



Analysis of the Impact of Artificial Intelligence-Driven Internet Celebrity Anchor Marketing on Consumer Psychological Health Through Big Data

Chao Zhang^{1*}, Di Jin² and Shuolin Zhang³

¹ School of Journalism and Communication, Henan University, Kaifeng, 475000, China

¹ School of Psychology, Henan University, Kaifeng, 475000, China, Steven@henu.edu.cn

² School of Journalism and Communication, Henan University, Kaifeng, 475000, China, DiJin202308@outlook.com

³ School of Information Engineering and Computer Science, Hebei Finance University, Baoding, 071000, China, ShuolinZhang202308@outlook.com

Corresponding author: Chao Zhang, Steven@henu.edu.cn

Abstract. To explore the influence of Internet celebrity marketing behavior on consumers' mental health, this paper improves the algorithm based on extensive social media data analysis. It proposes a D2D data distribution method based on user fusion social relations. In data shunting, this paper considers the influence of two relationships: the trust between users and the user's preference rating for data content. In addition, this paper divides the Community into a composite network. This paper compares five seed node selection methods to find the seed users suitable for data sharing within the Community. The experimental research shows that the extensive data analysis method based on social media can play a specific role in analyzing the influence of Internet celebrity marketing behavior on consumers' mental health. At the same time, from the data point of view, Internet celebrity marketing can effectively drive consumers' useless consumption and promote the emergence of impulsive consumption behavior, which also has a positive effect.

Keywords: social media; big data; Internet celebrity marketing; consumer mental health; Artificial Intelligence-Driven.

DOI: <https://doi.org/10.14733/cadaps.2024.S24.118-129>

1 INTRODUCTION

1.1 Related Work

Judging from the existing literature, in the past, researchers have done more research on impulsive consumption in the context of online live broadcasts, and most of them have been carried out from

the perspective of behavior. At the same time, more research must be done on impulse consumption willingness. The influence of the objective environment is involved, but the impact of the atmosphere cues of the live broadcast scene needs to be considered [2]. However, few studies have explored the psychological mechanism of impulsive consumption willingness in live broadcast scenarios from the perspective of psychology [6]. However, there are few related studies [3]

Literature [11] believes that carnival is a cultural phenomenon unique to the West, traces the history of carnival, and finds that carnival life occupies a considerable position in the medieval daily life of the West; the "carnival" theory of literature[9] has important implications for the study of carnival. The status must be addressed, especially since the research on carnival nature is of great significance. Literature [8] pointed out that dividing real life into two worlds and two lives is the premise of Bakhtin's "Carnival" theory.

The research[1] shows that consumers with high impulsiveness are more likely to immerse themselves in products and services, find exciting products and services, and quickly and without thinking about impulse purchase desires. Research[10] shows that people with high neuroticism and extraversion and people with low conscientiousness are more likely to become impulse buyers. Literature [5] conducted a study on impulse purchase intention in the e-commerce situation, possibly even stronger. In addition, studies have shown that pleasant and positive emotions positively affect impulsive consumption willingness[4]. Research on impulse spending and attitudes toward money has demonstrated that associating money with freedom, love, or power increases the likelihood of impulsive buying, and associating money with security decreases the possibility of impulsive buying[7]

1.2 Objectives

To explore the impact of Internet celebrity marketing behavior on consumers' mental health, this paper analyzes through social media big data technology. Moreover, this paper comprehensively analyzes the influence of Internet celebrity marketing behavior on e-commerce consumer groups, providing an effective method for the subsequent healthy development of e-commerce and improving consumers' mental health.

2 METHODOLOGY

2.1 D2D Data Distribution Method Based on User Fusion Social Relations

This paper describes the data transmission process of seed users based on the users' social relationships and intra-regional user mobility, as shown in Figure 1. Torrent users can periodically send data content to other users through D2D communication when they send access requests. The data traffic can be significantly reduced using this paper's method.



Figure 1: Seed user data transmission process.

We often use the Pearson correlation coefficient to weigh the correlation between variables. Its value is between 1 and -1. When the value is 1, the variables are positively correlated, -1 is the opposite, and 0 is irrelevant. We assume two sets of data, X and Y, each containing n elements; the method for calculating the covariance of the two can be written as:

$$S_{xy} = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y}) \quad (1)$$

The standard deviation is recorded as:

$$S_x = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2} \quad (2)$$

Among them, \bar{x} and \bar{y} represent the expectations of the two, the average. In the numerator, the numerator is positive when the values x and y are more significant than their mean values \bar{x} and \bar{y} , respectively. Moreover, the numerator is also positive when both are less than the mean value. The formula of the Pearson correlation coefficient is expressed as follows, wherein formula (3) represents the covariance, and formula (4) is the quotient of the variable standard deviation:

$$\frac{S_{xy}}{S_x S_y} = \frac{\sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^n (x_k - \bar{x})^2} \sqrt{\sum_{k=1}^n (y_k - \bar{y})^2}} \quad (3)$$

The Pearson coefficient can also be estimated by the mean of the relevant standard scores of the (x_k, y_k) sample points, and its equivalent expression is:

$$\frac{S_{xy}}{S_x S_y} = \frac{1}{n-1} \sum_{k=1}^n \left(\frac{x_k - \bar{x}}{S_x} \right) \left(\frac{y_k - \bar{y}}{S_y} \right) \quad (4)$$

The Pearson coefficient can also be estimated by the mean of the relevant standard scores of the (x_k, y_k) sample points, and its equivalent expression is:

We assume that the value of the data sample X, which contains n data contents $X = \{X_1, X_2, X_3, \dots, X_n\}$, has been given, and each object in the data sample is assigned m attributes of a dimension.

$$dis(X_i, C_j) = \sqrt{\sum_{t=1}^m (X_{it} - C_{jt})^2} \quad (5)$$

The formula of the K-means algorithm is as follows:

$$C_l = \frac{\sum_{X_i \in S_l} X_i}{|S_l|} \quad (6)$$

2.2 Integrate Social Relationship Model Construction

We assume that a mobile data distribution system involves data content and users, and $D = \{d_1, \dots, d_i\}$ and $U = \{u_1, \dots, u_m\}$ are data content and users, respectively. The user's personal attributes and data content items are given, and the user u's preference for data content D can be estimated. The parameter is set to be large enough, and the eigenvector corresponding to the largest eigenvalue can be taken as the transformation axis.

$$x_{ud} = \frac{a_{iu}^T u_d}{\|a_{iu}\| \|u_d\|} \quad (7)$$

$u = 1, \dots, m$ and $d = 1, \dots, i$. In this regard, the rating matrix M established by the user's preference for data content consists of $m \times n$ elements $\{m_{11}, m_{12}, \dots, \text{and } m_{mn}\}$. It is denoted as $[m_{ui}]_{n \times n}$, where m_{ui} represents the rating of user u to the i-th data content d_i , as shown in Table 1.

	d_1	d_2	d_3	d_4	d_5
u_1	2		4		
u_2		3		3	
u_3		2			2
u_4				5	
u_5			6		2

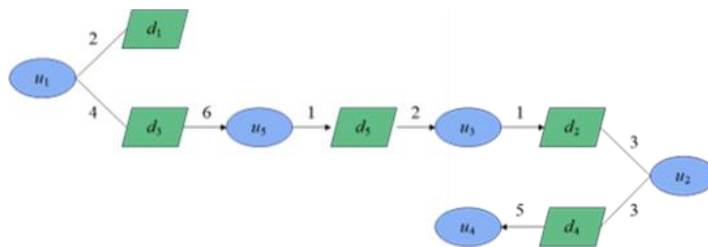
Table 1: User content rating matrix.

The relationship between users in a social network can be represented by a trust degree matrix $T [t_{uv}]_{m \times m}$, where t_{uv} means the closeness of the connection between user u and user v , reflecting the trust degree between users, as shown in Table 2.

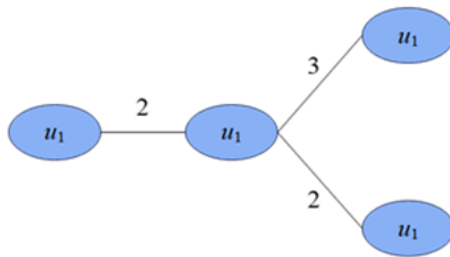
	u_1	u_2	u_3	u_4	u_5
u_1			1		
u_2					3
u_3		1			
u_4					2
u_5	2				

Table 2: User trust matrix.

Figure 2 shows the user content scoring network M_N and the user trust network T_N , where the user content scoring threshold M is set to 2, and the threshold T of the user trust network is also 2, as shown in the figure:



(a) User Content rating network



(b) User trust network

Figure 2: User content rating and trust network.

The graph nodes represent users and data content in the trajectory; the weight of the edge is based on the value of the social relationship indicator.

In the composite network, the attribute relationships between the sub-networks are integrated on the edges between nodes, and the user content rating network M_N in Figure 2(a) and the user trust network T_N in Figure 2(b) are fused into a composite network. We assume that the relationship strength parameter sf_1, sf_2 of r_1, r_2 is 1, the weight set of the M_N edge of the user content rating network is $\{1\}$, the weight set of the T_N edge of the user trust network is $\{0.5\}$, and the composite network of user content rating and trust degree is shown in Figure 3:

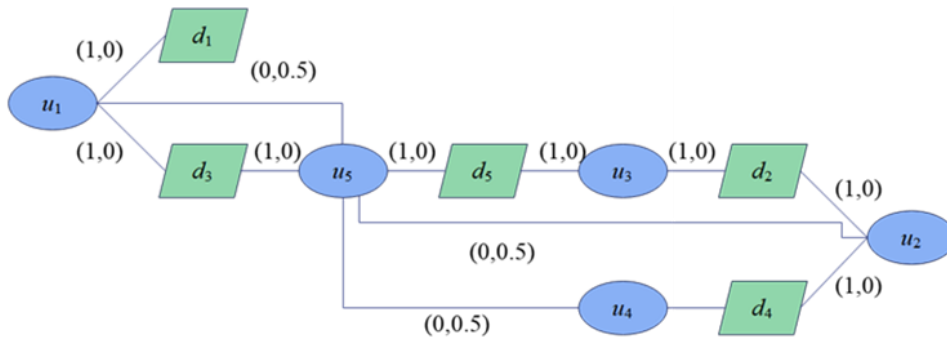


Figure 3: Composite network of user content rating and trustworthiness.

The user trust degree and the user preference similarity are combined in a particular proportion to obtain the user comprehensive similarity.

1. User trust degree

We assume that users u and v are closely related, and we believe that users u and v have higher preferences for the same data content type than others. This social relationship that promotes community division can be expressed as the degree of the user's trust. The user u and user v are represented by the vector X and the vector Y , respectively, and the Pearson similarity is used to measure the trust degree between the user u and the user v . The formula can be expressed as follows:

$$Deg_{\text{trust}}(u, v) = \frac{\sum_{i=1}^m (T_{ui} - \delta)(T_{vi} - \delta)}{\sqrt{\sum_{i=1}^m (T_{ui} - \delta)^2} \sqrt{\sum_{i=1}^m (T_{vi} - \delta)^2}} \tag{8}$$

2. Similarity of user content preference score

The user's rating of the data content also reflects the user's preference for the content to a certain extent. The higher the rating, the closer the user prefers the same content. The formula is as follows:

$$P(u, v) = \left(\frac{|D_u \cap D_v|(>R)}{|D_u|(>R)} + \frac{|D_u \cap D_v|(<R)}{|D_u|(<R)} \right) \times \left(\frac{|D_u \cap D_v|(>R)}{|D_v|(>R)} + \frac{|D_u \cap D_v|(<R)}{|D_v|(<R)} \right) \quad (9)$$

D_u and D_v represent the set of user u and user v 's ratings of data content, respectively. The user content preference score is fitted by Pearson similarity, characterized by $Deg_{\text{preference}}(u, v)$, S_{ud} represents the user u 's rating of data content d , and \bar{s}_u represents the average value of the user's rating on the data content. User content preference score similarity defined by Pearson similarity is as follows:

$$\begin{cases} Deg_{\text{preference_pearson}}(u, v) = \frac{\sum_{u=1}^m (S_{ud} - \bar{s}_u)(S_{vd} - \bar{s}_v)}{\sqrt{\sum_{u=1}^m (S_{ud} - \bar{s}_u)^2} \sqrt{\sum_{v=1}^m (S_{vd} - \bar{s}_v)^2}} \\ Deg_{\text{preference}}(u, v) = P(u, v) \times Deg_{\text{preference_pearson}}(u, v) \end{cases} \quad (10)$$

3. The integrated social relationship between user trust and user content preference score similarity

To achieve a better data diversion effect, this section integrates user trust and user content preference score similarity; the formula is as follows:

$$Deg(u, v) = (1 - \lambda) \times Deg_{\text{trust}} + \lambda \times Deg_{\text{preference}} \quad (11)$$

The parameter λ in the formula adjusts the fusion ratio of user trust degree and user content preference score similarity. Deg_{trust} is the trust degree of user u and user v , and $Deg_{\text{preference}}$ is the similarity degree of user u and user v content preference score. When λ is greater than $\frac{1}{2}$, the proportion of $Deg_{\text{preference}}$ is higher; however, when λ is smaller than $\frac{1}{2}$, the proportion of Deg_{trust} is higher.

2.3 Seed User Selection

Degree centrality has good utility in computing node centrality. Nodes are ranked according to the number of directly connected edges to identify the most popular nodes in the Community. Given a community G_k , the degree centrality of node u is as follows:

$$C_D(u) = \sum_{v \in V_k} C(u, v) \quad (12)$$

2.3.1 Closeness Centrality

Closeness centrality is used to describe the closeness of the relationship between a node and other nodes in the network. The shortest path between nodes determines it. The closer a node is to other nodes, the higher its closeness centrality. The formula is expressed as follows:

$$C_C(u) = \frac{1}{\sum_{v \in V_k} d_{G_k}(u, v)} \quad (13)$$

Among them, $d_{G_k}(u, v)$ represents the shortest path length in G_k from node u to node v in G_k .

2.3.2 Betweenness Centrality

The core concept of betweenness centrality refers to measuring the importance of the data content on the node pair to the content propagated throughout the network. When a node appears as an intermediate quantity on the shortest path connected to other nodes in the network, if a node has a

high degree of importance, high betweenness centrality will appear along with it, which is expressed as follows:

$$C_B(u) = \sum_{v_1, v_2 \neq u, v_1, v_2 \in V_k} \frac{g_{v_1, v_2}(u)}{g_{v_1, v_2}} \quad (14)$$

Among them, g_{v_1} and v_2 represent the shortest paths between nodes (v_1, v_2) , and $g_{v_1, and v_2}(u)$ denotes the shortest paths between node pairs $(v_1$ and $v_2)$ passing through node u .

2.3.3 PageRank Centrality

PageRank algorithm has a high timeline for calculating node centrality and can be used to measure the probability of nodes when connecting essential nodes in the Community. Given Community G_k , PageRank is:

$$PR(u) = 1 - \rho + \rho \sum_{v \in F(u)} \frac{PR(v)}{|F(v)|} \quad (15)$$

2.3.4 Eigenvector Centrality

Eigenvector centrality is often used to measure centrality. The number of nodes adjacent to a node in the network and its neighbors' importance can reflect the node's importance to a certain extent. Given a community G_k , there are a total of n_k nodes, and the eigenvector centrality formula is defined as:

$$EC(u) = x_u = \frac{1}{\lambda} \sum_{v \in E_k} a_{uv} x_v \quad (16)$$

Among them, x_u is the u -th element of the center vector x , and λ is a constant, representing the largest eigenvalue associated with the first eigenvector of the adjacency matrix A .

3 EXPERIMENTAL METHOD

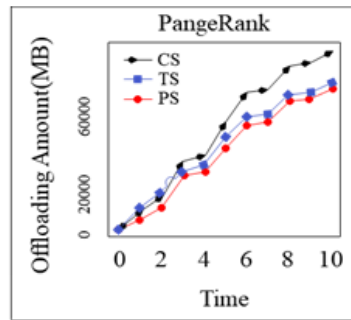
In this paper, the VisualStudio compilation environment is used, R language is used, and the Infocom dataset is selected for simulation experiments. The simulation duration is 337450 seconds.

This paper compares the following methods through experiments: (1) Cellular network data offloading strategy (Trust-based Scenario, TS) based on user trust to select seed nodes; (2) Cellular network data offloading strategy (Preference-based Scenario, PS) based on user content preference score to select seed nodes; (3) Cellular network data offloading strategy (Comprehensive-based Scenario, CS) based on user integration of social relations to select seed nodes.

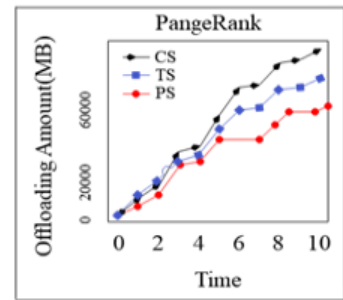
Based on the above research, this paper explores the influence of Internet celebrity marketing on consumers' mental health, mainly from three aspects: consumption impulse, consumption rationality, and useless consumption, and makes statistical evaluation results.

4 RESULTS

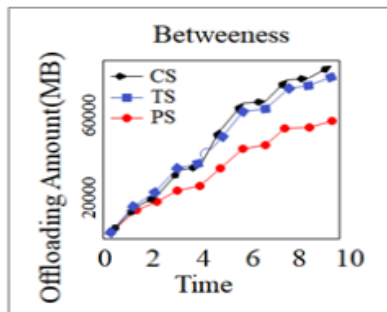
This paper investigates the time-varying data shunting of these three strategies under five different centrality properties, as shown in Figure 4.



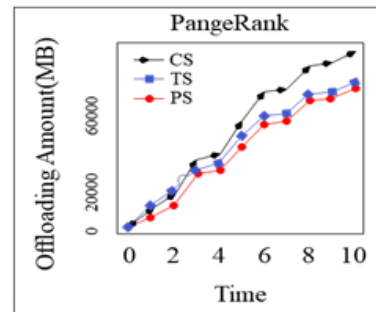
(a) Degree



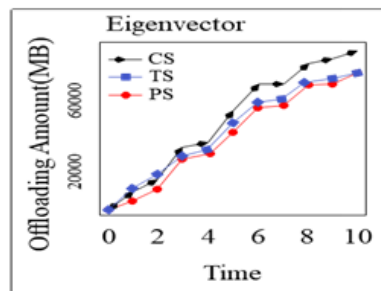
(b) Closeness



(c) Betweenness



(d) Page Rank



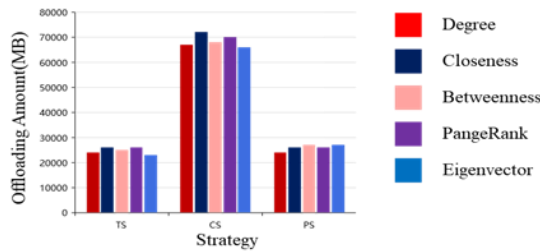
(e) Eigenvector

Figure 4: Performance comparison of three strategies under five attributes over time.

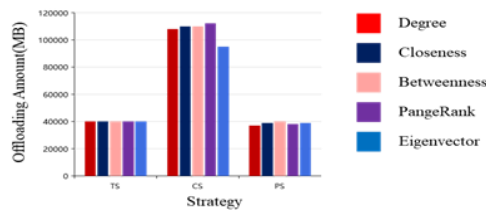
This part of the paper experimentally studies the number of seed nodes. The practical diversion effect cannot be achieved if the number of selected seed nodes is small. On the other hand, fewer seed nodes may increase network overhead and bring high costs. In this regard, four scenarios are constructed according to different seed numbers to find an appropriate number of seed nodes, as shown in Figure 5.

Through the above analysis, the D2D data diversion method (CS strategy) based on user fusion social relations is the best way to maximize data diversion. The following critical problem to solve is

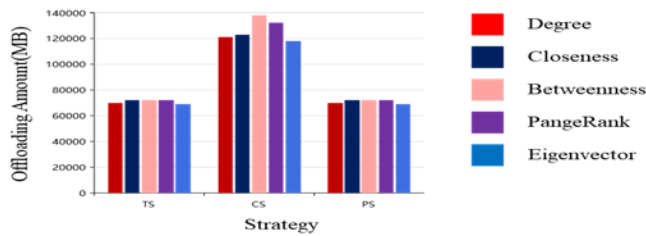
determining the number of seed nodes. Figure 6 shows the variation in the amount of data offloaded by the CS strategy with the number of seeds.



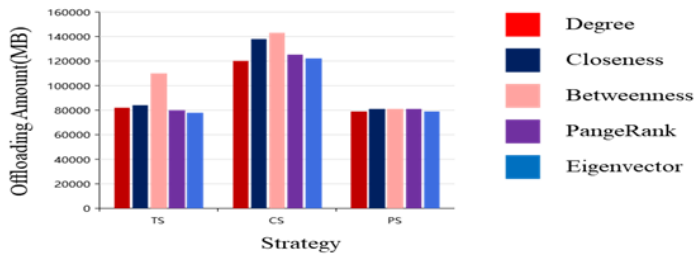
(a) Number of seeds 10



(b) Number of seeds 20



(c) Number of seeds 30



(d) Number of seeds 40

Figure 5: Performance Comparison of Five Centrality Measures in Three Strategies Under Different Numbers of Seed Nodes.

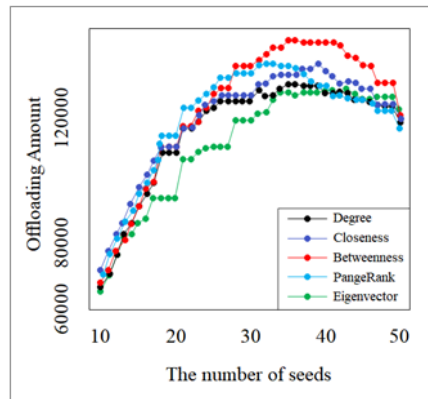


Figure 6: The Variation of the amount of data offloaded by the cs strategy with the number of seed nodes.

This paper explores the impact of Internet celebrity marketing on consumers' mental health, mainly from consumption impulse, consumption rationality, and useless consumption, as shown in Table 3.

N u m	<i>Consumption impulse</i>	<i>Consumption rationality</i>	<i>Useless consumption</i>	N u m	<i>Consumption impulse</i>	<i>Consumption rationality</i>	<i>Useless consumption</i>
1	59.68	27.83	44.89	15	62.93	28.70	52.30
2	51.94	27.52	42.76	16	46.63	23.82	53.29
3	51.26	29.32	43.32	17	44.70	29.43	50.79
4	62.89	29.70	54.69	18	42.03	24.44	51.46
5	57.48	18.87	46.36	19	58.22	19.50	51.04
6	45.83	17.74	49.23	20	60.22	20.18	44.21
7	54.23	28.75	49.13	21	59.26	20.46	55.69
8	57.72	14.31	50.24	22	52.52	21.61	46.08
9	59.86	26.44	61.39	23	53.22	20.52	56.70
10	43.16	26.09	43.87	24	62.74	18.64	51.09
11	46.82	23.97	49.55	25	44.81	24.73	56.16
12	49.71	15.81	60.51	26	53.82	19.83	61.76
13	62.34	20.00	49.69	27	58.44	21.62	47.56
14	53.18	25.40	45.68	28	48.79	24.30	61.04

Table 3: The influence of internet celebrity marketing behavior on consumer mental health based on extensive social media data analysis.

5 DISCUSSION

As shown in Figure 4, the seed node is initialized to 15. The cellular network data offloading (CS strategy) based on the seed node selection of user fusion social relations achieves the best data offloading effect among the three methods. The offloaded data volume is the largest. The

performance of close centrality, betweenness centrality, and PageRank centrality for seed node selection is better than several other methods in CS, and the maximum data unloading amount reaches nearly 100,000M. In comparison, the data distribution of the TS and PS strategies is similar to that of the CS strategies. Therefore, considering the D2D data offloading method based on user fusion social relations can minimize the data in the overloaded network.

As shown in Figure 5, from (a) to (d), the number of seeds is 10, 20, 30 and 40, respectively. As the number of seed nodes increases, the three strategies have apparent trends under different centrality measures, and it can be easily found that the CS strategy outperforms the other two strategies regarding data shunting. Tight centrality, betweenness centrality, and PageRank centrality have more advantages than degree centrality and eigenvector centrality of CS strategy. In addition, the data offloading effect of the TS strategy is comparable to that of the PS strategy.

Figure 6 shows that the CS strategy performs well in terms of close centrality, intermediate centrality, and PageRank centrality, which is consistent with previous results. These three attributes minimized the traffic load of 137000M, 145000M, and 138000M when the number of seeds is set to 31, 35, and 39, respectively. When the number of seed nodes is set between 30 and 40, the data distribution of the five strategies reaches the highest point and then slowly declines. Among these three properties, as the number of seeds increases, the performance of close centrality is no longer prominent, and betweenness centrality can be used to reduce the maximum amount of data. However, it requires a relative number of seeds to cooperate, and the cost is relatively high. In contrast, the PageRank attribute can be used to offload the data highly while selecting a small number of seeds. Therefore, the PageRank attribute is a better choice than the others.

Table 3 shows that the analysis method based on big data on social media can play a specific role in analyzing the impact of Internet celebrity marketing behavior on consumers' mental health. Combined with the data, Internet celebrity marketing can effectively drive consumers' useless consumption and promote the emergence of impulsive consumption behavior.

6 CONCLUSIONS

To explore the impact of Internet celebrity marketing behavior on consumers' mental health, this paper analyzes through social media big data technology. Moreover, this paper comprehensively analyzes the influence of Internet celebrity marketing behavior on e-commerce consumer groups. By building an intelligent system, the system's performance can be effectively improved. Moreover, the D2D data offloading method based on user fusion social relations can minimize the data in the overloaded network and use the PageRank attribute to offload the data height while selecting a small number of seeds. The analysis method based on social media big data can play a specific role in analyzing the influence of Internet celebrity marketing on consumers' mental health. According to the data, Internet celebrity marketing can effectively drive consumers' useless consumption and promote the emergence of impulsive consumption behavior.

Through the big data of social media, this paper analyzes and verifies that Internet celebrity marketing behavior can have a particular impact on consumers' psychology, and it can be effectively guided by intelligent means to eliminate harmful guiding consumption and improve consumers' rationality of consumption. The effect of AI-driven influencer marketing on consumer psychological health through extensive data analysis is vital. Ethical practices should guide these endeavors to ensure a balance between marketing effectiveness and consumer well-being in the evolving digital landscape.

Chao Zhang, <https://orcid.org/0000-0002-6930-6505>

Di Jin, <https://orcid.org/0009-0003-7951-7482>

Shuolin Zhang, <https://orcid.org/0009-0005-0995-725X>

ACKNOWLEDGEMENT

General Project of Humanities and Social Sciences Research in Henan Universities, "Research on the Construction of Henan's Language Image in the Context of Chinese Modernisation" [2024-ZZJH-310].

REFERENCES

- [1] Brown, L. J.; Jones, G. M.; Bond, M. J.: E-health: Psychosocial Challenges for South Australian Rural Mental Health Consumers, *Rural and Remote Health*, 19(3), 2019, 5103-5103. <https://doi.org/10.22605/RRH5103>
- [2] Cutler, N. A.; Sim, J.; Halcomb, E.; Moxham, L.; Stephens, M.: Nurses' Influence on Consumers' Experience of Safety in Acute Mental Health Units: A Qualitative Study, *Journal of Clinical Nursing*, 29(21-22), 2020, 4379-4386. <https://doi.org/10.1111/jocn.15480>
- [3] Easter, M. M.; Swanson, J. W.; Robertson, A. G.; Moser, L. L.; Swartz, M. S.: Impact of Psychiatric Advance Directive Facilitation on Mental Health Consumers: Empowerment, Treatment Attitudes and the Role of Peer Support Specialists, *Journal of Mental Health*, 30(5), 2021, 585-593. <https://doi.org/10.1080/09638237.2020.1714008>
- [4] Harris, S.; Farnworth, L.; Mynard, L.: Experiences of Disclosure for Vocational Occupations by Forensic Mental Health Consumers, *Journal of Vocational Rehabilitation*, 52(3), 2020, 291-301. <https://doi.org/10.3233/JVR-201078>
- [5] Mullen, A.; Isobel, S.; Flanagan, K.; Harman, K.: Involving Mental Health Consumers in Nursing Handover: A Qualitative Study of Consumer Perspectives, *Issues in Mental Health Nursing*, 42(8), 2021, 730-735. <https://doi.org/10.1080/01612840.2020.1853288>
- [6] Neary, M.; Schueller, S. M.: State of the Field of Mental Health Apps, *Cognitive and Behavioral Practice*, 25(4), 2018, 531-537. <https://doi.org/10.1016/j.cbpra.2018.01.002>
- [7] O'Brien, A. J.; Abraham, R. M.: Evaluation of Metabolic Monitoring Practices for Mental Health Consumers in the Southern District Health Board Region of New Zealand, *Journal of Psychiatric and Mental Health Nursing*, 28(6), 2021, 1005-1017. <https://doi.org/10.1111/jpm.12729>
- [8] Sacchetto, B.; Ornelas, J.; Calheiros, M. M.; Shinn, M.: Adaptation of Nussbaum's Capabilities Framework to Community Mental Health: A Consumer-Based Capabilities Measure, *American Journal of Community Psychology*, 61(1-2), 2018, 32-46. <https://doi.org/10.1002/ajcp.12221>
- [9] Scholz, B.; Bocking, J.; Happell, B.: Improving Exchange with Consumers within Mental Health Organizations: Recognizing Mental Ill Health Experience as a Sneaky, Special Degree, *International Journal of Mental Health Nursing*, 27(1), 2018, 227-235. <https://doi.org/10.1111/inm.12312>
- [10] Shpigelman, C. N.; HaGani, N.: The Impact of Disability Type and Visibility on Self-Concept and Body Image: Implications for Mental Health Nursing, *Journal of Psychiatric and Mental Health Nursing*, 26(3-4), 2019, 77-86. <https://doi.org/10.1111/jpm.12513>
- [11] Visser, K.; Slattery, M.; Stewart, V.: Help or Hinder? An assessment of the Accessibility, Usability, Reliability, and Readability of Disability Funding Website Information for Australian Mental Health Consumers, *Health & Social Care in the Community*, 29(5), 2021, 1378-1390. <https://doi.org/10.1111/hsc.13192>