang, X. (Shane), & Lu, S. (2023). Does Topic Consistency Matter? A Study of Critic and User Reviews in the Movie Industry. Journal of Marketing, 87(3), 428–450. https://doi.org/10.1177/00222429221127927

[10]. Li L., Wang Z., Zhang Q. & Wen H. (2020). Effect of anger, anxiety, and sadness on the propagation scale of social media posts after natural disasters. Information Processing & Management, 57(6), Article 102313.

[11]. Li Jing , Shiqi Zhang, Wenting Ao. Why is instant messaging not instant? Understanding users’ negative use behavior of instant messaging software, Computers in Human Behavior, Volume 142,2023,107655.

[12]. Liu Huiting , Chen Yi , Li Peipei , Zhao Peng & Wu Xindong. Enhancing review-based user representation on learned social graph for recommendation. Knowledge-Based Systems,Volume 266,2023,110438.

[13]. Mao Rui , Lin Chenghua & Guerin F.. (2018). Word Embedding and WordNet Based Metaphor Identification and Interpretation. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Volume 1, 1222–1231, Melbourne, Australia. Association for Computational Linguistics.

[14]. Peng Chih-Hung , Yin Dezhi & Zhang Han. (2020). More than Words in Medical Question-and-Answer Sites: A Content-Context Congruence Perspective. Information Systems Research.https://doi.org/10.1287/isre. [31].

[15]. Qiao Dandan, Huaxia Rui (2022) Text Performance on the Vine Stage? The Effect of Incentive on Product Review Text Quality.Information Systems Research. https://doi.org/10.1287/isre.1146.

[16]. Stieglitz Stefan & Linh Dang-Xuan. (2013). Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. Journal of Management Information Systems, 29(4): 217–247.

[17]. Thaler, R. H., & Sunstein, C. R. (2009). Nudge: Improving decisions about health, wealth, and happiness. Penguin.

[18]. Wang, X., Zhang, M.,Fan, W., & Zhao, K. (2022). Understanding the spread of COVID-19 misinformation on social media: The effects of topics and a political leader's nudge. Journal of the Association for Information Science and Technology, 73(5). 726-737.

[19]. Watts Duncan, J., & Dodds, Peter Sheridan (2007). Influentials, networks, and public opinion formation. Journal of Consumer Research, 34(4), 441–458.

[20]. Wang, Ru, Rho, Seungmin, Bowei, Chen, & Wandong, Cai (2017). Modeling of large-scale social network services based on mechanisms of information diffusion: Sina Weibo as a case study. Future Generation Computer Systems, 74, 291–301.

[21]. Xu, Jiuping, Tang, Weiyao, Zhang, Yi, & Wang, Fengjuan (2020). A dynamic dissemination model for recurring online public opinion. Nonlinear Dynamics, 99, 1269–1293.

[22]. Yang Yi & Subramanyam Ramanath(2023).Extracting Actionable Insights from Text Data: A Stable Topic Model Approach.MIS Quarterly, 47(3), 923-954.

[23]. Yin Dezhi , Triparna de Vreede, Logan M. Steele, Gert-Jan de Vreede (2022) Decide Now or Later: Making Sense of Incoherence Across Online Reviews. Information Systems Research.https://doi.org/10.1287/isre.2022.1150.

[24]. Zhong Ning & David A. Schweidel (2020). Capturing Changes in Social Media Content: A Multiple Latent Changepoint Topic Model. Marketing Science, 39 (4), 827–46

[25]. Zhang Yi, Lu Jie, Liu Feng, Liu Qian, Porter Alan, Chen Hongshu & Zhang Guangquan. Does deep learning help topic extraction? A kernel k-means clustering method with word embedding. Journal of Informetrics, 12(4), 1099-1117.

## Appendix 1 Figures & Table in the present study

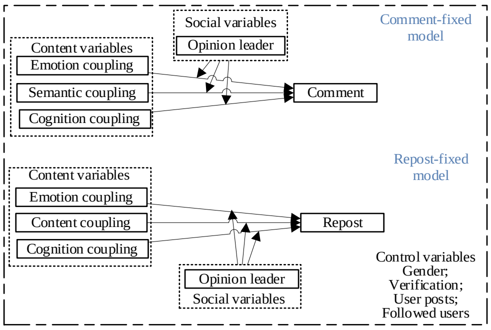


**Figure 1:** Conceptual model of the present study

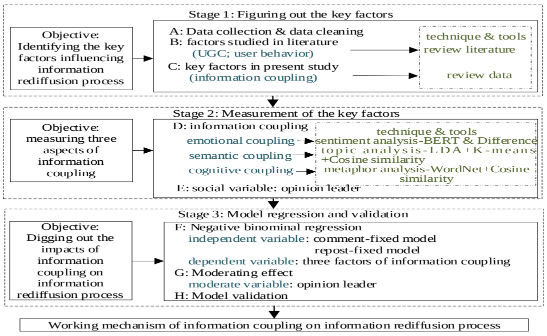


**Figure 2:** research framework of the present study

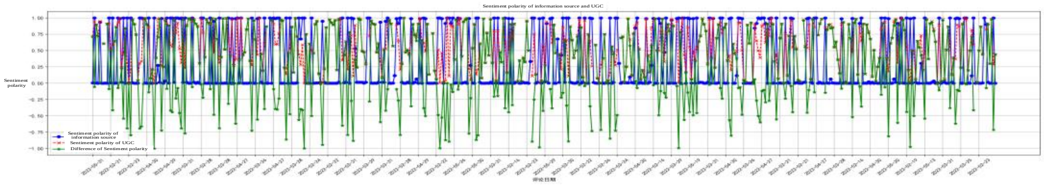

**Figure 3:** Emotional fluctuation in time span

**Table 2** Mean, standard error and correlation variables in comment-fixed effect model

| variables | M | SD | Comment-fixed models | | | | |
|---|---|---|---|---|---|---|---|
| | | | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
| Emotional coupling | 1.084 | 0.557 | | 1.007** | 1.210** | 1.014** | 1.001** |
| Semantic coupling | 1.033 | 0.034 | | | 0.667*** | 0.698*** | 0.699*** |
| Cognitive coupling | 1.401 | 0.505 | | | | 0.637*** | 0.658*** |
| Opinion leader | 3.706 | 0.007 | | | | | 0.227* |
| Gender | 0.800 | 0.201 | 0.450** | 0.417** | 0.415** | 0.454** | 0.421** |
| Verification | 1.462 | 0.211 | 0.599*** | 0.554*** | 0.534*** | 0.522*** | 0.535*** |
| User posts | -9.895 | 1.105 | -1.122*** | -1.145*** | -1.146*** | -1.136*** | -1.131*** |
| Followed users | -2.566 | 0.001 | 0.487*** | 0.424*** | 0.402*** | 0.467*** | 0.435*** |
| $R^2$ | | | 0.645 | 0.786 | 0.782 | 0.784 | 0.788 |

Note: * p < .05. ** p < .01. *** p < .001.

**Table 3** Mean, standard error and correlation variables in repost-fixed effect model

| variables | M | SD | Repost-fixed models | | | | |
|---|---|---|---|---|---|---|---|
| | | | Model 6 | Model 7 | Model 8 | Model 9 | Model 10 |
| Emotional coupling | 1.084 | 0.557 | | 0.946** | 0.958** | 0.954** | 0.967** |
| Semantic coupling | 1.033 | 0.034 | | | 0.417*** | 0.535*** | 0.447*** |
| Cognitive coupling | 1.401 | 0.505 | | | | 0.668*** | 0.674*** |
| Opinion leader | 3.706 | 0.007 | | | | | 3.388** |
| Gender | 0.800 | 0.201 | 0.857** | 0.842** | 0.756** | 0.631** | 0.817** |
| Verification | 1.462 | 0.211 | 2.345** | 2.398** | 2.452** | 2.354** | 2.315** |
| User posts | -9.895 | 1.105 | -0.475*** | -0.425*** | -0.397*** | -0.545*** | -0.465*** |
| Followed users | -2.566 | 0.001 | 0.035*** | 0.041*** | 0.042*** | 0.038*** | 0.048*** |
| $R^2$ | | | 0.771 | 0.782 | 0.781 | 0.786 | 0.788 |

Note: * p < .05. ** p < .01. *** p < .001.

**Table 4** The moderated mediation effect of opinion leader on comment and repost

| Variables | Model 11 (comment) | Model 12 (repost) |
|---|---|---|
| Emotional coupling × opinion leader | 2.317*** | 0.389*** |
| Semantic coupling × opinion leader | 0.532*** | 2.359*** |
| Cognitive coupling × opinion leader | 2.304*** | 2.707*** |
| gender | 0.454** | 0.631** |
| Verification | 0.522*** | 2.354** |
| User posts | -1.136*** | -0.545*** |
| Followed users | 0.467*** | 0.038*** |
| $R^2$ | 0.527 | 0.642 |