

# A Study on Social Network Analysis and Information Dissemination Based on Big Data

Sitong Chen<sup>1,a,\*</sup>, Yanbin Long<sup>1,b</sup>

<sup>1</sup>University of Science and Technology Liaoning, Anshan, China  
<sup>a</sup>2934788312@qq.com, <sup>b</sup>1034182681@qq.com  
\*Corresponding author

**Abstract:** In this paper, we utilize big data technology to study the information dissemination phenomenon in social networks. Through collecting, processing and analyzing massive social network data, we reveal the basic laws of information dissemination, identify the key factors affecting the dissemination effect, and propose targeted information dissemination optimization strategies. The results show that these strategies can effectively improve the efficiency and scope of information dissemination in social networks, and provide new perspectives and methods for research and practice in related fields.

**Keywords:** big data; social network analysis; information dissemination; influencing factors

## 1. Introduction

With the rapid development of the Internet and the widespread popularization of smart devices, social networks have penetrated into every aspect of people's daily lives, becoming an important platform for information acquisition, communication and interaction, and entertainment. Every day, hundreds of millions of users post, share and comment all kinds of information on social networks, forming a huge and complex information dissemination network. Such information not only reflects users' personal interests, emotions and opinions, but also shapes social opinions and public topics to a certain extent<sup>[1]</sup>.

However, the information dissemination process in social networks is not simple and linear, but is influenced and constrained by a variety of factors, showing complex and variable characteristics. How to accurately portray and deeply understand this process, so as to optimize the information dissemination strategy and improve the efficiency and effect of information dissemination has become a common concern of academia and industry<sup>[2]</sup>.

The rise of big data technology provides a new opportunity to solve this problem. By collecting, storing, processing and analyzing huge amount of social network data, we can have a more comprehensive and in-depth understanding of user behavior, social relationship and information dissemination mechanism. Based on this, this paper aims to make use of big data technology to conduct in-depth analysis and research on information dissemination in social networks, reveal the basic laws of information dissemination, explore the key factors affecting information dissemination, and put forward corresponding information dissemination optimization strategies. This not only helps to improve the service quality and user experience of social network platforms, but also provides new perspectives and methods for research and practice in related fields<sup>[3-5]</sup>.

## 2. Related work

Before examining social network analytics and information dissemination based on big data, it is crucial to understand and review related work. This section will detail the underlying theories of social network analysis, the current state of research on information dissemination modeling, and the application of big data technologies in these areas.

Social network analysis is a multidisciplinary field that integrates theories and methodologies from sociology, mathematics, computer science, and statistics. A social network comprises a set of interconnected individuals or organizations, and social network analysis aims to uncover the patterns, laws, and relationships inherent in these structures. For instance, centrality measures (such as degree centrality and closeness centrality) illuminate the influence of key nodes within the network, while community detection identifies tightly knit groups. These fundamental theories furnish crucial tools and

frameworks for comprehending and analyzing information dissemination within social networks<sup>[6]</sup>.

The information dissemination model is a pivotal approach for studying the mechanisms and processes of information spread. Traditional models, like the SIR (Susceptible-Infected-Recovered) and SIS (Susceptible-Infected-Susceptible) models, rely primarily on probability theory and statistical principles to simulate and analyze information dissemination within networks. However, recent advancements in complex network theory have shifted the focus toward examining the impact of network structure on information transmission. Consequently, a range of network-based information transmission models have emerged, including the independent cascade model and the linear threshold model. These models consider not only the propagation probabilities between individuals but also the influence of network topology, node attributes, and other factors on information flow<sup>[7]</sup>.

The emergence of big data technology has presented new opportunities for social network analysis and information dissemination research. Big data technology encompasses a suite of techniques for storing, processing, and analyzing vast and diverse datasets. In the context of social network analysis, big data technology facilitates the collection and processing of large-scale social network data, encompassing user attributes, social relationships, user behaviors, and more. In information dissemination research, big data technology enables real-time monitoring and analysis of information dissemination processes, including propagation speed, reach, and pathways. Furthermore, the integration of big data technology with machine learning, deep learning, and other artificial intelligence techniques holds promise for predicting and modeling user behavior and information dissemination within social networks.

Relevant studies demonstrate the significant theoretical foundations and practical implications of social network analysis, information dissemination modeling, and big data technology in this domain. By delving into these areas, we can gain a deeper understanding and analysis of information dissemination phenomena within social networks, leveraging the power of big data<sup>[8]</sup>.

### **3. A big data-based approach to social network analysis**

Under the background of big data, social network analysis methods have been greatly expanded and deepened. Social network analysis based on big data not only involves data collection, storage and processing, but also covers data mining, analysis and visualization. The main steps and key technologies of this method are described in detail below<sup>[9]</sup>.

#### **3.1 Data collection**

The word "big" of big data is first reflected in the amount of data. Therefore, data collection is the first step of social network analysis based on big data. This is usually achieved through web crawlers, API interfaces, or third-party data providers. Web crawlers can automatically traverse social networking sites to capture user information, social relationships, published content and other data. API interface provides a more structured and standardized data acquisition method. In addition, in order to obtain more comprehensive and accurate data, it may also need to integrate and supplement with multiple data sources.

#### **3.2 Data storage and processing**

A large amount of raw data collected needs to be effectively stored and processed for subsequent analysis and mining. Distributed storage systems (such as Hadoop's HDFS) and NoSQL databases (such as MongoDB and Cassandra) are common technologies for handling big data storage. These technologies can store petabytes of data and provide high throughput data access. Data processing includes data cleaning, de duplication, conversion and formatting to ensure data quality and consistency<sup>[10]</sup>.

#### **3.3 Data mining and analysis**

After data preprocessing, the next key step is data mining and analysis. This includes identifying key nodes in social networks, community structure, information transmission paths, etc. Graph analysis algorithms (such as PageRank and HITS) and machine learning algorithms (such as classification, clustering and regression analysis) play an important role here. In addition, in order to better understand user behavior and information dissemination mechanism, it may also be necessary to combine natural language processing (NLP) technology to conduct emotional analysis and topic extraction of text content.

### **3.4 Data visualization**

Data visualization is the key link to present the analysis results in an intuitive and easy to understand way. Through visualization tools and technologies (such as D3.js, Tableau, Gephi, etc.), complex social network structures and information dissemination modes can be transformed into graphics, charts, or animations, helping analysts gain a faster insight into the laws and trends in data.

### **3.5 Real-time analysis and stream processing**

Unlike traditional batch processing analysis, social network analysis based on big data also emphasizes real-time. With the rapid update and dissemination of information on social networks, real-time analysis and stream processing technologies (such as Apache Kafka and Spark Streaming) have become critical. These technologies can process and analyze the newly generated data stream in real time, and provide near real-time analysis results and responses.

The social network analysis method based on big data is a complex process involving multiple links and technologies. By effectively combining these techniques and methods, we can gain a deeper understanding of the structure and function of social networks, reveal the internal mechanism and law of information dissemination, and thus provide powerful support for the optimization and development of social network platforms.

## **4. Analysis of the Laws of Information Dissemination and Influencing Factors**

Information dissemination in social networks follows certain patterns and is influenced by a variety of factors. In this part, we will discuss these rules and factors in detail, with a view to understanding in depth the mechanism of information dissemination in social networks.

### **4.1 Laws of Information Dissemination**

#### **4.1.1 Cascade Effect**

The propagation of information within social networks typically exhibits a cascade effect. That is, information originates from one node and successively spreads to its adjacent nodes, which then further disseminate it to their own neighbors, and the process continues. This dissemination mechanism facilitates the rapid spread of information across the entire network.

#### **4.1.2 Small-World Phenomenon**

Social networks are characterized by the small-world phenomenon, which indicates that the average distance between any two nodes is surprisingly short, irrespective of the network's overall size. This feature enables information to efficiently reach any part of the network with minimal hops.

#### **4.1.3 Community Structure**

The structural organization of communities within social networks profoundly impacts information dissemination. Information tends to circulate more rapidly among members of the same community, while its transmission between distinct communities may encounter obstacles.

#### **4.1.4 Time Decay Effect**

The freshness of information significantly influences its communication impact. Generally, recently emerged information is more apt to capture users' attention and spread widely, while over time, interest in that information gradually diminishes.

### **4.2 Analysis of Influencing Factors**

#### **4.2.1 User Characteristics**

Individuals' unique features, such as age, gender, and personal interests, shape their inclination to receive and disseminate particular types of information. For instance, younger users may be more predisposed to sharing content related to entertainment and fashion trends, whereas older individuals often gravitate towards health and wellness topics.

#### **4.2.2 Strength of Social Relationships**

The nature of social connections between users plays a pivotal role in information dissemination. Information exchange within close-knit relationships (like that of family members or close friends) tends to be more trustworthy, whereas interactions among weaker ties (strangers or casual acquaintances) can introduce greater diversity in information exchange.

#### **4.2.3 Quality of Information Content**

The intrinsic value and originality of information content determine its likelihood of being noticed and shared by users. Engaging, useful, and innovative content holds greater appeal and is more prone to being propagated, whereas inferior quality information may go unnoticed or be actively avoided.

#### **4.2.4 External Environmental Factors**

Circumstances beyond the social network, such as societal hotspots and timely events, can influence information flow within the network. When specific issues captivate widespread societal attention, related information tends to gain traction and spread more rapidly through social media channels.

In summary, information dissemination in social networks follows certain patterns and is influenced by many factors. A deeper understanding of these rules and factors can help us better understand the mechanism of information dissemination and provide powerful support for optimizing information dissemination strategies.

### **5. Strategies for optimizing information dissemination**

Based on the above analysis of information dissemination laws and influencing factors, we can propose a series of targeted information dissemination optimization strategies to improve the efficiency and scope of information dissemination in social networks.

#### **5.1 Enhancing the quality of information content**

High-quality information content is the key to attract users' attention and dissemination. Therefore, the first strategy for optimizing information dissemination is to improve the quality of information content. This includes ensuring the accuracy, usefulness and novelty of the information, as well as using catchy headlines and visual elements to enhance the appeal of the information. By providing valuable, interesting and easy-to-understand content, users' willingness to share and forward can be increased, thereby expanding the reach of the message.

#### **5.2 Utilizing social relations to enhance communication**

Social relationships play an important role in information dissemination. One of the strategies to optimize information dissemination is to leverage users' social relationships to enhance dissemination. This can be done by encouraging users to share information with their friends, family or followers. In addition, social network algorithms can be used to identify and influence key nodes (e.g., opinion leaders or influential users) to amplify the effect of information dissemination.

#### **5.3 Optimizing the timing of the release of information**

The time of information release has a significant impact on the communication effect. In order to optimize information dissemination, it is necessary to carefully select the time point of information release. This can determine the best publishing time by analyzing users' online behavior and active time. For example, publishing information in a time period with high user activity can increase the opportunities for information to be seen and interacted by users, thus improving the communication effect.

#### **5.4 Cross-platform communication strategies**

Posting and promoting information on multiple social platforms at the same time can expand the scope of information dissemination. Therefore, one of the strategies to optimize information dissemination is to adopt a cross-platform dissemination strategy. This involves adapting the content to different social platforms and customizing the communication strategy to take advantage of the characteristics and user groups of each platform. By establishing a consistent brand image and voice on

different platforms, users' awareness and memory of the message can be enhanced, thus improving the communication effect.

### ***5.5 Interaction and feedback mechanisms***

Encouraging users to interact with information and provide feedback is one of the key strategies for optimizing information dissemination. This can be achieved by setting up interactive functions such as commenting, liking and sharing, as well as responding to users' comments and questions on a regular basis. By establishing a positive interactive relationship with users, it can enhance their sense of participation and belonging to the information, thus increasing their willingness to disseminate the information. At the same time, collecting and analyzing user feedback can also help to improve information content and dissemination strategies to further enhance the effectiveness of dissemination.

In summary, strategies to optimize the dissemination of information include improving the quality of information content, using social relationships to enhance dissemination, optimizing the timing of information release, cross-platform dissemination strategies and establishing interaction and feedback mechanisms. These strategies can be used individually or in combination to improve the efficiency and scope of information dissemination in social networks.

## **6. Experimentation and analysis**

In order to verify the effectiveness of the above proposed information dissemination optimization strategy, we designed and implemented a series of experiments and conducted an in-depth analysis of the experimental results.

### ***6.1 Experimental design***

#### ***6.1.1 Dataset Preparation***

We have gathered authentic information dissemination data from multiple renowned social platforms, encompassing user details, social connections, posted content, dissemination pathways, and more. This comprehensive dataset serves as a rich source of information and a solid foundation for our experiments.

#### ***6.1.2 Control Group Establishment***

To precisely gauge the efficacy of our optimization strategy, we have established a control group experiment. In this group, information is published and disseminated in its conventional manner, without the application of any optimization techniques.

#### ***6.1.3 Strategy Implementation***

Within the experimental group, we have implemented the aforementioned optimization strategies. These strategies encompass enhancing the quality of informational content, leveraging social relationships to bolster dissemination, optimizing the timing of information release, facilitating cross-platform dissemination, and establishing an interactive feedback mechanism.

#### ***6.1.4 Evaluation Metrics***

To quantitatively assess the effectiveness of our strategy, we have devised a set of evaluation metrics. These metrics include dissemination speed, reach, and user engagement levels, among others. These comprehensive indicators enable us to accurately measure the efficacy and efficiency of information dissemination, providing a holistic view of our strategy's performance.

### ***6.2 Experimental results and analysis***

#### ***6.2.1 Propagation Speed and Range***

The experimental results demonstrate that the experimental group employing optimization strategies significantly outperforms the control group in terms of both information propagation speed and range. Notably, strategies aimed at enhancing the quality of information content and leveraging social relationships to strengthen communication have been particularly effective in improving the overall communication impact.

### **6.2.2 User Participation**

By introducing interaction and feedback mechanisms, user engagement in the experimental group has increased significantly compared to the control group. Users exhibit a greater willingness to interact with high-quality information content and provide valuable feedback, highlighting the effectiveness of these strategies in enhancing user participation.

### **6.2.3 Synergy of Strategy Combination**

Our findings also reveal that combining multiple optimization strategies can yield powerful synergistic effects. For instance, the combination of improving information content quality and implementing a cross-platform dissemination strategy can further amplify the reach and influence of information dissemination.

### **6.2.4 Limitations and Future Directions**

While the experimental results are generally promising, we have observed that certain strategies may have varying effects on different types of information or specific user groups. This underscores the importance of adapting and optimizing strategy combinations based on specific contexts and requirements in practical applications. Future research could explore the nuances of strategy effectiveness across diverse information types and user demographics to further refine optimization approaches.

## **6.3 Conclusions and Implications**

Through this experiment and analysis, we verify the effectiveness of the proposed information dissemination optimization strategies. These strategies are important for improving the efficiency and scope of information dissemination in social networks. Meanwhile, the experimental results also provide us with valuable insights: in future research, we can further explore the combination effect between different strategies and their applicability in different scenarios; in addition, we can consider introducing more advanced machine learning and artificial intelligence technologies to further enhance the effectiveness and efficiency of information dissemination.

## **7. Realize information dissemination optimization strategy through Java**

Implementing information dissemination optimization strategies in Java usually involves multiple steps, including data processing, algorithm design and implementation, and visualization or quantitative evaluation of results. The following is a simplified framework that shows how to use Java to implement these strategies.

### **7.1 Data model definition**

First, a data model is defined to represent users, information and social relationships in a social network.

```
class User {  
    String id;  
    List<User> followers; // Follower list  
    // Other attributes, such as name, interests, etc  
}  
class Post {  
    String id;  
    User author;  
    String content;  
    DateTime postTime;  
    List<User> likers; // Like user list  
    List<User> sharers; // Share user list  
    // Other attributes, such as comments, tags, etc  
}
```

### 7.2 Information dissemination simulation

Create a simulation function to model the propagation process of information. This function can be adapted to different optimization strategies.

```
class SocialNetwork {
    List<User> users;
    List<Post> posts;
    public void simulateSpread(Post post, int maxIterations) {
        Queue<User> toProcess = new LinkedList<>();
        toProcess.addAll(post.author.followers); // Initially add authors' followers to the queue
        int iteration = 0;
        while (!toProcess.isEmpty() && iteration < maxIterations) {
            User currentUser = toProcess.poll();
            If (currentUser.sharesPost (post)) { //Suppose this method checks whether the user shares
the post
                for (User follower : currentUser.followers) {
                    if (!post.sharers.contains(follower)) {
                        post.sharers.add(follower);
                        toProcess.offer(follower);
                    }
                }
            }
            iteration++;
        }
    }
}
```

### 7.3 Optimization strategy implementation

Implement specific optimization strategies, such as improving content quality, leveraging social connections, and optimizing posting times.

```
class OptimizedSocialNetwork extends SocialNetwork {
    // rewrite or add methods to implement optimization strategies
    @Override
    public void simulateSpread(Post post, int maxIterations) {
        //Optimization logic can be added here. For example, the quality of post content can be
evaluated first
        // Or, for example, adjusting the release time according to the user's online time
        super.simulateSpread(post, maxIterations);
    }
    public void optimizeContentQuality(Post post) {
        // Implementing the logic of content quality optimization
    }
    public void leverageSocialRelationships(Post post) {
        // Realizing the logic of leveraging social relationships
    }
    public void optimizePostingTime(Post post) {
        // Implementing logic to optimize release times
    }
}
}
```

### 7.4 Experimentation and analysis

Finally, experiments were conducted to compare the effects of different strategies.

```
public class Experiment {
    public static void main(String[] args) {
        OptimizedSocialNetwork socialNetwork = new OptimizedSocialNetwork();
        // create some users and posts
    }
}
```

```

    User userA = new User();
    User userB = new User();
    // ... Initialize user and post data
    Post post = new Post();
    // Initialize post data
    // Benchmarking: information dissemination without optimization strategies
    socialNetwork.simulateSpread(post, 100);
    System.out.println("Spread without optimization: " + post.sharers.size());
    // applying optimization strategies
    socialNetwork.optimizeContentQuality(post);
    socialNetwork.leverageSocialRelationships(post);
    socialNetwork.optimizePostingTime(post);
    // Optimized dissemination of information
    socialNetwork.simulateSpread(post, 100);
    System.out.println("Spread with optimization: " + post.sharers.size());
    // Analyzing the results, e.g. comparing the range, speed, etc., of the two propagations
}
}

```

Please note that the above code is a very simplified example to illustrate the basic concepts of how to implement information dissemination optimization strategies in Java. In practical applications, the information dissemination model will be more complex, involving more user behavior, network structure, time dynamics and other factors. In addition, the implementation of optimization strategies will be more specific and detailed, and may require the use of advanced technologies such as machine learning, graph algorithms, and statistical analysis.

## 8. Conclusion

Through experiments and analysis, this study deeply discusses the effectiveness of information dissemination optimization strategies. The experimental results show that the effectiveness and efficiency of information dissemination have been significantly improved through the implementation of optimization strategies, including improving the quality of information content, using social relations to strengthen communication, optimizing information release time, cross platform communication, and establishing interaction and feedback mechanisms. Specifically, the experimental group using the optimization strategy is superior to the control group in key indicators such as propagation speed, propagation range and user participation, which verifies the practical value of the proposed strategy.

These strategies are not only important for understanding the mechanism of information dissemination in social networks, but also provide powerful tools and methods for practitioners of information dissemination, which can help to improve the visibility, accessibility and influence of information. Meanwhile, the experimental results also reveal the applicability and effectiveness of different strategies in different scenarios and different types of messages, which provide a basis for the selection and adjustment of strategies.

## References

- [1] Hou Shuqi & Cheng Xuejun. *The absence and improvement of legal protection of personal credit information rights in the era of big data* [J]. *Credit Information*, 2022 (09).
- [2] Shan Jianjun. *Research on personal credit information collection in the context of big data era* [J]. *Research on Regional Finance*, 2022 (02).
- [3] Zhang Chen & Wan Xiangyu. *Construction of personal credit evaluation system and evaluation model under the background of big data* [J]. *Credit Information*, 2019 (10).
- [4] Wu Xuli. *Personal credit information protection in the era of big data - taking the improvement of personal credit reporting system as an opportunity* [J]. *Journal of Xiamen University (Philosophy and Social Sciences Edition)*, 2019 (01).
- [5] Han Peng. *Research on the Construction of China's Personal Credit Information System* [J]. *Journal of Hebei University of Economics and Trade*, 2016 (06).
- [6] Kong Dechao. *A preliminary study of big data credit reporting - based on the perspective of personal credit reporting* [J]. *Modern Management Science*, 2016 (04).
- [7] Wu Jingmei. *Look at big data credit reporting from the connotation and composition of credit* [J].



*Journal of Capital Normal University (Social Science Edition), 2015 (06).*

[8] Wang Mingyang, Jia Chongchong & Yang Donghui. *Social friend recommendation mechanism based on three degrees of influence [J]. Computer Application, 2015 (07).*

[9] Chen Shi. *International Comparison of Personal Credit Reporting System and Its Enlightenment [J]. Financial Forum, 2012 (10).*

[10] Chen Dongyu, Li Weijun, Peng Zhongli & Xu Yun. *Discussion on the necessity of introducing third-party personal credit investigation into online lending [J]. Credit, 2012 (01).*