Postprint: Core User Discovery in Social Resource Recommendation Based on k-core Collapse Sequence

Authors: Wu Huijuan, Jia Tina Du, Sun Hongfei, Jannatul Fardous

Date: 2017-11-08T00:00:00+00:00

Abstract

【Purpose】Through analyzing user behavior on social networking platforms, this study aims to identify core users within social niche groups, providing references for social resource recommendation services. 【Method】We collected 1,208 tags from Douban Reading users, constructed a tag co-occurrence matrix for the top 100 ranked tags, analyzed the K-core network structure of users, and investigated the fluctuation patterns of users' K-core collapse sequences. 【Results】Compared with methods such as degree centrality and minimum K-core depth values, the K-core collapse sequence-based method discovered new core users within social niche groups. 【Limitations】The sample data scale is relatively small and confined to a specific domain; the ranking problem cannot be effectively resolved, necessitating further improvement of the K-core analysis method. 【Conclusion】This study assists managers of social networking platforms in formulating or improving new resource recommendation strategies, thereby promoting the better development of social networking platforms.

Full Text

Identifying Core Users in Social Resource Recommendation Systems with K-Shell Collapse Sequences

Wu Huijuan[1,2], Jia Tina Du[2], Sun Hongfei[1], Jannatul Fardous[2]
[1](School of Economics & Management, Northeast Dianli University, Jilin 132012, China)
[2](Information Technology and Mathematical Sciences, University of South Australia, Adelaide 5001, Australia)

Abstract

[**Objective**] This study aims to identify core users within social minority groups by analyzing user behavior on social networking platforms, thereby providing insights for social resource recommendation services. [**Methods**] We collected 1,208 tags from Douban Reading users and constructed a tag co-occurrence matrix for the top 100 tags. We then analyzed the K-shell network structure of users and examined the fluctuation patterns of their K-shell collapse sequences. [**Results**] Compared with methods such as degree centrality and minimum K-shell depth values, our approach based on K-shell collapse sequences successfully identified new core users within social minority groups. [**Limitations**] The sample size was relatively small and limited to a single domain. Additionally, the ranking problem could not be adequately resolved, requiring further improvements to the K-shell analysis method. [**Conclusions**] This research can help social networking platform administrators formulate or improve resource recommendation strategies, thereby promoting better development of social networking platforms.

Keywords: Core user; Social resource recommendation; Social network analysis; K-shell collapse sequences
Classification Number: G250

Introduction

The problem of big data on the Internet is becoming increasingly severe, with daily data generation equivalent to filling approximately 168 million DVDs. This data contains rich information about user behaviors, particularly in the form of "tags," which together with users and resources constitute the primary objects of study in social recommendation research. How to deeply mine user relationship networks established through tags to provide users with precise resources that match their interest preferences represents a key research focus in both information science and management fields, and will remain a crucial direction for information services over the next decade.

In social networking services, user preferences for resources exhibit a long-tail distribution compared to mainstream information needs, tending toward minority interests. Users require personalized information, yet traditional filtering-based social resource recommendations primarily focus on popular mainstream content, making it difficult to identify needs within niche communities. On one hand, users need access to deep-level personalized information; on the other hand, as members of preference communities, users share similar needs with others in their community. Consequently, the relationship structure between users and resources forms and exhibits minority structural characteristics, creating what we term "niche communities." In virtual social communities with multi-dimensional semantic relationships (such as browsing, commenting, and forwarding), every user has the potential to become a niche core user within a specific domain. Therefore, a critical challenge in this field is how to leverage social platform resources and employ social network and complex network analysis methods to hierarchically identify these niche core users.

Literature Review

Related research on core users both domestically and internationally primarily focuses on "opinion leaders" and "niche experts." For instance, Momtaz et al. utilized social network analysis to identify opinion leaders by comprehensively considering centrality, structural holes, and indegree. Zhang et al. designed a clustering algorithm based on time series for community extraction and opinion leader mining, demonstrating its feasibility through empirical analysis of the Tianya community. Gnambs et al. established a moderation model for the knowledge and characteristics of opinion leaders. Wang et al. and Gu et al. analyzed opinion leader identification in the context of emergency events. Li et al. used MetaFilter as a data source to refine niche experts from the perspectives of betweenness centrality and clustering coefficients, identifying user characteristics and roles of niche experts across different periods. Chen et al. studied the influence mechanisms of opinion leaders in public opinion events from the perspective of their guiding role.

Overall, when using social network analysis methods for social resource recommendation, the key challenge lies in identifying core users within relationship networks. Many scholars have analyzed this problem from traditional perspectives such as degree centrality, betweenness centrality, and closeness centrality, while some have approached it from the K-shell value perspective. For example, He et al. studied personalized information recommendation in microblogs by leveraging the characteristic that higher K-values indicate stronger propagation capabilities. Kitsak et al. argued that users with high propagation efficiency in communities generally exist in K-shell decompositions. Zhou et al. proposed an importance evaluation matrix to identify the most important users in networks and analyzed the propagation capabilities of users with small K-shell values. Ren et al. analyzed the propagation capabilities of users with small K-shell values using neighbor user K-shell information. However, few scholars have analyzed core users from the perspective of K-shell collapse sequences. This paper primarily utilizes K-shell collapse sequences to identify the overall fragmentation of user relationship networks, discover core users and their niche communities, and recommend the resources they master to other users.

Methodology

3.1 Basic Approach In the social resource recommendation process, we constructed a social resource recommendation model as shown in [Figure 1: see original paper]. The approach involves first building a user relationship network and performing K-shell decomposition to construct users' K-shell collapse sequences. By analyzing the overall fragmentation of the network, we can identify core users if fragmentation exists. Next, we screen for high-density subgroups to discover niche communities. Finally, we recommend information mastered by core users to other members of these niche communities, completing the social resource recommendation process.

3.2 Constructing K-Shell Network Structure K-shell is a highly effective method for studying the hierarchical structure of complex networks, capable of revealing cohesive subgroups. It represents a measurement standard based on degree. Seidman argued that minimum degree criteria could be applied to study component structures to distinguish between high and low cohesion domains. Analysis of a graph' s "K-shell" structure serves as an important supplement to density measurement. A K-shell is a maximal subgraph where each user connects to at least K other users, meaning each user has a degree of at least K. A simple component constitutes a "1-shell" where all users are connected with a degree of at least 1. A "2-shell" is formed by removing all users with degree 1 and examining the relational structure among remaining users, consisting of remaining associated users with degree 2, with subsequent shells following this pattern.

For an undirected graph G=(V, E), where V represents users and E represents the tag co-occurrence set between users, the largest subgraph Hk=(W, E W) in the set W V is the K-shell. That is, for any V C, degree PH(v)≥K, indicating that a K-shell exists in the network. The K-shell decomposition process analyzes network structure layer by layer, extending from outer to inner layers. It recursively removes all users with degree values less than or equal to K, revealing structural characteristics and hierarchical properties of the network. The minimum K-shell users represent the outermost layer of the network. [Figure 2: see original paper] illustrates the K-shell decomposition of a user relationship network.

In [Figure 2: see original paper], n users can be partitioned into four groups with degrees of 2, 3, 5, and 6 respectively, where the 6-shell is the largest connected subgraph. Users in the 6-shell occupy the core area of the core-periphery structure, with each user connected to at least six other users in the graph. From the largest 6-shell to the 5-shell, 3-shell, and finally the smallest 2-shell, larger shells are subgraphs of smaller shells, meaning all users contained in larger shells can be found completely within smaller shells. During the clustering process from smaller to larger shells, residual users may emerge at each level.

3.3 Analyzing K-Shell Collapse Sequences While K-shell represents a relatively cohesive region within an entire graph, it is not necessarily the most cohesive subgraph, as some regions may be loosely connected yet highly cohesive, indicating overall network fragmentation. Seidman employed core collapse sequences to estimate a network' s overall fragmentation. The core collapse sequence focuses on residual users generated during each level of clustering. Points in a K-shell can be divided into two sets: those in K+1 and those not in that shell. Seidman termed the latter group the K-residual set. The proportion of points disappearing from the shell each time K increases by one unit can be arranged as a vector (a simple row of values) that describes the local density structure within components. If values in the vector continue increasing to relatively high K values, the network structure demonstrates consistency. If the vector shows zero values persistently appearing after lower K values, the network contains multiple high-density regions.

illustrates the K-shell collapse process. As the K-shell gradually collapses with K values increasing from 0 to 6, many residual users are generated, producing a core collapse sequence of: (0.05, 0.10, 0.15, 0.00, 0.10, 0.15, 0.45). The specific sequence changes are shown in [Figure 3: see original paper].

In [Figure 3: see original paper], when K ranges from [0, 2], vector values increase from 0.05 to 0.15. Then, when K equals 3, the vector value becomes 0, causing minor fluctuations in the collapse sequence. As vector values generally increase gradually, high-density subgroups emerge when K equals 4 and 5, centered around users (U4, U15, U1, Un-1) respectively, forming subgroups (U4, U15, U6, U13) and (U1, U10, U12, U13, U17, Un-1). These represent niche communities discovered beyond the common 6-shell subgroup (U12, U13, U9, U10, U6, U7, U8, U16, U17).

For the first subgroup, resource preferences of U4 and U15 can be recommended to other users in the niche community. Similarly, for the second subgroup, resource preferences of U1 and Un-1 can be recommended.

Data Collection

4.1 Data Acquisition We randomly crawled user data from Douban's "Reading" page, selecting samples from the list of users each person followed, and expanded our sample using a snowball sampling approach. We collected data from 35 users, assigning each a unique identifier for statistical convenience. Each user's tags were counted after removing duplicate annotations, resulting in a total of 1,208 tags across all 35 users.

We sorted all tags from the 35 users by frequency in descending order and selected the top 100 tags, with partial data shown in . The tag "literature" appeared most frequently, indicating that many of the 35 users favored literary works.

We constructed a tag co-occurrence matrix for the 35 users according to their assigned identifiers. First, we sorted the tag collection by usage frequency and selected the top 37 tags as sample data. We then analyzed how these 35 users utilized these 37 tags, calculating the number of times each pair of users used the same tags. If two users used the same tag three times, their tag co-occurrence value was set to 3. The diagonal of the matrix was set to a specific value (e.g., 0) to represent the relationship between a user and themselves, resulting in the final tag co-occurrence matrix.

Results Analysis

4.2 Results Analysis The K-shell analysis results for Douban data in Ucinet are shown in [Figure 4: see original paper]. The figure reveals that the 35 users can be partitioned into seven groups with degrees of 3, 4, 5, 7, 8, 9, and 10. The 10-shell is the largest connected subgraph, with its users occupying the core area of the core-periphery structure, where each user connects to at least ten other users. From the largest 10-shell to the 9-shell, 8-shell, 7-shell, 5-shell, 4-shell, and finally the smallest 3-shell, larger shells are subgraphs of smaller shells, meaning all users in larger shells can be found within smaller shells.

During the clustering process from smaller to larger shells, residual users may emerge at each level. shows the K-shell collapse for Douban data. As K values increase from 0 to 10, many residual users are generated, producing a core collapse sequence of: (0.06, 0.09, 0.00, 0.11, 0.11, 0.14, 0.00, 0.03, 0.03, 0.03, 0.40). The sequence changes are illustrated in [Figure 5: see original paper].

In [Figure 5: see original paper], when K ranges from [0, 5], vector values first decrease from 0.06 to 0, then increase to 0.14. Despite minor fluctuations, the vector values generally increase gradually with relatively small magnitude, making collapse sequence variations in this interval negligible. However, when K ranges from [5, 10], vector values first decrease from 0.14 to 0, then increase to 0.03, causing noticeable fluctuations. Finally, when K equals 10, the vector value suddenly increases to 0.40, representing a significant change in the collapse sequence. Consequently, when K equals 7, 8, and 9, the network generates three high-density subgroups. The first two represent niche communities: the first niche community (U30, U1, U26, U27, U2, U10) centers on U30 and U1, while the second niche community (U25, U26, U34, U12, U16) centers on U25. The third high-density subgroup (U6, U7, U8, U9, U10, U12, U13, U16, U17, U19, U22, U26, U28) centers on U26 and U28. Core users can recommend resources they master to other users in their respective niche communities.

(1) Comparison with Degree Centrality

In social network analysis, higher user centrality indicates a more central network position. Some users with relatively high degree centrality are shown in . The table reveals that U26 has the highest degree centrality, indicating it is a core user in the network, followed by U28, U10, U13, etc. However, this table does not identify the core status of U1 and U30. From the perspective of K-shell collapse sequences, however, we discovered that U1, U25, and U30 are all core users in niche communities.

(2) Comparison with Minimum K-Shell Depth

In social network analysis, numerous users have small K-shell values and are generally considered peripheral users. By examining the maximum K-shell value (depth) from the neighbor set of such nodes, core users can be identified. shows the depth of minimum K-shell values (generally 0 or 1) for Douban data. Since U21 has a depth of 28, meaning the maximum K-shell value in its neighbor set is 28, we can identify neighbor U26 as a core user when combined with [Figure 4: see original paper]. However, this method fails to identify the core status of U1, U25, U28, and U30.

Therefore, compared with degree-based and minimum K-shell depth methods, the K-shell collapse sequence approach demonstrates certain advantages in identifying core users using social network analysis.

Conclusion

This paper proposes utilizing K-shell collapse sequences to identify core users in social network groups and provide niche recommendations to their communities. Using sample data extracted from Douban' s reading social platform, our empirical analysis demonstrates that the K-shell collapse sequence method is both feasible and superior compared to degree centrality, betweenness centrality, and minimum K-shell depth approaches. However, this study has limitations. First, the research only examined Douban reading users, resulting in a relatively small sample size. Second, when K-shell values are identical, we cannot effectively rank other users within the niche community containing the core user. Future work will expand the sample size and further improve the K-shell collapse sequence ranking algorithm to provide better social niche information resources for users.

References

[1] Guy I, Zwerdling N, Ronen I, et al. Social Media Recommendation Based on People and Tags [C]. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2010: 194-201.
[2] Wang J, Clements M, Yang J, et al. Personalization of Tagging Systems [J]. Information Processing & Management, 2010, 46(1): 58-70.
[3] Guy I, Carmel D. Social Recommender Systems [C]. In: Proceedings of the 20th International Conference Companion on World Wide Web. ACM, 2011: 283-284.
[4] Wortham J. Search Takes a Social Turn [EB/OL]. [2016-04-02]. http://premiumknowledge.typepad.com/files/_{r1}.pdf.
[5] Hu Jiming, Zhang Mandi. Research on the Social Minority Recommendation Model Based on User-Resources Association [J]. Information Studies: Theory & Application, 2014, 37(4): 123-126, 118.
[6] Jean Maisonneuve. Group Dynamics [M]. Translated by Yin Shicai, Sun Zhaotong. Beijing: The Commercial Press, 1997.
[7] Li Gang, Ye Guanghui. Research on Information Fusion for Multiple-sensor Expert Features [J]. New Technology of Library and Information Service, 2014(4): 27-33.
[8] Momtaz N J, Aghaie A, Alizadeh S. Identifying Opinion Leaders for Marketing by Analyzing Online Social Networks [J]. International Journal of Virtual Communities and Social Networking, 2011, 3(3): 19-34.

[9] Zhang W, He H, Cao B. Identifying and Evaluating the Internet Opinion Leader Community Based on K-clique Clustering [J]. Neural Computing & Applications, 2014, 25(3): 595-602.
[10] Gnambs T, Batinic B. The Roots of Interpersonal Influence: A Mediated Moderation Model for Knowledge and Traits as Predictors of Opinion Leadership [J]. Applied Psychology, 2013, 62(4): 597-618.
[11] Wang Guohua, Zhang Jian, Bi Shuaihui. Study on Opinion Leaders of Emergencies in Network Opinion Evolution: A Case Study of Yao Jiaxin Event [J]. Journal of Intelligence, 2011, 30(12): 1-5.
[12] Gu Pinhao, Jiang Guan. Analysis on Network Opinion Leaders in Public Emergencies—A Case Study of YANG Dacai Event [J]. Journal of Intelligence, 2013, 32(5): 20-24.
[13] Li Gang, Ye Guanghui, Zhang Yan. Feature Recognition of Niche Expert —Empirical Analysis Based on MetaFilter Dataset [J]. New Technology of Library and Information Service, 2015(6): 71-77.
[14] Chen Fuji, Chen Ting. Research on Public Opinion Emergencies Evolution: Based on the Perspective of Opinion Leaders Guiding Role [J]. Information and Documentation Services, 2015(2): 23-28.
[15] He Y, Tan J. Study on SINA Micro-blog Personalized Recommendation Based on Semantic Network [J]. Expert Systems with Applications, 2015, 42(10): 4797-4804.
[16] Kitsak M, Gallos L K, Havlin S, et al. Identification of Influential Spreaders in Complex Networks [J]. Nature Physics, 2010, 6(11): 888-893.
[17] Zhou Xuan, Zhang Fengming, Li Kewu, et al. Finding Vital Node by Node Importance Evaluation Matrix in Complex Networks [J]. Acta Physica Sinica, 2012, 61(5): 05020101-05020107.
[18] Ren Zhuoming, Liu Jianguo, Shao Feng, et al. Analysis of the Spreading Influence of the Nodes with Minimum K-shell Value in Complex Networks [J]. Acta Physica Sinica, 2013, 62(10): 10890201-10890206.
[19] Seidman S B. Network Structure and Minimum Degree [J]. Social Networks, 1983, 5(3): 269-287.
[20] Zong Gang, Zhao Xiaodong. Construction of Boolean Bipartite Network for Chinese Beer Brands Based on K-Core Analysis [J]. Journal of Beijing University of Technology, 2013, 39(6): 936-940.

Author Contributions

Wu Huijuan: Conceptualized the research, designed and implemented the methodology, wrote the manuscript; Jia Tina Du: Provided revision suggestions; Sun Hongfei: Conducted data analysis; Jannatul Fardous: Collected data.

Conflict of Interest Statement

All authors declare no conflict of interest.

Supporting Data

The supporting data is self-archived by the authors and available upon request at E-mail: 413170720@qq.com.
[1] Wu Huijuan, Jia Tina Du, Sun Hongfei, Jannatul Fardous. usercollection.xls. User data collection.
[2] Wu Huijuan, Jia Tina Du, Sun Hongfei, Jannatul Fardous. tagmatrix.xls. Tag co-occurrence matrix.
[3] Wu Huijuan, Jia Tina Du, Sun Hongfei, Jannatul Fardous. networkdensity.doc. Network density diagram.
[4] Wu Huijuan, Jia Tina Du, Sun Hongfei, Jannatul Fardous. structure.doc. Core-periphery structure diagram.

Manuscript Received: 2016-04-15
Revised Manuscript Received: 2016-05-09

Note: Figure translations are in progress. See original paper for figures.