Analysing Customer Interaction Networks in E-Commerce - 2024 Topic 2 (PG20)

Fatuma Mohamoud
Electronics Engineering and Computer Science
Queen Mary University of London
London, UK
ah20327@qmul.ac.uk

Sharanyak Podder
Electronics Engineering and Computer Science
Queen Mary University of London
London, UK
ec23493@qmul.ac.uk

Zeet Chowdhury
Electronics Engineering and Computer Science
Queen Mary University of London
London, UK
ec23731@qmul.ac.uk

Muskan Dasturia
Electronics Engineering and Computer Science
Queen Mary University of London
London, UK
ec23824@qmul.ac.uk

***Abstract*—In the dynamic landscape of e-commerce, deciphering customer interactions offers a pathway to refine user engagement and drive strategic growth. This study leverages network science to dissect a bipartite network of customer-product relationships, drawn from a UK-based online retailer. By examining network centrality and community patterns, alongside the temporal sway of consumer activities, we reveal how seasonal trends influence shopping behaviors. A focused comparison between PageRank and collaborative filtering illuminates the nuanced role of network topology in shaping recommendation effectiveness. Our insights advocate for a blended approach, harnessing both graph-based and traditional techniques to elevate recommendation systems. This convergence heralds advanced personalisation strategies, fostering a more intuitive and consumer-aligned e-commerce environment.**

***Index Terms*—network analysis, centrality measures, e-commerce, customer interactions, louvain algorithm, pagerank algorithm**

I. Introduction

E-commerce platforms have developed from simple online stores to intricate networks rich in information about user interactions and transactions. These platforms are now thought of as dynamic networks where users connect, choose, and create a rich web of online behavior. With the goal of unraveling the complex patterns of consumer behavior that inform choices and exert influence in the digital marketplace, this study delves into the core of these networks.

A. Context

The shift in worldwide consumer behaviour towards online purchasing has accelerated the expansion of e-commerce while highlighting the necessity for careful examination of these virtual spaces. Understanding the online shopper's journey, which is influenced by a combination of complex algorithms and social interactions, is a challenge for traditional consumer behaviour models. Our study offers a new perspective on digital consumer dynamics by navigating the maze-like world of e-commerce interactions to uncover underlying trends.

B. Research Problem and Motivation

At the core of our exploration is the endeavor to model and analyse e-commerce interactions, with an eye towards predicting behaviors, enhancing user satisfaction, and boosting sales. Our focus narrows on identifying influential network nodes—be they standout products or pivotal consumers—and dissecting the web of interactions that define the e-commerce experience. The drive behind this inquiry is clear: to convert complex data insights into actionable intelligence that can elevate the online shopping experience and offer businesses a leg up in a competitive landscape.

C. Challenges to Address

This research navigates several hurdles, from managing vast e-commerce datasets to distilling actionable insights amidst the digital noise. Employing network science within a traditionally analytics-driven field introduces its own set of challenges, not least of which is safeguarding consumer privacy in the process. These obstacles underscore the pioneering nature of our work and its potential to chart new territory in the understanding of digital marketplaces.

D. Structure of the Paper

Following this introduction, the paper unfolds as follows: Section III delves into our dataset and network construction, shedding light on our methods and the visualisation of the e-commerce network. Section IV presents network statistics, paving the way for a deeper analysis, where we explore network dynamics and key influencers. Section V is results and discussion. We conclude with reflections on our findings and their broader implications, considering future avenues for research in the dynamic landscape of e-commerce in Section VI.

II. Related Work

Exploring the digital marketplace's customer interaction networks touches on diverse fields such as network analysis, digital consumer behavior, and data mining. This section distills significant literature, laying the groundwork for our inquiry into the nuanced interplay within e-commerce platforms.

A. Clarity and Relevance of Literature

Pioneering studies, notably Barabási's exploration of scale-free networks [3], introduce a framework pivotal for understanding complex network structures, hinting that e-commerce interactions likely share these scale-free traits. Additionally, Kumar et al. [4] illuminate the architecture of online shopping networks, revealing consumer behavior patterns that echo established network theories.

B. Depth and Breadth

This field's research spectrum ranges from theoretical models to practical applications in e-commerce, including refining recommendation systems and segmenting customers. Zhou et al. [5] delve into using network analysis to fine-tune product recommendations. Similarly, Oestreicher-Singer and Sundararajan [6] probe the intricacies of recommendation networks, providing valuable context for our study's focus on enhancing e-commerce platforms.

C. Critical Analysis

Despite the solid foundation existing literature offers, there's a noticeable gap in applying these insights to the dynamic and expansive realm of modern e-commerce. Previous research often remains anchored in static network models or smaller datasets, bypassing the fluidity of consumer interactions and the evolving digital marketplace landscape. Our investigation seeks to bridge this gap, leveraging current interaction data from a prominent online retailer to capture the real-time dynamics of network behavior.

D. Relation to the Problem and Present Work

Leveraging this foundational work, our research dives into the dynamic world of e-commerce. With an eye on real-time data, we aim to shed light on the evolving nature of customer interaction networks, striving to reveal consumer behavior insights that static analyses have overlooked. This novel approach promises not only to deepen our understanding of digital marketplaces but also to equip online retailers with the knowledge to enhance customer engagement and tailor personalisation strategies more effectively.

III. Dataset and Network Presentation

A. Description of the Data and Its Collection

Our study utilises the "Online Retail II" [1] dataset , a comprehensive collection of transactions from a UK-based online retail platform. This dataset encompasses detailed records of customer purchases over the period from 1 December 2009, to 9 December 2011. It includes crucial information such as invoice numbers, stock codes, descriptions, quantities purchased, invoice dates, prices, customer IDs, and country of purchase.

B. Link with Research Problem

The chosen dataset is inherently linked to our research problem, which aims to dissect and understand the complex network of interactions within an e-commerce platform. By analysing these transaction records, we seek to identify patterns of customer behaviour, influential products, and the overall structure of the consumer-product network, directly addressing our research objectives.

C. Description of Networks and Their Construction

To analyse the dataset, we constructed a bipartite network, where one set of nodes represents customers and the other set represents products. Edges in this network are drawn from customers to products, with each edge signifying a purchase.

D. Network Visualisations

We have visualised the dataset, and the top 1000 nodes are shown in the figure.

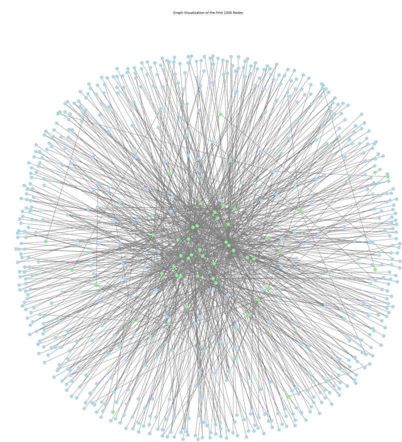

Figure 1. Network visualisation highlighting customer-product interactions.

E. Network Statistics (Task 1)

Here is a comprehensive overview of the network's statistics which is given in Task 1:

- The **node degree** is the number of neighbours of a node. In the given cyclic bipartite graph, every node (C1, C2, C3, C4, C5, P1, P2, P3, P4, P5) has a degree of 2.
- The **clustering coefficients** for each node are calculated by finding the ratio of actual edges amongst neighbours to the total possible edges. Because it is a cyclic graph, this results in a clustering coefficient of 0 when no direct connections exist amongst associated products.
- **Modularity** allows us to decide if a particular community partition is better than some other one. The range of modularity is between 0.34 to 0.38, which suggests a moderate level of community structure within the network.
- The **betweenness centrality** captures how much a given node is in-between others. Each node in the network had

a betweenness centrality of 0.2222. This value indicates that all nodes are equally central within the network in terms of the number of shortest paths that pass through them.

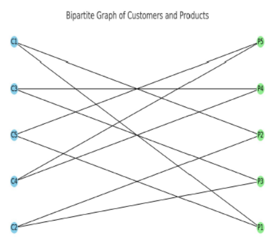

Figure 2. Bipartite graph

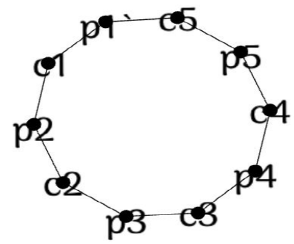

Figure 3. Customers and Product nodes

IV. Network Analysis Methodology

Our network analysis methodology is separated in task 2, 3 and 4. Each task has a detailed analysis with it.

A. Task 2:

1) Objectives: The primary objective of Task 2 is to construct a bipartite network from the given dataset, representing interactions between customers and products. The task focuses on detecting communities that signify groups of products frequently purchased together and groups of customers with similar buying patterns.

2) Methodology: The methodology to achieve the objectives involves several steps, we have implemented it through Python programming with the use of libraries such as *Pandas*, *NetworkX* [12], *Matplotlib*, and the community detection library `community_louvain` [11]. The steps are as follows:

1) **Data Preparation**: The dataset is loaded into a Pandas DataFrame, ensuring the 'Target' and 'Source' columns are present. This dataset forms the basis for constructing the bipartite network. From the big dataset, we have filtered out only the positive values and only the country `UK`. From now we will use this modified dataset only.
2) **Bipartite Network Construction**: A bipartite graph is initialized using *NetworkX*, where nodes represent customers and products, and edges denote the purchasing interactions between them.
3) **Community Detection**: The `community_louvain` method is applied to the product projection of the bipartite graph to identify communities within the network. This projection focuses on product nodes, with edges representing shared customer interactions, hence facilitating the detection of product groups commonly purchased together.
4) **Community Analysis**:The analysis includes evaluating the size of each community and identifying top products within each community based on purchasing frequency.

3) Justification: The rationale behind the approach selection centres on its ability to reveal patterns and structures inside intricate networks that are not easily discernible using traditional analysis techniques. We find that the bipartite network model captures the dual structure of our dataset well, treating both consumers and products as related entities. Through the use of community detection algorithms—more especially, the Louvain method—we are able to partition the network into groups that exhibit a high degree of internal consistency with respect to consumer behaviour. Targeted analysis of customer preferences and product attractiveness within different communities is made possible by this segmentation.

These findings support the effectiveness of network analysis in revealing latent patterns in data and highlight the significance of community detection in granularly comprehending consumer behaviour.

B. Task 3:

1) Objectives: Task 3 focuses on exploring the dynamic aspects of the customer-product interaction network. The objectives are threefold:

- Perform a Temporal Analysis to track changes in customer-product interactions over time, with a particular focus on identifying the impact of promotional events or seasonal trends.
- Conduct an Influence Analysis to identify key nodes (customers or products) based on their centrality and assess their impact on the network's structure and community dynamics.
- Execute Network Change Modelling to observe and interpret changes in the network's properties, such as centrality measures and community structures, over time.

2) Methodology: The methodology applied in Task 3 incorporates data manipulation, network construction, and complex network analysis techniques, utilising Python libraries such as `pandas`, `networkx`, and `matplotlib` for visualisation. This method came from studies in social network analysis [7] and bipartitte network studies [8]. The approach is detailed as follows:

- **Temporal Analysis:** Customer-product interactions were visualised over time using transaction data. Monthly networks were constructed to analyse variations in transactions, highlighting significant trends during promotional events and seasonal periods.
- **Influence Analysis:** Degree centrality was calculated for every node to identify the most influential customers and products. The top five nodes by degree centrality were analysed to gauge their impact on the network.
- **Network Change Modelling:** Changes in network properties were tracked over time. This included plotting average degree centrality monthly and observing the evolution of community numbers to discuss the implications on customer behavior.

3) Justification: Temporal analysis revealed how interactions are influenced by seasonal trends and promotional events, providing a basis for strategic planning. Influence analysis pinpointed key actors within the network, offering a potential focus for targeted marketing strategies. Network change modelling offered a macro view of how the network evolves, reflecting shifts in customer purchasing behavior and product popularity. The combination of these analyses provides a comprehensive understanding of the network's dynamics, which is crucial for informed decision-making in marketing and product management.

C. Task 4:

1) Objectives: Task 4 aims to enhance the understanding of the network structure through the lens of graph-based recommendation algorithms. The objectives include:

- Implementing a graph-based recommendation process using the PageRank algorithm to suggest products.
- Comparing the performance of the PageRank algorithm against traditional recommendation systems, such as collaborative filtering.
- Discussing the influence of the customer-product network structure on the recommendation systems and providing insights for optimising graph-based recommendations.

2) Methodology: This methodology combined a graph theory application, which employs [9] and [10]. The methodology employed to address these objectives comprises the following components:

1) **Graph-Based Recommendation Implementation**: Utilising the PageRank algorithm modified to factor in previous customer transactions, enabling personalised product recommendations.
2) **Collaborative Filtering Implementation**: Creating a user-item interaction matrix and employing cosine similarity to identify similar users and recommend products based on their preferences.
3) **Performance Comparison**: Evaluating the effectiveness of both recommendation methods using the same customer ID for comparison and analysing the results.

3) Justification: The PageRank algorithm, renowned for its ability to rank items based on their importance within a network, provides a unique perspective by considering the interconnectedness of customers and products. On the other hand, collaborative filtering offers a traditional yet robust approach to understanding user preferences through similarity metrics. By comparing these methodologies, we aim to uncover the strengths and limitations of each approach within the context of our specific customer-product network.

V. Results and Discussion

A. Task 2:

1) Results: The community detection algorithm identified four distinct communities within the network, with sizes varying significantly. Community 0, being the largest, and Community 3, being the smallest.

2) Illustrations:

Table I
COMMUNITY SIZES

Community	Size
0	1581
2	1157
1	649
3	476

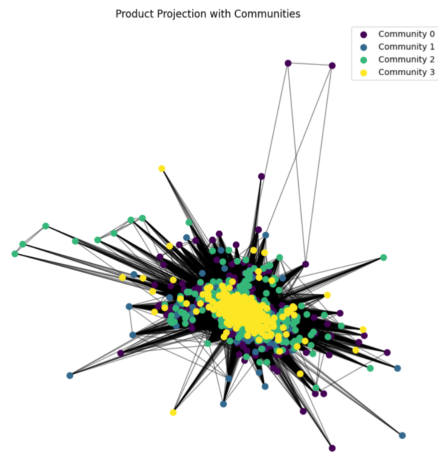

Figure 4. Communities within the Bipartite network

3) Discussion & Limitations: The results underline the network's complex structure, where both broad appeal and specialised interests coexist. This diversity in communities and their top products provides valuable insights into market segmentation and customer preferences. It also suggests that targeted marketing strategies could be developed to cater to the distinct needs of each community. Top product in community 0 with 2240 purchases is **WOOD S/3 CABINET ANT WHITE FINISH**.
One limitation of this analysis is its static nature, which does not account for temporal dynamics in customer-product interactions. Additionally, the method's sensitivity to the network's density and the chosen algorithm parameters could influence the identified communities and their interpretation.

Table II
TOP PRODUCTS IN EACH COMMUNITY

Community 0 (Size: 1581)		**Community 2 (Size: 1157)**		**Community 1 (Size: 649)**		**Community 3 (Size: 476)**	
Source	Count	Source	Count	Source	Count	Source	Count
17850	2240	17841	4506	16549	2269	14606	3557
13089	1500	12748	2548	15311	1742	14081	649
14527	1373	13081	1255	16782	1643	16072	533
17448	1117	17589	1090	15039	1241	14535	515
15768	986	17920	922	15005	1210	16550	493

B. Task 3:

1) Results:

- **Temporal Analysis:** Figure 5 The amount of transactions appears to be quite stable until the summer. The minor rise in purchases around July can be attributed to mid-year promotions, such as a summer sale. After August, there is a significant increase in sales during the rest of the year, which may be attributed to holidays such as Halloween in October and the Christmas season in December, as well as promotional events such as Black Friday. If the sharp drop at the end of the year is not caused by data cutoffs or collection difficulties, it might be a natural fall following the Christmas rush.

- **Influence Analysis:** Degree centrality measures how well a node is related to other nodes. A node's high degree centrality may indicate that it is a network hub. Customers 17841, 12748, 14606, and 22423 are central in this network, indicating that they shop here regularly or have a wide range of interests. This gives it influence on the remainder of the network. Product **85123A(WHITE HANGING HEART T-LIGHT HOLDER)** has the highest degree centrality, indicating that it is the most popular item or top seller. Knowing the most popular products may help the business maximise profits by ensuring they are constantly available for sale.
- **Network Change Modelling:** The increase in average degree centrality Figure 7 may be attributed to holiday seasons and promotional events such as Christmas, Halloween, and Black Friday, however the significant decline towards the end of the year can be attributed to a decrease in the number of people purchasing goods. The growth in communities Figure 6 can be attributed to people diversifying their purchases, as observed throughout the summer and Christmas seasons. Perhaps the rise is due to the introduction of new items into the firm, which would result in the formation of new communities. The declines might be explained by small communities banding together, possibly because customer purchases become more similar. Or perhaps since the items have been withdrawn, the communities within which they exist are progressively disappearing.

2) Illustrations:

- **Temporal Analysis:**

Figure 5. customer product interactions over time.

- **Influence Analysis:** Table III
- **Network Change Modelling:** Figure 6 and Figure 7

3) Discussion & Limitations: The fluctuation in transactions over time provides critical insights into how external events influence customer purchasing behavior. The role of influential nodes in shaping network dynamics underscores the

Table III
TOP 5 INFLUENTIAL NODES BY DEGREE CENTRALITY

Node	Degree Centrality
17841	0.18389
12748	0.16927
'85123A'	0.13578
14606	0.13526
'22423'	0.09854

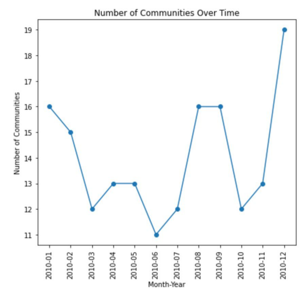

Figure 6. No. of communities over time.

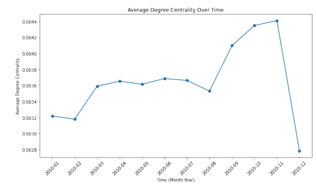

Figure 7. Average degree centrality over time.

importance of key customers and products in driving sales and network engagement.

The analysis might not fully capture the causal relationships between external events and network dynamics. Additionally, the degree centrality measure, while useful, may oversimplify the complexity of influence within the network.

C. Task 4:

1) Results: To compare performance, we utilised the same customer ID (13821). For the pagerank algorithm, I compared prior customer transactions to those recommended and discovered that two out of five had been previously purchased by that consumer. As a result, this algorithm takes into account both prior customer purchases and those that the consumer may be interested in. On the other hand, the collaborative filtering suggestions included only goods that the consumer had never purchased before. By using user similarities, our algorithm is extremely good at discovering new things for customers. Many of the recommended goods began with "JUMBO BAG". Perhaps this is because the buyer had previously purchased other jumbo bags, and the algorithm suggested further varieties of this sort of bag.

2) Illustration:

- **Graph-Based Recommendation System**
  - Utilising algorithms like PageRank to determine the influence or popularity of items based on a network of interactions. There is a column "Score" i.e. pagerank score.
- **Collaborative Filtering System**
  - Traditional recommendation systems like collaborative filtering generate suggestions based on the past behavior of users.

Table IV
TOP 5 RECOMMENDED PRODUCTS BY PAGERANK

Code	Description	Score
85099B	JUMBO BAG RED WHITE SPOTTY	0.003298
21931	JUMBO STORAGE BAG SUKI	0.002642
84671B	CROCHET BEAR WITH BLUE STRIPES	0.002620
85123A	WHITE HANGING HEART T-LIGHT HOLDER	0.002591
48194	DOORMAT HEARTS	0.002393

Table V
COLLABORATIVE FILTERING RECOMMENDED PRODUCTS

Code	Description
20711	JUMBO BAG TOYS
22386	JUMBO BAG PINK WITH WHITE SPOTS
85099C	JUMBO BAG BAROQUE BLACK WHITE
21930	JUMBO STORAGE BAG SKULLS
48185	DOOR MAT FAIRY CAKE

3) Discussion & Limitations: The network topology has a significant impact on the page rank method. Influential products and things that were vital to the customer's prior transaction are highlighted. This approach is excellent at determining this customer's preferences but not so good at proposing new items. Whereas collaborative filtering recommends new goods based on consumer commonalities. This is excellent at proposing products similar to prior purchases; but, it falls short when it comes to suggesting items that are not as comparable to previous purchases.

To improve graph-based recommendation systems, we may adjust the personalisation vector or the weights based on how recent the purchase was. The more recent the purchase, the heavier the weight. This update may make the recommendations more relevant to shifting preferences.

VI. CONCLUSIONS AND PERSPECTIVES

A. Summary of Problem and Methodology

This study investigated the intricate network dynamics of e-commerce platforms, utilizing advanced network analysis to understand the complex interactions between consumers and products. Employing the ”Online Retail II” dataset, we constructed a bipartite network to dissect these interactions systematically. Our methodology involved detailed data preparation, network construction using tools like NetworkX, and community detection through algorithms like Louvain to map out the transactional relationships within the network.

B. Summary of Main Results and Insights

Our findings shed light on several key aspects:

- **Community Dynamics:** We identified distinct communities within the network, highlighting clusters of products frequently purchased together and groups of customers with similar purchasing patterns.
- **Influential Nodes:** The analysis pinpointed key nodes that significantly influence the network's structure, providing insights into which products or consumers hold pivotal roles within the marketplace.
- **Temporal Dynamics:** Temporal analysis revealed how consumer interactions fluctuate with seasonal trends and promotional events, offering strategic insights for timing marketing efforts.

C. Perspectives on Future Work

Looking forward, the study opens several avenues for deeper exploration and application:

- **Real-Time Data Analysis:** Integrating real-time transaction data could enhance the responsiveness of network models, allowing for more dynamic and timely insights into consumer behavior.
- **Cross-Platform Analysis:** Expanding the analysis to include multiple e-commerce platforms could provide a broader view of consumer behavior across different digital ecosystems.
- **Algorithm Enhancement:** Further refining the recommendation algorithms used in the study could improve their accuracy and relevance, particularly by incorporating machine learning techniques to better adapt to consumer preferences over time.

REFERENCES

[1] “Online Retail II Data Set,” UCI Machine Learning Repository, 2010. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Online+Retail+II

[2] A.-L. Barabási, *Network Science*, Cambridge University Press, 2016. Available at: http://networksciencebook.com/chapter/1

[3] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509-512, 1999.

[4] R. Kumar, J. Novak, and A. Tomkins, “Structure and evolution of online social networks,” in *Link Mining: Models, Algorithms, and Applications*, pp. 337-357, Springer, 2010.

[5] Y. Zhou, D. Wilkinson, R. Schreiber, and R. Pan, “Solving the apparent diversity-accuracy dilemma of recommender systems,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 10, pp. 4511-4515, 2010.

[6] G. Oestreicher-Singer and A. Sundararajan, “Recommendation networks and the long tail of electronic commerce,” *MIS Quarterly*, vol. 36, no. 1, pp. 65-83, 2012.

[7] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, 1994. Available at: https://psycnet.apa.org/record/1995-97740-000

[8] M. Latapy, C. Magnien, and N. Del Vecchio, “Basic notions for the analysis of large two-mode networks,” *Social Networks*, vol. 30, no. 1, pp. 31-48, 2008.

[9] L. Page, S. Brin, R. Motwani, and T. Winograd, “The PageRank Citation Ranking: Bringing Order to the Web,” presented at The Web Conference, 1999. Available at: https://api.semanticscholar.org/CorpusID:1508503

[10] J. L. Herlocker, J. A. Konstan, and J. Riedl, “Explaining collaborative filtering recommendations,” in *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pp. 241-250, 2000.

[11] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.

[12] A. A. Hagberg, D. A. Schult, and P. J. Swart, “Exploring network structure, dynamics, and function using NetworkX,” in *Proceedings of the 7th Python in Science Conference (SciPy2008)*, Gäel Varoquaux, Travis Vaught, and Jarrod Millman, Eds., Pasadena, CA, USA, pp. 11-15, Aug 2008.