BMC Public Health

RESEARCH

Open Access

Health promotion campaigns using social media: association rules mining and co-occurrence network analysis of Twitter hashtags

Atousa Ghahramani[1*], Maria Prokofieva[1] and Maximilian Pangratius de Courten[2]

Abstract

Background Social media hashtags play a significant role in increasing the visibility of health information by making it easier for people to explore health-related content. Health promotion campaigns use campaign-specific hashtags to disseminate health-related messages, enabling individuals to access accurate and timely resources and updates. The study aims to discover patterns of connection between hashtags and identify the most influential hashtags used on Twitter in the American Heart Month campaigns.

Method We collected a total of 73,288 tweets containing #AmericanHeartMonth between January 2019 and March 2023 and retrieved 18,143 original tweets, 42,930 retweets, 2,519 quotes, and 20,846 likes related to the past five campaigns. We adapted co-occurrence network analysis to explore the patterns of relationships between hashtags and association rules mining to assess the quality and strength of association between the co-occurred hashtags.

Result While #AmericanHeartMonth, #OurHearts, and #HeartMonth play central roles in all hashtag co-occurrence networks, the results of association rules mining indicate a significant association of #OurHearts within the networks. The highest density of hashtags has been observed in the quoted tweets, introducing a new range of hashtags such as #GoRedForWomen, #WearRedDay, #HeartDisease, and #HeartHealth by Twitter users, indicating the positive correlation between co-occurring hashtags and users' engagement. The results of quality measurements of association rules (Lift > 1) indicate positive relationships between the co-occurred hashtags in the top 5 rules in all data subsets.

Conclusion We employed co-occurrence network analysis and association rules mining as powerful techniques to identify influential hashtags that may have a central role in health-related discussions and drive engagement within the co-occurrence hashtag network. In conclusion, we recommend additional hashtag structures in conjunction with heart health-related topics to improve community building and the effectiveness of disseminating messages in future heart health promotion campaigns. The study contributes to knowledge and practice by offering a structured and data-driven approach and providing practical guidance for public health practitioners, professionals, and organisations to optimise content, targeting, and messaging to reach and engage a broader audience with health-related information.

Keywords Association rules mining, Co-occurrence networks, Hashtag analysis, Health promotion, Twitter

*Correspondence:
Atousa Ghahramani
atousa.ghahramani@live.vu.edu.au
Full list of author information is available at the end of the article

Background

An estimated 17.9 million people died from cardiovascular disease in 2019, constituting 32% of all global deaths. Of these deaths, 85% were attributed to heart attacks and strokes [1]. Raising awareness of heart health appears crucial in encouraging individuals to make changes toward a healthy lifestyle [2]. Health organisations leverage the user-friendly features of Twitter [3] to communicate health messages to a broad range of audiences, with the ultimate goal of influencing individuals and public lifestyles to bring about changes in health behaviours [4, 5]. Studies have highlighted the need for health organisations to exert efforts in establishing credible health information in online environments [3]. Additionally, they should carefully and regularly evaluate the performance of campaigns, making necessary adjustments to reach and engage a broader audience. Such efforts can significantly impact public health awareness and facilitate the process of changing health-related behaviours [4, 6].

The use of social media for health promotion

Social media plays a multifaceted and significant role in health promotion interventions. The influential and wide-reaching nature of social media allows health professionals and organisations to communicate with diverse audiences, ultimately influencing health behaviour [7]. Analysis of user engagement and the monitoring of specific health behaviours on social media can be used to suggest systems and procedures to stakeholders to improve strategies for future health promotion campaigns [8]. Users increasingly search for and share health-related information on social media channels, creating rich social and diverse cultural data available for exploration in the health domain [9–11]. Such data, along with interactions between social media users, provides rich data sources that support the discovery of latent patterns of users' health-related behaviours [11]. In the context of promoting health-related behaviours on social media platforms such as Twitter, tweets, retweets, quotes, and likes to play important roles in disseminating health messages and engaging with target audiences [12]. There is a wealth of opportunities to use social media for health promotion through targeted messages, the ability to interact with the public, targeting hard-to-reach groups, and creating dynamic campaigns [13, 14].

Evidence shows that health promotion initiatives could increase the reach, exposure, impression, impact, and engagement of social media users and impact their health-related behaviours [15]. Prior research analysed the potential of social media campaigns to improve user engagement and enhance people's behaviour towards a healthy lifestyle [9]. They addressed specific methodological approaches, which vary widely in focus, target population, theoretical foundations, mode of delivery, functionality, and usability. This variation makes it difficult to ascertain what works and how, complicating efforts to compare approaches [11].

The primary feature of Twitter is the capability to provide communication loops in Twitter networks and increase engagement through retweeting, liking, and quoting a tweet [14], allowing users to communicate a topic of interest within a community [3]. Twitter users retweet when they want to share information with their network and wider audiences and value other users' opinions and thoughts [16]. Despite growing public interest in interactive communication and dialogue on social media, the majority of health-related messages remain predominantly one-way, focusing on information dissemination rather than fostering engagement or conversation [17]. The numbers of retweets, likes, replies, and quotes on a tweet posted by a health promotion organisation represent other users' thoughts about the shared messages. It can be identified as a key to attract people and improve perceptions and credibility of the message [16].

Thematic use of hashtags on Twitter helps categorise and relate content, making them one of the most commonly employed components in tweets. Recent studies have shown that by following, tweeting, and retweeting specific hashtags, users actively engage in communication within broader online communities, fostering discussions around shared topics [15]. Public health organisations create and promote particular hashtags or use popular hashtags that can facilitate easy search and discovery of health-related information by users on social media [3]. Hashtag analysis has been broadly used as a tool in research that allows different quantifications [18, 19]. Fuster-Casanovas et al. [20] applied network analysis related to #VaccinesWork, promoted by UNICEF, to detect the optimal time to launch the campaigns and investigate the shared main messages and involvement of key leaders that had a significant influence on the dissemination of health information on Twitter [20]. Uwins et al. [21] employed hashtag network analysis of the World Gynaecologic Oncology Day awareness campaign to assess the impact and reach of the campaign on Twitter and revealed that a single, relevant, memorable, and easy-to-spell hashtag would be the most effective hashtag used in a campaign [21]. Prior research using hashtag analysis to investigate public engagement with discussed health-related topics on social media aims to understand users' health-related behaviour in various health promotion campaigns [22, 23].

This study aims to investigate association rules between used health-related hashtags on Twitter to discover patterns of relationship and connection between hashtags and identify the hashtags that played central roles in original tweets and retweets, quotes, and likes in the American Heart Month campaigns. We reviewed worldwide heart-health promotion campaigns and selected the "American Heart Month" campaign to comply with methodological and sample size requirements. Each February, NHLBI and The Heart Truth® celebrate "American Heart Month" by motivating Americans to adopt healthy lifestyles to prevent heart diseases. The research questions answered in this study are: What is the overall structure of the co-occurrence network of used hashtags in the American Heart Month campaigns? What are the top frequency used hashtags? Which hashtags tend to co-occur most frequently with other hashtags on Twitter? What are the central hashtags that play significant roles in connecting different heart health-related discussions on the Twitter network? Do co-occurring hashtags correlate with user engagement, such as retweets, quotes, and likes?

Method

In this study, we employed a computer programming language and visual techniques for statistical computing and graphics to analyse Twitter data. We leveraged the advantages of applying Association Rules Mining and Co-occurrence Network Analysis to four data subsets (Original tweets, Retweets, Quotes, and Likes) to explore the trends of used hashtags in the conversations on Twitter during the American Heart Month campaigns. The data was analysed in RStudio Version 4.3.1, using R packages and libraries. The methodology flowchart is illustrated in Fig. 1.

Sampling and data collection

The Twitter data used in this study were collected retrospectively for the period between January 2019 and March 2023 using the Twitter REST API. This approach enabled the collection of historical data relevant to the American Heart Month campaigns. Retrospective data collection was chosen to allow for the examination of longitudinal trends and the relationships among hashtags over multiple campaign cycles. To ensure the reliability and relevance of the dataset, specific filtering criteria were applied, such as including only tweets containing campaign-related hashtags (e.g., #AmericanHeartMonth), limiting to English-language tweets, and excluding private or restricted content. This process

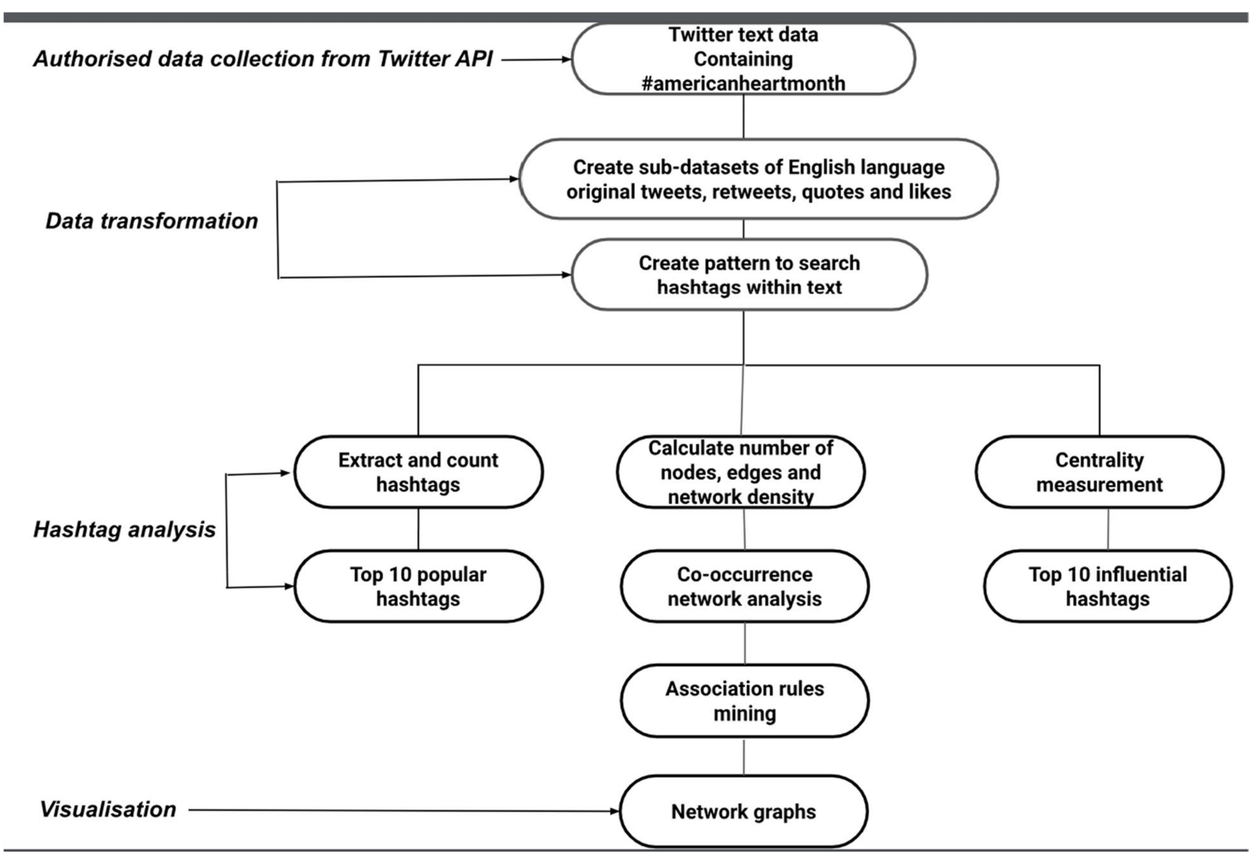

Fig. 1 The proposed framework for hashtag network analysis

provided a robust foundation for conducting co-occurrence network analysis and association rule mining.

The study employed a case study approach, analysing 73,288 tweets from the American Heart Month campaigns between 2019 and 2023. The dataset comprises four subsets: 18,143 original English-language tweets, 42,930 retweets, 2,519 quotes, and 20,846 likes. Data collection performed using the Twitter REST API, which provides access to extensive real-time public data and supports researchers in analysing user activity without requiring informed consent. The data were filtered based on specific criteria to ensure relevance and quality and were collected and processed using RStudio Version 4.3.1 with appropriate R packages and libraries.

Campaign and platform selection

The period from 2019 to 2023 witnessed significant technological advancements and shifts in user behaviour on social media platforms starting with the COVID-19 pandemic, including on Twitter [24]. These changes have notably influenced hashtag usage and engagement in public health campaigns underscoring the importance of examining social media interactions within the context of technological and behavioural shifts during this period [25]. The chosen period captures multiple iterations of the campaign, which likely include variations in public engagement.

Case studies are particularly effective for investigating complex, real-world phenomena within their natural contexts [26]. This campaign was selected due to its consistent use of heart-health related hashtags over a multi-year timeframe, providing a robust foundation for examining engagement patterns and hashtag dynamics. The selection aligns with Yin's criteria for case studies, as it enables an in-depth exploration of public engagement with heart health-related hashtags on social media [26].

The dataset, spanning a five-year period, offers a longitudinal perspective that captures shifts in public engagement and hashtag usage trends. This approach adheres to Yin's [26] recommendation to use diverse and robust datasets to enhance the validity and generalisability of findings. This study underscores the importance of integrating both qualitative and quantitative data in case studies for effective theory-building, a principle that underpins the methodology of this research [27]. This approach is further supported by best practices in social media research for health promotion, as demonstrated in prior studies [21].

Twitter is widely recognised as a supplementary tool for enhancing engagement and strengthening social networks in various health promotion campaigns. Existing research has utilised Twitter as a source of public and searchable data to monitor trends in health-related behaviors [8]. In the United States, where the American Heart Month campaign primarily targets audiences, Twitter's popularity and extensive user base make it a valuable platform for such initiatives. In 2019, the platform reported approximately 330 million monthly active users, with significant growth observed through 2023 (Statista, 2024). This substantial user base underscores Twitter's broad reach and influence, making it a pivotal medium for campaigns like American Heart Month. Additionally, Twitter's dynamic environment fosters active exchanges between users through mechanisms such as retweets, replies, and likes, which facilitate user engagement and the dissemination of health-related messages [28]. Its hashtag system allows for the categorisation and easy retrieval of campaign-related content, enabling the analysis of co-occurrence patterns and relationships between hashtags. Furthermore, Twitter provides excellent accessibility for researchers through its APIs, offering open access to publicly available data [29]. This feature allows for collecting large datasets, such as the 73,288 tweets analysed in this study, providing a robust foundation for conducting co-occurrence network analysis and association rule mining. The ability to process extensive datasets over multiple years enhances the reliability of findings and offers valuable insights into longitudinal trends in hashtag usage. These factors collectively position Twitter as the most suitable platform for analysing the structural relationships among health-related hashtags and their role in engaging the public during the American Heart Month campaigns. Its accessibility, popularity, and real-time interaction features were instrumental in achieving the objectives of this research [30].

The selection of the American Heart Month campaign was based on specific methodological and sample size requirements designed to ensure the robustness and relevance of the analysis. To identify a suitable campaign, we reviewed several worldwide heart-health promotion campaigns that leveraged social media platforms, particularly Twitter. The criteria for selection included: consistent hashtag usage, sufficient tweet volume, relevance to public health and temporal scope. The American Heart Month campaign met all these requirements. It is a well-established annual initiative with consistent social media activity, particularly on Twitter, where hashtags such as #AmericanHeartMonth and #OurHearts were heavily promoted. Its focus on heart health, a critical public health issue, further justified its selection for this study.

Other campaigns were excluded because they did not meet one or more of these criteria. For instance, some campaigns lacked a primary hashtag or had inconsistent usage, resulting in datasets that were too fragmented for meaningful analysis. Others generated an insufficient volume of tweets or did not span a multi-year period, limiting their utility for longitudinal analysis.

Co-occurrence network analysis

Co-occurrence network analysis can provide valuable insights into the content and often identify communities of users discussing similar topics and helps to understand the dynamics of conversations on Twitter [20]. In the co-occurrence network analysis of hashtags, we explored the relationships between hashtags based on their co-occurrence patterns in a dataset. This analysis helps identify which hashtags frequently appear together in the same tweets. Co-occurrence network analysis is frequently applied to Twitter data to uncover relationships between specific words and hashtags that commonly appear together in tweets. This approach has been widely used to study various topics, events, and trends [12, 22–24], particularly in the field of health [31, 32]. We applied a corpus-based co-occurrence network analysis and measured the size of the network, including the number of nodes (hashtags), edges (co-occurrences), and the proportion of actual connections to possible connections (density). We visualised the co-occurrence and relationships of used hashtags. To enhance co-occurrence network analysis, we adjusted the visualisation to make it more informative by filtering the network to extract the most relevant hashtags and calculated centrality measurements to identify influential hashtags.

Co-occurrence Matrix

The co-occurrence matrix M represents the number of times each pair of hashtags appears together in the dataset. If n is the number of hashtags, then the matrix M is an n × n matrix.

$$Mxy = Number\ of\ \ times\ hashtags\ x\ and\ y\ co\ occur$$

Co-occurrence Probability

The co-occurrence probability P(x,y) represents the likelihood that hashtags x and y appear together relative to their individual occurrences. It is normalised by the total number of occurrences.

$$\mathrm{P(x,y)} = \frac{\mathrm{Mxy}}{\text{Total number of occurrences of hashtags x and y}}$$

Co-occurrence network analysis involves detecting hashtags that play a significant role in connecting different hashtags and influencing heart health-related conversations in the Twitter network [25]. We calculated the Eigenvector Centrality for each hashtag (node). Centrality analysis of hashtags measures the influence of a hashtag in the network, considering the influence of its connected hashtags, based on the quality and quantity of connections in the network [26].

$$\text{Centrality (x)} = \frac{1}{\lambda}\ \sum \text{y Adjacency Matrix (x, y) x Centrality (y)}$$

A higher centrality indicates that a hashtag is connected to other highly central hashtags. Visualisation of the network and inspection of the layout can help to identify central hashtags. Central hashtags are likely to be located in the centre of the network.

Association rules mining

We employed association rules mining, a machine learning technique, to discover patterns and relationships between used hashtags within a large dataset collected from Twitter. This approach provides valuable insights into the co-occurrence of hashtags used in five years of American Heart Month campaigns [27]. While we aim to investigate the frequent sets of used hashtags, this information is often presented as a collection of if–then rules, referred to as "Association Rules" [21]. The form of an association rule is {X—> Y}, where {X} is a set of hashtags, and {Y} is a hashtag. This association rule implies that if all the items in {X} appear in some hashtag sets, then {Y} is "likely" to appear in that hashtag set as well [33].

Association measures

In association rules mining, the following key components of metrics are measured to assess the quality and strength of identified association rules. Visualisation aids in the interpretation and communication of complex association rules, helping users understand the relationships between hashtags and make informed decisions based on the mined patterns. The choice of visualisation depends on the specific goals of the analysis and the characteristics of the data. The metrics can be used to evaluate the significance of the relationships between a set of hashtags [2]:

1. Support: measures the frequency of occurrence of each hashtag. Higher support suggests that a hashtag is more common and frequently occurs. The metric support explains how popular a set of hashtags is. In general, for a set of hashtags X, with n = number of all co-occurrences [33]:

$$\text{Support (X)} = \frac{\text{Frequency (X)}}{\text{n}}$$

2. Confidence: measures the likelihood of co-occurrence of two hashtags. The higher confidence values indicate a stronger association between the two hashtags. Confidence explains how likely a hashtag (Y) is occurring given that hashtag (X) has occurred which is expressed as {X—>Y}. It is measured by the proportion of co-occurrences with hashtag (X), in which hashtag (Y) also occurred. The confidence can be interpreted as an estimate of the probability P(Y|X). Support $(X \cup Y)$ means the support of the union of the hashtags in X and Y. The confidence of a rule is defined as [33]:

$$\text{Confidence }(X- > Y)\frac{\text{Support }(X \cup Y)}{\text{Support }(X)} = \frac{P\ (X\ \cap\ Y)}{P\ (X)} = P\ (Y/X)$$

3 Lift: assesses the strength of an association rule. It's the ratio of the observed confidence to the expected confidence if two hashtags were independent. Lift values greater than 1 indicate that the occurrence of a hashtag increases the likelihood of the occurrence of another hashtag, suggesting a positive relationship between the occurrence of hashtags. Lift explains how likely hashtag Y occurs when item X occurs, while controlling for how popular hashtag Y and X are. It measures how many times more often hashtags X and Y occur together than expected if they were independent statistically. Lift measures how many times more often hashtags X and Y occur together than expected if they were independent [33]. A lift value of 1, implies no association between hashtags, if greater than 1 means that hashtag Y is likely to occur if hashtag X has occurred, if less than 1 means that hashtag Y is unlikely to occur if hashtag X has occurred.

$$Lift\ (X- > Y)\ = \frac{P\ (X\ \cap Y)}{P(X)\ \times\ P(Y)}$$

A-Priori Algorithm

The main idea behind the A-Priori Algorithm is to reduce the number of hashtag sets that we examine. We eliminate hashtag sets by selecting smaller sets and detecting that a large set cannot be frequent unless all its subsets are frequent. If a hashtag set is infrequent, then all its subsets must also be infrequent [33]. The list of popular hashtag sets is obtained through these steps. The A-Priori Algorithm is used to identify hashtag sets with high support and hashtag associations with high confidence or lift [2, 21]. We apply the A-Priori Algorithm to a particular set of co-occurrences to find all rules with support greater than or equal to the support threshold and confidence greater than or equal to the confidence threshold. Support and Confidence Thresholds are used to set minimum support and confidence thresholds to filter out uninteresting or weak associations. This helps focus on the most meaningful rules [33].

Relative hashtag frequency

Relative item frequency is one of the key metrics in association rule mining that typically explores the most frequently used hashtags within the collected dataset from Twitter [18]. It measures how often a specific hashtag set occurs relative to the total number of co-occurrences. We set a minimum support threshold to identify association rules involving hashtags that occur frequently enough to be considered relevant within the hashtag set. Relative item frequency is a measure used to assess the significance or prevalence of a specific hashtag within the dataset [28].

$$\text{Relative hashtag Frequency }(A) = \frac{\text{Number of cooccurrences containing hashtag set A}}{\text{Total number of cooccurrences}}$$

Results

Descriptive user contribution analysis

We conducted a descriptive analysis, focusing on the distribution and usage frequency of top used hashtags by individual contributors and organisations during the American Heart Month campaign. The data was categorised by contributors, grouped by hashtags, and summarised to identify key patterns and trends in engagement. The visual representation highlights the relative contributions of top entities and the diversity of hashtags they employed (Fig. 2).

The results indicate that "Contributor 1" actively utilised multiple hashtags, demonstrating a broad engagement strategy and emphasising the role of influencers in disseminating health-related information. Similarly, "Organisation 1" exhibited diverse hashtag usage, likely reflecting its role as a central stakeholder in the campaign. In contrast, entities such as "Organisation 5" primarily focused on specific hashtags (e.g., #HeartMonth), suggesting a targeted messaging approach to engage niche audiences.

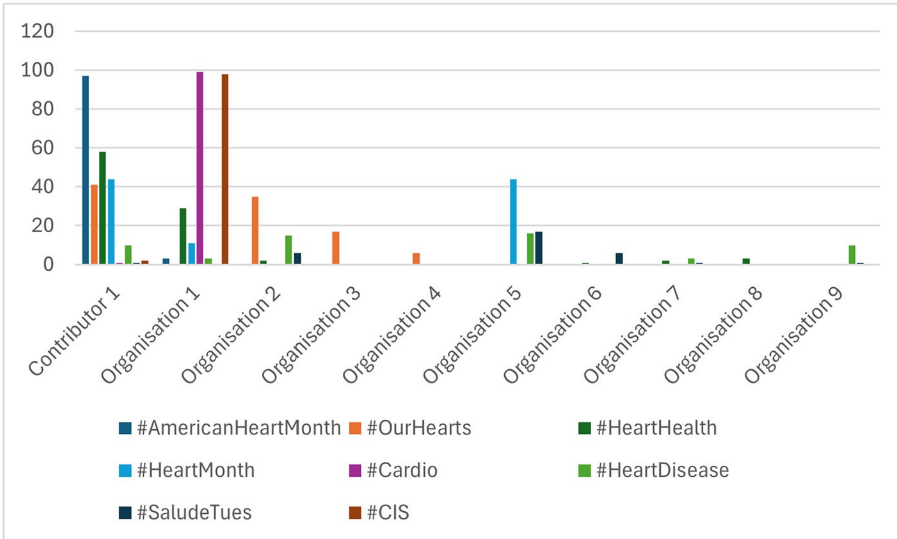

Fig. 2 Hashtag contributions by top contributors and organisations in the American Heart Month Campaign

Co-occurrence network analysis of hashtags

Co-occurrence network analysis is used in studies to uncover the patterns of hashtag networks to track and identify emerging trends in real-time [34]. Co-occurrence relationships (Edges) connect pairs of hashtags that co-occur more frequently with each other in tweets. The results of co-occurrence network analysis of used hashtags in English language original tweets (Fig. 3), Retweets (Fig. 4), Quotes (Fig. 5) and Likes (Fig. 6) represent the complex patterns of relationships between

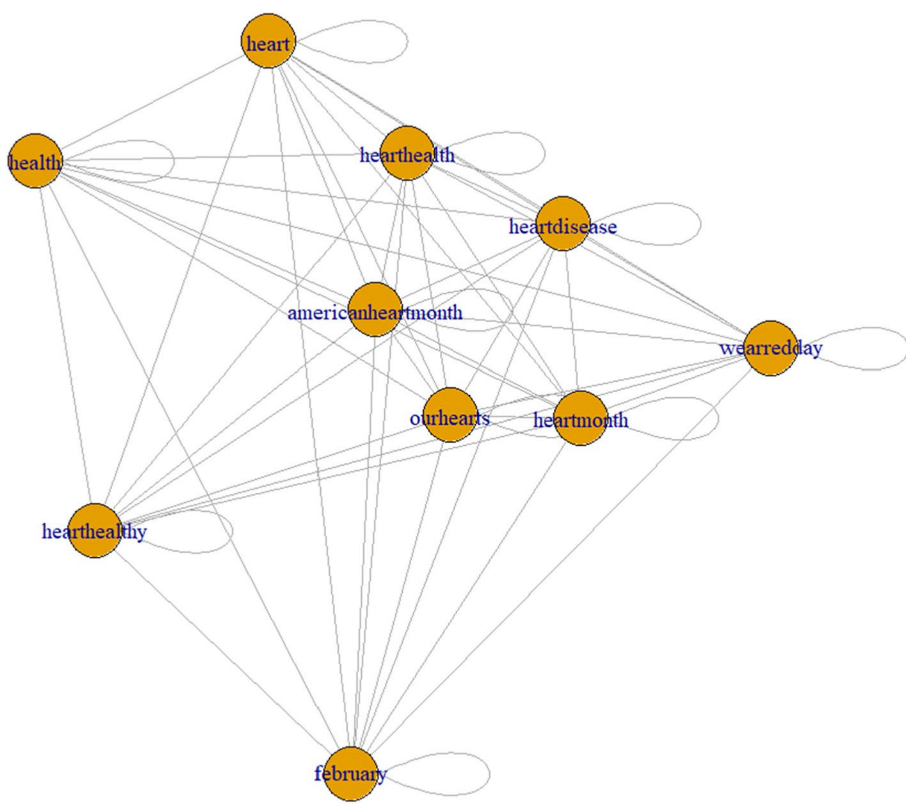

Fig. 3 Visualisation of co-occurrence hashtag network in English language original tweets related to American Heart Month campaigns in the past five years

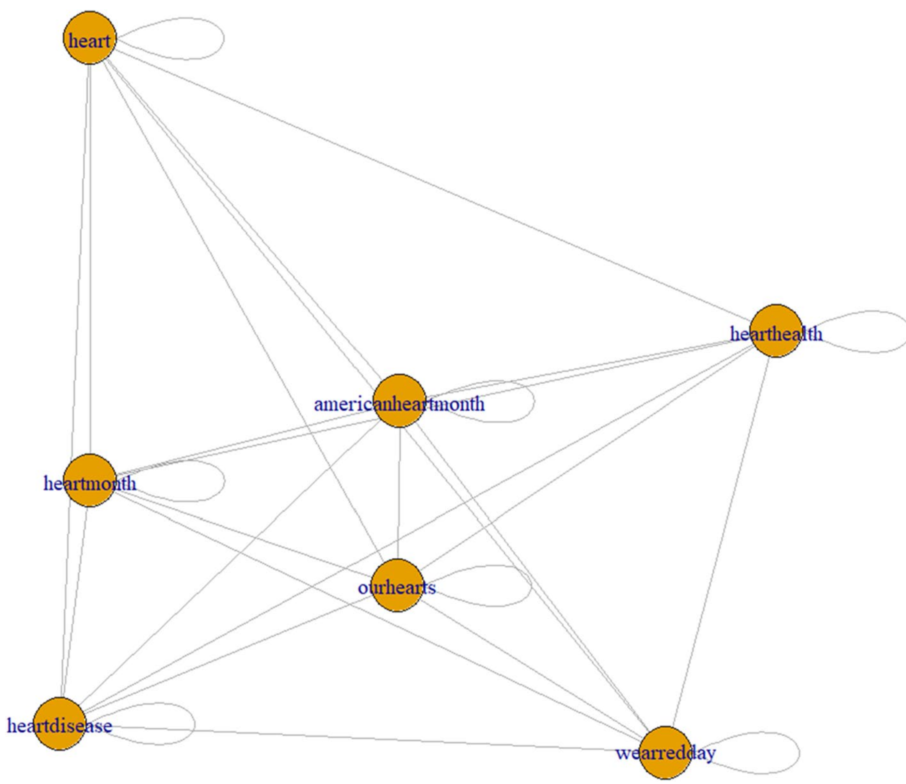

Fig. 4 Visualisation of co-occurrence hashtag network in the English language retweets related to American Heart Month campaigns in the past five years

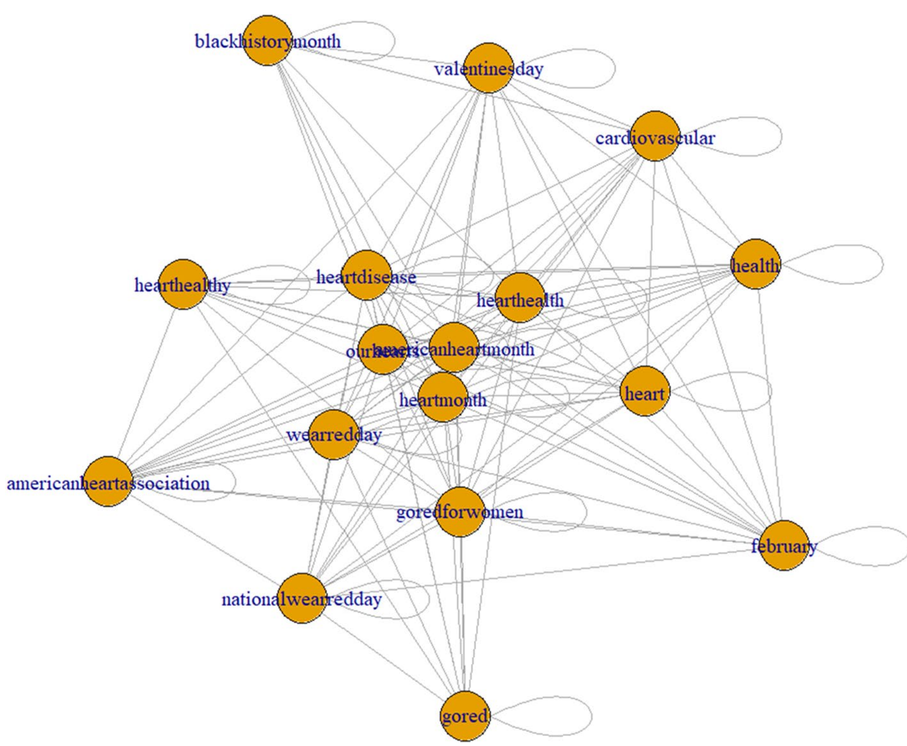

Fig. 5 Visualisation of co-occurrence hashtag network in English language quotes related to American Heart Month campaigns in the past five years

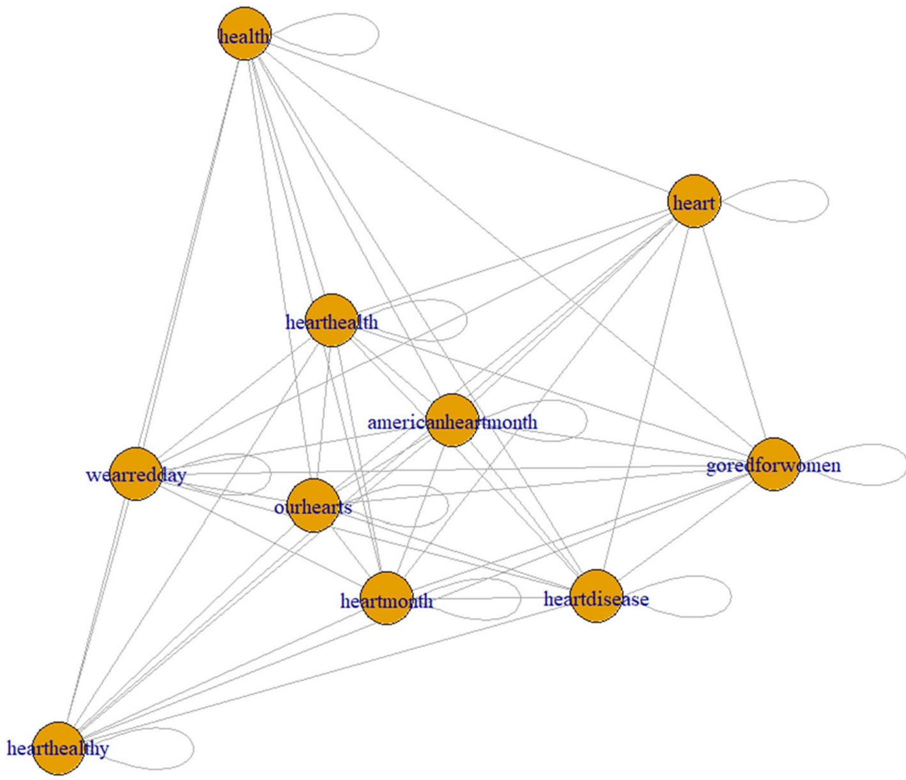

Fig. 6 Visualisation of co-occurrence hashtag network in English language likes related to American Heart Month campaigns in the past five years

the hashtags. Quotes with a minimum number of nodes ($n = 1661$) and edges ($n = 1624$), obtain a maximum network density (118 X 10^{-5}). Likes with a maximum number of nodes ($n = 9101$) and edges ($n = 9004$), obtain a minimum network density (22 X 10^{-5}). #americanheartmonth and #ourhearts play significant roles in all hashtag networks. We summarised metrics such as number of nodes, number of edges and density of each network in Table 1 to provide quantitative measures of the network's structure and characteristics (see Appendix A).

Association rules mining

The results of association rules mining of hashtag networks visualised the rules between hashtag networks to identify the association rules between hashtags in a clear and interpretable manner. We detect the top 5 rules within the hashtag co-occurrence networks with the highest confidence that indicates a strong conditional probability and the highest lift that indicates a significant association between the co-occurred hashtags.

The results of association rules mining in English language original tweets (*Confidence = 0.62—0.83 and Lift = 0.90—2.86*) indicate a strong probability of co-occurrence between (#OurHerats and #Heartmonth) and (#AmericanHeartMonth and #heartdisease, #HeartHealth, #heart) (Fig. 7).

The results of association rules mining in English language retweets (*Confidence = 0.66—0.79 and Lift = 1.78—3.98*) indicate a strong probability of co-occurrence between (#OurHerats and #SaludTues) and (#AmericanHeartMonth and #heartdisease, #HeartHealth, #heart) (Fig. 8).

The results of association rules mining in English language quotes (*Confidence = 0.61—0.99 and Lift = 1.00—2.58*) indicate a strong probability of co-occurrence between (#OurHerats and #SaludTues, #HeartDisease, #HeartMonth, #WearRedDay). Rule 5 stands out as an outlier rule that indicates no co-occurrence with #AmericanHeartMonth within the hashtag network in quotes which requires further investigation. (Fig. 9) The results of association rules mining in English language likes (*Confidence = 0.62—0.72 and Lift = 0.99—2.24*) indicates strong probability of co-occurrence between (#OurHerats and #HeartDisease, #HeartMonth) and (#AmericanHeartMonth and #heartdisease, #HeartHealth) (Fig. 10).

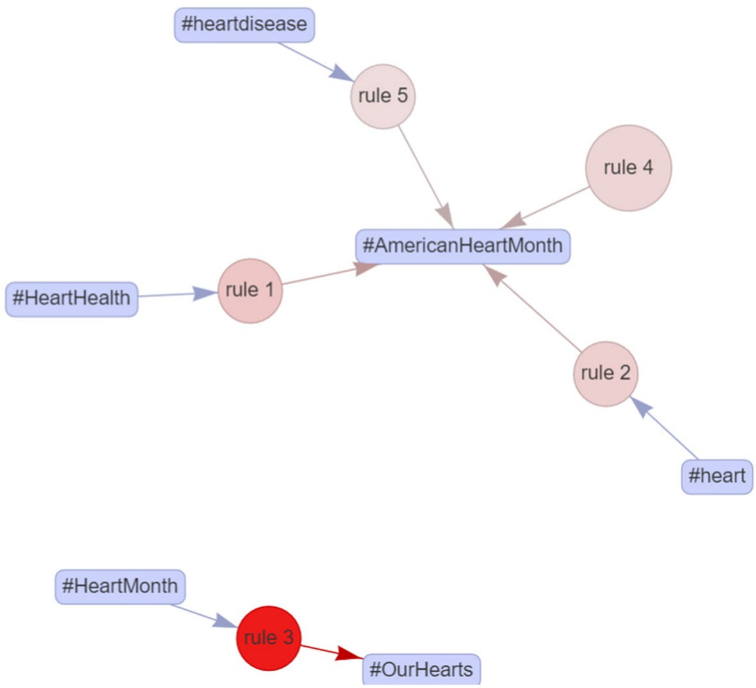

Fig. 7 Visualisation of top five association rules in co-occurrence of hashtags in English language original tweets related to American Heart Month campaigns in the past five years

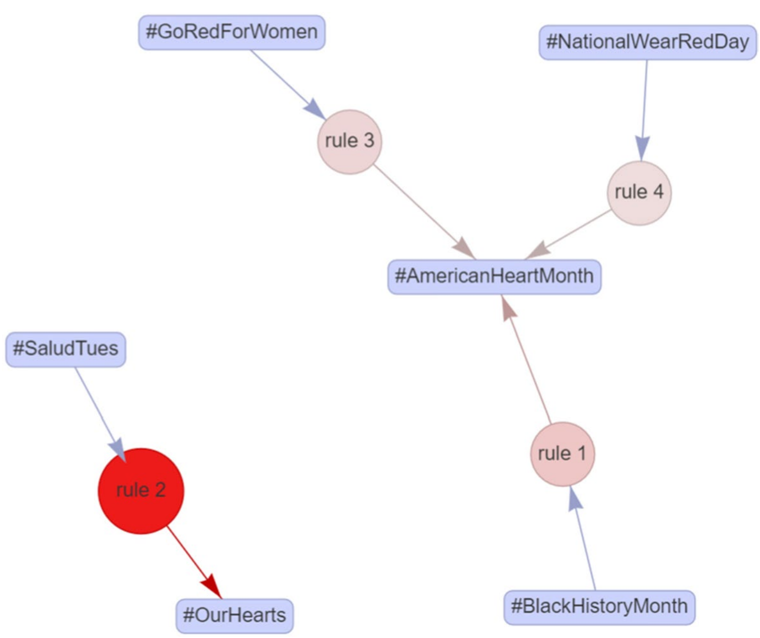

Fig. 8 Visualisation of top five association rules in co-occurrence of hashtags in English language retweets related to American Heart Month campaigns in the past five years

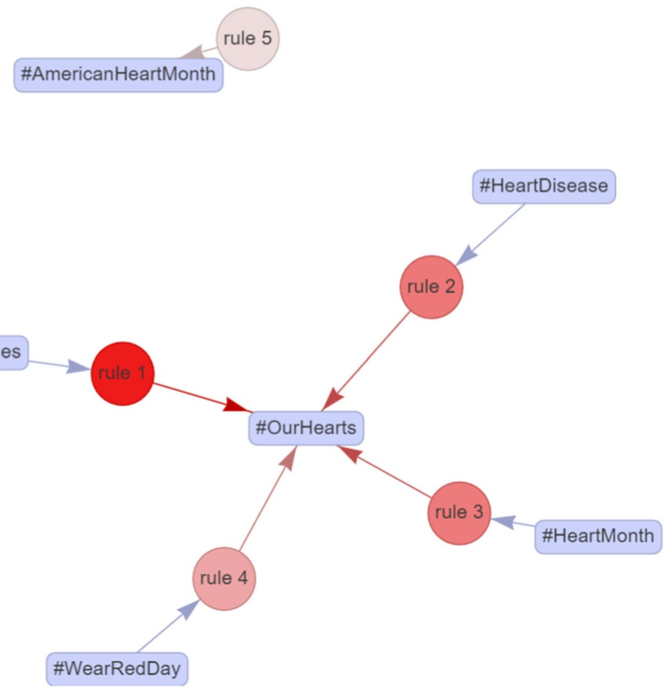

Fig. 9 Visualisation of top five association rules in co-occurrence of hashtags in English language quotes related to American Heart Month campaigns in the past five years

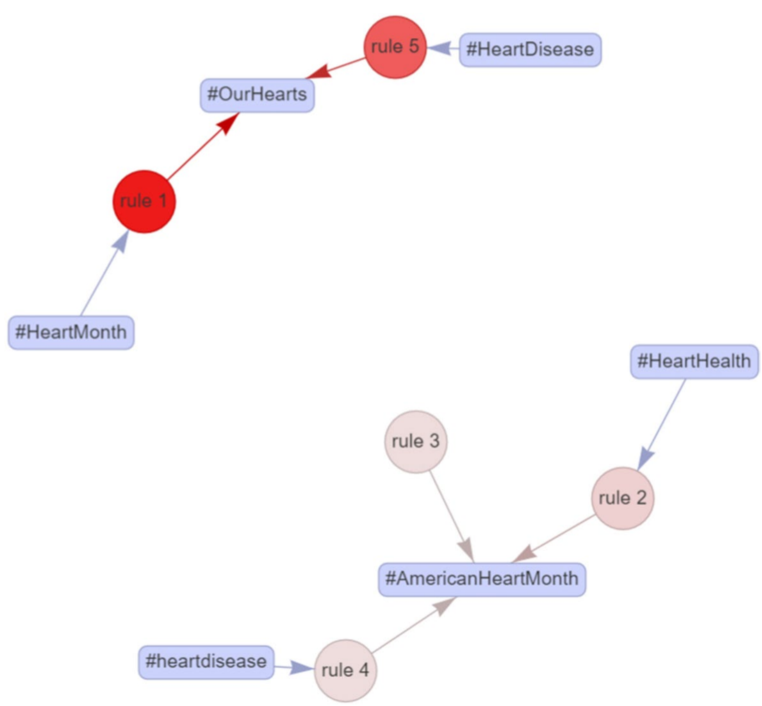

Fig. 10 Visualisation of top five association rules in co-occurrence of hashtags in English language likes related to American Heart Month campaigns in the past five years

We also summarised the measured parameters used to obtain association rules in hashtag networks related to American Heart Month campaigns in the past five years in Table 2 (see Appendix B).

Frequency distribution of hashtag occurrences

In the context of association rule mining, relative frequency often refers to the support of the co-occurrence of a hashtag within the network [28]. The results of the frequency of the top 10 used hashtags in English language original tweets, retweets, likes and quotes identify #AmericanHeart with the highest frequency within the hashtag networks (Fig. 11).

The relative hashtag frequencies in original tweets, retweets, likes, and quotes indicate that #AmericanHeartMonth is relatively common and occurs more frequently than the total occurrences in these datasets. However, #AmericanHeartMonth obtained significant engagement from target audiences by choosing easy-to-remember and readable hashtags in tweets related to the campaign [21]. While #AmericanHeartMonth, #OurHearts and #HeartMonth play a central role in all hashtag networks, the results of associate rules mining indicate a significant association of #OurHearts within hashtag networks. The highest density of hashtags has been observed in the quoted tweets through the introduction of a new range of hashtags such as #GoRedForWomen, #WearRedDay, #HeartDisease and #HeartHealth by the Twitter users which indicates the positive correlation between co-occurring hashtags with users' engagement. The results of quality measurements (Lift > 1) indicate positive relationships between the co-occurred hashtags in the detected association rules in all datasets [28]. (Table 2) (Appendix B).

Discussion

Heart-health organisations need to stand out as credible sources by providing up-to-date and accurate heart-health-related information through their social media channels for sustained engagement of the public in health-related issues and to connect people with similar health concerns and interests [3]. Evidence shows that Twitter health promotion campaigns that promote dedicated hashtags are highly effective in promoting health-related behaviours [29]. Using hashtags in tweets facilitates finding similar health-related topics for users [29]. People interested in health-related topics can simply search and follow tweets with particular hashtags [30, 31]. Promoting specific hashtags can help campaign organisers monitor and track the interaction of people with heart-health-related messages [32]. Comparing the frequency of users' engagement with each hashtag used during the campaign can easily guide the identification of high-performance hashtags to be used to increase engagement in future campaigns [33].

The hashtags analysed in this study were utilised by diverse audiences, including individuals and organisations [32]. A longer time frame can contribute to a more robust dataset for association rule mining. Understanding the historical use of hashtags can guide planning for future campaigns. When conducting association rules mining, the comprehensive dataset derived from a wide time frame enhances the potential for discovering meaningful associations and patterns that may be missed in

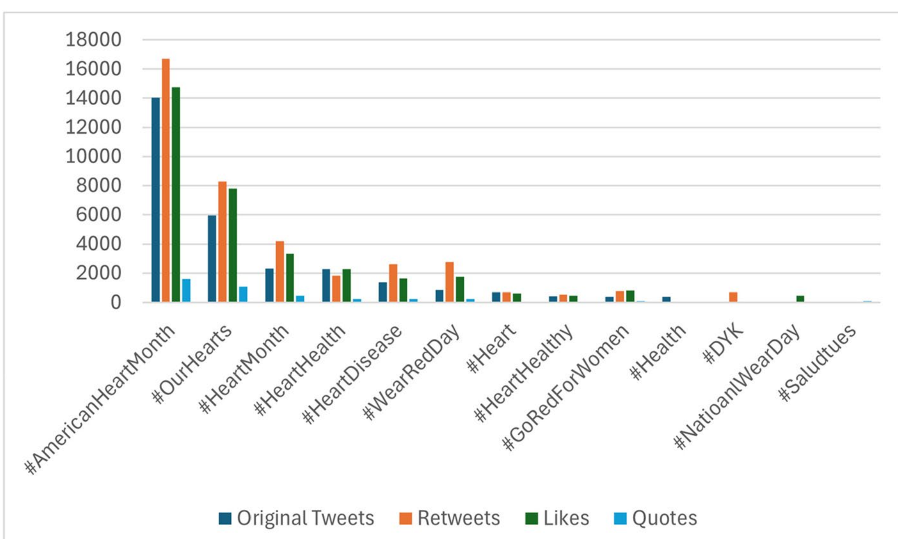

Fig. 11 Frequency distribution of top hashtags across Original Tweets, Retweets, Likes, and Quotes during the campaign in five years

a shorter time. It also provides a more nuanced understanding of the dynamics surrounding American Heart Month campaigns [2]. Insights obtained from association rules mining can guide the selection of effective hashtags, messaging strategies, and targeted outreach for upcoming American Heart Month campaigns. Tracking hashtags over an extended period may reveal shifts in public behaviour, awareness, or sentiment related to heart health [34]. Changes in hashtag usage patterns can be indicative of evolving attitudes and behaviours within the target audience [35, 36].

We analysed the co-occurrence of hashtags in original tweets, retweets, quotes, and likes to compare the variations and differences in the patterns of used hashtags in different datasets. We assume that most original tweets are shared on Twitter by the American Heart Month organisers and retweets, quotes, and likes are the results of users' engagement with heart health-related information. The most important benefit of co-occurrence network analysis of hashtags is to detect hashtags that play a central role in connecting different topics. These influential hashtags often drive the engagement of users on social media platforms to disseminate information and can guide the improvement of social media strategies [37]. Furthermore, researchers can compare and predict the co-occurrence patterns across different periods and user groups to identify variations in the use of hashtags [38]. The analysis of the main contributors in top hashtag usage offers valuable insights into the roles of key stakeholders, such as influencers and organisations, in driving engagement and amplifying campaign messages. These efforts are instrumental in fostering awareness and promoting sustainable health behaviors among target audiences [39].

Data mining techniques such as Association Rule Mining are widely used for extracting and summarising information from social media data that potentially represent topics or themes discussed by social media users [39]. Exploring association rules within hashtag networks is a valuable method to develop health-related content on social media [40]. By identifying associations between hashtags, we can suggest related or complementary hashtags to increase the discoverability and engagement of health-related messages among social media users [41].

Future heart health promotion campaigns need to enhance strategies on social media to achieve a higher rate of reach and engagement of target audiences by choosing simple, relevant, and easy-to-remember hashtags in tweets [21]. The American Heart Month campaigns promoted #OurHearts as the unique campaign hashtag and other hashtags such as #AmericanHeartMonth and #HeartMonth which allows social media users to easily search and follow heart health-related discussions communicated by the campaigns. Observing a range of new hashtags (such as #GoRedForWomen, #WearRedDay, #HeartDisease, and #HeartHealth) in the hashtag co-occurrence network of quoted tweets indicates the positive correlation between co-occurring hashtags with Twitter users' engagement who played a significant role in the dissemination of heart-health related messages among the community. Hashtag recommendations [36] based on association rules mining help discover relevant content and support campaigners to expand content reach and improve users' engagement on social media platforms. To recommend a combination of hashtags to be used during future campaigns, we target the top frequently used hashtags filter the related association rules and select the top 5 rules. #OurHearts indicates a positive association within hashtag networks and is recommended to be used with #AmericanHeartMonth, #HeartMonth, #GoRedForWomen, #HeartDisease, and #HeartHealth in the tweets to increase the engagement rate of users with the heart health-related tweets before, and during, and after the campaigns [36].

Prior studies revealed that engagement with tweets related to a health-related topic can impact users' health-related behaviours [42]. Evidence shows that the higher engagement of users who communicate health-related content on social media actually attempts to change their health-related behaviours in real life [43].

Limitations and future research directions

This study acknowledges several limitations related to technical constraints and tweet collection. Firstly, the data were collected using the Twitter REST API, which relies on specific hashtags for data retrieval. This approach may exclude relevant tweets that do not explicitly use the predefined hashtags, potentially limiting the dataset's comprehensiveness. Future research could integrate additional keyword-based retrieval methods or advanced natural language processing (NLP) techniques to capture a broader range of health-related content.

Secondly, the study focuses solely on Twitter, which, while valuable, represents only one social media platform. User engagement patterns and health-related behaviours may vary significantly across platforms. Future studies could adopt a cross-platform perspective, incorporating other social media platforms to provide a more holistic understanding of social media's role in health promotion. Thirdly, while this study measures user engagement through metrics such as retweets, likes, and quotes, these do not necessarily reflect real-life behavioural changes. Future research could explore methodologies that link social media interactions with tangible health outcomes, such as surveys or interventions that track behavioural changes influenced by health campaigns.

Additional limitations include the reliance on the Streaming Twitter API, which collects data based on specific hashtags but does not use random sampling. This could limit the representativeness of the dataset and the generalisability of the findings [44]. Moreover, the sample is restricted to publicly available tweets, excluding private data that may provide further insights. Twitter users themselves are not necessarily representative of the broader population [45], and metrics such as retweets and likes do not fully capture users' real-life behaviours [13]. The accuracy, validity, and credibility of information disseminated on Twitter remain critical concerns when analysing social media data. The longer data collection period provides a more comprehensive dataset, enabling the capture of variations and trends in hashtag usage over multiple campaigns. While this contributes to a broader analysis, the reliability of the patterns observed in association rule mining depends not only on the duration of data collection but also on factors such as dataset diversity, volume, and the analytical thresholds applied (e.g., support, confidence, and lift). Future research could validate these patterns through additional datasets or by testing them in real-world contexts [45]. Furthermore, expanding data mining techniques, such as incorporating machine learning models, could help predict user behaviour with greater precision and improve the validation of human behaviour analysis [46]. Addressing these limitations in future research will enhance the generalisability, reliability, and practical impact of findings in the domain of social media-based health promotion.

The findings of this study highlight the potential for historical hashtag usage to inform future campaign strategies by identifying patterns of co-occurrence and centrality in hashtag networks. While our analysis focused on the relationships between hashtags and their engagement patterns, the deliberate orchestration of hashtag relationships by campaign organisers or specific ties of individual participants to campaigns were not explicitly explored. These aspects represent significant opportunities for further research. Future campaigns could benefit from understanding whether certain hashtag relationships are strategically planned by organisers or emerge organically through participant contributions. For example, examining campaign planning documents or conducting interviews with organisers could provide insights into the intentional use of hashtags to target specific audiences or amplify campaign messages. Additionally, our findings suggest that monitoring historical hashtag performance, such as the success of specific combinations of hashtags in driving engagement, could support campaign preparation. Organisers might use this information to optimise the selection and pairing of hashtags, ensuring they resonate with the target audience while fostering broader reach. We recommend that future studies incorporate mixed-method approaches, combining quantitative hashtag analysis with qualitative insights from campaign organisers and participants. This would provide a more comprehensive understanding of how hashtag strategies influence engagement and campaign outcomes.

While this study provides valuable insights into the structural relationships among hashtags and their co-occurrence patterns, it does not include a detailed analysis of user behaviour or demographics. This omission restricts the ability to fully understand the diversity of audiences engaging with the campaign or the specific strategies that may have influenced user behaviour.

Future research could address this gap by incorporating user-centric analyses, such as examining user demographics, behaviour patterns, or sentiment.

Conclusion

By employing co-occurrence network analysis and association rule mining, this study highlights the structural relationships between hashtags and identifies those central to driving engagement. Hashtags such as #AmericanHeartMonth and #OurHearts consistently played pivotal roles in connecting users and amplifying campaign messages. These findings offer practical implications for future public health campaigns. Campaign organisers can leverage these insights to design hashtag strategies that enhance visibility and engagement. For example, combining high-impact hashtags identified in this study with emerging ones could foster greater audience interaction and message dissemination. Furthermore, understanding co-occurrence patterns can help organisers target specific user groups and optimise their outreach efforts. While this study provides valuable technical insights, it also acknowledges its limitations in exploring the motivations and demographics of users driving hashtag engagement. Future research should incorporate user-centric analyses, such as demographic profiling, sentiment analysis, or qualitative approaches, to deepen our understanding of audience behaviours and their implications for campaign design.

In conclusion, this research contributes to the field of social media-based health promotion by offering a structured approach to understanding hashtag dynamics. These findings serve as a foundation for refining health campaign strategies, bridging technical insights with actionable outcomes, and fostering broader public engagement with health-related content. Specifically, this study provides insight into the use of Twitter in health promotion interventions to amplify the dissemination of health-related messages to a broader audience. By evaluating the performance of the American Heart Month campaign, this research extends its principles to the domain of health promotion, identifying opportunities for using social media in interventions.

This study exemplifies how health promotion campaigns can benefit from Twitter's networking capabilities as a highly efficient online environment to disseminate credible health-related information and attract a larger audience. The visual techniques and analytical methods used here can be extended by researchers employing new technologies to collect, analyse, and build predictive models with real-time data extracted from social media, seeking patterns in public health-related behaviour [47]. Public health practitioners can use these findings to enhance the performance of health promotion campaigns, achieving a sustained impact on behaviour change. The results also provide guidance for amplifying the dissemination of health information, increasing reach, and improving engagement with a broader audience. Public health organisations must evaluate the performance of campaigns to optimise the sharing of health-related information on social media and maximise user engagement.

Although social media provides opportunities to engage users with health-related messages, this interaction is not one-sided; further research is required to understand the key actors disseminating campaign messages, the viewers, and the individuals engaging with heart health-related content [19]. Future campaigns must enhance social media strategies to facilitate meaningful interactions among Twitter users, increase engagement with heart health-related topics, and sustain interest before, during, and after campaigns [48]. Sustained engagement can be achieved by developing comprehensive social media strategies, such as adopting specific heart health-related hashtags, to promote heart health behaviours and establish lasting habits that reduce the burden of morbidity and mortality from cardiovascular diseases [49, 50].

Further research should aim to predict and evaluate actual health-related behaviour changes influenced by the American Heart Month campaigns [20]. Future studies should also focus on assessing the effectiveness of health-related messages on individuals' lifestyles, as well as the credibility and usefulness of information disseminated by heart health-promotion organisations. Our team plans to conduct further research to uncover hidden semantic structures in the text retrieved from Twitter using topic modelling to identify the dominant themes discussed during the American Heart Month campaign [49, 50].

Ethical Considerations

This study collected data from users who consented in the privacy policy on Twitter to disclose their data. Therefore, no ethical approval was required from our institution as we did not interfere with any human data in measuring the internet activity of Twitter users, and the users' accounts remained anonymous [51].

Appendix A

Summary of Key Metric Measurements

Table 1 Summary of key metrics for co-occurrence network analysis of hashtags related to the American Heart Month campaigns in the past five years, including influential hashtags

Type of tweet	Number Nodes (hashtags)	Number of edges (co-occurrences)	Density of network X10^{-5}	Influential hashtags
Original Tweets	7736	7642	25	# americanheartmonth # ourhearts # heartmonth
Retweets	5899	5781	33	# americanheartmonth # ourhearts
Quotes	1661	1624	118	# americanheartmonth # ourhearts # heartmonth # hearthealth # goredforwomen # wearredday # heartdisease
Likes	9101	9004	22	# americanheartmonth # ourhearts # heartmonth

Appendix B

Parameters used for association rules and quality measures

Table 2 Grid search parameters used to obtain association rules andquality measures in the likelihood of co-occurrence in hashtag networks related to American Heart Month campaigns in the past five years

Type of tweets	Support value Min Max	Confidence value Min Max	Final support threshold	Final confidence threshold	Lift Min Max	Number of Rules
Original tweets	*0.022 0.68*	*0.62 0.83*	*0.02*	*0.6*	*0.90 2.86*	*5*
Retweets	*0.0064 0.0085*	*0.66 0.79*	*0.006*	*0.6*	*1.78 3.98*	*5*
Quotes	*0.031 0.61*	*0.61 0.99*	*0.03*	*0.6*	*1.00 2.58*	*5*
Likes	*0.023 0.65*	*0.62 0.72*	*0.02*	*0.6*	*0.99 2.24*	*5*

Acknowledgements
All authors express their sincere gratitude to the reviewers and the co-editors for the constructive advice.

Authors' contributions
AGH, MP and MDC contributed to the interpretation and analysis of the findings as well as the development of the manuscript. All authors read and approved the final manuscript.

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability
The dataset for this study was obtained from publicly accessible posts on Twitter via the REST API. Details of data collection and preprocessing are provided in the paper. Due to restrictions imposed by Twitter's terms of use, the raw datasets cannot be shared publicly.

Declarations

Consent for publication
The data used in this study were collected from publicly accessible posts on Twitter in compliance with Twitter's terms of use. No identifiable personal data were collected or used, and all data have been anonymised to ensure privacy. As the data are publicly available, explicit consent to publish is not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Institute for Sustainable Industries & Liveable Cities, Business School, Victoria University, 370 Little Lonsdale St, Melbourne VIC 3000, Australia. [2]Victoria University, Institute for Health and Sport and Australian Health Policy Collaboration, Melbourne, Australia.

Received: 13 December 2023 Accepted: 30 December 2024
Published online: 07 January 2025

References

1. www.who.int. (n.d.). Cardiovascular diseases (CVDs) Fact Sheet. [online] Available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
2. Leskovec J, Rajaraman A, Ullman JD. Mining of massive data sets. Cambridge university press; 2020 Jan 9.
3. Park H, Reber BH, Chon MG. Tweeting as health communication: health organizations' use of Twitter for health promotion and public engagement. J Health Commun. 2016;21(2):188–98.
4. Yom-Tov E, Shembekar J, Barclay S, Muennig P. The effectiveness of public health advertisements to promote health: a randomized-controlled trial on 794,000 participants. NPJ digital medicine. 2018;1(1):24.
5. Grajales FJ III, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res. 2014;16(2):e2912.
6. Zhao X. Health communication campaigns: A brief introduction and call for dialogue. Int J Nurs Sci. 2020;10(7):S11–5.
7. Highfield T, Leaver T. A methodology for mapping Instagram hashtags. First Monday. 2015;20(1):1–1.
8. Ghahramani A, de Courten M, Prokofieva M. The potential of social media in health promotion beyond creating awareness: an integrative review. BMC Public Health. 2022;22(1):1–3.
9. Appel G, Grewal L, Hadi R, Stephen AT. The future of social media in marketing. J Acad Mark Sci. 2020;48(1):79–95.
10. Fudolig MI, Alshaabi T, Arnold MV, Danforth CM, Dodds PS. Sentiment and structure in word co-occurrence networks on Twitter. Appl Netw Sci. 2022;7(1):1–27.
11. Vrontis D, Makrides A, Christofi M, Thrassou A. Social media influencer marketing: A systematic review, integrative framework and future research agenda. Int J Consum Stud. 2021;45(4):617–44.
12. Ju R, Jia M, Cheng J. Promoting mental health on social media: A content analysis of organizational Tweets. Health Commun. 2023;38(8):1540–9.
13. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res. 2021;23(5):e17917.
14. Kim HM, Saffer AJ, Liu W, Sun J, Li Y, Zhen L, Yang A. How public health agencies break through COVID-19 conversations: A strategic network approach to public engagement. Health Commun. 2022;37(10):1276–84.
15. Omena JJ, Rabello ET, Mintz AG. Digital methods for hashtag engagement research. Social Media+ Society. 2020;6(3):2056305120940697.
16. Engelmann I, Kloss A, Neuberger C, Brockmann T. Visibility through information sharing: The role of tweet authors and communication styles in retweeting political information on Twitter. Int J Commun. 2019;12(13):20.
17. Nerghes A, Mulder B, Lee JS. Dissemination or participation? Exploring scientists' definitions and science communication goals in the Netherlands. PLoS ONE. 2022;17(12):e0277677.
18. Singla RK, De R, Efferth T, Mezzetti B, Uddin MS, Ntie-Kang F, Wang D, Schultz F, Kharat KR, Devkota HP, Battino M. The International Natural Product Sciences Taskforce (INPST) and the power of Twitter networking exemplified through# INPST hashtag analysis. Phytomedicine. 2023;1(108):154520.
19. Lawes-Wickwar S, Ghio D, Tang MY, Keyworth C, Stanescu S, Westbrook J, Jenkinson E, Kassianos A, Scanlan D, Laidlaw L, Howlett N. A rapid systematic review of public responses to health messages encouraging vaccination against infectious diseases in a pandemic or epidemic.
20. Fuster-Casanovas A, Das R, Vidal-Alaball J, Lopez Segui F, Ahmed W. The# VaccinesWork hashtag on Twitter in the context of the COVID-19 pandemic: network analysis. JMIR Public Health Surveill. 2022;8(10):e38153.
21. Uwins C, Yilmaz Y, Bilir E, Bhandoria GP. World Gynecologic Oncology Day: the use of Twitter to raise awareness of gynecologic cancers. AJOG Global Reports. 2022;2(3):100079.
22. Si Y. Public Discourse on SDG7: Hashtag Co-Occurrence Networks and Social Justice Messaging on Twitter. InAcademy of Management Proceedings 2023 (Vol. 2023, No. 1, p. 15412). Briarcliff Manor, NY 10510: Academy of Management.
23. Kondo M, Nishikawa S, Uchida O, Utsu K. Text analysis of tweets with rescue request hashtags posted during 2018 Japan floods. In2019

International Conference on Information and Communication Technologies for Disaster Management (ICT-DM) 2019 Dec 18 (pp. 1–2). IEEE.
24. Stokel-Walker C. Twitter changed science-what happens now it's in turmoil? Nature. 2023;613(7942):19–21.
25. Nawaz FA, Riaz MM, Tsagkaris C, Faisal UH, Klager E, Kletecka-Pulker M, Kimberger O, Willschke H, Khan N, Sultan MA, Atanasov AG. Impact of# PsychTwitter in promoting global psychiatry: a hashtag analysis study. Front Public Health. 2023;22(11):1065368.
26. Hollweck T. Robert K. Yin.(2014). Case Study Research Design and Methods . Canadian Journal of Program Evaluation. 2015;30(1):108–10.
27. Eisenhardt KM. Building theories from case study research. Acad Manag Rev. 1989;14(4):532–50.
28. Aldamen Y, Abdallah A. Dialogical Health Communication via Twitter (X) During COVID-19 in African Countries: Ghana as a Case Study. World. 2024;5(4):962–80.
29. Chen K, Duan Z, Yang S. Twitter as research data: Tools, costs, skill sets, and lessons learned. Politics Life Sci. 2022;41(1):114–30.
30. Xu WW, Chiu IH, Chen Y, Mukherjee T. Twitter hashtags for health: applying network and content analyses to understand the health knowledge sharing in a Twitter-based community of practice. Qual Quant. 2015;49:1361–80.
31. Puerta P, Laguna L, Vidal L, Ares G, Fiszman S, Tarrega A. Co-occurrence networks of Twitter content after manual or automatic processing. A case-study on "gluten-free". Food Quality and Preference. 2020;86:103993.
32. Loukianov A, Burningham K, Jackson T. The patterning of the discursive space in search for the# goodlife: A network analysis of the co-occurrence of Instagram hashtags. Inf Soc. 2023;39(1):62–78.
33. Jan Kirenz 2020, Introduction to Association Rule Mining in R, Jan Kirenz, viewed 19 December 2023, <https://www.kirenz.com/post/2020-05-14-r-association-rule-mining/>
34. Lane JM, Habib D, Curtis B. Linguistic methodologies to surveil the leading causes of mortality: scoping review of Twitter for public health data. J Med Internet Res. 2023;12(25):e39484.
35. Cheng IK, Heyl J, Lad N, Facini G, Grout Z. Evaluation of Twitter data for an emerging crisis: an application to the first wave of COVID-19 in the UK. Sci Rep. 2021;11(1):19009.
36. Chakrabarti P, Malvi E, Bansal S, Kumar N. Hashtag recommendation for enhancing the popularity of social media posts. Soc Netw Anal Min. 2023;13(1):21.
37. Cruickshank IJ, Carley KM. Characterizing communities of hashtag usage on twitter during the 2020 COVID-19 pandemic by multi-view clustering. Appl Netw Sci. 2020;5:1–40.
38. Kumar N, Qiu L, Kumar S. A hashtag is worth a thousand words: An empirical investigation of social media strategies in trademarking hashtags. Inf Syst Res. 2022;33(4):1403–27.
39. Gupta S, Dash SB, Mahajan R. The role of social influencers for effective public health communication. Online Inf Rev. 2022;46(5):974–92.
40. Katz MS, Anderson PF, Thompson MA, Salmi L, Freeman-Daily J, Utengen A, Dizon DS, Blotner C, Cooke DT, Sparacio D, Staley AC. Organizing online health content: developing hashtag collections for healthier internet-based people and communities. JCO Clin Cancer Inform. 2019;3:1.
41. Bardus M, El Rassi R, Chahrour M, Akl EW, Raslan AS, Meho LI, Akl EA. The use of social media to increase the impact of health research: systematic review. J Med Internet Res. 2020;22(7):e15607.
42. Campi A, Palese C. Twitter Association Rule Mining using Clustering and Graph Databases. In2021 the 5th International Conference on Information System and Data Mining 2021 May 27 (pp. 90–95).
43. Neiger BL, Thackeray R, Van Wagenen SA, Hanson CL, West JH, Barnes MD, Fagen MC. Use of social media in health promotion: purposes, key performance indicators, and evaluation metrics. Health Promot Pract. 2012;13(2):159–64.
44. Joseph K, Landwehr PM, Carley KM. Two 1% s don't make a whole: Comparing simultaneous samples from Twitter's streaming API. InSocial Computing, Behavioral-Cultural Modeling and Prediction: 7th International Conference, SBP 2014, Washington, DC, USA, April 1–4, 2014. Proceedings 7 2014 (pp. 75–83). Springer International Publishing.
45. Kapoor KK, Tamilmani K, Rana NP, Patil P, Dwivedi YK, Nerur S. Advances in social media research: Past, present and future. Inf Syst Front. 2018;20:531–58.
46. Reyes-Menendez A, Saura JR, Alvarez-Alonso C. Understanding# WorldEnvironmentDay user opinions in Twitter: A topic-based sentiment analysis approach. Int J Environ Res Public Health. 2018;15(11):2537.
47. Watson, K., 2019. Predictive Analytics in Health Care: Emerging Value and Risks. Deloitte Development LLC: New York, NY, USA.
48. Gonsalves CA, McGannon KR, Schinke RJ. Social media for health promotion: What messages are women receiving about cardiovascular disease risk by the Canadian Heart and Stroke Foundation? J Health Psychol. 2020;25(8):1017–29.
49. Gupta V, Hewett R. Real-time tweet analytics using hybrid hashtags on twitter big data streams. Information. 2020;11(7):341.
50. Ozemek C, Laddu DR, Lavie CJ, Claeys H, Kaminsky LA, Ross R, Wisloff U, Arena R, Blair SN. An update on the role of cardiorespiratory fitness, structured exercise and lifestyle physical activity in preventing cardiovascular disease and health risk. Prog Cardiovasc Dis. 2018;61(5–6):484–90.
51. Ahmed W, Bath PA, Demartini G. Using Twitter as a data source: An overview of ethical, legal, and methodological challenges. The ethics of online research. 2017;12(2):79–107.

Publisher's Note