Social Network Analysis and Mining (2024) 14:238
https://doi.org/10.1007/s13278-024-01404-9

ORIGINAL ARTICLE

Social network and linguistic analysis of the #nutrition discourse on the social network platform X, formerly known as Twitter

Cassandra H. Ellis[1,3] · J. Bernadette Moore[1] · Peter Ho[1] · Wasim Ahmed[2] · Charlotte E. L. Evans[1]

Received: 21 August 2024 / Revised: 26 November 2024 / Accepted: 12 December 2024 / Published online: 20 December 2024

Abstract

Social network analysis (SNA) of social media content allows information transfer to be visualised, identifies influential actors, and reveals public opinion. However, to date no research has investigated content related to nutrition on X. This study examined the #nutrition conversations on X (formerly Twitter) utilising SNA and linguistic methods. NodeXL Pro was used for network, semantic and sentiment analyses on English language posts including '#nutrition' collected between 1 and 21 March 2023. The #nutrition network included 17,129 vertices (users) with 26,809 edges (relationships). NodeXL Pro was used to assess the structure of the network and the actors involved by calculating the network metrics. The results show a low density, dispersed network (graph density = 0.001) with most users communicating heavily with a small number of other users. These subgroup community cluster structures restrict information flow outside of the subgroups (modularity = 0.79). These network structures rely on influential users to share information (betweenness centrality range, 0 to 23,375,544). Notably, influential users were typically from both personal and not-for-profit accounts. Semantic analysis identified 97,000 word-pair edges with the most frequently discussed topics related to *health, healthy lifestyle and diet,* with a positive sentiment found across the network. By using SNA, semantic, and sentiment analyses, this study found a dispersed X network with a high proportion of unconnected users who did not have relationship with other users in the network. The findings reveal a publicly driven debate focused on healthy diets and lifestyle, with information primarily propagated through reposting.

1 Introduction

Social media allows people to connect with each other, share information and build networks. With many people engaging with social media daily, the spread of information is ubiquitous and rapid (Toraman et al. 2022). Networks are an important aspect of social media, as virtual communities can be built creating a space to share news and information relating to a common interest, and followers able to follow conversations using hashtags (Erz et al. 2018). Hashtags permit networks to be built with brands, organisations, and influencers; and such virtual communities have changed the way the public access health and nutrition information (Eaton et al. 2023). Therefore, examining social media can provide unique insights into the nutrition and diet information, reaching and influencing large segments of the general population (Harris et al. 2014).

The social media platform X (formerly Twitter) has enabled the creation of networks that go beyond sharing news, and consequently, has been used to freely and quickly engage in advocacy and lobbying (Hunt 2021). Social network analysis (SNA) facilitates investigation of such networks by applying mathematical network and graph theory to visualise information transfer as relational networks of connected nodes. The theory posits that individuals are part of an extensive network of social interactions that can be visualised as a flow of behaviour, influence, or ideas (Wasserman 1994). Measuring node connectivity (centrality) permits the identification of 'influencers' within the network and provides insight into individuals' motivations to engage with other users and share content.

SNA has been utilised to examine wide range of topics including political views (Chakraborty and Mukherjee 2023), gender difference in climate activism (Holmberg and Hellsten 2015), and the impact of food poverty (Eskandari

✉ Cassandra H. Ellis
fsce@leeds.ac.uk

[1] School of Food Science and Nutrition, University of Leeds, Leeds LS2 9JT, UK

[2] Hull University Business School, University of Hull, Hull HU6 7RX, UK

[3] The Nutrition Society, London W6 7NJ, UK

et al. 2022). Previous research has also considered the impact of X networks in public health promotion, in the context of promoting breast feeding practices (Moukarzel et al. 2021), disease outbreaks (Vijaykumar et al. 2018), and vaccination uptake (Yousef et al. 2022). SNA has also been a useful tool to monitor and understand public opinion on policy debates, including front of pack labelling (Septia Irawan et al. 2022), mask wearing to stop the transmission of infectious disease (Ahmed et al. 2020), and fiscal policies (Bridge et al. 2021). In addition to SNA, sentiment analysis of online discourse is also a useful tool as it provides insight into public opinion. For example, investigation into COVID-19 vaccination found a predominately negative discussion (66%) (Yousef et al. 2022), which increased vaccine hesitancy (Sussman et al. 2023). However, to our knowledge, no research has specifically examined nutrition networks and the sentiment of the discourse leaving a gap in the literature. It is important to understand how the general public are engaging with nutrition content as the evidence from related health subjects suggests non-experts are leading the debate (Engel et al. 2024). This is problematic as it perpetuates misinformation, negatively impacting dietary choices (Vijaykumar et al. 2021). By developing a more complete understanding of who is leading the debate, the type of content that gains attention, and the language popular posts use, this research will add to the wider body of literature, and we hope empower nutritionists to lead the debate with evidence-based nutrition information.

1.1 Influencers

Influencers have become a key component of social media networks creating personas that generate trust and the illusion of an interpersonal relationship. Research has identified a variety of reasons to explain the following of influencers. These include increasing knowledge (Alhothali and Aljefree 2023), self-improvement (Aaminah Zaman et al. 2023), and entertainment (Croes and Bartels 2021). Notably, X users that discuss healthy food have been found to be strongly influenced by influencers (Pilař et al. 2021). Conversely, while dietitians who use social media to communicate dietary advice want to have influence, they do not consider themselves to be influencers as they feel influencers lack authenticity (Marauri-Castillo et al. 2024). Yet research by Lee and colleagues found authenticity one of the key reasons people follow influencers (Lee et al. 2022), and the 'influencer' community was the second largest community in a social network analysis on discussions on vegan food (Pilař et al. 2021). However, more recent research analysing the motivation to engage on X, found that semantic content was more important than the author (Appel et al. 2020). While SNA allows the identification of key influencers in a network and to measure the level of their influence, semantic and sentiment analysis facilitates an understanding of the content shared.

1.2 Aims and objectives

This study builds on our previous research (Ellis et al. 2024) that investigated the quality of nutrition information available online and shared on Twitter (now known as X). Specifically, the results showed that poorer quality nutrition information was more likely to be re-shared and highlighted the importance of understanding user networks to predict motivations of sharing (Ellis et al. 2024). To our knowledge, previous work examining nutrition information shared on X have been limited to content analysis with no research to date that has assessed the flow of information through social networks. While more focussed SNA has been used to investigate sugar tax (Bridge et al. 2021), breastfeeding practices (Moukarzel et al. 2020), and vegan diets (Pilař et al. 2021), none have examined the broader topic of nutrition. Therefore, using SNA, semantic and sentiment analyses, along with data visualisation, the aims of this study were to: 1. investigate the characteristics of the #nutrition network, 2. identify the influential actors and characterise how they influence the flow of information, 3. investigate how nutrition debates are portrayed on X, 4. identify the most frequently discussed topics in the #nutrtion network, and 5. measure the sentiment of the #nutrition network.

2 Methods

X was selected for this study as it was designed to encourage information and opinion sharing. In addition, X is also used for professional networking and used by NGOs, companies and organisations, as well as personal users, to disseminate information, thereby is a rich source of data for researchers. Lastly, when the study was initiated, real time network information could be collected. It is important to note that the data collected for this study was collected prior to July 2023 before the rebranding of Twitter to X, but for consistency we refer herein to the platform as X.

2.1 Data retrieval

Posts were directly imported, analysed and visualised using the Microsoft Excel plugin, NodeXL Pro (Network Overview for Discovery and Exploration in Excel; version 1.0.1.510) (Smith et al. 2010). The Twitter Search application programming interface (API, v1.1), as it was known then before the rebrand to X, was used to gather data. NodeXL Pro has been used previously to explore social media networks across a range of research including COVID-19 (Eskandari et al. 2022), climate change (Yuan et al. 2024), and professional health networks (Probst and Peng 2019). The NodeXL Pro was selected over currently available open-source software as it allows for advanced network analysis as well as detailed content, semantic, and sentiment analyses.

All English language posts, and associated metadata (likes, reposts and mentions) that included #nutrition, or were posted in response to a post that included the hashtag, were strategically collected daily in longitudinal fashion from 2023-03-01 00:00:00 to 2023-03-21 23:59:59 (Greenwich Mean Time). Hashtags were used for the search term as hashtags add significance beyond the initial act of posting, allowing users to link their posts to broader issues and campaigns (Ahmed 2018). Hashtags also allow for the creation of networks which can be computationally investigated to provide a snapshot of the debate relating to a campaign or trending topic (Bridge et al. 2021). The use of hashtags enables unconnected users to view and comment on messages that have included the hashtag, therefore including this in the search term ensured the network was representative of the debate on X during the selected time period. There were no geographical restrictions on the search query.

There were several reasons for the selected data collection period. Initially, March was identified based on research investigating Google Trends that established that March falls outside of season peaks of public interest in nutrition (Passos et al. 2020), such as religious holidays, seasonal trends (Palomo-Llinares et al. 2021) and World Nutrition Day (28 May). At the time of data collection, NodeXL Pro was limited by Twitter's Search API (removed by X at the end of March 2023) that constrained returns to 18,000 posts on each request run daily for the 3 weeks. From pilot work giving us a sense of #nutrition posting volume (daily retrievals on average exceeding the Twitter Search API), and in light of Smith's argument for SNA of shorter chunks of time (Smith et al. 2009), we rationalised that a longitudinal design of repeated daily retrievals over 3 weeks in combination with our use of the hashtag would permit us to investigate the characteristics, influential actors, topics and sentiment of the #nutrition network on X at the time.

The data gathered from X included vertices (or nodes, i.e., users; see Table S1 for definitions of network theory terminology and metrics), edges (connections between vertices), metadata about the post and the user interactions, and metrics that allow the assessment of user influence within a network. These metrics include the centrality measures: betweenness, closeness and eigenvector centrality. In brief, high betweenness identifies nodes between other nodes suggesting high influence (Freeman 1977), while high closeness centrality indicates close proximity to other nodes and thus high influence (Sabidussi 1966). Whereas high eigenvector centrality identifies nodes that are connected to many nodes who themselves have high scores, again suggesting high influence within the network (Newman 2018). An overview of the data collection and analysis methods is illustrated in Fig. 1.

2.2 Data cleaning

Data were manually cleaned removing any posts that were not written in English. As this was a large data set, further steps were taken to prepare the data and make it more manageable. Duplicate edges (where the same two users are discussing the same thing) were counted and merged, and after sorting the edges by date, old posts (pre-2023) were removed. Finally, edges with an edge weight of less than 5 were removed to focus on individuals with stronger, more frequent ties and influence.

2.3 Data analysis

2.3.1 Social network analysis

SNA was used to investigate user relationships and flows of information within the X communities discussing #nutrition. SNA is a multi-step process, automated through NodeXL Pro, which calculates graph metrics and visualises the network permitting exploration of connections and patterns (see Smith et al. 2009 for a detailed methodology of NodeXL (Smith et al. 2009)). Within the network visualizations, vertices (i.e., the X users) were grouped by cluster using the Clauset–Newman–Moore cluster algorithm (Clauset et al. 2004). The layout of the graph was generated using the Harel–Koren Fast Multiscale layout algorithm (Harel and Koren 2004). For clearer visual analysis, the top 20 groups within the network were illustrated. The ten largest groups were further annotated to provide insight into the main types of discussions taking place, identified during the textual

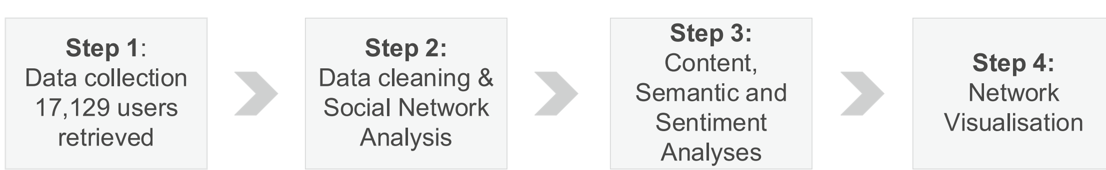

Fig. 1 Overview of research methods in NodeXL Pro

analysis run in parallel to the network analysis (i.e., each cluster's conversation was textually analysed).

Influence, as measured by SNA, does not necessarily relate to the number of followers or posts. Instead, it is measured by the betweenness centrality metric, a measure of influence within the network defined by the shortest number of paths that pass through it. The influential vertices (users) within a network act as a bridge between different clusters within the network (i.e., nodes between nodes). The higher the betweenness centrality number, the quicker the propagation of information through the network is, and therefore the greater the influence on information flow.

2.3.2 Textual analysis: word pairs, semantic and sentiment analyses

Analysing word pairs facilitates understanding of the discourse across the network. The most common word pairs were placed in one cluster, and word pairs that appeared less frequently were placed in separate clusters (Singh et al. 2016). Once the top word pairs from the posts were identified, a new data set was created using the word pairs, and a semantic network visualisation was generated. Semantic networks are composed of linked words that reveal the relationships between ideas embedded in the network. Visualising semantic networks can reveal the most central ideas in a corpus and identify how ideas cluster together. There is no gold standard for selecting cut-off values for semantic visualisations and different studies have used different cut-offs (Ferra and Nguyen 2017). With more than 97,000 word-pair edges identified in our #nutrition network, here only word-pairs with a frequency of more than 10 occurrences were considered to reduce processing noise as recommended in the literature and to increase the likelihood that only genuinely semantically related words were represented (Eskandari et al. 2022; Bruzzese et al. 2022). Subsequently, sentiment analysis was conducted using NodeXL Pro drawing upon the Opinion Lexicon's list of positive and negative words (Hu and Liu 2004), which provided the number of words that matched either positive or negative words, excluding those that fit into neither category. In generating the semantic network visualisation, the top five clusters were shown to ensure clear visualisation. The larger and more prominent the word, the more repeatedly it was found within the network.

3 Results

Figure 2 shows the different types of posts published during the collection period. There were 26,809 posts published. Of these, there was more re-shared content which included reposts, mentions and replies, than original content.

3.1 Social network analysis

The #nutrition network included 17,129 users (vertices), with 26,809 relationships (edges), graph metrics are summarised in Table 1. Only 6 in 100 X users were mutually connected, which implies users prefer to share nutrition content, but not engage in debate (reciprocated relationships ratio, 0.064; Table 1). Similarly, the graph density was low (0.001; Table 1), suggesting that most users only communicated heavily with a few other users. The number of X users who did not have any relationships was high (n = 3874, approximately 20%; Table 1). There was also a large group that contained 'self-loops' (n = 5815; Table 1), where users were not connected to other users; these were not removed to show the number of users that share posts without mentioning or replying to another user.

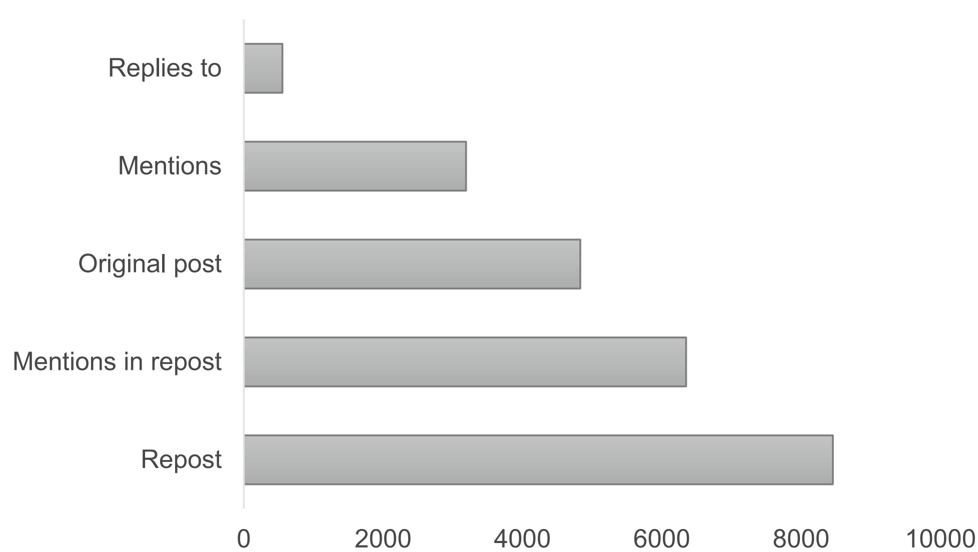

Fig. 2 Number of posts by published category

Table 1 Summary of the #nutrition network debate taking place on X

Graph metric	Value
X users (vertices)	17,129
Total relationships (edges)	26,809
Reciprocated X user pair ratio	0.0330
Reciprocated relationships ratio	0.0640
Graph density*	0.001
Isolated X user	3874
Self-loops (edge that connects to itself)	5815
Maximum geodesic distance (diameter)	20
Average geodesic distance	5.26
Average shortest path	5.257
Modularity (the density of connections between nodes)	0.7797

Data were harvested between 1–21 March 2023 and analysed using NodeXL Pro

*Calculated by comparing the ratio of the number of relationships present in the network and the number of possible relationships

Overall, these data suggest a directed network with relationships and information flows that may not be reciprocated (Hansen et al. 2010).

X users with lower geodesic distances (the shortest distance path between two users) have a greater influence on the debate because they are typically more central within the network. In this network, the maximum geodesic distance (that is, the greatest distance between the two users that are farthest from each other) was 20 (Table 1). The graph (Fig. 3), and the associated network data, show that the online nutrition conversations were dispersed (average geodesic distance, 5.26). This network pattern suggests that, while information moved quickly within the groups that contain influential users, dissemination across the wider network was likely slower. This was further reinforced by our finding that there was less cross-group communication between smaller clusters (modularity 0.78).

The #nutrition community on X was dispersed with groups having isolated conversations with little between group crosstalk as visualised in Fig. 3. The largest group (G1) was an 'isolates' group where users conversed without mentioning each other (that is, individual posts), and the remainder of the group represented a range of large, medium, and small communicative clusters. The network had a high modularity score of 0.78 (Table 1) suggesting that the groups were well-defined. Semantic analysis of the groups revealed that a range of discussions took place, ranging from fitness and lifestyle, family & child nutrition, nutrition & diet awareness, and nutrition and advocacy awareness (Fig. 3). There was a level of interaction between larger clusters (e.g., G2, G3 and G4; Fig. 3) within the network, but smaller groups (e.g., G5, G9 and G10) did not interact at the same level.

3.2 Influencers (betweenness centrality)

Exploring the network graph metrics helps to identify who the influential actors were and how the information flows through the network since it is possible to identify the groups of actors across the network and identify which actors influence the debate. In this network, betweenness centrality ranged from 0 to 23,375,543.67 (indicates how much a user acts as a bridge across the network with higher scores indicating greater influence). The top 10 influencers in this network (Table 2) were from personal accounts, online blogs, and not-for-profit organisations (NGOs) according to self-reported user profiles. These influential users had low closeness centrality (Table 2), which suggests they are also central to the network. The most important X user in this network, as measured by eigenvector centrality (0.257; Table S2), was from a personal account with high betweenness centrality showing they are also an influential user.

3.3 Semantic analysis

Semantic analysis showed that health was the top word associated with nutrition in this network; while healthy lifestyle, fitness, and food were also commonly associated words (Table 3). The semantic network map illustrates the overall pattern of word pairs communicated in the network, with the most frequent pairs shown in groups one and two on the left of the map (Fig. 4).

Similar to the semantic analysis of the individual groups in Fig. 3, analysis of the word-pairs (n = 97k) data collectively further established that health, lifestyle and well-being were prominent conversations within the discussion. The relationship between the words also demonstrates the interconnected topics such as diet, wellness, and clinical nutrition. Not least, this analysis shows how users draw upon hashtags within their vocabulary, as shown by the largest cluster containing many hashtags connected to other words across the network.

3.4 Sentiment analysis

Positive words were more prominent in this network (Table 4), suggesting that the discourse relating to nutrition was more likely to be positive than negative. The posts included those that were raising awareness and linking health to general wellness, which were more likely to be positive than negative.

4 Discussion

This study is the first to visualise the #nutrition conversation taking place on a large social media platform known as X (formerly Twitter). The key findings in this study show

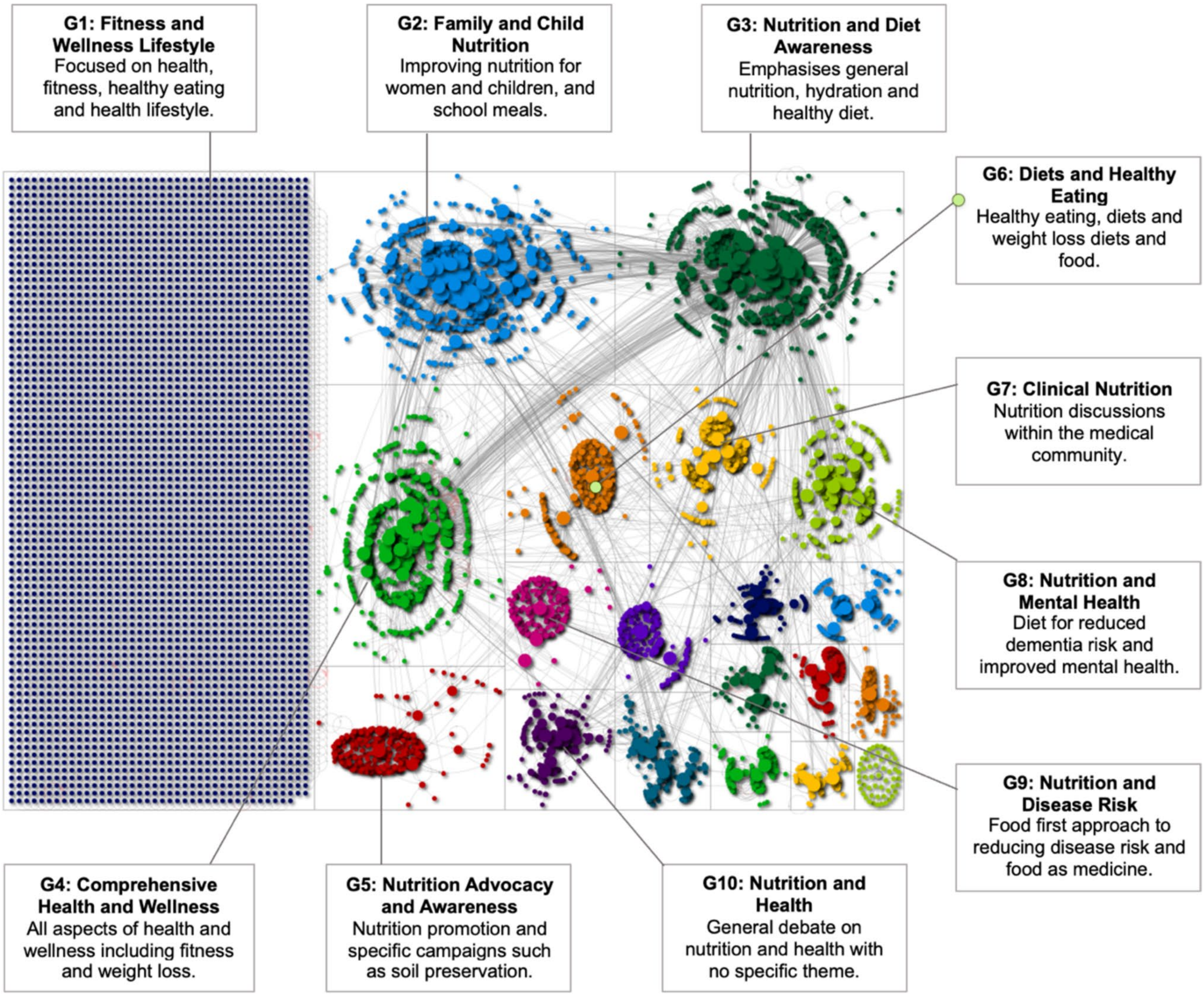

Fig. 3 Network visualisation of the #nutrition debate on X. Data were harvested between 1 and 21 March 2023, analysed and visualised using NodeXL Pro. Users (n = 17,129 vertices) that interacted (n = 26,809 edges) through mentions or reposts are clustered together as groups. Each group is visualised, with the largest group of users in the upper left corner of the map (Group 1 annotated as G1) and the smallest in the lower right corner

that the network was dispersed without polarisation with the largest group being unconnected users who did not have relationship with other users in the network. Concerningly, the most influential actors were public users without nutrition training or expertise. The online nutrition community was more likely to propagate information through reposting, as opposed to posting original content. The most frequently discussed topics in this network was; healthy diet, food, fitness, and healthy lifestyle, with a discourse that was largely positively framed.

Our #nutrition network map illustrates community clusters of multiple smaller groups with their own influence and sources of information (Smith et al. 2014), and suggests a low-density network with a number of isolated users and high modularity. This type of network structure can restrict the flow of information, meaning individual groups of users rely on just a few influential users in that cluster to share information on a specific topic relating to nutrition (Bruzzese et al. 2022). This structure has been termed 'bazaars' (Bridge et al. 2021). Although there is little engagement between groups and information flow is restricted, because there is no single source of information, multiple conversations can occur concurrently within the network (as in a market bazaar), with each group having its own view (Himelboim et al. 2017).

As with other studies investigating social networks (Lynn et al. 2020), the #nutrition network showed a wide range of actors influencing the discourse. However, the most

Table 2 Top 10 influential users in the #nutrition network ranked by their betweenness centrality score

Rank	User classification[a]	Betweenness centrality	Closeness centrality[b]	Eigenvector centrality[c]
1	Personal	23,375,544	0.163	0.257
2	Personal	10,254,555	0.151	0.203
3	Company	9,715,469	0.160	0.238
4	News	9,045,685	0.160	0.243
5	Blog	8,066,282	0.153	0.173
6	NGO	7,472,289	0.135	0.064
7	Personal	6,731,353	0.159	0.239
8	NGO	6,042,642	0.150	0.098
9	Personal	5,650,452	0.134	0.088
10	Personal	5,207,999	0.150	0.157

[a]Classification is self-selected by X users and not verified

[b]Position within the network

[c]Influence within the network

Table 3 Top 10 hashtags, words and word pairs

Rank	Top hashtags*	Count	Top words	Count	Top word pairs	Count
1	Nutrition	865	#nutrition	18,317	#health, #nutrition	785
2	Savesoil, nutrition	446	#health	4,223	#nutrition, #health	722
3	Health, healthylifestyle, nutrition, regime	324	food	3,174	#healthylifestyle, #nutrition	573
4	Medicine, health, nutrition	222	more	2,612	#health #healthylifestyle	530
5	Healing, hope, gratitude, love, meditation, letters, nutrition, goals, chakras,	211	nutrition	2,562	#diet, #nutrition	513
6	Nutrition, foodheroes, schoolmeals	145	health	2,476	nutritious, food	512
7	Anganwaditeachers anganwadi, nutrition, ruraldistress, telangana	87	#diet	1,845	bone, health	454
8	Nutrition health diet	75	#healthylifestlye	1,726	roar, #nutrition	433
9	Doctor farmer healthy	65	#fitness	1,612	#savesoil, voice	433
10	Meded medtwitter foamed gitwitter nutrition dietitian	58	#food	1,604	want, continue	432

Hashtags usually relate to trends or themes and may not appear to be English, for example, #gitwitter is used by Gastroenterologist when tagging content

influential users in this network were public accounts that did not mention any nutrition education or training in their profiles. Notably, this contrasts with studies looking at the discourse on diets, which found that users with health or science backgrounds were the most influential (Eaton et al. 2023); and also with discourse on climate change, where international organisations are central to the debate (Yuan et al. 2024). Probst and colleagues (2019) found that X is being used more frequently by nutritionists as a professional network and a tool to disseminate research findings for professional development however their voices do not appear to be cutting through network noise (Probst and Peng 2019). This is a particularly concerning finding for nutrition as the general population turn to social media often for dietary advice (Alhothali and Aljefree 2023) and our study suggests they may be receiving poor quality advice from unqualified users.

Social network research looking at how debates are portrayed suggests that strong ties, such as those between personal contacts, have the greatest influence over the debate (Bridge et al. 2021). However, weak ties, such as those with unknown users, play an equally important role as they enable information diffusion. Similar to other health communication networks, the ratio of reposted to original content was higher (Lynn et al. 2020).

Our study finds individuals' posts were overwhelmingly associated with health, heathy diet, fitness and healthy lifestyles with #health, #healthylifestyle and #diet the most commonly associated hashtags across our network. This gives an indication of the public narrative on X, and suggests that X users may be using the platform for advice on healthier diets. These findings support research that compared the USA food pyramid to X conversations, and found 'healthy food' and 'healthy diets' in the top 10 associated

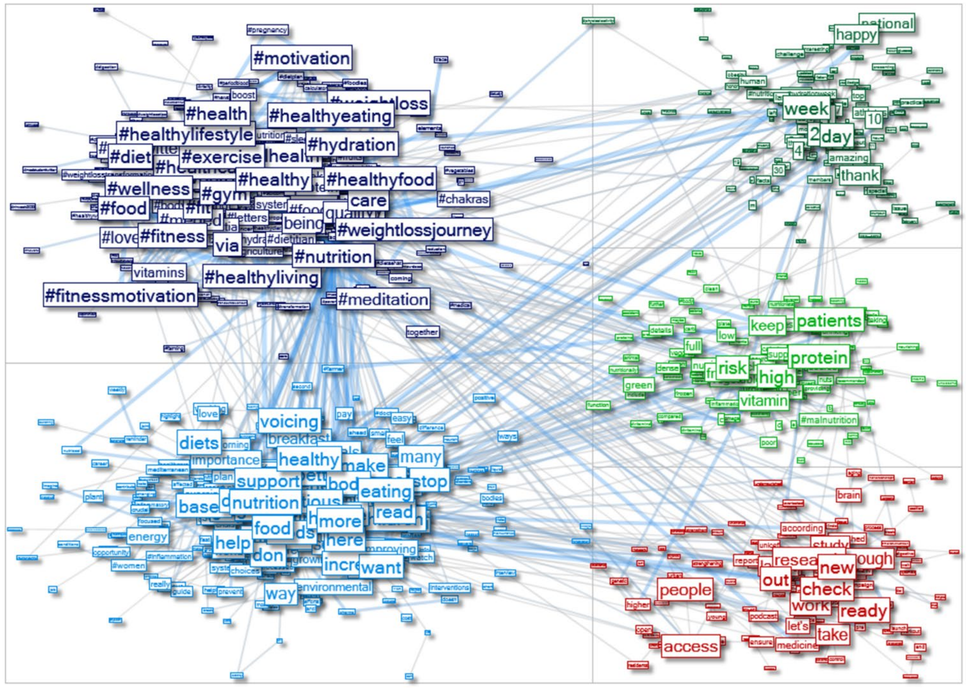

Fig. 4 Semantic network map of word pairs. The overall patterns of word pairs communicated in the top 5 clusters in our network, based on word-pair (n = 97K) frequency

Table 4 Sentiment analysis of the network

Sentiment	Count	Salience (%)	Example words
Positive words	22,218	47	Support, healing, better, great, well, improve
Negative words	6,795	15	Out, want, promoting, stop, time, weight
Total words	46,8191	100	

words (Saura et al. 2020). A separate study looking at conversations using #healthyfood also found #healthylifestyle and #healthyliving was most commonly associated with the 'healthy food' search term (Pilař et al. 2021).

The sentiment used in posts across this nutrition network was broadly positive, although it appeared to be topic dependant. These findings support other positively framed topics that have focused on healthy diets (Saura et al. 2020), organic foods, and veganism, which has shifted overtime to a more positive narrative as it has increased in popularity (Shamoi et al. 2022). A systematic review looking at data science methods to examine sentiment in food and cooking social media content, found similar results to ours with a more positive discourse overall (Molenaar et al. 2023).

However, our findings differ from more emotive topics such as front-of-pack labelling (Septia Irawan et al. 2022), ultra-processed foods (Saura et al. 2020), and fiscal policies (Bridge et al. 2021), which are negatively framed. Research investigating discourse on disease also tends to be more negative (Perez-Perez et al. 2019). Similarly, findings from other disciplines have shown the online debate on climate change to be negative, and sometimes aggressive (Yuan et al. 2024), with attempts to discredit climate change scientists (Getson et al. 2021). Likewise, fear is a common emotion in political networks (Chakraborty and Mukherjee 2023). Nutrition is therefore better placed than climate change or politics, as the discussion is generally positive, although the public appear to be leading the debate. Nutrition professionals are best placed to clearly communicate evidence-based nutrition advice on social media but were under-represented voices. Our research suggests, increasing the use of social media by credentialled Nutrition professionals could have far reaching benefits given the publics reliance on social media for nutrition advice.

This study had several limitations, most notably the short data collection period which only provides a snapshot of the discourse. Additionally, there is no comparison with other periods so it is not possible to tell whether the discourse and public opinion on nutrition changed over time or in response to real time events. This study only considered the discourse on X. As user accounts are open, X was deemed the most suitable platform carry out public opinion research, however this limits the generalisability of results. Other social media platforms may yield different results and should be considered in future research. While this research is representative of the nutrition discourse on X during the collection period, the very use of the hashtag could bias data collection as those using the hashtag will likely have an interest in nutrition. Therefore, the debate investigated in this study may only reflect views of a specific cohort of users, rather than X users generally.

A further limitation is that no consideration was given to automated accounts known as “bots” and how they might affect the network. It is estimated that between 9% (Varol et al. 2017) and 29% (Weng and Lin 2022) of social media activity is attributable to bots. Bots generate fewer replies, reposts, and mentions from human users and tend to repost more than humans (Lynn et al. 2020). Based on this, it is likely that our dataset included a number of bots as 20% of X users were not connected to other users, and the network had a high repost rate. Bots are of particular concern as they can perpetuate misinformation and disproportionately spread articles from low-credibility sources (Shao et al. 2018). Therefore, future research should include a plan to mitigate the negative impact of bots. Furthermore, only English language posts were collected and analysed. With around 40% of X posts written in English, a large proportion of nutrition related posts are not considered in this research, and the findings are not necessarily applicable to non-English speaking users. X user geolocation was not used in the analysis as it is not reliably captured and is not mandatory when creating an account. Finally, as with all social media analysis, the results are difficult to replicate due to changes in users, public mood, political landscape and data collection periods (Bruzzese et al. 2022).

Nonetheless, there are notable strengths to our study. The main strength is that, to our knowledge, it is the first study that has attempted to investigate the public discourse relating to nutrition on social media and to map the network. In addition to the novel topic, it adds to the wider body of literature looking at social networks for public health communication and can contribute methodology to other mixed methods papers.

Moreover, it is also the first study to analyse what the public are discussing in relation to nutrition and to identify the key actors. The user analysis showed that there are many actors involved in the nutrition debate but more importantly, this study was able to identify the influencers in this network, and to measure how they propagate nutrition information online. This study identified that public X users without formal nutrition training were the key actors and they appeared to be particularly interested in health and healthy diets and lifestyles. This is important for understanding who may be influencing the public debate. These findings could be used to encourage those with nutrition/dietitian training to use X more to influence the debate. It could also help inform the language professionals should use to engage the public when disseminating nutrition messages on X and highlights the strength of using high profile accounts to amplify nutrition messages. Finally, the results could be used to discourage the public from sharing nutrition information as they could be propagating misinformation and poor-quality information that is not evidence based.

4.1 Future research

Future research should consider how the nutrition debate is portrayed over a longer time period. If the data collection period was long enough, this could include seasonal changes when interest in nutrition peaks (Passos et al. 2020; Palomo-Llinares et al. 2021). Given the high number of public X users discussing nutrition in our research, future research could also analyse popular dietary patterns or compliance to public health advice. This could support policy makers by informing on public opinion, the sentiment towards dietary guidelines and public acceptance of new initiatives. Not least, future research should also use communication theories to better understand what motivates users to partake in the debate on X. Motivations for reposting, liking and commenting should also be investigated as these were more common than posting original content in this research. This could also support nutrition professions to share evidence-based advice and engage in X debates in the future.

5 Conclusions

By using SNA, semantic and sentiment analysis, this study provides novel insight into the public #nutrition debate on X. The findings revel a publicly driven debate which focuses on healthy diets, fitness and lifestyle. These findings are important for nutritionists and healthcare professionals as it provides understanding into how the public may be accessing nutrition information, and our study provides insight into the language that should be used to engage the public. These findings should be used to encourage nutrition professionals to engage in the online debate and be the leading voices.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13278-024-01404-9.

Authors contribution CHE researched literature and conceived the study. JBM and CELE were involved in study design and protocol development. WA contributed to the statistical and data analysis. CHE wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Aaminah Zaman M, Thapa S, Paswan AK (2023) Social media influencer (SMI) as a human brand—a need fulfillment perspective. J Prod Brand Manag 32:173–190

Ahmed W (2018) Public health implications of #ShoutYourAbortion. Public Health 163:35–41

Ahmed W, Vidal-Alaball J, Segui FL, Moreno-Sanchez PA (2020) A social network analysis of tweets related to masks during the COVID-19 pandemic. Int J Environ Res Public Health 17:8235

Alhothali GT, Aljefree NM (2023) Young adults' sought gratifications from, and perceptions of food advertising by, social media influencers: a qualitative approach. J Health Popul Nutr 42:103

Appel G, Grewal L, Hadi R, Stephen AT (2020) The future of social media in marketing. J Acad Mark Sci 48:79–95

Bridge G, Flint SW, Tench R (2021) A mixed-method analysis of the #SugarTax debate on Twitter. Public Health Nutr 24:3537–3546

Bruzzese S, Ahmed W, Blanc S, Brun F (2022) Ecosystem services: a social and semantic network analysis of public opinion on Twitter. Int J Environ Res Public Health 19:15012

Chakraborty A, Mukherjee N (2023) Analysis and mining of an election-based network using large-scale twitter data: a retrospective study. Soc Netw Anal Min 13:74–74

Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. Phys Rev E Stat Nonlinear Soft Matter Phys 70:066111–066211

Croes E, Bartels J (2021) Young adults' motivations for following social influencers and their relationship to identification and buying behavior. Comput Hum Behav 124:106910

Eaton MC, Probst YC, Smith MA (2023) Characterizing the discourse of popular diets to describe information dispersal and identify leading voices, interaction, and themes of mental health: social network analysis. JMIR Infodemiol 3:e38245

Ellis CH, Ho P, Moore JB, Evans CEL (2024) Content quality versus sharing practices on social media: a cross-sectional analysis of nutrition information on Twitter. medRxiv: 2024.08.15.24312059

Engel E, Gell S, Heiss R, Karsay K (2024) Social media influencers and adolescents' health: a scoping review of the research field. Soc Sci Med 340:116387

Erz A, Marder B, Osadchaya E (2018) Hashtags: motivational drivers, their use, and differences between influencers and followers. Comput Hum Behav 89:48–60

Eskandari F, Lake AA, Butler M (2022) COVID-19 pandemic and food poverty conversations: social network analysis of Twitter data. Nutr Bull 47:93–105

Ferra I, Nguyen D (2017) #Migrantcrisis: "tagging" the European migration crisis on Twitter. J Commun Manag 21:411–426

Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry 40:35–41

Getson JM, Sjöstrand AE, Church SP, Weiner R, Hatfield JL, Prokopy LS (2021) Do scientists have a responsibility to provide climate change expertise to mitigation and adaptation strategies? Perspectives from Climate Professionals. Public Underst Sci 30:169–178

Hansen D, Shneiderman B, Smith MA (2010) Analyzing social media networks with NodeXL: insights from a connected world. Morgan Kaufmann Publishers Inc.

Harel D, Koren Y (2004) A fast multi-scale method for drawing large graphs. In: Graph algorithms and applications 3. World Scientific

Harris JK, Moreland-Russell S, Tabak RG, Ruhr LR, Maier RC (2014) Communication about childhood obesity on twitter. Am J Public Health 104:62

Himelboim I, Smith MA, Rainie L, Shneiderman B, Espina C (2017) Classifying Twitter topic-networks using social network analysis. Soc Media + Soc 3(1)

Holmberg K, Hellsten I (2015) Gender differences in the climate change communication on Twitter. Internet Res 25:811–828

Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery, Seattle, pp 168–177

Hunt D (2021) How food companies use social media to influence policy debates: a framework of Australian ultra-processed food industry Twitter data. Public Health Nutr 24:3124–3135

Lee JA, Sudarshan S, Sussman KL, Bright LF, Eastin MS (2022) Why are consumers following social media influencers on Instagram? Exploration of consumers' motives for following influencers and the role of materialism. Int J Advert 41:78–100

Lynn T, Rosati P, Santos GL, Endo PT (2020) Sorting the healthy diet signal from the social media expert noise: preliminary evidence from the healthy diet discourse on Twitter. Int J Environ Res Public Health 17:8557

Marauri-Castillo I, Rodríguez-Gonzalez MDM, Marín-Murillo F (2024) Disseminators, not influencers: communication of dietitians on social networks. Vivat Academia. 157. https://doi.org/10.15178/va.2024.157.e1495

Molenaar A, Jenkins EL, Brennan L, Lukose D, McCaffrey TA (2023) The use of sentiment and emotion analysis and data science to assess the language of nutrition-, food- and cooking-related content on social media: a systematic scoping review. Nutr Res Rev 37(1):43–78

Moukarzel S, Rehm M, del Fresno M, Daly AJ (2020) Diffusing science through social networks: the case of breastfeeding communication on Twitter. PLoS ONE 15:e0237471

Moukarzel S, Caduff A, Rehm M, del Fresno M, Pérez-Escamilla R, Daly AJ (2021) Breastfeeding communication strategies, challenges and opportunities in the twitter-verse: perspectives of influencers and social network analysis. Int J Environ Res Public Health 18:6181

Newman M (2018) Networks. Oxford University Press

Palomo-Llinares R, Sánchez-Tormo J, Wanden-Berghe C, Sanz-Valero J (2021) Trends and seasonality of information searches carried out through google on nutrition and healthy diet in relation to occupational health: infodemiological study. Nutrients 13:4300

Passos JA, Vasconcellos-Silva PR, Santos L (2020) Cycles of attention to fad diets and internet search trends by Google trends. Cien Saude Colet 25:2615–2631

Perez-Perez M, Perez-Rodriguez G, Fdez-Riverola F, Lourenco A (2019) Using Twitter to understand the human bowel disease community: exploratory analysis of key topics. J Med Internet Res 21:e12610

Pilař L, Stanislavská LK, Kvasnička R (2021) Healthy food on the Twitter social network: vegan, homemade, and organic food. Int J Environ Res Public Health 18:3815

Probst YC, Peng Q (2019) Social media in dietetics: insights into use and user networks. Nutr Diet 76:414–420
Sabidussi G (1966) The centrality index of a graph. Psychometrika 31:581–603
Saura JR, Reyes-Menendez A, Thomas SB (2020) Gaining a deeper understanding of nutrition using social networks and user-generated content. Internet Interv Appl Inf Technol Ment Behav Health 20:9
Septia Irawan A, Shahin B, Njuguna DW, Nellamkuzhi NJ, Quoc TB, Mahrouseh N, Varga O (2022) Analysis of content, social networks, and sentiment of front-of-pack nutrition labeling in the European Union on Twitter. Front Nutr 9:846730
Shamoi E, Turdybay A, Shamoi P, Akhmetov I, Jaxylykova A, Pak A (2022) Sentiment analysis of vegan related tweets using mutual information for feature selection. PeerJ Comput Sci 8:e1149
Shao C, Ciampaglia GL, Varol O, Yang K-C, Flammini A, Menczer F (2018) The spread of low-credibility content by social bots. Nat Commun 9:1–9
Singh M, Bansal D, Sofat S (2016) Behavioral analysis and classification of spammers distributing pornographic content in social media. Soc Netw Anal Min 6:1–18
Smith M, Shneiderman B, Milic-Frayling N, Rodrigues EM, Barash V, Dunne C, Capone T, Perer A, Gleave E (2009) Analyzing (social media) networks with NodeXL. ACM, pp 255–264
Smith M, Ceni A, Milic-Frayling N, Shneiderman B, Mendes Rodrigues E, Leskovec J, Dunne C (2010) NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010/2013/2016, from the Social Media Research Foundation: https://www.smrfoundation.org
Smith MA, Rainie L, Shneiderman B, Himelboim I (2014) Mapping Twitter topic networks: from polarized crowds to community clusters. https://www.pewresearch.org/internet/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-tocommunity-clusters/. Accessed Aug 2024
Sussman KL, Bouchacourt L, Bright LF, Wilcox GB, Mackert M, Norwood AS, Altillo BSA (2023) COVID-19 topics and emotional frames in vaccine hesitation: a social media text and sentiment analysis. Digit Health 9:20552076231158308
Toraman C, Şahinuç F, Yilmaz EH, Akkaya IB (2022) Understanding social engagements: a comparative analysis of user and text features in Twitter. Soc Netw Anal Min 12:47
Varol O, Ferrara E, Davis C, Menczer F, Flammini A (2017) Online human-bot interactions: Detection, estimation, and characterization. In: Proceedings of the international AAAI conference on web and social media, pp 280–289
Vijaykumar S, Nowak G, Himelboim I, Jin Y (2018) Virtual Zika transmission after the first U.S. case: who said what and how it spread on Twitter. Am J Infect Control 46:549–557
Vijaykumar S, McNeill A, Simpson J (2021) Associations between conflicting nutrition information, nutrition confusion and backlash among consumers in the UK. Public Health Nutr 24:914–923
Wasserman S (1994) Social network analysis: methods and applications. Cambridge University Press, Cambridge
Weng Z, Lin A (2022) Public opinion manipulation on social media: social network analysis of Twitter bots during the COVID-19 pandemic. Int J Environ Res Public Health 19(24):16376
Yousef M, Dietrich T, Rundle-Thiele S (2022) Actions speak louder than words: sentiment and topic analysis of COVID-19 vaccination on Twitter and vaccine uptake. JMIR Form Res 6(9):e37775
Yuan S, Chen Y, Vojta S, Chen Y (2024) More aggressive, more retweets? Exploring the effects of aggressive climate change messages on Twitter. New Media Soc 26(8), 4409-4428