Journal of Engg. Research Vol. 8 No. (4) December 2020 pp. 185-205 https://doi.org/10.36909/jer.v8i4.8369

Opinion-Based Co-Occurrence Network for Identifying the Most Influential Product Features

Ashok Kumar J* and Abirami S

Department of Information Science and Technology, Anna University, CEG Campus, Chennai, 600028, India
**Corresponding Author: jashokkumar83@auist.net*

Submitted: 27/06/2019
Revised: 28/12/2019
Accepted: 06/01/2020

ABSTRACT

Nowadays, social networking sites such as Facebook, Twitter, LinkedIn, YouTube, and other e-commerce websites produce a large number of text reviews. These text reviews mostly describe the product features and their opinions, which are the most important to the product developers, launchers, or buyers for business development and decision-making processes. Therefore, we present an opinion-based co-occurrence network for product reviews. The main aim of this research is to identify the popularity of product features or popular terms, the number of connections of a term, the strong relationship between terms, grouping the product terms, and the sentiment polarity links between terms in both positive sentiment and negative sentiment. Also, we employed the Harel-Koren fast multiscale layout algorithm and CNM (Clauset-Newman-Moore) algorithm for visualizing and grouping the network. We then measured the overall graph metrics and vertex metrics to characterize the network. Additionally, the experimental result shows the ranked product features and their social strength between product features and sentiments.

Keywords: Social network analysis; Opinion mining; Co-occurrence network; Information network; Grouping product features; Network visualization.

INTRODUCTION

Opinion mining computes the reviewers (or authors) moods, attitudes, and emotions, which are expressed in social media contents such as posts, reviews, comments, and tweets. In particular, it computes two types of information, namely, the objective information and subjective information. The objective information is factual information, and the subjective information reflects the author's personal feelings on products, services, events, or organizations. Opinion mining is performed on documents, sentences, or aspects (Tubishat et al., 2018). It is widely used in the application of business and government intelligence, financial markets, political science, product manufacturing and service, and decision support systems (Qiu et al., 2018). The products features extracted from the comments or reviews in social media are often used to rate the products. These reviews may help other customers decide which products to buy (Yang et al., 2019 & Yang et al., 2018). More specifically, the aspect based sentiment analysis is focused on product reviews. An aspect (also called a feature) can be an entity, attribute, or event in the reviews. However, the sentiment polarity can be determined for each review using sentiment dictionaries, sentiment analysis tools, or human annotations. Then, it can be categorized into different classes, such as positive sentiment, negative sentiment, and neutral sentiment. For instance, the review "*the photo quality is amazing, and I know I am going to have fun with all the features. However, the only legitimate criticism from the online reviews is that fact that the lens does obstruct the viewfinder*" discusses two entities, namely, photo quality and viewfinder. The review expresses a positive opinion about photo quality and negative opinion about the viewfinder. In natural language processing, most of the existing works focused on sentiment classification and clustering tasks. Sentiment classification generally classifies the reviews based on the derived sentiment features (Kauer et al., 2016).

Sentiment clustering approach categorizes the documents into groups, where it does not indicate positive or negative (Cheng et al., 2017). Even though these methods are most popular in the scientific community, the product users, buyers, developers, or launchers want to understand the most prominent product features in positive sentiment and negative sentiment groups. To overcome the problem of identifying product features in both positive and negative sentiment groups, we apply social network analysis (SNA) method. A social network describes the social structure between actors, groups, organizations, villages, communities, computers, regions, information entities, and so on (Chang, 2018). It indicates the connection between various social relationships or interactions. SNA investigates or characterizes the social structure pattern (actors, nodes, people, or things) distribution and relational ties (edges, lines, or links). For instance, the social structures like transitional network (Wissink et al., 2018), traffic network (Hughes et al., 2017), risk network (Ongkowijoyo et al., 2018), weighted network (Zhou et al., 2017), reconciliation network (Zhang et al., 2017), and co-occurrence network (Li et al., 2016; Radhakrishnan et al., 2017) are visualized through a sociogram. It is the graphical representation of nodes and lines. Sometimes the graphical representation can be sufficient for the decision-making process. Therefore, SNA has gained an important role among practitioners and researchers over more than two decades in network applications (Li et al., 2016; Abdelsadek et al., 2018). In the era of Industry 4.0 and Web 3.0, the effectiveness of social network is studied for business development using social networking sites and customer-related services (Chang, 2018).

In this paper, we propose an opinion-based co-occurrence network for product reviews to identify the prominent features. Co-occurrence network method analyzes and characterizes the strength of social relationships between product features and their popularities. The networks are visualized and grouped using the Harel-Koren fast multiscale layout algorithm and CNM (Clauset-Newman-Moore) algorithm, respectively.

The rest of this paper is organized as follows. Section 2 describes the related works in the field of social network analysis. Section 3 presents the objective of the research work. Section 4 presents the sentiment based co-occurrence network analysis for product reviews. In Section 5, we report and discuss the experimental results using social network analysis metrics. Finally, conclusions and future research directions are presented in Section 6.

RELATED WORKS

In this section, we briefly review the existing approaches in the field of social network analysis (SNA). The SNA methods are widely used in different applications such as topic identification (Qiu et al., 2018), ranking keywords (Zhao et al., 2018), healthcare, big data (Chang, 2018), disaster management (Kim et al., 2018), information propagation network (Li et al., 2018), ranking community (Roy et al., 2018), telecommunication services (Farasat et al., 2016), emergency services, collaborative learning experiences (Claros et al., 2016), automobile insurance, recommendation systems (Razghandi et al., 2017), resume verification process (Shin et al., 2017), risk analysis (Colladon et al., 2017; Ongkowijoyo et al., 2018), project selection, citations (Amplayo et al., 2018), gender ratios (Hayat et al., 2017), migration and policy implementations (Wissink et al., 2018), the product management and diversification (Hughes et al., 2018), and sentiment analysis (Kauer et al., 2016; Tubishat et al., 2018). In particular, a co-word analysis is one of the most popular methods in the domain of information science (Zhao et al., 2018). It determines the relationship between words and significant topics in the subject. Online social networking sites produce a large amount of data in the form of long and short texts (reviews, comments, posts, and tweets). This textual information can be used to identify hidden messages by constructing co-occurrence networks for business intelligence. On Twitter, a user's interest is identified explicitly or implicitly in a given time interval using a graph-based link model.

This model is investigated to infer users' implicit interests based on the influence of different factors such as explicit user contributions, connections, and the relatedness on topics (Zarrinkalam et al., 2018). Qiu et al. (2018) proposed a dynamic social network topic model for clustering users in short texts. This model incorporates dynamic topic distributions and user social relations. They also employed a new collapsed Gibbs sampling algorithm to deal with the sparsity of short text. Further, the domain knowledge was explored quantitatively for iMetrics or information

metrics research using co-word analysis (Khasseh et al., 2018). The authors' results reveal the higher frequency of keyword categories in Web of Science and the updated intellectual structure of iMetrics for papers published in Scientometrics, Journal of Informetrics, and another six journals. Also, the co-word analysis is used with cluster analysis and social network analysis methods to examine the research patterns and trends of the recommendation system in China from 2004 to 2013 (Hu et al., 2015). The keywords with high correlations are grouped into the same cluster. Then, the behavior of the word co-occurrence network was measured using SNA metrics. Kulig et al. (2017) addressed the punctuation marks and words in text corpora using two approaches, namely, Zipf rank-frequency distributions and word-adjacency networks. The authors treated the punctuation marks like words and compared the result of these punctuation marks with ordinary words. This study indicates that punctuation marks can be included effectively in linguistic studies. Moreover, Garg et al. (2017) identified key-phrases using word co-occurrence networks. These key-phrases are ranked in three categories (headline, detailed description, and candidate keywords) using an Analytic Hierarchy Process (AHP). The AHP is an evaluation measure, which calculates the edge strength, strength of difference, phrase degree, and degree computation of the word co-occurrence network.

In Amplayo et al. (2018), the authors present a network-based approach to extract features in scholarly literature.

This approach introduces two graphs, namely, macro-level graph and micro-level graph. The macro-level graph represents the authors and documents as nodes, and the micro-level graph represents the words, keywords, and topics as nodes. Then, the authors used an autoencoder neural network to detect novelty of a research paper. Their results reveal that the keyword-level graph features perform the best using regression analysis. Forss et al. (2016) defined an algorithm to rank companies from news to company networks based on sentiment polarization. The sentiment network identifies the companies' influence and their relations in the news. In Cheng et al. (2017), the authors propose a SignedSenti framework to group text posts into k different clusters. The method incorporates the signed social relations and sentiment signals, which includes posts-textual terms relation, user-text relation, user-user relation, text-item relation, text-sentiment cluster matrix, and term-sentiment matrix. They showed that the proposed SignedSenti significantly outperforms the baseline methods. Radhakrishnan et al. (2017) proposed a keyword co-occurrence network (KCN) to undertake systematic reviews of the scholarly literature. This approach identifies the co-occurrence patterns and most frequently occurring terms in nano-related fields. It also reduces the time and effort for a systematic review.

Furthermore, the researchers have constructed the co-occurrence based scientific knowledge map to characterize and visualize the current research topics in the domain of corporate network public opinion (Li et al., 2018). Li et al. (2016) integrated the annual articles co-keyword networks and keywords co-occurrence networks to analyze the topological features on a given topic, complex networks. The authors also measured the innovation coefficient of the networks. Yang et al. (2016) introduced the author keyword coupling analysis method to analyze the intellectual structure of a field of information science. This method was compared with the authors' bibliographic coupling analysis method.

The relations are represented by the same references in the keyword coupling analysis and the same keywords in the bibliographic coupling analysis. They showed that the proposed method provides a complete visualization of a disciplined structure in both first- and all-author counting. Li et al. (2017) proposed a utility-based approach to understand the formation of co-author networks by considering the latent meeting sequence. The authors also developed a double Markov Chain Monte Carlo (MCMC) algorithm to estimate the preference parameters. Their results showed that the proposed approach outperforms in predicting the co-author network formation. The work of Tubishat et al. (2018) is the latest source to describe the aspect extraction methods in both supervised and unsupervised learning using the co-occurrence relation. These methods extract implicit features based on the aspects and their opinion words by calculating the Pointwise Mutual Information (PMI). The existing approach in information science extracts the implicit features in association with a set of aspects and a set of sentiment words and determines the relationship between words and topics. However, there is no detailed study in social network analysis method to

address the sentiment-based individual product features and their influences in text corpora. Therefore, we introduce an opinion-based co-occurrence network for product reviews to identify the most prominent product features.

OBJECTIVES

Text reviews describe opinion about the products and their features. The features influence the buyers, product developers, or launchers in the market. Therefore, a product can be decided based on features. The main objective of this work is to identify the popularity of product features in both positive sentiment and negative sentiment reviews. Through the co-occurrence network, we intend to visualize, group, and measure the product features and their sentiment groups in text reviews. We also intend to rank the product features and their opinions based on network analysis metrics. A unique feature list is prepared by combining all product features. This feature list is used to extract only the useful features in product reviews.

Then, the features are ranked and compared with other products. Moreover, it helps identify the reason for the sale of a product, defect of a product, or movement of a product in the market.

THE PROPOSED METHOD

Context

This study examines the key term-based sentiment analysis for product reviews using social network analysis. Sentiment analysis is an important area to understand people's opinion about the products, elections, services, organizations, terror attacks, government policies, and so on. In social network analysis, the product reviews can be analyzed to understand the most influential or popular term, the number of connections of a term, the strong relationship between terms, grouping the product terms, and the sentiment polarity links between terms. The results reported in this paper are part of a large number of product domains (De Brún et al., 2018). Therefore, the data highlights the importance of SNA in product domains, which explains the practical information from the data analysis. Further, this study encourages the product developers or launchers to standardize and improve the quality of the product and motivates the users or buyers for the decision-making process. The general architecture of the proposed method is shown in Figure 1.

Data set

In this paper, we used a feature specific sentiment analysis dataset (Mukherjee et al., 2012). The dataset consists of three parts, namely, the dataset 1, dataset 2, and dataset 3. Dataset 1 contains 1257 annotated reviews with a feature specific sentiment polarity (positive or negative) for different domains such as camera, antivirus, iPod, music player, and mobile. Similarly, dataset 2 and dataset 3 contain 3834 and 425 annotated reviews, respectively. The data are represented as triplet <*F, SP, R*>, i.e., a feature, sentiment polarity, and text review.

In particular, we used three camera domain reviews in dataset 1 (c_canon, c_canon2, and c_canon3), which are used in the data analysis using social network analysis. In addition, we created a new data set by combining these three product reviews. The product datasets c_canon, c_canon2, c_canon3, and combined data contain 30, 17, 31, and 78 feature specific reviews, respectively. Then, we split the sentences, which contains "but", "yet", and "although", for feature-based analysis. For instance, the review "*This is a great camera for you! But this camera has a major design flaw*" is split into "*This is a great camera for you!*" and "*This camera has a major design flaw*". Finally, the dataset contains 75, 45, 68, and 188 reviews, respectively.

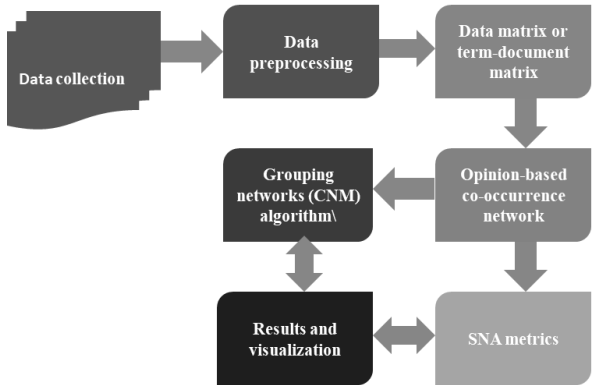


Figure 1. Opinion-based product information network.

Data preprocessing

Data preprocessing is an important task to prepare the quality of data for analysis. It reduces the redundant and irrelevant information present in the data (Ingo Feinerer, 2017; Ovidiu Serban et al., 2018). The principal focus of this research work is to identify the product features using co-occurrence network. Therefore, we apply various steps to prepare the quality of data. The preprocessing steps include converting text to lower case, removing numbers, removing stop words, removing punctuations, and lemmatization (Alam et al., 2016).

Convert text to lower case used to avoid the distinction between same words in the entire reviews like "camera" and "Camera". The numbers are not contributing a lot to this data analysis, and hence, it is removed. In English, the commonly used stop words like "a", "an", "to", and "the" are removed to provide results that are more accurate. Further, punctuations like apostrophes (‘), question marks (?), dots (.), exclamation marks (!), colons (:), dashes (-), commas (,), and other typography marks are removed to get the entire reviews with no punctuations. Lemmatization is applied to convert a word into its root form or common form using a WordNet tag. For example, the words "pictures" and "cameras" are tagged as plural nouns (NNS) using part-of-speech (PoS) tag and then lemmatized into the WordNet dictionary form "picture" and "camera".

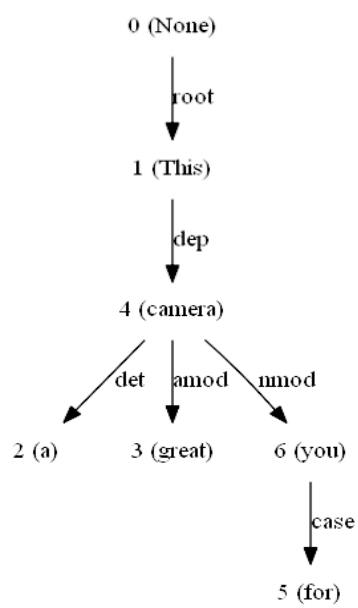


Figure 2. The dependency tree with Stanford parser.

Product features and opinion identification

In this section, the challenging task is to identify product features and their opinions in a sentence. Stanford dependencies are used to extract relations between words (De Marneffe et al., 2008). This dependency is designed in the form of triplets <relation, governor, dependent>. Moreover, Stanford dependencies provide more than 50 relations to deal with English sentences.

For the sentence, "*This is a great camera for you*", the dependency tree is designed as shown in Figure 2. In particular, the triplets < *('camera', 'NN'), 'amod', ('great', 'JJ')* > were extracted to identify product features and sentiments. Then, MeaningCloud API was used to identify sentiments for the extracted dependency relations (http://www.meaningcloud.com). The product features were identified into five categories: strong positive (P+), positive (P), neutral (NEU), negative (N), strong negative (N+), and none. For the above example, the product "camera" is identified as a strong positive. In this paper, we ignored the neutral and none categories for co-occurrence network analysis.

Construction of data matrix

The preprocessed data were converted into a term-document matrix, where it describes the frequency of terms that occur in the document. Let $A = m \times n$ be a two-dimensional matrix, where m rows are the terms and n rows are the documents. The individual elements of a matrix (*A)* are denoted by $a_{i,j}$, where max i = m, max j = n (Richard G. Brereton, 2017). For example, the extracted product features and sentiments *D1 = <Camera, strongPostive>, D2 = <Quality, negative>, and D3 = <viewfinder, positive>* are represented as a term-document matrix, and then they are converted into a Boolean matrix as shown in Table 1. In the Boolean table, if an element occurs greater than or equal to one, then it is assigned as one, otherwise, zero. Further, a word-word co-occurrence matrix was constructed for every pair of word features (Chen et al., 2018; Cheng et al., 2017; Ingo Feinerer, 2017; Li et al., 2016). The rows and columns in this matrix represent word features, and the values represent the frequent occurrence of two words (called concurrence or coincidence) from text documents in a certain order. For this representation, the Boolean-based term-document matrix (*A*) is multiplied by the transpose of the Boolean-based term-document matrix (A^T), where the frequency of a single word is not encoded in the diagonal.

Table 1. Data matrix.

Terms	Term-Document Matrix			Boolean Matrix		
	D1	D2	D3	D1	D2	D3
camera	1	0	0	1	0	0
strong positive	1	0	0	1	0	0
negative	1	1	0	1	1	0
quality	0	1	0	0	1	0
positive	0	0	1	0	0	1
viewfinder	0	0	1	0	0	1

Table 2. Co-occurrence matrix.

	camera	Strong positive	negative	quality	positive	viewfinder
camera	0	2	0	0	0	0
strong positive	2	0	0	0	0	0
negative	0	0	0	1	0	0
quality	0	0	1	0	0	0
positive	0	0	0	0	0	1
viewfinder	0	0	0	0	1	0

Construction of co-occurrence network

A co-occurrence network visualizes a potential relationship between nodes, entities, concepts, people, or organizations within written documents. The co-occurrence network plays a vital role in the field of information retrieval, text mining, sentiment analysis, and biomedical for knowledge discovery. The co-occurrence network can be constructed based on the interconnection of terms and their presence in the text documents (Li et al., 2016; Tang et al., 2015; Cheng et al., 2017; Hirsch et al., 2016; Forss et al., 2016; Li et al., 2018). The number of occurrences of the two terms represents the links between terms. A term without links to other terms represents no occurrence. For example, the terms "quality" and "negative" occur only once in the document (Table 2).

Therefore, a pair of co-occurring terms is called neighbors, and they can be grouped into neighborhoods based on their relations. The sentiment polarity for each review can be distinguished based on their occurrences. For example, the sentiment polarity labels "strong positive", "positive", "negative", and "strong negative" are associated with neighbors based on their co-occurrences. The reviews in the feature-specific dataset mostly discuss product features.

However, the co-occurrence social network is drawn to the preprocessed data matrix using the Harel-Koren fast multiscale layout algorithm. This algorithm elaborates two parts: multiscale graph representation and locally nice layout (Algorithm. 1). The multiscale graph *G(N, L)* is represented as a sequence of graphs based on locally preserving k-clustering problem concerning the radius. In this problem, the actors (*N*) are partitioned into k clusters for minimizing the longest distance between two actors. The local nice layout part constructs a new representation of the graph by considering every pair of actors (Koren et al., 2002). In this section, an undirected co-occurrence social network is constructed for the products c_canon, c_canon2, c_canon3, and combined data shown in Figure 3. These networks consist of 43 actors (terms) and 44 lines (links), 27 actors and 28 lines, 28 actors and 29 lines, and 72 actors and 88 lines. The green color represents a link with positive, the lime color represents a link with strong positive, the red color represents a link with negative, and the blue color represents a link with the strong negative.

Grouping networks

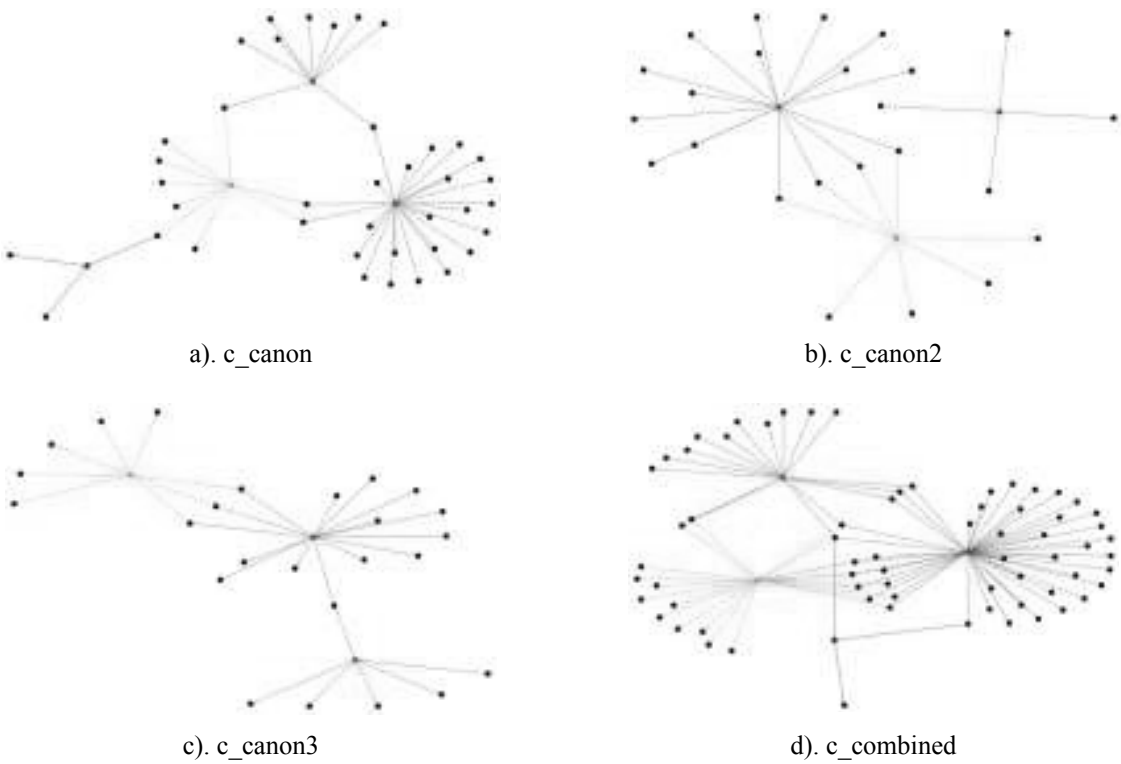
Grouping (also called clustering) is a collection of individual actors with dense link patterns internally and sparse link patterns externally (Bello-Orgaz et al., 2016; Mishra et al., 2007; Qiu et al., 2017;). Newman and Girvan (NG) presented a clustering algorithm initially using modularity that measures the quality of grouping network.

Later, the CNM (Clauset-Newman-Moore) algorithm proposed to find groups (or to analyze various subgroups in the co-occurrence network (Cristian Zanoci and Jim Andress, 2016; Wakita et al., 2007).

Algorithm 1: A fast multiscale layout drawing algorithm

Input: A graph *G* with *N* actors and *L* lines
Output: Visualization of a multiscale layout L

1 Determine: Rad, Iterations, Ratio, and Minsize
2 Compute the shortest path for all-pairs ($d_{v\times v}$)
3 Set up a random layout (L)
4 $k \leftarrow MinSize$
5 **While** $k \leq |N|$ **do**
6 $centers \leftarrow \boldsymbol{K-Centers}(G(N,L),k)$ # % Goal: $S \subseteq N$
7 $k = Max_{v\in centers}\, Min_{v\in centers}\{d_{uv}\}.rad$
8 $LL(d_{centers\times centers}, L(centers), radius, iterations)$
9 **for** each actor, $n \in N$ **begin**
10 $L(v) \leftarrow L\big(center(v)\big) + rand$
11 $k \leftarrow k.ratio$
12 **Return** L



a). c_canon b). c_canon2

c). c_canon3 d). c_combined

Figure 3. Visualization of co-occurrence network for product features and sentiments.

The modularity is formulated to compare the different groups in the same network. Let *G = (N, L)* be an undirected social network. Let A be a co-occurrence matrix. Then, the relations are represented as follows.

$$A_{mn} = \begin{cases} 1 & (m,n) \in L \\ 0 & Otherwise. \end{cases}$$

The total number of relations and the degree of an actor is defined as follows.

$$l = \sum_{m,n \in N} \frac{A_{mn}}{2}$$

$$k_m = \sum_{n \in N} A_{mn}$$

A graph *G* can be partitioned into a set of groups or subsets:

$$C = \{c_1, c_2, c_3, \ldots\}, \; c_i \cap c_j = \emptyset (i \neq j), \bigcup_{c_i \in C} c_i = N$$

$$r_{ij} = \sum_{m,n \in N} \frac{A_{mn}}{2}$$

$$a_i = \sum_{n \in N} A_{mn}$$

where r_{ij} refers to the proportion of relation that links members of the group c_i and c_j, and a_i refers to the proportion of c_i's relation. Therefore, the modularity is defined as follows.

$$Q(G, C) = \sum_i (e_{ij} - a_i^2)$$

CNM algorithm works the same as the NG clustering algorithm, but it incorporates two data structures, namely, balanced binary tree or heap tree and max heap or priority heap to find a group pair repeatedly. The improved modularity is formulated as follows.

$$\Delta Q_{c_i, c_j}^{C} = Q\left(G, C - c_i - c_j + (c_i \cup c_j)\right) - Q(G, C)$$

The algorithm gives the maximum group pair values ΔQ and merges them into a new group. The ΔQ values of groups that adjoin the new group need to be updated. Therefore, the algorithms stop when there is no group pair to merge. Wakita et al. (2007) replaced a doubly linked list instead of baa lanced binary tree and max heaps. The algorithm works the same as CNM algorithm, but it controls the growth of groups by incorporating three kinds of heuristics consolidation ratios. These heuristics measure the group size in terms of degree, both candidate group pair and degree, and the number of actors.

Social network analysis metrics

The SNA metrics are the standard measurement to study the characteristic of a network. It can be classified into two categories, namely, the overall metrics and vertex metrics. First, the overall metrics summarize the key points or properties of the entire network, which includes graph type, unique edges, edges with duplicates, total edges, self-loops (number of nodes that connects itself), vertices, and graph density. The graph density is the proportional ratio between the number of actual connections and the number of possible (or potential) connections (Scott et al., 2005; Kim et al., 2018; Peng et al., 2018; Zhang et al., 2017;). Mathematically, it is defined as follows.

$$Graph\ Density = \frac{Number\ of\ actual\ connections}{Number\ of\ possible\ connections}\ \ or$$

$$\Delta = \frac{L}{n(n-1)/2} = \frac{2L}{n(n-1)}$$

where

$$\text{Number of possible connections} = \frac{\text{n(n} - 1)}{2}$$

Second, the vertex metrics can be performed at the individual node-level, dyad-level, triad-level, group-level, or network level. In this metric, each node value directly relates to one of the nodes in the graph. The vertex metrics include centrality and prestige. Centrality measures high involvement of an actor in the undirected graph. Prestige measures the popularity of an actor in the directed graph. In this paper, we describe various centrality measures for the undirected graph, namely, Degree Centrality (DC), Betweenness Centrality (BC), Closeness Centrality (CC), and Eigenvector Centrality (EC). These metrics are presented as follows.

Degree Centrality

Degree centrality (also called degree) is the simplest form of network measure. It assigns an important score to an actor based on the number of ties or relationships in a network. The highest score of an actor plays a central role in the network (Varlamis et al., 2010; Kim et al., 2018; Peng et al., 2018; Roy et al., 2018). It is used to interpret the network information flow, or transactions take place with other actors. Graph G contains a set of nodes ($N = \{n_1, n_2, n_3, \dots, n_g\}$) and a set of lines ($L = \{l_1, l_2, l_3, \dots, l_g\}$).

Therefore, the degree centrality on the node level is defined as

$$C_D(n) = \deg(n)$$

where *deg(n)* refers to the number of links on the node *n*. The standardized or normalized degree centrality is calculated by dividing the maximum possible connections as follows:

$$C'_D(n) = \frac{\deg(n)}{(\text{g} - 1)}$$

The node-level definition can be extended to the entire graph G, called graph centralization (Arif, 2015). Let $n*$ be the highest degree in graph G. Let $A := (B, C)$ be the $|B|$ connected graph that is maximized as follows:

$$H = \sum_{i=1}^{|B|} [C_D(b*) - C_D(b_j)]$$

where $b*$ refers to the highest degree centrality in A. The graph centralization is defined as follows:

$$C_D(G) = \frac{\sum_{i=1}^{|N|} [C_D(n*) - C_D(n_i)]}{H}$$

Betweenness Centrality

In a network, an actor that lies between other actors is called betweenness. The number of times an actor lies between other actors on the shortest path is called betweenness centrality. It reflects the indirect connectivity of actors through direct links or ties. Betweenness centrality can be used to find the actors who influence the information flow in the system. The actor's high betweenness count indicates authority or control over other actors (Colladon et al., 2017;

Kim et al., 2018; Peng et al., 2018; Roy et al., 2018; Arif, 2015). Mathematically, the actor betweenness centrality is defined as the sum of the proportions between all pairs of actor *j* and actor *k*.

$$C_B(n_i) = \sum_{j<k} \frac{g_{jk}(n_i)}{g_{jk}}$$

where g_{jk} denotes the total number of shortest paths from actor *j* to actor *k* and $g_{jk}(n_i)$ denotes the number of paths that pass through actor i.

Closeness Centrality

Closeness centrality measure scores an individual actor based on their closeness to all other actors within the network either directly or indirectly. This measure calculates the shortest paths between the actor and all other actors in the network. Then, it assigns a score to each actor based on the sum of the shortest paths. Closeness centrality can be used to find the best actor to access information or influence the whole network. In a highly connected graph, all actors may have a similar score. In that case, the influences can be identified within a single cluster using closeness (Colladon et al., 2017; Varlamis et al., 2010; Peng et al., 2018; Roy et al., 2018; Arif, 2015). Let *G* be a graph with *n* actors. The closeness centrality is defined mathematically as

$$C_C(n_i) = \frac{n-1}{\sum_{j=1}^{n} d(n_i, n_j)}$$

where $C_c(n_i)$ refers to the closeness centrality of the actor n_i and $d(n_i, n_j)$ denotes the shortest path (or geodesic distance) between the actor n_i and n_j.

Eigenvector centrality

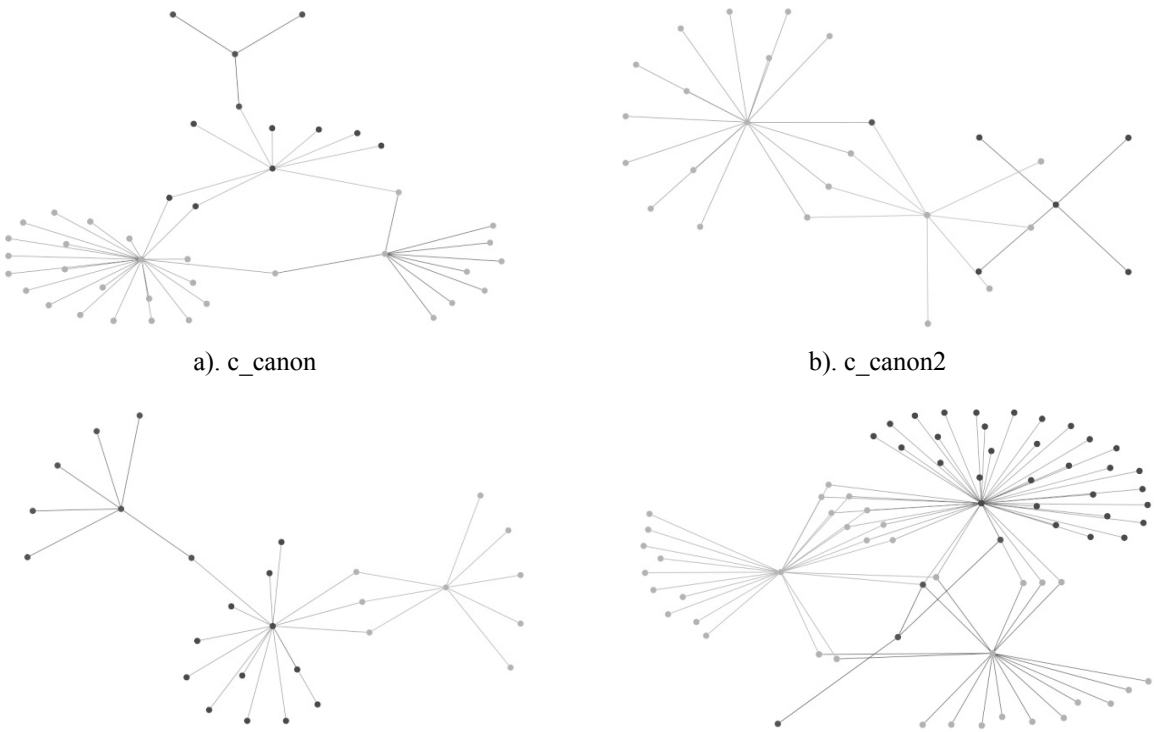
Eigenvector centrality (or eigencentrality) is a measure like degree centrality, but it measures an actor's popularity based on the number of links and the quality of those links within the network (Kim et al., 2018; Peng et al., 2018; Arif, 2015). Eigencentrality assigns a relative score to all actors based on the connections and the score of an actor. Moreover, an actor that is connected to many actors who themselves have a high score is called a high eigenvector score. Let $G = (N, L)$ be a graph with $|N|$ actors. Let $A = (a_{n,t})$ be the adjacency matrix.

Then, *A* can be defined as $a_{n,t} = 1$ if an actor *n* is connected to the actor *t*, and $a_{n,t} = 0$ otherwise. Therefore, the eigencentrality scores of an actor *n* can be defined as follows.

$$C_E(n) = \frac{1}{\lambda} \sum_{t \in M(n)} x_n = \frac{1}{\lambda} \sum_{t \in G} a_{n,t} x_n$$

where $M(n)$ refers to a set of neighbors of the actor *n* and λ is a constant. The eigencentrality can also be written in vector notation as follows:

$$AX = \lambda X$$



a). c_canon b). c_canon2

c). c_canon3 d). Combined

Figure 4. Grouping the product reviews.

RESULTS AND DISCUSSION

Key findings

In this section, we applied social network analysis metrics to analyze the product features and their relationships from co-occurrence networks (Smith et al., 2009; Garg et al., 2017; Forss et al., 2016; Zhao et al., 2018).

The co-occurrence networks were constructed for three product reviews (c_canon, c_cannon2, and c_canon3) individually. Then, we constructed one more co-occurrence network by combining all three product reviews. These reviews were preprocessed and identified product specific features and their corresponding sentiments: strong positive, positive, negative, and strong negative. CNM algorithm was employed to cluster the individual actors to form subgroups using modularity in the constructed networks. The network groups were identified with the same vertex color and size. For instance, consider the product c_canon co-occurrence social network. This network contains 43 terms (or actors), which are clustered into four subgroups (Table 3). The subgroups are assigned with a name, vertex color, and vertex shape. The groups C52, C73, C88, and C82 contained 4, 8, 10, and 21 terms, respectively. The groups represent the disk shape with four different colors, namely, red (255, 0, 0), blue (0, 0, 255), orange (255, 172, 0), and lime (0, 255, 0). Similarly, the grouping method is applied to the product reviews c_canon2, c_canon3, and combined data. The groups and their links are shown in Figure 4.

Further, the SNA metrics like Graph density, degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality are explored to identify the role of product features. The overall graph metrics are calculated for all products, as shown in Table 4. It includes the number of actors, unique edges, graph density, and the number of clusters. The graph density was calculated to identify a product information flow. It increases by increasing the number of actors. The product *c_canon2* has a higher density than *c_canon* and *c_canon3*. Vertex metrics were calculated

for all actors in the networks, but some of the actors are not useful for products. Therefore, we selected the most prominent product features based on the predetermined product's features, which is unique and made by combining all product features. Table 5 and Table 6 show 20 selected features based on degree centrality. Degree centrality measured the most popular or influential terms in the co-occurrence network.

Table 3. List of subgroups.

Cluster	Vertex color	Vertex shape	Group terms
C52	255, 0, 0	Disk	Strong negative, camera, strap, time
C73	0, 0, 255	Disk	Strong positive, Picture, quality, review, photo, meter, zoom, fine
C78	255, 172, 0	Disk	Negative, casing, flaw, hand, purpose, scene, thing, battery, distortion, viewfinder
C82	0, 255, 0	Disk	Positive, shot, heft, mode, price, lens, direction, feature, hope, photo bugs, box, flash, amaze, canon, control, computer, addition, semiserious, use, look, bag

Table 4. Overall graph metrics.

Metric	Datasets			
	c_canon	c_canon2	c_canon3	combined
Graph Type	Undirected	Undirected	Undirected	Undirected
Unique Edges	44	28	29	88
Edges with Duplicates	-	-	-	-
Self-Loops	-	-	-	-
Vertices	43	27	28	72
Graph Density	0.049	0.080	0.077	0.034
No. of Clusters	4	4	3	4

The sentiment polarity terms are strong positive, positive, negative, and strong negative, and the product feature terms like camera, picture, quality, and viewfinder play a vital role in the product c_canon.

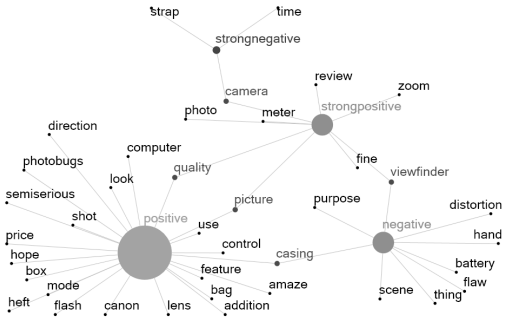
Betweenness centrality is calculated to identify the number of times an actor lies between other actors in the shortest distance. The terms like positive, strongly positive, negative, camera, casing, quality, and pictures indicate a high betweenness count for the product c_canon. Therefore, it has more authority or control over the network.

Closeness centrality measured the shortest distance between all terms and then assigned a score to each term. The individual terms strong negative, battery, purpose, distortion, flaw, use, feature, flash, and viewfinder are the best places to influence the product network c_canon.

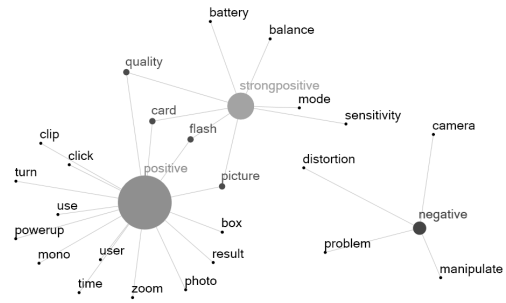
Table 5. Calculation of vertex metrics for c_canon and c_canon2 networks.

c_canon					c_canon2				
Vertex	DC	BC	CC	EC	Vertex	DC	BC	CC	EC
positive	23	1.000	1.976	0.960	positive	16	1.000	1.429	0.883
negative	9	0.429	2.643	0.077	Strong positive	8	0.435	2.190	0.366
Strong positive	9	0.543	2.500	0.140	negative	4	0.034	1.000	0.000
Strong negative	3	0.125	4.214	0.007	picture	2	0.092	1.905	0.092
viewfinder	2	0.141	3.048	0.011	quality	2	0.092	1.905	0.092
casing	2	0.283	2.524	0.051	card	2	0.092	1.905	0.092
quality	2	0.178	2.429	0.054	flash	2	0.092	1.905	0.092
picture	2	0.178	2.429	0.054	photo	1	0.000	2.381	0.065
camera	2	0.181	3.333	0.007	user	1	0.000	2.381	0.065
battery	1	0.000	3.619	0.004	time	1	0.000	2.381	0.065
scene	1	0.000	3.619	0.004	use	1	0.000	2.381	0.065
thing	1	0.000	3.619	0.004	power up	1	0.000	2.381	0.065
purpose	1	0.000	3.619	0.004	turn	1	0.000	2.381	0.065
distortion	1	0.000	3.619	0.004	mono	1	0.000	2.381	0.065
flaw	1	0.000	3.619	0.004	click	1	0.000	2.381	0.065
hand	1	0.000	3.619	0.004	result	1	0.000	2.381	0.065
use	1	0.000	2.952	0.047	zoom	1	0.000	2.381	0.065
feature	1	0.000	2.952	0.047	clip	1	0.000	2.381	0.065
hope	1	0.000	2.952	0.047	battery	1	0.000	3.143	0.027
flash	1	0.000	2.952	0.047	sensitivity	1	0.000	3.143	0.027
*DC-Degree centrality, BC-Betweenness centrality, CC-Closeness centrality, EC-Eigenvector centrality									

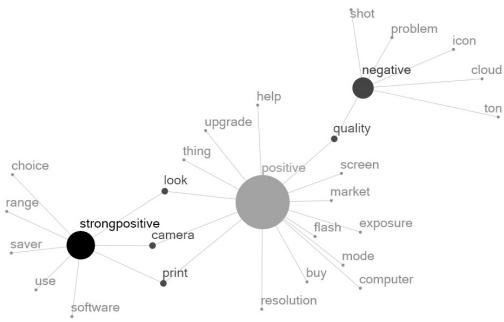
Eigencentrality is calculated to identify the influence of a term in the entire network. In the product c_canon, the terms positive, strongly positive, negative, casing, picture, quality, feature, and flash show the highest relative score. These terms indicate more contribution to the whole network than low scoring terms. The vertex metrics were also calculated for c_canon2, c_canon3, and combined data.



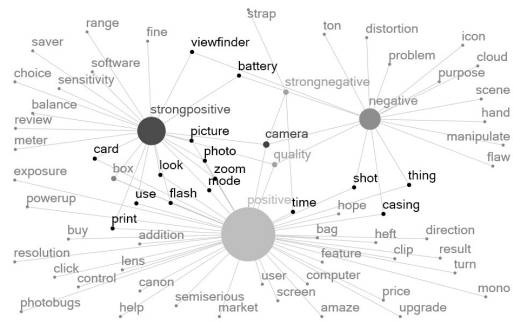
a). c_canon



b). c_canon2



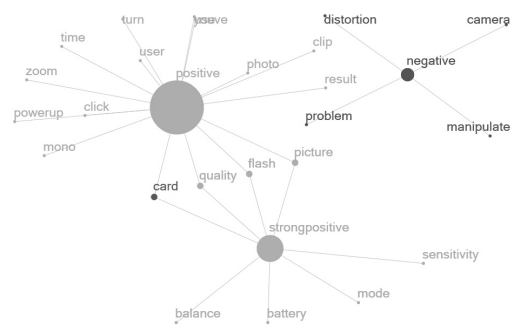
c). c_canon3



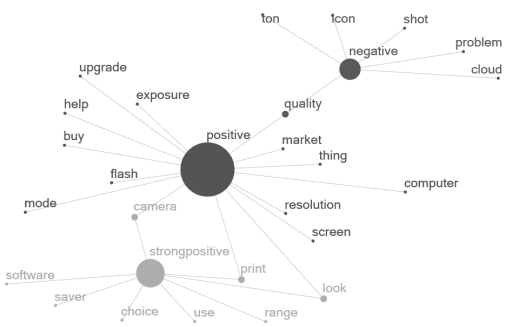
d). combined

Figure 5. Visualization of the product features based on degree centrality.

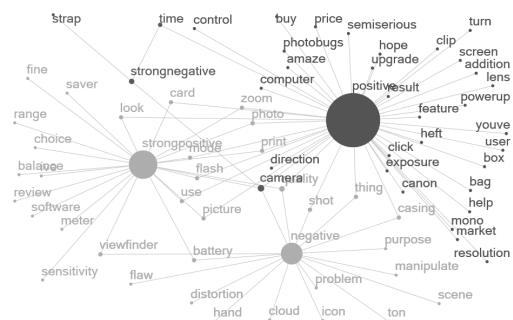
a). c_canon



b). c_canon2



c). c_canon3



d). combined

Figure 6. Visualization of the network categorization based on degree centrality.

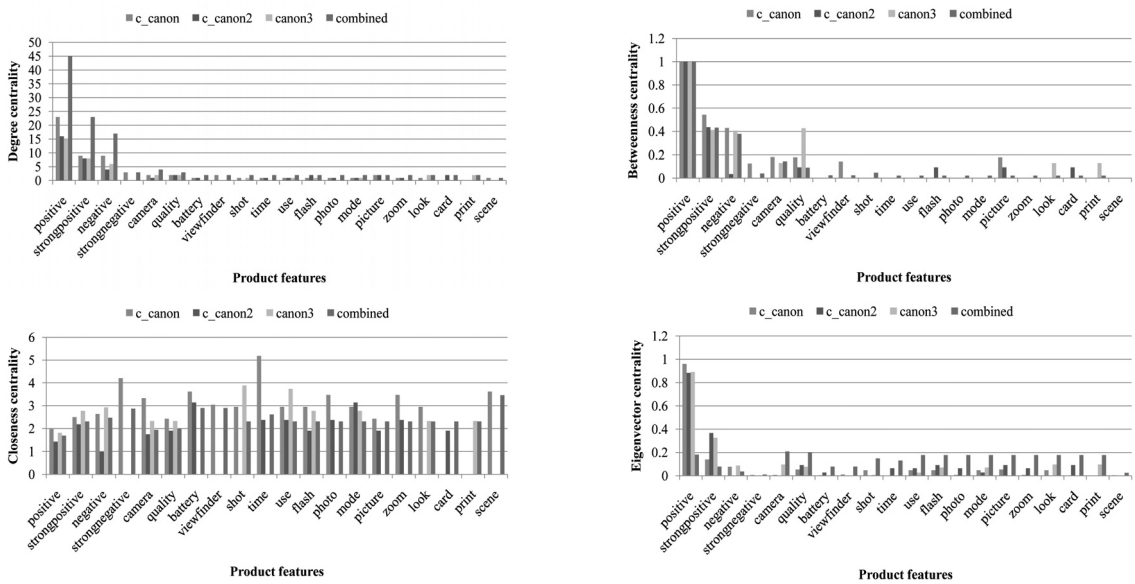


Figure 7. The comparison of product features based on vertex metrics.

Table 6. Calculation of vertex metrics for c_canon3 and combined networks.

c_canon3					c_combined				
Vertex	DC	BC	CC	EC	Vertex	DC	BC	CC	EC
positive	15	1.000	1.815	0.891	positive	45	1.000	1.690	0.182
Strong positive	8	0.411	2.778	0.327	Strong positive	23	0.431	2.310	0.078
negative	6	0.406	2.926	0.088	negative	17	0.378	2.479	0.037
camera	2	0.129	2.333	0.096	camera	4	0.143	1.944	0.209
look	2	0.129	2.333	0.096	quality	3	0.090	2.000	0.202
print	2	0.129	2.333	0.096	Strong negative	3	0.040	2.873	0.010
quality	2	0.426	2.333	0.077	battery	2	0.023	2.901	0.078
buy	1	0.000	2.778	0.070	viewfinder	2	0.023	2.901	0.078
choice	1	0.000	3.741	0.026	shot	2	0.046	2.310	0.149
cloud	1	0.000	3.889	0.007	casing	2	0.046	2.310	0.149
computer	1	0.000	2.778	0.070	thing	2	0.046	2.310	0.149
exposure	1	0.000	2.778	0.070	time	2	0.021	2.620	0.131
flash	1	0.000	2.778	0.070	use	2	0.021	2.310	0.177
help	1	0.000	2.778	0.070	flash	2	0.021	2.310	0.177

icon	1	0.000	3.889	0.007	photo	2	0.021	2.310	0.177
market	1	0.000	2.778	0.070	mode	2	0.021	2.310	0.177
mode	1	0.000	2.778	0.070	picture	2	0.021	2.310	0.177
problem	1	0.000	3.889	0.007	zoom	2	0.021	2.310	0.177
range	1	0.000	3.741	0.026	look	2	0.021	2.310	0.177
resolution	1	0.000	2.778	0.070	card	2	0.021	2.310	0.177
*DC-Degree centrality, BC-Betweenness centrality, CC-Closeness centrality, EC-Eigenvector centrality									

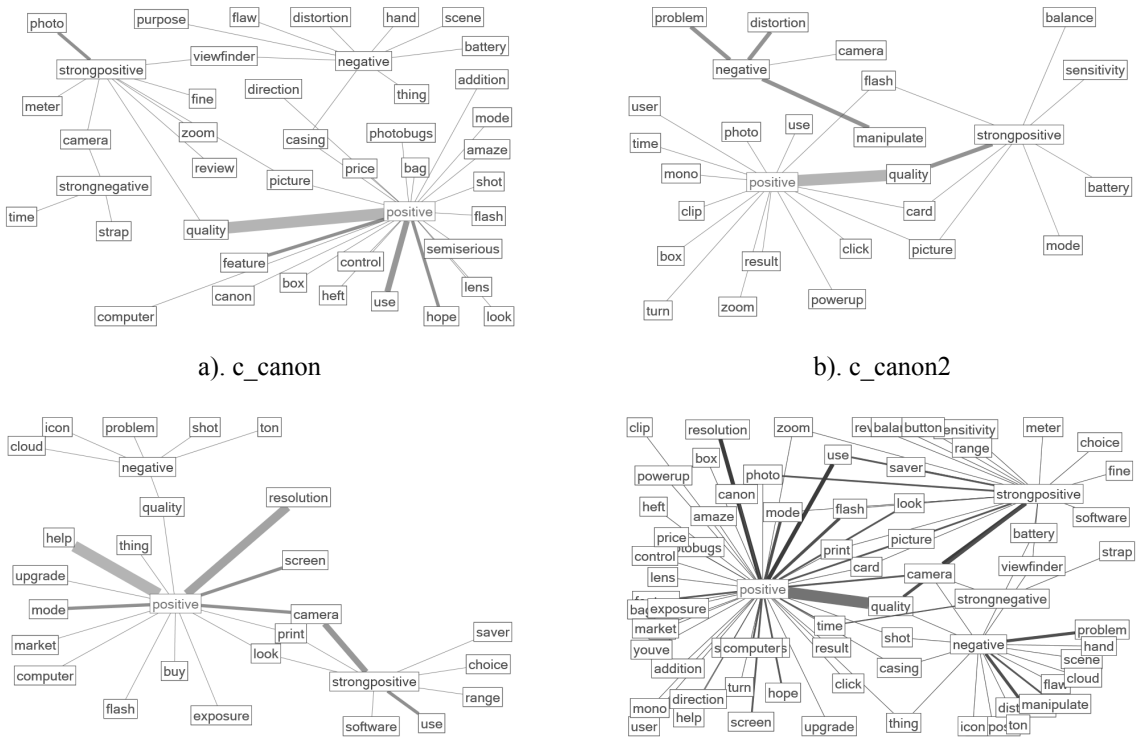


Figure 8. Visualization of social strength between product features and sentiment polarity.

A combined network helps understand the overall influence of the product features from all products. Moreover, the networks are grouped and visualized for all products based on degree centrality, as shown in Figure 5. The terms with the same degree value are categorized into different groups, which represent the same color, shape, and size (Figure 6) in the categorized networks. Similarly, the betweenness centrality, closeness centrality, and eigencentrality are performed for all products. The comparison of top product features from three products reviews and combined reviews is shown in Figure 7. It helps the business launcher or buyer decide products based features and sentiments. In particular, there are several positive features in c_canon than in other products. The degree of combined data represents an overall influence of features in all products.

Also, the strength of their relationship between other terms is highlighted and visualized based on their edge weight (Figure 8). There is a strong relationship for the terms: quality and use with the positive sentiment in the product c_canon. Moreover, the proposed research extends the application of network science to the identification of product features for sentiment analysis. The findings focused on textual reviews with specific product features. It can be applied to any product domains, events, or organizations. In particular, the proposed approach is not discussing a semantic category or meaning of text reviews. It focused only on the co-occurrence of product features and sentiments.

Theoretical and practical implications

The proposed method adopts a sentiment-based co-occurrence network approach to identify the influence of product features in text corpora. This method provides an example to construct a co-occurrence network for product reviews. In result, the sentiment-based co-occurrence network analysis provides an idea to identify the popularity of product features and their labeled sentiment polarities. Therefore, the proposed method significantly contributes to related works in several ways. In theoretical perspective, first, the method identifies product terms and their sentiments, which were used to construct the co-occurrence network using the fast multiscale layout algorithm. Second, the CNM algorithm was employed to investigate the hidden message in the product reviews. Third, the method examines the network metrics to share knowledge with the product developers, launchers, and sellers. Finally, the method examines the strength of the social relationship between product features and sentiments. In particular, the results show that the sentiment based co-occurrence network analysis provides a valuable framework to identify the popularity of product features. From a practical perspective, the network metrics are employed to characterize the product information network. The proposed method also identifies individual product features and their influences in text reviews. However, the proposed method is too general. It can be employed in any text reviews in the real world.

Therefore, the proposed method can be used to identify the popularity of product features and their relationship with sentiments.

CONCLUSION

In this paper, we presented an opinion-based co-occurrence network for product reviews using social network analysis to identify the most influential product features in positive sentiment and negative sentiment. The product data were collected from a feature specific sentiment analysis dataset. Co-occurrence networks were constructed for product reviews after preprocessing data and constructing a data matrix. The Harel-Koren fast multiscale drawing algorithm was employed to visualize the networks. Further, the CNM algorithm was applied to group the actors in the co-occurrence information network. Additionally, we measured the graph density, degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality to identify the importance of actors in the network. Moreover, we ranked only the top 20 actors based on centrality measures. The results revealed that the product features have different roles in positive and negative sentiment. The strong social relationship between product features and sentiment polarity is the most influential in the network. In the future, this work will be performed and compared using n-gram features and big data with different sentiment lexicons such as Sentistrength, Sentiment140, SentiWordNet, VADER, AFINN, SenticNet, and ANEW.

ACKNOWLEDGMENT

This work was supported by the University Grants Commission (UGC) National Fellowship (F1-17.1/201617/RGNF-2015-17-SC-TAM-4711).

REFERENCES

Abdelsadek, Y., Chelghoum, K., Herrmann, F., Kacem, I. & Otjacques, B. 2018. Community extraction and visualization in social networks applied to Twitter. Information Sciences, **424**: 204-223.

Alam, M.H., Ryu, W.J. & Lee, S. 2016. Joint multi-grain topic sentiment: modeling semantic aspects for online reviews. Information

Sciences, 339, 206-223.

Amplayo, R.K., Hong, S. & Song, M. 2018. Network-based approach to detect novelty of scholarly literature. Information Sciences, **422**: 542-557.

Arif, T. 2015. The mathematics of social network analysis: metrics for academic social networks. International Journal of Computer Applications Technology and Research, **4**(12): 889-93.

Bello-Orgaz, G., Jung, J.J. & Camacho, D. 2016. Social big data: Recent achievements and new challenges. Information Fusion, **28**: 45-59.

Chang, V. 2018. A proposed social network analysis platform for big data analytics. Technological Forecasting and Social Change, **130**: 57-68.

Chen, J., Liu, Y., Yang, G. & Zou, M. 2018. Inferring tag co-occurrence relationship across heterogeneous social networks. Applied Soft Computing, **66**: 512-524.

Cheng, K., Li, J., Tang, J. & Liu, H. 2017. Unsupervised sentiment analysis with signed social networks. In Thirty-First AAAI Conference on Artificial Intelligence.

Claros, I., Cobos, R. & Collazos, C.A. 2015. An approach based on social network analysis applied to a collaborative learning experience. IEEE Transactions on Learning Technologies, **9**(2): 190-195.

Colladon, A.F. & Remondi, E. 2017. Using social network analysis to prevent money laundering. Expert Systems with Applications, **67**: 49-58.

Cristian Zanoci & Jim Andress, 2016. The Times They Are a Changin': Evolving Communities in a Musician Network. Standford Network Analysis Project. http://snap.stanford.edu/class/cs224w-2016/projects/cs224w-70-final.pdf.

De Brún, A. & McAuliffe, E. 2018. Social Network Analysis as a methodological approach to explore health systems: A case study exploring support among senior managers/executives in a hospital network. International journal of environmental research and public health, **15**(3): 511.

De Marneffe, M.C. & Manning, C.D. 2008. Stanford typed dependencies manual. Technical report, Stanford University, 338-345

Farasat, A., Gross, G., Nagi, R. & Nikolaev, A.G. 2016. Social network analysis with data fusion. IEEE Transactions on Computational Social Systems, **3**(2): 88-99.

Forss, T. & Sarlin, P. 2016. From news to company networks: Co-occurrence, sentiment, and information centrality. In 2016 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1-8). IEEE.

Garg, M. & Kumar, M. 2018. Identifying influential segments from word co-occurrence networks using ahp. Cognitive Systems Research, **47**: 28-41.

Hayat, T.Z., Lesser, O. & Samuel-Azran, T. 2017. Gendered discourse patterns on online social networks: A social network analysis perspective. Computers in Human Behavior, **77**: 132-139.

Hirsch, L. & Andrews, S. 2016. Visualising text co-occurrence networks. In CEUR Workshop Proceedings (Vol. 1637, pp. 19-27). Tilburg University.

Hu, J. & Zhang, Y. 2015. Research patterns and trends of Recommendation System in China using co-word analysis. Information processing & management, **51**(4): 329-339.

Hughes, C.E., Bright, D.A. & Chalmers, J. 2017. Social network analysis of Australian poly-drug trafficking networks: How do drug traffickers manage multiple illicit drugs?. Social Networks, **51**: 135-147.

Ingo Feinerer. 2017. Introduction to the tm Package Text Mining in R. https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf.

Kauer, A.U. & Moreira, V.P. 2016. Using information retrieval for sentiment polarity prediction. Expert Systems with Applications, **61**: 282-289.

Khasseh, A.A., Soheili, F., Moghaddam, H.S. & Chelak, A.M. 2017. Intellectual structure of knowledge in iMetrics: A co-word analysis. Information processing & management, **53**(3): 705-720.

Kim, J. & Hastak, M. 2018. Social network analysis: Characteristics of online social networks after a disaster. International Journal of Information Management, **38**(1): 86-96.

Koren, D.H.Y. 2002. A fast multi-scale method for drawing large graphs. Journal of graph algorithms and applications, **6**(3): 179-202.

Kulig, A., Kwapień, J., Stanisz, T. & Drożdż, S. 2017. In narrative texts punctuation marks obey the same statistics as words. Information Sciences, **375**: 98-113.

Li, H., An, H., Wang, Y., Huang, J. & Gao, X. 2016. Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network. Physica A: Statistical Mechanics and its Applications, **450**: 657-669.

Li, L., Zhang, Q., Tian, J. & Wang, H. 2018. Characterizing information propagation patterns in emergencies: A case study with Yiliang Earthquake. International Journal of Information Management, **38**(1): 34-41.

Li, Y., Tu, Y. & Li, X. 2018. Study on Enterprises' Internet Public Opinion Area Hotspots Based on Social Network Analysis.

Li, Y., Zhang, D., Luo, P. & Jiang, J. 2017. Interpreting the formation of co-author networks via utility analysis. Information Processing & Management, **53**(3): 624-639.

MeaningCloudTM (http://www.meaningcloud.com) have been used for Text Analytics purposes in the development/testing/validation of this research/prototype/software.

Mishra, N., Schreiber, R., Stanton, I. & Tarjan, R. E. 2007. Clustering social networks. In International Workshop on Algorithms and Models for the Web-Graph (pp. 56-67). Springer, Berlin, Heidelberg.

Mukherjee, S. & Bhattacharyya, P. 2012. Feature specific sentiment analysis for product reviews. In International Conference on Intelligent Text Processing and Computational Linguistics (pp. 475-487). Springer, Berlin, Heidelberg.

Ongkowijoyo, C.S. & Doloi, H. 2018. Understanding of Impact and Propagation of Risk based on Social Network Analysis. Procedia engineering, **212**: 1123-1130.

Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J. & Jia, W. 2018. Influence analysis in social networks: a survey. Journal of Network and Computer Applications, **106**: 17-32.

Qiu, J., Liu, C., Li, Y. & Lin, Z. 2018. Leveraging sentiment analysis at the aspects level to predict ratings of reviews. Information Sciences, **451**: 295-309.

Qiu, Z. & Shen, H. 2017. User clustering in a dynamic social network topic model for short text streams. Information Sciences, **414**: 102-116.

Radhakrishnan, S., Erbis, S., Isaacs, J.A. & Kamarthi, S. 2017. Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. PloS one, **12**(3): e0172778.

Razghandi, M. & Golpaygani, S.A.H. 2017. A context-aware and user behavior-based recommender system with regarding social network analysis. In 2017 IEEE 14th International Conference on e-Business Engineering (ICEBE)(pp. 208-213). IEEE.

Richard G. Brereton. 2017. Basic matrix algebra. Journal of Chemometrics, **31**: 1-4. DOI: 10.1002/cem.2833.

Roy, S., Dey, P. & Kundu, D. 2017. Social Network Analysis of Cricket Community Using a Composite Distributed Framework: From Implementation Viewpoint. IEEE Transactions on Computational Social Systems, **5**(1): 64-81.

Scott, J., Tallia, A., Crosson, J.C., Orzano, A.J., Stroebel, C., DiCicco-Bloom, B., ... & Crabtree, B. 2005. Social network analysis as an analytic tool for interaction patterns in primary care practices. The Annals of Family Medicine, **3**(5): 443-448.

Shin, K.Y. & Lee, J.H. 2017. A job applicants' résumé verification method using a social network analysis Using Facebook like as Linkedin for a recruiting. In 2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA) (pp. 1-5). IEEE.

Smith, M.A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., ... & Gleave, E. 2009. Analyzing (social media) networks with NodeXL. In Proceedings of the fourth international conference on Communities and technologies (pp. 255-264). ACM.

Tang, J., Qu, M. & Mei, Q. 2015, August. Pte: Predictive text embedding through large-scale heterogeneous text networks.

In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1165-1174). ACM.

Tubishat, M., Idris, N. & Abushariah, M.A. 2018. Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges. Information Processing & Management, **54**(4): 545-563.

Varlamis, I., Eirinaki, M. & Louta, M. 2010. A study on social network metrics and their application in trust networks. In 2010 International Conference on Advances in Social Networks Analysis and Mining (pp. 168-175). IEEE.

Wakita, K. & Tsurumi, T. 2007. Finding community structure in mega-scale social networks. In Proceedings of the 16th international conference on World Wide Web (pp. 1275-1276). ACM.

Wissink, M. & Mazzucato, V. 2018. In transit: Changing social networks of sub-Saharan African migrants in Turkey and Greece. Social Networks, **53**: 30-41.

Yang, B., Liu, Y., Liang, Y. & Tang, M. 2019. Exploiting user experience from online customer reviews for product design. International Journal of Information Management, **46**: 173-186.

Yang, H.L. & Lin, Q.F. 2018. Opinion mining for multiple types of emotion-embedded products/services through evolutionary strategy. Expert Systems with Applications, **99**: 44-55.

Yang, S., Han, R., Wolfram, D. & Zhao, Y. 2016. Visualizing the intellectual structure of information science (2006–2015): Introducing author keyword coupling analysis. Journal of Informetrics, **10**(1): 132-150.

Zarrinkalam, F., Kahani, M. & Bagheri, E. 2018. Mining user interests over active topics on social networks. Information Processing & Management, **54**(2): 339-357.

Zhang, Z., Gu, Q., Yue, T. & Su, S. 2017. Identifying the same person across two similar social networks in a unified way: Globally and locally. Information Sciences, **394**: 53-67.

Zhao, W., Mao, J. & Lu, K. 2018. Ranking themes on co-word networks: Exploring the relationships among different metrics. Information Processing & Management, **54**(2): 203-218.

Zhou, F., Qu, Q. & Toivonen, H. 2017. Summarisation of weighted networks. Journal of Experimental & Theoretical Artificial Intelligence, **29**(5): 1023-1052.