

基于社区发现的农产品共同购买网络可视化分析

孙瑞志^{1,2} 侯曼曼¹ 赵凯艺¹

(1. 中国农业大学信息与电气工程学院, 北京 100083;

2. 农业农村部精准农业技术集成科研试验基地(畜牧业), 北京 100083)

摘要: 在实际的农产品生产销售场景中,消费者共同购买行为所形成的网络结构非常复杂且多变。虽然社区发现算法可以有效挖掘出共同购买行为背后隐藏的信息,但存在分析结果不易理解、支持决策条件不足的问题。为此,考虑到社区发现算法在共同购买网络分析中的广泛应用及可视化技术对分析结果的呈现能力,提出一种基于社区发现的共同购买网络可视化分析方法。该方法首先利用社区发现算法 Clauset - Newman - Moore (CNM) 划分共同购买网络,其次对网络结构的不同社区中农产品数量、共同购买行为频数和农产品价格众数所占百分比进行分析,然后对各社区共同购买农产品的消费者信息进行交互式分析,最后将分析结果进行交互式可视化展示,根据可视化界面得出共同购买的一些行为规律,进而深度挖掘消费者的消费模式。为了更好地呈现该分析方法,通过设计实现的可视化分析界面,交互式探索分析了一组在青岛地区的农产品动态销售数据。

关键词: 共同购买网络; 社区发现; 可视化; Clauset - Newman - Moore

中图分类号: TP391.1 **文献标识码:** A **文章编号:** 1000-1298 (2021) S0-0482-07

Visual Analysis of Co-purchase Network for Agricultural Products Based on Community Discovery

SUN Ruizhi^{1,2} HOU Manman¹ ZHAO Kaiyi¹

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Scientific Research Base for Integrated Technologies of Precision Agriculture (Animal Husbandry), Ministry of Agriculture and Rural Affairs, Beijing 100083, China)

Abstract: In the real production and sales scenarios of agricultural products, the network structure formed by consumers' co-purchase behavior is very complex and changeable. Although the community discovery algorithm can effectively dig out the hidden information behind the co-purchase behavior, there are problems that the analysis results are not easy to understand, and the supporting decision-making conditions are insufficient. Because of the widespread application of community discovery algorithms in the analysis of co-purchase networks, and the ability of visualization technology to present the analysis results, a visual analysis method for the co-purchase networks of agricultural products based on community discovery was proposed. Firstly, the method used the community discovery algorithm Clauset - Newman - Moore (CNM) to divide the agricultural product co-purchase network. Secondly, the quantity of agricultural products in different communities in the network structure, the frequency of co-purchase behaviors, and the proportion of the price mode of agricultural products were analyzed, and then interactive analysis on the information of customers who co-purchase a certain agricultural product in each community was conducted. Finally, the analysis results were displayed interactively and visually. According to the visual interface, some behavioral rules of co-purchases were obtained, and then their consumption rules were deeply explored. In order to better present the visual analysis method, a set of dynamic sales data of agricultural product in Qingdao area were interactively explored and analyzed through the design of a visual analysis interface, and the sales model found can not only inspire the improvement and optimization of the manufacturers' marketing methods, but also can help consumers to better choose agricultural products that suit them.

Key words: co-purchase network; community discovery; visualization; Clauset - Newman - Moore

收稿日期: 2021-07-18 修回日期: 2021-09-08

基金项目: 国家重点研发计划项目(2016YFB0501805)和综合数据服务系统之基础平台建设项目(JZNYYY001)

作者简介: 孙瑞志(1964—),男,教授,博士生导师,主要从事复杂网络和社区发现研究,E-mail: sunruizhi@cau.edu.cn

文献[10]通过分析共同购买网络,挖掘共同购买网络定义的用户社区中用户行为模式的变化,发现产品购买遵循“长尾”原则,即其中很大一部分购买属于很少出售的商品。文献[11]研究了共同购买网络结构如何影响电子商务的需求,以及这种网络结构如何通过改变电子商务网页之间的流量分布来改变需求模式,发现了大多数人所需要的产品类别受网络结构的影响很大^[12-14]。共同购买网络也可以利用社区发现算法把一些节点划分成社区,从而可依据社区内部与社区间的结构得出感兴趣的信息^[15-18]。文献[19]使用 Amazon 共同购买网络作为基准数据,提出了一种基于最大模块度与层次划分的 CNM 社区检测方法,可以从网络中提取有意义的社区。学者们通过对共同购买网络的研究,在帮助用户选择更好的产品上有显著优势,但是这些算法处理结果缺少一种直观的方式帮助用户理解与分析。可视化是把数据信息转换为人眼能识别的可视符号,是人机交互、虚拟现实应用的关键技术^[20]。可视化图是信息可视化领域的一个重要分支,又称

为关联网、网络或关联图,是关联数据分析过程中的重要部分^[21-26]。因此,基于交互式探索的可视化模型能更好地实现共同购买网络的分析过程。

针对共同购买网络分析中社区发现算法的分析结果不易理解和不够直观的问题,本文设计一种交互式可视化模型用于对共同购买的社区网络结构进行分析展示。另外,每一个共同购买网络社区的形成往往不是受单一因素影响,因此本文将社区内被共同购买农产品的一些属性及社区内消费者的地理位置信息也进行可视化展示,并建立与社区网络结构可视化模型之间的交互方式。通过基于社区发现的共同购买网络可视化模型,对一个区域内的农产品销售规律和消费者购买行为进行深度挖掘,以期为企业清楚地描绘农产品的销售规律,帮助消费者选购更适合自己的农产品。

1 数据与算法

1.1 研究对象

共同购买网络表示连接共同购买农产品的无向图,在这个网络中,节点用来表示不同的农产品,边用来表示共同购买的关系。网络规模越大,越难找到感兴趣的信息,也越有可能出现错误的分析。

本文基于山东省青岛市茶叶的真实销售数据进行实验研究,包括 2018 年 1—4 月份每月的销售数据,所分析的信息包括农产品信息、消费者信息和购买订单信息。图 1 所展示的网络即 2018 年 4 月的农产品共同购买网络,图中节点的标注是农产品编号。

1.2 社区发现算法

Clauset – Newman – Moore (CNM) 社区发现算

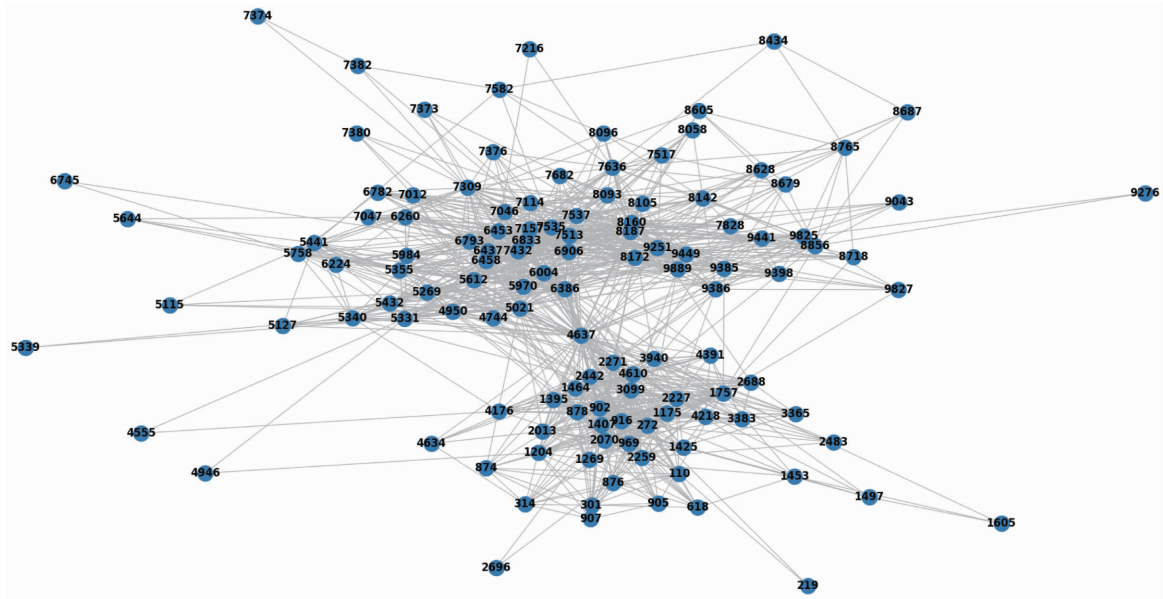


图 1 共同购买网络数据图
Fig. 1 Co-purchase network data

法^[19]是在 NewmanFastGN^[27]算法的基础上提出的一种贪心算法,其采用堆的数据结构来计算和更新网络的模块性。这个贪心算法并不是通过初始的连接矩阵计算模块度的增量 ΔQ ,而是直接构造一个模块度的增量矩阵 ΔQ ,然后通过对它的元素进行更新来得到模块度最大的一种社区结构。如果合并两个不相连的社区,模块度 Q 的值是不会变的。因此,只需要存储那些有边相连的社区 i 和 j 相应的元素,从而达到节省存储空间、降低算法复杂度的目的。

2 可视化基本模型

图2是本文设计实现的动态交互式可视化方法的框架图。该可视化方法,能用来分析不同时间段内、不同社区中共同购买农产品及对应消费者的购买行为模式。首先,将实际生产数据接入,包括消费者资料、产品信息、销售记录和数据字典等。其次,构建实时的农产品共同购买网络,利用CNM算法进行社区发现,其结果可以与其他信息关联,进而分析产品社区的空间与时间属性,挖掘消费者购买行为背后的深层知识。然后,映射获得的知识,形成分析报告,根据设计的视觉模型,绘制、渲染出易于理解的可视化展示界面。最后,将可视化结果在任意屏幕上显示,使用户能够快速、便捷地获得最新的知识。在这个过程中,可视化基本模型的设计可以辅助用户更好地获取相应的信息。因此,本文主要从可视化基本模型的设计展开来论述,也描述了社区发现算法与关联分析方法在其中的作用。

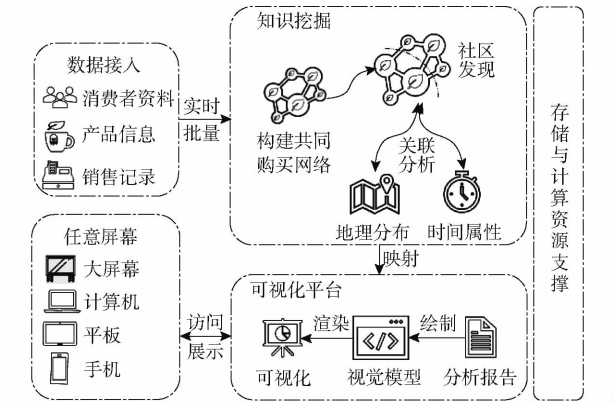


Fig. 2 Dynamic interactive visualization framework

2.1 共同购买网络的社区发现及可视化

首先,利用被广泛使用的CNM算法来寻找某一时刻共同购买网络的社区结构。然后,交互式可视化方法中的时间轴默认显示当前时刻的社区结构,但可以动态调整为不同时刻。在这个社区关系的动态可视化图中,节点表示农产品,边为农产品间的共

同购买关系。从可视化结果来看,用户可以清晰地看出网络中的社区结构。

图3为2018年4月的共同购买网络社区结构的可视化,通过不同的颜色显示了3个大小不同的社区,社区2里的共同购买农产品数量明显少于社区0和社区1。

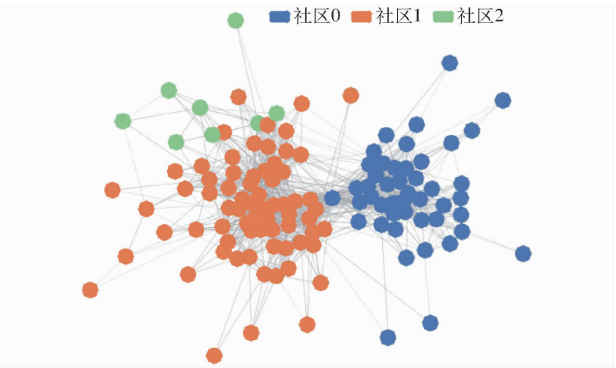


图3 社区结构划分及可视化
Fig. 3 Community structure division and visualization

本文可视化中的一个重要任务是动态网络分析以及跟踪社区的演化。分析方法可以通过图3社区发现里的交互式机制实现,虽然默认显示全部社区,但显示的社区可自由地进行多选实现,当在上方点击任意社区的按钮时,可以改变该社区显示状态。

2.2 农产品信息可视化

随着销售记录中时间轴的演化,某一时刻下的共同购买的农产品信息将通过可视化平台中的扇形图进行动态交互式可视化分析展示。如图4所示,从左到右的扇形图依次显示2018年4月这一时刻下不同社区中的农产品数量、共同购买行为频数和农产品价格众数所占百分比。当用户的交互操作关注在扇形图的不同扇区时,即可显示该时刻下相应的社区信息。

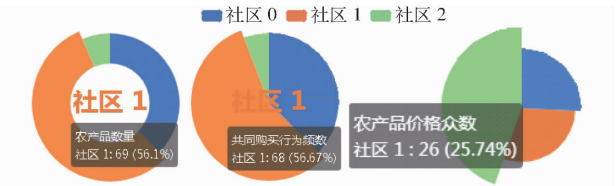


图4 共同购买的农产品信息交互式可视化
Fig. 4 Interactive visualization of co-purchase product information

图4中,由第3个扇形图可看出黄色区域所对应社区1里农产品的价格众数所占百分比最低,仅为25.74%,而在左侧两个扇形图中该社区下农产品数量和共同购买行为频数所占百分比最大,分别是56.1%和56.67%。蓝色区域所对应社区0的农产品价格众数所占百分比次低,而蓝色区域在左侧

两个图中所占百分比为次高,绿色区域所对应社区 2 里农产品的价格众数最高,在左侧两个图中社区 2 的农产品数量和共同购买行为频数所占百分比都为最低。

在对应时刻里,本文发现当某一社区的农产品价格众数占有所有社区的农产品价格众数比例越低,这个社区里的农产品数量所占比例越大,共同购买行为越频繁,代表这类社区的农产品类型是消费者共同购买较多的种类。因此,大部分消费者倾向于共同购买一些价格较低的农产品,反之,农产品数和共同购买行为频数所占比例较小的社区表示这类社区的农产品价格只能被少部分人接受。据此可以得出结论,在一定范围内农产品价格众数越低,农产品数和共同购买行为频数则越多。

2.3 消费者信息可视化

为了直观分析在当前时刻里各地区顾客倾向于购买哪一类社区里的农产品,对每个时间段里各社区共同购买某农产品的顾客信息进行交互式分析。如图 5 所示,利用分组柱状图显示青岛市各区县购买不同社区农产品的顾客数。当用户关注某一区县下共同购买农产品社区时,则动态显示出对应的区县里共同购买农产品社区的消费者数量。

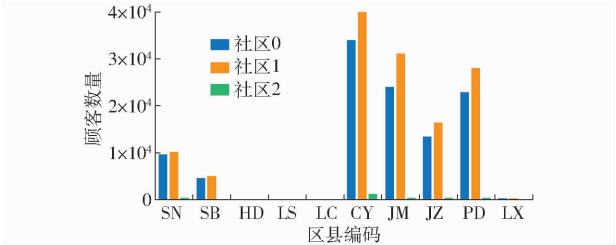


图 5 不同社区中消费者信息的交互式可视化
Fig.5 Interactive visualization of consumer information in different communities

通过交互式可视化分析,观察出每个区县在对应的时刻里,总有两个社区在共同购买农产品中对应的消费者数量相近。而且,这两个社区价格众数相近,且都低于另一个消费者数量不相近的社区价格众数。如图 5 所示,在 2018 年 4 月,每个区县里蓝、黄两种颜色对应的社区 1、2,消费者共同购买数量相近,绿色对应的社区 0 消费者共同购买数量则很少。由上述规律可知,消费者一旦购买某种农产品,可能会共同选购两类社区价格众数较低且相近的农产品,而非单独某一社区里的农产品,表明消费者愿意在自己能接受的价格范围内,多尝试几种农产品类型。

另外,为了分析购买各社区农产品消费者的地理位置分布规律,本文关联了农产品社区的消费者位置信息,并通过不同颜色代表相应社区里的消费

者进行地图可视化。如图 6 所示,基于地图对消费者的空间分布进行交互式可视化分析,综合 4 个时间段里消费者的地理位置,本文发现地区 HD(黄岛区)的消费者在前两个时间段里热衷于购买所有社区的农产品,地区 CY(城阳区)的消费者在后两个时间段里倾向于购买所有社区里的农产品,而地区 HD 的消费者在后两个时间段则不愿意购买这类农产品。

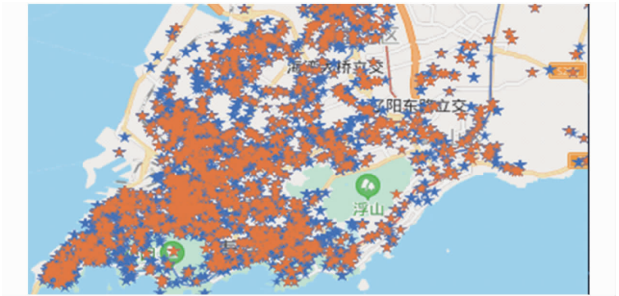


图 6 共同购买社区的消费者空间分布地图
Fig.6 Interactive visualization of consumer location

通过上述分析,结果证实消费者在一段时间内热衷于对某一类农产品的购买,而并非长时间倾向于购买某类农产品。另外一方面,这种农产品购买者主要集中在 HD(黄岛区)、CY(城阳区)、JZ(胶州区)、PD(平度区)、JM(即墨区)等经济发展程度相对较低的地区。反之,经济发展程度相对较高的 LS(崂山区)、LX(莱西市)、SB(市北区)、LC(李沧区)地区购买此类农产品的消费者较少。对此规律进行分析,可以认为大多数发达地区对这类农产品中的低端类型需求较小,高端类型销量也不大。

3 交互设计和分析

本文利用 ECharts、百度地图 JavaScript API GL 和 D3 等开源可视化库,设计开发了面向农产品共同购买网络的交互式可视化分析模型。基于 jQuery 技术,建立了有效的可视化基本模型间的交互机制,以实现多方面(消费者、产品、订单)、多维度(时间、空间)动态分析农产品共同购买网络的目的。

3.1 共同购买网络的交互式设计

本文基于社区发现算法,动态划分了每个时间段中的共同购买网络,获得了相应的社区结构。图 7 是用户与可视化模型的交互示意图。该交互式可视化模型允许用户对时间轴的時刻进行点击,也可以对社区进行多选从而得到不同時刻的共同购买网络社区结构的演化过程。另外,该可视化模型也可以通过缩放显示整体划分或详细显示社区结构内的连接,也可以单击查看某一结点的具体信息。在选定的时间与社区下,右侧的扇形图和左侧上方的

分组柱状图,分别显示对应的农产品信息和消费者区县信息,可通过鼠标悬停显示具体信息。左侧下方

方的地图可通过拖拽或者缩放显示出不同的层次和分析粒度。

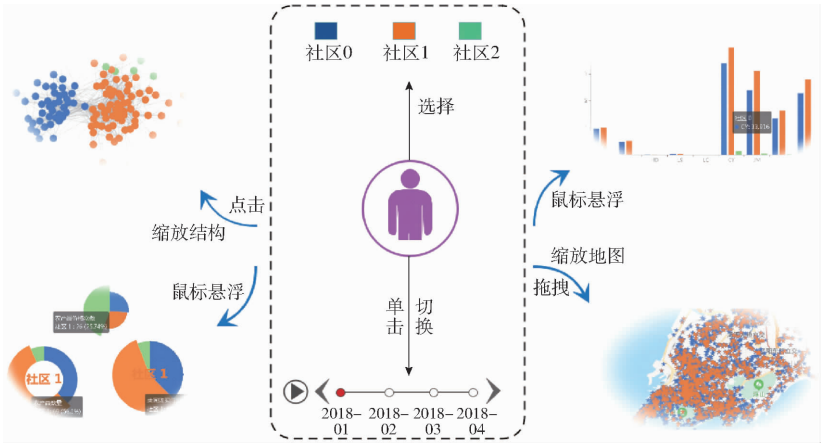


图 7 用户交互示意图

Fig. 7 Schematic of user interaction

在中下方的时间轴上选择了 2018 年 4 月,得到共同购买网络可视化模型如图 8 所示。中间的社区网络结构与左右两侧信息对应,可分析得出对应社

区农产品价格众数越低,共同购买的农产品频数和对应的消费者数越多,且这些消费者大多集中在欠发达地区。

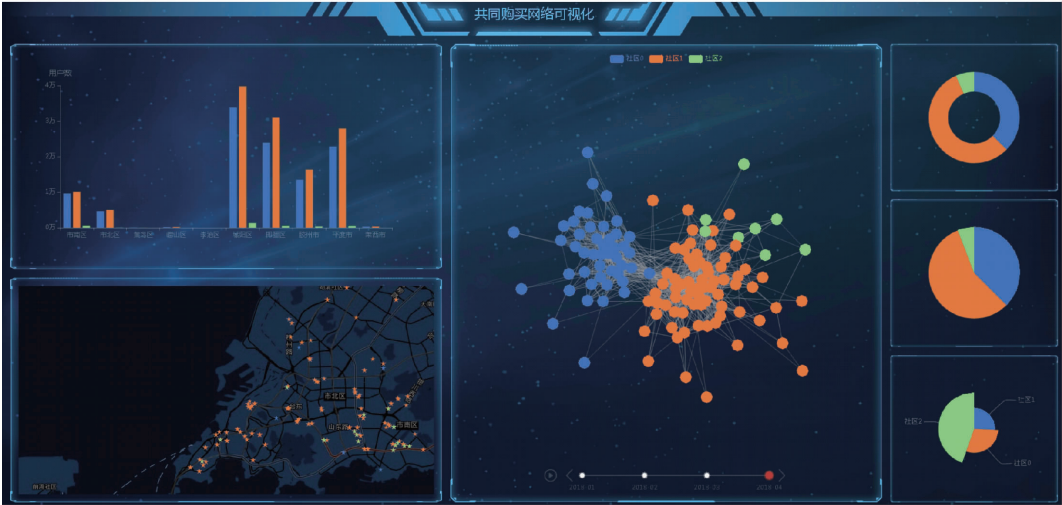


图 8 共同购买网络可视化模型

Fig. 8 Co-purchase network visualization model

3.2 分析与讨论

通过对山东省青岛市市南区、市北区、台东区、四方区、李沧区、黄岛区、崂山区、胶州市、即墨市、平度市、胶南市、莱西市的茶叶真实销售数据进行实验研究,包括 2018 年 1—4 月份每月的销售数据,构建共同购买网络可视化模型,可实时观察每个时刻下社区结构的整体划分情况,动态交互式对每个时间段里的社区结构、共同购买农产品信息和消费者信息进行分析,得到以下分析结果:

(1)虽然这些农产品的平均价格相差不大,但是消费者无论在哪个时间段都更倾向于购买价格相对较低的农产品,且购买这些农产品的大多数为发展程度相对较低地区的消费者。

(2)在同一时刻里会同时热销两个社区里的农产品,表明消费者愿意在自己能接受的农产品价格范围内尝试较多类型的农产品。

(3)其中黄岛区、城阳区对这类农产品的喜爱有明显的阶段性,表明消费者对某一类农产品的购买行为受时间影响,并非长时间倾向于购买某类农产品。

综合上述分析,对此类农产品可作出以下推测:在发展程度相对较低地区,消费者会对社区中价格较低的一些农产品有显著的共同购买趋势。

4 结束语

本文设计实现了交互式可视化模型,用于对共同购买网络进行深度挖掘与分析。首先基于社区发

现算法 CNM 对共同购买网络进行探索。然后通过共同购买网络可视化模型,动态交互式分析共同购买网络结构的多方面信息,包括不同社区内农产品信息和消费者信息。最后,基于可视化展示结果,发现不同地区的消费者共同购买行为的一些规律,进

而深度挖掘其消费规律。基于社区发现的共同购买网络可视化模型,揭示了消费者对此类农产品的购买趋势,给商家清楚地描绘出不同地区消费者对不同价格农产品的购买模式,以便厂家选择更适合的农产品营销策略。

参 考 文 献

- [1] AL-SADI A, AL-AYYOUB M. Identifying influential positively perceived users in co-purchase networks [C] // 2018 9th International Conference on Information and Communication Systems (ICICS). IEEE, 2018:78–83.
- [2] BEMO B C S, GABARDO A C, HATTORI L T, et al. A framework for analyzing book covers and co-purchases using object detection and data mining methods [C] // 2019 IEEE Latin American Conference on Computational Intelligence (LA – CCI). IEEE, 2019:1–6.
- [3] CHEN H M. Do online recommendations matter? —a multimodal investigation of Amazon’s co-purchase network [J]. Journal of Digital Information Management, 2015, 13(3): 176–184.
- [4] GHOSH S, MITRA A, BASUCHOWDHURI P, et al. Analysis of online product purchase and predicting items for co-purchase [C] // Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics. Springer, New Delhi, 2016: 581–591.
- [5] COLIZZA V, FLAMMINI A, SERRANO M A, et al. Detecting rich-club ordering in complex networks [J]. Nature Physics, 2006, 2(2): 110–115.
- [6] BLLACKHURST J P, CALMAN M A, BONDESEN L C, et al. Credit source recommendation based on product level data analysis; U. S. Patent Application 13/953,264 [P]. 2015–01–29.
- [7] DELLAROCAS C, ZHANG X M, AWAD N F. Exploring the value of online product reviews in forecasting sales; the case of motion pictures [J]. Journal of Interactive Marketing, 2007, 21(4): 23–45.
- [8] LI X, LAI J. Analysis of the dynamic co-purchase network based on image shape feature [C] // International Conference on Genetic and Evolutionary Computing. Springer, Singapore, 2017: 93–100.
- [9] BEDI P, SHARMA C. Community detection in social networks [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2016, 6(3): 115–135.
- [10] LESKOVEC J, ADAMIC L A, HUBERMAN B A. The dynamics of viral marketing [J]. ACM Transactions on the Web, 2007, 1(1): 5–es.
- [11] OESTREICHER-SINGER G, SUNDARARAJAN A. Linking network structure to ecommerce demand: theory and evidence from Amazon.com’s Copurchase Network [C] // Com’s Copurchase Network. TPRC, 2006.
- [12] RAMADAN Z B, FARAH M F, KASSAB D. Amazon’s approach to consumers’ usage of the dash button and its effect on purchase decision involvement in the U. S. market [J]. Journal of Retailing and Consumer Services, 2019, 47: 133–139.
- [13] LUO F, WANG J Z, PROMISLOW E. Exploring local community structures in large networks [J]. Web Intelligence and Agent Systems, 2008, 6(4): 387–400.
- [14] HILDRUM K W, YU S L. Focused community discovery [C] // Fifth IEEE International Conference on Data Mining (ICDM’05). IEEE, 2005: 4.
- [15] NEWMAN M. Detecting community structure in networks [J]. European Physical Journal B, 2004, 38(2): 321–330.
- [16] GOMEZ S, JENSEN P, ARENAS A. Analysis of community structure in networks of correlated data [J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2009, 80(1): 16114.
- [17] YAN C, ZHANG Y. Research of community discovery algorithm based on user tags [J]. Science Technology and Engineering, 2011, 11(6): 1237–1240.
- [18] PARTHASARATHY S, RUAN Y, SATULURI V. Community discovery in social networks: applications, methods and emerging trends [M] // Social Network Data Analytics. Springer, Boston, MA, 2011: 79–113.
- [19] CLAUSET A, NEWMAN M E J, MOORE C. Finding community structure in very large networks [J]. Physical Review E, 2005, 70(6): 066111.
- [20] 陈为, 沈则潜, 陶煜波. 数据可视化 [M]. 北京: 电子工业出版社, 2013.
- [21] 孙扬, 蒋远翔, 赵翔. 网络可视化研究综述 [J]. 计算机科学, 2010, 37(2): 12–18.
SUN Yang, JIANG Yuanxiang, ZHAO Xiang. Survey on the research of network visualization [J]. Computer Science, 2010, 37(2): 12–18. (in Chinese)
- [22] 唐卫东, 刘振文, 刘冬生, 等. 基于形态重构的叶片性状特征可视化表达方法 [J]. 农业机械学报, 2019, 50(8): 249–256, 212.
TANG Weidong, LIU Zhenwen, LIU Dongsheng, et al. Visual expression method of leaf traits based on morphological reconstruction [J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(8): 249–256, 212. (in

Chinese)

[23] 任磊,杜一,马帅,等. 大数据可视分析综述[J]. 软件学报, 2014,25(9): 1909 – 1936.
REN Lei, DU Yi, MA Shuai, et al. Visual analytics towards big data[J]. Journal of Software, 2014, 25(9): 1909 – 1936. (in Chinese)

[24] 王戈,于强,YANG Di,等. 基于复杂网络分析法的层级生态网络结构研究[J]. 农业机械学报, 2019, 50(7):258 – 266,312.
WANG Ge, YU Qiang, YANG Di, et al. Hierarchical ecological network structure based on complex network analysis[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(7):258 – 266,312. (in Chinese)

[25] 杨璐,张馨月,郑丽敏. 挖掘数据关系的食品抽检数据可视化分析图研究[J]. 农业机械学报, 2019, 50(6): 272 – 279.
YANG Lu, ZHANG Xinyue, ZHENG Limin. Visual analysis graph research of food sampling data based on mining data relationship[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(6): 272 – 279. (in Chinese)

[26] 黄河燕. 在线社交网络的可视化分析[J]. 中国科学院院刊, 2015(2): 107 – 115.
HUANG Heyan. Visual analysis of online social networks[J]. Bulletin of the Chinese Academy of Sciences, 2015(2): 107 – 115. (in Chinese)

[27] NEWMAN M, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2): 26113.