

Review

Not peer-reviewed version

Unveiling Dietary Complexity: A Scoping Review of Network Analysis in Dietary Pattern Research and a Methodological Roadmap for Future Research

[Rebecca M.J. Taylor](#) , Jack A. Moore , [Alecia L. Cousins](#) , Amy R. Griffiths , [Hayley A. Young](#) *

Posted Date: 5 August 2025

doi: 10.20944/preprints202508.0362.v1

Keywords: dietary pattern analysis; network analysis; guiding principles



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Review

Unveiling Dietary Complexity: A Scoping Review of Network Analysis in Dietary Pattern Research and a Methodological Roadmap for Future Research

Rebecca Taylor, Jack Moore, Amy Griffiths, Alecia Cousins and Hayley A Young *

Swansea University, Wales, SA2 8PP, United Kingdom

* Correspondence: h.a.young@swansea.ac.uk

Abstract

Background/Objectives: Dietary patterns play a crucial role in health, yet most research examines foods individually, overlooking how they interact. This approach provides an incomplete picture of how diet influences health outcomes. Network analysis (e.g., Gaussian Graphical Models, Mutual Information Networks, Mixed Graphical Models) offers a more comprehensive way to study food co-consumption by capturing complex relationships between dietary components. However, while researchers have applied various network algorithms to explore food co-consumption, inconsistencies in methodology, incorrect application of algorithms, and varying results have made interpretation challenging. To address this, the aim of this study was to review existing research and establish guiding principles for future studies. **Methods:** Using PRISMA-ScR criteria, a scoping review identified 18 relevant studies across different populations and health outcomes. **Results:** Gaussian Graphical Models were the most frequently used, often paired with regularisation techniques (e.g., graphical lasso) to improve clarity. However, several methodological challenges were identified, including the use of cross-sectional data, which limits the ability to determine cause and effect, reliance on centrality metrics in unbounded networks, and difficulties in handling non-normal data. Only a few studies addressed these limitations, such as using semiparametric extensions of Gaussian Graphical Models to manage non-normal data. **Conclusions:** To improve the reliability of network analysis in dietary research, this review proposes five guiding principles - model justification, design-question alignment, transparent estimation, cautious metric interpretation, and robust handling of non-normal data - that future studies can adopt. Overall, this review highlights the potential of network analysis to uncover hidden relationships between dietary components and enhance our understanding of how diet influences health. This review was preregistered <https://doi.org/10.17605/OSF.IO/R5VE6>.

Keywords: dietary pattern analysis; network analysis; guiding principles

1. Introduction

Dietary patterns have been associated with a variety of health outcomes. For example, the Mediterranean diet, characterised by a high consumption of fruits, vegetables, whole grains, and healthy fats, has been linked to the prevention of cardiovascular disease [1,2], better cognitive performance [3], and a longer life expectancy [4]. In comparison, the Western diet consisting of a high intake of red and processed meat, refined grains, sugars, fats, and fast food with a low intake of fruit and vegetables has been associated with higher rates of obesity [5,6] and an increase in cancer risk [7].

Currently, most of the research looking into nutrition and its effect on health has focused on analysing foods and nutrients separately from each other [8,9] or has produced 'a priori' diet quality scores or 'data-driven' composite scores [10]. These traditional dietary assessment methods do not

expose food synergies, which may lead to an incomplete understanding of dietary patterns and their health implications [11,12]. Therefore, it is crucial to research not only what foods are consumed but also how foods are consumed in combination. For example, a recent study found that garlic may counteract some of the detrimental effects associated with red meat consumption, including a reduced risk of cardiovascular disease from a high intake of red meat [13]. This finding emphasises the need to examine the interactions between different foods to fully understand their health impacts.

The network approach offers a promising and more holistic way to analyse the co-consumption of foods. It enables the exploration of dietary patterns by using advanced statistical techniques to map and analyse the connections between various dietary components [12]. By capturing the co-consumption patterns and their associated outcomes, the network approach can reveal insights into the relationship between nutrition and health that traditional methods may have previously overlooked.

This review aims to determine whether network approaches can offer a more comprehensive analysis of dietary intake. By evaluating the effectiveness of these methods, this review aims to contribute to the development of more accurate dietary assessment tools, potentially leading to improved dietary recommendations and interventions which are tailored to human eating behaviours. First, the traditional approaches to dietary pattern analysis are examined, highlighting their limitations. Next, a comprehensive scoping review of studies which have applied network analysis to explore dietary patterns is presented, demonstrating its potential to address these challenges. Finally, guiding principles are offered to enhance methodological rigor and advance this innovative field.

1.1. Dietary Patterns and Health

Using dietary analysis in nutritional research is essential for the development of dietary interventions aimed at improving both physical and psychological wellbeing. Accurate dietary analysis allows researchers to identify dietary patterns and nutrients that influence health outcomes, enabling the formulation of targeted interventions. For instance, the Dietary Approach to Stop Hypertension (DASH) diet was developed after observational research found that a carbohydrate rich diet with fruits, vegetables and low-fat dairy products were associated with lowered blood pressure [14,15]. Notably, randomised controlled trials of individual nutrients such as magnesium, potassium, calcium, and fibre, had produced inconsistent results [14]. One explanation was that nutrients from dietary supplements may not benefit health as effectively as those obtained from whole foods due to the synergistic interactions between nutrients and other components present in the diet. Unfortunately, due to the limitations of conventional dietary pattern analysis, most nutritional interactions remain undiscovered. However, recent computational advances may help unveil non-additive and non-linear interactions [16], thereby improving dietary recommendations and the development of multi-component functional foods and supplements that benefit health [8].

1.2. Traditional Dietary Pattern Analysis

The traditional methods used for dietary pattern analysis include principal component analysis (PCA), cluster analysis, and a priori composite scores. These techniques are widely used to investigate the relationships between diet and health outcomes by summarising dietary intake data into meaningful patterns. Details of these traditional methods can be found in Table 1.

Table 1. Traditional methods for dietary analysis.

Method	Algorithm	Linear/Nonlinear	Assumptions	Strengths/Limitations
Principal Component Analysis (PCA)	Eigenvalue decomposition	Linear	Assumes normally distributed data, linear relationships between variables, uncorrelated components.	Identifies what dietary patterns exist in a population. Can determine which foods are consumed together in a diet but does not reveal interactions between those foods.
Factor Analysis	Factor extraction	Linear	Assumes normally distributed data, linear relationships, data can be grouped into latent factors.	Can identify the underlying dietary factors that explain variations in food intake. However, does not provide information about how particular food interact.
Cluster Analysis	k-means, hierarchical clustering	Nonlinear	Assumes defined clusters with similar characteristics and independent observations.	Groups individuals based on their dietary patterns. Useful for segmenting consumers based on dietary patterns. Can handle nonlinear associations between variables. Assumes pairwise similarity or proximity but does not explicitly capture direct or indirect interdependencies among multiple variables.
Dietary Index/Scores	Predefined scoring	Linear	Assumes each score represents healthfulness, often based on a reference diet. Each component is typically weighted (sometimes equally), ignoring potential interactions between components. Requires prior knowledge.	Can identify how closely an individual's diet aligns with a healthy/reference dietary pattern.

PCA is a statistical method which identifies underlying dimensions of food consumption by grouping food items based on the degree to which they correlate [17]. For example, PCA may reveal that certain foods like fruits, vegetables, and whole grains appear together in diets to form a healthy eating component [18]. This method has been used to link dietary patterns to various health outcomes, such as an association between a Western diet and increased prevalence of obesity [5], and between a Mediterranean diet and reduced occurrence of cardiovascular disease [1].

Cluster analysis groups individuals with similar diets into homogenous subgroups [19]. This method groups individuals who consume similar types and amounts of foods, which can be based on food frequency, standardised nutrient intakes, or a combination of dietary and biochemical measures. Further interpretation of these clusters involves comparing dietary profiles and health outcomes across groups, providing the insights into how specific dietary patterns influence health [17].

A priori composite scores, such as the Mediterranean diet score or the DASH score, are based on predefined dietary guidelines and recommendations [20,21]. These scores evaluate adherence to specific dietary patterns known to be beneficial for health, such as assessing the intake of key components of the Mediterranean diet, which has been linked to a reduced risk of cardiovascular disease [20] and improved cognitive function [3].

While these traditional dietary pattern analysis methods provide valuable insights into the associated health outcomes, there are limitations. One significant limitation is that they are unable to fully capture the complex interactions between different dietary components [22]. The beneficial effect of certain nutrients may be enhanced or inhibited by other dietary components, but these synergies are often hidden in these traditional analyses [23]. When dietary patterns are reduced to unidimensional scores, their multidimensional nature is disregarded. While these patterns may capture some synergies, this is only possible when interactions are explicitly recognised and incorporated during score development which is rare [24]. Moreover, in research focusing on individual nutrients or foods, interactions are often implicitly assumed to be non-existent in the model design [24]. Finally, these methods often assume dietary patterns are relatively static, ignoring potential changes in diet over time due to aging, economic changes, or health conditions [25]. These incorrect assumptions about interactions, or assumptions of staticity in model design can result in obscured or false associations and biased effect estimates.

In contrast the methods usually used to quantify dietary patterns, more prescriptive approaches such as linear optimisation have been instrumental in designing diets that meet specific targets for health, sustainability, or cultural appropriateness. Whilst beyond the scope of this review, these methods are fundamentally “knowledge-based,” meaning they can only optimise for the variables they already know about - the handful of macro- and micronutrients that have been well-characterised. This represents the “known knowns” of nutrition, which constitute less than 1% of the thousands of distinct bioactive compounds and phytochemicals present in our food chain. This inherent limitation means that such models are blind to the vast “nutritional dark matter” and the complex food synergies that are crucial for health.

To overcome the limitations of traditional dietary analysis, network approaches have emerged as a promising alternative, representing a superior, bottom-up alternative to knowledge-based prescriptive models like linear optimisation. Unlike the traditional methods which focus on individual nutrients and patterns in isolation [26], network analysis does not require comprehensive prior knowledge of every bioactive compound. Instead, it is a data-driven approach that learns directly from real-world eating behaviours. While also data-driven, this provides a key advantage over methods like PCA or cluster analysis; instead of reducing diet to composite scores or groups, network analysis explicitly maps the web of interactions and conditional dependencies between individual foods [12].

Methods such as Gaussian Graphical Models and Mutual Information Networks enable researchers to visualise and analyse these intricate relationships within a diet [27]. By mapping the

connections between foods and nutrients, these methods reveal how they collectively influence health outcomes and allow for the discovery of beneficial food combinations and protective synergies that emerge rather than a pre-defined biochemical model. Furthermore, dynamic or time-varying networks can model how dietary patterns change over time within individuals or populations, turning the complexity of our diet from a limitation into a source of discovery [28].

A variety of network algorithms have been developed, although not all have hitherto been applied to diet (Tables 2 and 3). Gaussian Graphical Models (GGMs) are probabilistic models that use partial correlations to identify conditional independence between variables. These models are particularly useful for exploring linear relationships in dietary data, offering insights into how one nutrient interacts with others while accounting for the broader dietary context. For example, GGMs can reveal whether the intake of saturated fats and sodium is conditionally independent given calorie consumption. This could help identify whether their relationship is direct or merely a byproduct of consuming high-calorie foods. This makes them valuable for understanding direct and indirect nutrient associations within diets. A limitation is that GGMs assume linear relationships, making them unsuitable for capturing the non-linear interactions that are often present in dietary data. For example, the effect of salt on hypertension may be moderated by the potassium and sugar content of the diet [29]. Additionally, GGMs are sensitive to non-normal distributions, which can distort the results in datasets with significant deviation [30].

Related to GGMs, Mixed Graphical Models (MGMs) accommodate datasets containing both continuous variables (e.g., nutrient intake) and categorical variables (e.g., demographic characteristics) [31]. This versatility is particularly useful for dietary studies that integrate diverse types of information. For example, MGMs can explore how continuous measures of dietary intake correlate with categorical socioeconomic factors such as education or income. By modelling these mixed data types jointly, MGMs expand the applicability of graphical models to more complex nutritional datasets, potentially yielding deeper insights into diet-health relationships. However, MGMs share several limitations with GGMs, including sensitivity to non-normal distributions for continuous variables [30].

Mutual Information (MI) Networks measure the amount of information shared between pairs of dietary components, capturing both linear and nonlinear associations [32]. This can uncover hidden patterns and relationships which may not have been found using traditional correlation-based methods [33]. For instance, by modelling nonlinear patterns, MI Networks can explore how sugar and fat intake interact to disproportionately influence obesity or cardiovascular risk, identifying subtle dependencies such as threshold effects that might be missed by simpler models. However, a significant limitation is that MI algorithms usually give rise to denser networks reducing interpretability and the ability to tease apart direct and indirect dependencies.

Bayesian Networks (BN) are probabilistic graphical models that represent the relationships between variables through directed acyclic graphs, enabling the identification of potential causal pathways [34]. BNs have not yet been applied to dietary pattern analysis however, unlike other traditional correlation-based methods, BNs provide insights into causality [34] which may make them a powerful tool to explore how changes in dietary components may influence one another. One possible application of BNs is to model the fat-sugar seesaw phenomenon, where reducing fat tends to lead to an increase of sugar in diet [35]. One advantage of BNs is their ability to incorporate prior knowledge into the model structure [36], potentially enhancing the interpretability and plausibility of the derived dietary network. There are limitations to BNs, in particular the computational difficulty of exploring a previously unknown network [37].

Dynamic Networks incorporate time-varying dependencies [38], enabling researchers to observe how dietary patterns and meal compositions evolve over time. This approach allows the study of how diets respond to external factors, such as seasonal changes or price increases, and how interventions might alter established dietary habits [39]. One possible application of Dynamic Networks is to understand how patterns of co-consumption may change following dietary interventions. For example, if a person tends to consume meat with vegetables and an intervention

reduces meat consumption, Dynamic Networks could reveal whether this intervention also leads to an unintended reduction in vegetable intake. By modelling these changes, the unintended consequences of public health policies can be predicted, such as how promoting plant-based diets may inadvertently decrease the intake of other beneficial food groups. A limitation of Dynamic Networks is the need for detailed longitudinal monitoring of diet over time; this often involves resource intensive data collection methods, such as repeated dietary recalls or food diaries over extended periods of time.

Hypergraphs extend traditional graph theory by allowing edges, known as hyperedges, to connect more than two nodes [40], making it possible to represent group-level interactions or clusters. Standard graph models only consider the pairwise interactions, such as edges between nutrients or foods, while hypergraphs can account for higher-order interactions. This ability to model interactions involving multiple nodes would be particularly useful in dietary pattern research as multiple nutrients often work together with shared function to influence health outcomes. For example, a hyperedge in a hypergraph could represent a meal containing protein, fats and carbohydrates where the combined impact on health emerges from the interplay between these nutrients and cannot be explained by pairwise interactions alone. One limitation of hypergraphs is that their high computational demand can make them resource-intensive [41], especially for large datasets. Another limitation is that the complexity of hypergraphs often makes them hard to interpret [42], reducing their accessibility for researchers who are less familiar with advanced network methods.

Multilayered Graphs represent systems with multiple interconnected layers, each capturing distinct but related interactions [43]. For example, one layer may represent nutrient interactions, another food-level relationships, and another may represent food context. These graphs allow for a comprehensive exploration of both intra-layer and inter-layer connections [43]. Analysing cross-domain relationships is valuable in nutrition research as the impact of food context on nutrient intake can be examined. However, Multilayered Graphs are computationally demanding, especially for large datasets spanning multiple domains. They are also complex and can be challenging to interpret [43].

The choice of method depends on the research question, data type, and available computational resources. By leveraging these diverse tools, researchers can gain deeper insights into dietary patterns, overcoming some of the limitations of traditional dietary pattern analysis (Table 2).

The purpose of this paper is to provide an overview of the existing studies using network approaches for dietary pattern analysis. Specifically, the objectives of this scoping review were to (1) provide a clearer understanding of the strengths and limitations of traditional dietary assessment methods, (2) identify the potential advantages of network approaches, (3) determine whether network approaches can offer a more nuanced analysis of dietary intake data, (4) create guiding principles for the use of network approaches in dietary pattern analysis.

Table 2. Network methods for dietary analysis.

Method	Algorithm	Linear/Nonlinear	Assumptions	Strengths/Limitations
Gaussian Graphical Models (GGM)	Inverse covariance matrix estimation	Linear	Assumes normally distributed data, linear relationships, sparsity.	Measures the conditional dependencies between different foods. Reveals how certain foods are commonly consumed together, or how foods may displace each other in the diet. Can increase understanding how variables (e.g., foods, nutrients) directly interact, independent of others in the context of the whole diet. Relies on partial correlation matrix and is sensitive to non-normally distributed data.
Mixed Graphical Models (MGM)	Combination of GGM and discrete modelling techniques	Both	Assumes mixed data types can be represented in a joint network, requires sparsity.	Can identify direct relationships while accommodating diverse variable types. Standard MGMs assume linear relationships but with extensions such as kernel methods non-linear models can be developed.
Mutual Information Network	Information-theoretic methods	Nonlinear	No strict distributional assumptions, assumes mutual information represents dependence.	Uses entropy-based measures to quantify shared information. Reveals how certain foods are commonly consumed together, even in non-linear relationships (e.g., nutrient thresholds or diminishing returns). Similar to GGM but without normality assumption. Does not differentiate direct and indirect associations.
Bayesian Networks (BN)	Directed acyclic graphs	Both	Assumes probabilistic relationships between variables.	Provides insights into causality and allows the exploration of causal pathways. Can incorporate prior knowledge for enhanced interpretability. Computationally intensive when discovering unknown networks.
Dynamic Networks	Time-varying graph algorithms	Both	Requires longitudinal data with high temporal resolution.	Models time-varying dietary patterns and tracks changes in diet over time. Useful for predicting unintended consequences of interventions.

				Requires resource-intensive longitudinal data collection for accurate analysis.
Hypergraphs	Hyperedge-based graph algorithm	Both	Assumes interactions can involve more than two nodes.	Captures higher-order interactions. Useful for modelling the combined health impact of foods/nutrients which are unable to be explained by pairwise interactions. Computationally demanding and resource intensive. Complexity may affect interpretability.
Multilayered Graphs	Layered network construction	Both	Assume information is shared between all layers.	Enables analysis of intra- and inter-layer connections. Valuable for cross-domain analysis. Computationally demanding and complex. Challenging to interpret for large datasets.

Table 3. Characteristics of the eligible studies.

Author (Year)	Population	Dietary Assessment	Network Model	Aims	Findings
Slurink et al. (2023) [61]	74132 participants (59.7% female) Lifelines cohort study	Flower-FFQ	Mixed graphical model	To investigate associations of total dairy and dairy types with incident prediabetes. To assess how dairy intake is linked with metabolic risk factors, lifestyle behaviours, and foods, as potential explanations for these associations.	Low fat milk intake associated with higher prediabetes risk. High-fat yogurt intake had nonsignificant inverse association with prediabetes risk. Heterogenous associations by dairy type and fat content may be due to confounding caused by behaviours and food intake related to dairy intake.
Schwedhelm et al. (2021) [52]	365 women, 12 weeks gestation Chapel Hill Healthcare System, North Carolina	3 x Automated Self-Administered 24-hour dietary recalls	Semiparametric Gaussian copular graphical model (SGCGM)	To investigate food networks across meals in pregnant women. To explore differences by overall diet quality classification.	Food combinations differed by meal and between dietary quality tertiles. Meal-specific patterns which differed between diet quality tertiles: <ul style="list-style-type: none"> Intake of vegetables, whole-grain bread, cooked grains and nuts at

					breakfast in high diet quality group.
					<ul style="list-style-type: none"> Sugar sweetened beverages, sandwiches, and fried potatoes at all main meals in low diet quality group.
Felicetti et al. (2022) [56]	424 participants with MS (67% female) 165 healthy controls (68% female) Sant'Andrea Hospital, Rome	MeDi adequacy questionnaire	Mutual information	To investigate food networks across meals in people with multiple sclerosis (PwMS) and healthy controls (HC). To explore differences by overall diet quality classification.	Fruit, vegetables, cereal, and fish were identified as hubs in PwMS. Meat and alcohol identified as hubs in HC. PwMS showed overall healthier dietary pattern than HC. Vegetables and fish intake associated with disability outcomes; higher disability status, lower vegetable and fish intake.
Samieri et al. (2020) [33]	1522 participants (73.7% female, 209 with dementia) 3C study	FFQ	Mutual information	To use network science to model complex diet relationships a decade before onset of dementia in a large French cohort.	Food networks substantially differed between cases and controls. Cases had charcuterie as the main hub, with connections to foods typical of French southwestern diet and snack foods. Controls had several disconnected subnetworks reflecting diverse and healthier food choices.
Jayedi et al. (2021) [50]	850 participants (69% female) Tehran, Iran	FFQ	Gaussian graphical model (GGM)	To describe dietary networks identified by GGM, representing patterns of dietary intake in a sample of Iranian adults.	Identified 3 dietary networks – healthy, unhealthy, saturated fats. Cooked vegetables, processed meats, and butter central to networks, respectively.

				To investigate the potential associations of these dietary patterns with general and abdominal adiposity.	Top tertile of saturated fats network score associated with higher likelihood of central obesity by waist-to-hip ratio. No association between dietary network scores and general obesity.
Iqbal et al. (2016) [55]	27120 participants (60% female) EPIC cohort FFQ		GGM (results confirmed through SGCGM)	To apply GGMs to derive sex-specific dietary intake networks representing consumption patterns in a German adult population.	Men – 1 major dietary network consisting of red meat, processed meat, cooked vegetables, sauces, potatoes, cabbage, poultry, legumes, mushrooms, soup, and whole-grain breads Women – similar network with addition of fried potatoes.
Schwedhelm et al. (2018) [51]	814 participants (49.5% female) EPIC cohort 3 24-hour recalls		SGCGM	To estimate and describe meal and habitual dietary networks derived through SGCGMs. To compare relations found in meal networks to the ones present in the habitual network.	Breakfast network – 5 communities of food groups Lunch and afternoon snacks network – higher variability in food consumption, 6 communities in each networks Dinner network – 2 networks with 5 communities Meal-specific dietary network only partly reflected in habitual network; analysing food consumption on habitual level did not exactly reflect meal level intake.
Gunathilake et al. (2022) [59]	397 participants with cancer (61.5% female), 7477 participants without cancer (63% female) FFQ		GGM (also used PCA and RRR)	To investigate the association between dietary communities identified by a GGM and cancer risk.	GGM identified 17 and 16 dietary communities for total and matched populations. For each one-unit increase in SD of community-specific score of community

	Cancer screening Cohort, South Korea				composed of dairy products and bread, there was a reduced cancer risk. Matched population – poultry, seafood, bread, cakes and sweets, and meat byproducts showed significantly reduced risk of cancer.
Iqbal et al. (2019) [60]	22245 participants (61% female) EPIC-Potsdam cohort	FFQ	GGM (also used PCA)	To investigate the association between previously identified GGMs food intake networks and risk of major chronic diseases as well as intermediate biomarkers in the EPIC-Potsdam cohort.	Higher adherence to GGM Western-type pattern associated with increased risk of type 2 diabetes in women. Adherence to high-fat dairy pattern associated with lower risk of type 2 diabetes in both men and women.
Hoang et al. (2021a) [47]	1049 participants with cancer (76% female), 9728 participants with no cancer (64% female) Korea Cancer Screening cohort	FFQ	GGM	To identify major dietary patterns of Korean adults using a GGM. To examine the associations between DP scores and prevalence of self-reported cancer.	Identified 4 networks – principal, oil-sweet, meat, and fruit. Odds of moderate and high consumption of foods in oil-sweet DP for cancer patients were 25% and 34% lower than those with no reported cancer diagnosis. Odds of meat DP consumption was 29% for cancer patients. Increase in odds of fruit DP consumption observed for cancer patients.
Jahanmiri et al. (2022) [49]	850 participants (69% female) Tehran, Iran	FFQ	GGM	To derive dietary networks and assess their association with metabolic syndrome.	3 networks – healthy, unhealthy and saturated fats. Adherence to saturated fats network with centrality of butter associated with higher odds of having metabolic syndrome and higher odds of having hyperglycaemia.

					No significant association observed between healthy and unhealthy networks with metabolic syndrome, hypertension, hypertriglyceridemia, and central obesity.
Aguirre-Quezada and Aranda-Ramírez (2024) [45]	230 students Azogues, Ecuador	FFQ	GGM	To apply GGMs to derived specific networks for groups of healthy and unhealthy obese individuals that represent the nutritional, psychological, and metabolic patterns in an Ecuadoran population.	Higher carbohydrate intake is associated with lower protein intake. Intake of fibre, proteins, carbohydrates, and fats showed positive relationship with BMI for metabolically unhealthy obese individuals.
Hoang et al. (2021b) [48]	7423 participants (35% female) Cancer Screening Examination cohort	FFQ	Mixed Graphical Model	To elucidate the complex interrelatedness among dietary intake, demographics, and risk of comorbidities.	Normal and heavy eating significantly associated with increases of at least 20% in the risks of elevated BP, hypertension, and mild kidney impairment.
Landaeta-Díaz et al. (2023) [58]	1242 participants (76.6% female) Chile	FFQ	GGM	To explore food networks in the Chilean adults sample and in people with anhedonia symptoms.	Intake of fruits, vegetables, and fast foods has central role in sample of Chilean adults. Fruit consumption positively associated with vegetables, negatively associated with fast food. Direction of association maintained in those with anhedonia. Stronger association and central place in network for “pasta, rice & potatoes” and “bread” for anhedonia network
Xia et al. (2020) [53]	2043 matched controls (31% female) for 2043 newly	FFQ	GGM	To construct dietary networks from network science.	Two major networks in case group. One major network in control group.

	diagnosed non-alcoholic fatty liver disease (30% female)			To explore the associations between complex dietary networks and non-alcoholic fatty liver disease.	Results suggest dietary structures are different between case and control groups.
Fereidani et al. (2021) [46]	134 women with breast cancer, 266 hospital controls Tehran, Iran	FFQ	GGM	To compare food intake networks derived by GGMs for women with and without breast cancer to better understand how foods are consumed in relation to each other according to disease status.	On both principal networks, vegetables, fruits, sweets and fried potatoes were central food groups. For cases, main network consisted of 9 central food groups. For controls, main network consisted of 5 central food groups. Network of cases showed more conditional dependencies between intakes of food groups compared to controls.
Gunathilake et al. (2020) [54]	415 gastric cancer cases (35% female), 830 controls (35% female) NCC Hospital, Korea Cancer Screening cohort	FFQ	GGM	To apply GGMs to identify dietary patterns. To investigate the associations between dietary patterns and gastric cancer risk in a Korean population.	Vegetable and seafood network and fruit network associated with decreased risk of GC for whole study population. Those in highest tertile of vegetable and seafood network-specific score had a reduced risk of GC compared to those in lowest tertile.
Gunathilake et al. (2021) [57]	268 patients with GC (36% female), 288 healthy controls (37% female) NCC Hospital, Korea	FFQ	GGM	To observe the combined effects of GGM-derived dietary patterns and the gastric microbiome on the risk of gastric cancer in a Korean population.	Vegetable and seafood pattern may interact with dysbiosis to attenuate the risk of GC in males. Dairy pattern may interact with dysbiosis to reduce GC risk in females.

2. Methods

2.1. Search Strategy and Selection Criteria

This scoping review aims to identify and critically evaluate literature on the application of networks to dietary data. Studies were included if they used network analysis to analyse dietary data for human participants. In particular, the network models searched for included Gaussian graphical models, mutual information networks, multilayered networks, hypergraphs, and mixed graphical models. Studies were excluded if network analysis was used for metabolics or systems biology.

A comprehensive search for studies written in the English language and published up to 7th March 2025 was carried out using PubMed, Scopus, and PsycINFO. The following search terms were used: “network approach” or “network analysis” or “network method” or “graphical model” and “gaussian graphical model” or “GGM” or “mutual information network” or “mixed graphical model” and “dietary analysis” or “dietary data” or “nutrition analysis” or “food intake” or “diet” or “nutrition.” Google Scholar was used to obtain additional articles identified by journal hand searching. After deduplication, titles and abstracts were read to assess eligibility based on predefined inclusion criteria. Irrelevant articles or duplicates were excluded. Full texts of the remaining articles were read to verify their suitability. Reference lists from the articles deemed suitable were checked for additional studies. This systematic search was conducted independently by two reviewers (RT and JM); any disagreements were resolved by discussion.

2.2. Data Extraction

Data was extracted by RT, and checked by JM, from the eligible articles using a standardised spreadsheet: author and year of publication, study aims, participant information (age, gender, ethnicity), dietary assessment used, approach to derive dietary patterns, network model used, and appropriateness of the network model.

Consistent with PRISMA-ScR and JBI guidance, we did not appraise risk of bias because our aim was to map methodological characteristics, not to evaluate intervention effects.

3. Results

3.1. Search and Selection of Network Studies

The search conducted in March 2025 identified 171 studies potentially relevant to the review. After deduplication and relevance screening based on language, dietary data collection methods, and network analysis approaches, 19 studies met the eligibility criteria for a full-text review. One study did not have an English translation so was not included in the review [44]. The remaining 18 articles were read, and all were deemed relevant for this scoping review. The flow of studies from identification to final inclusion is represented in Figure 1.

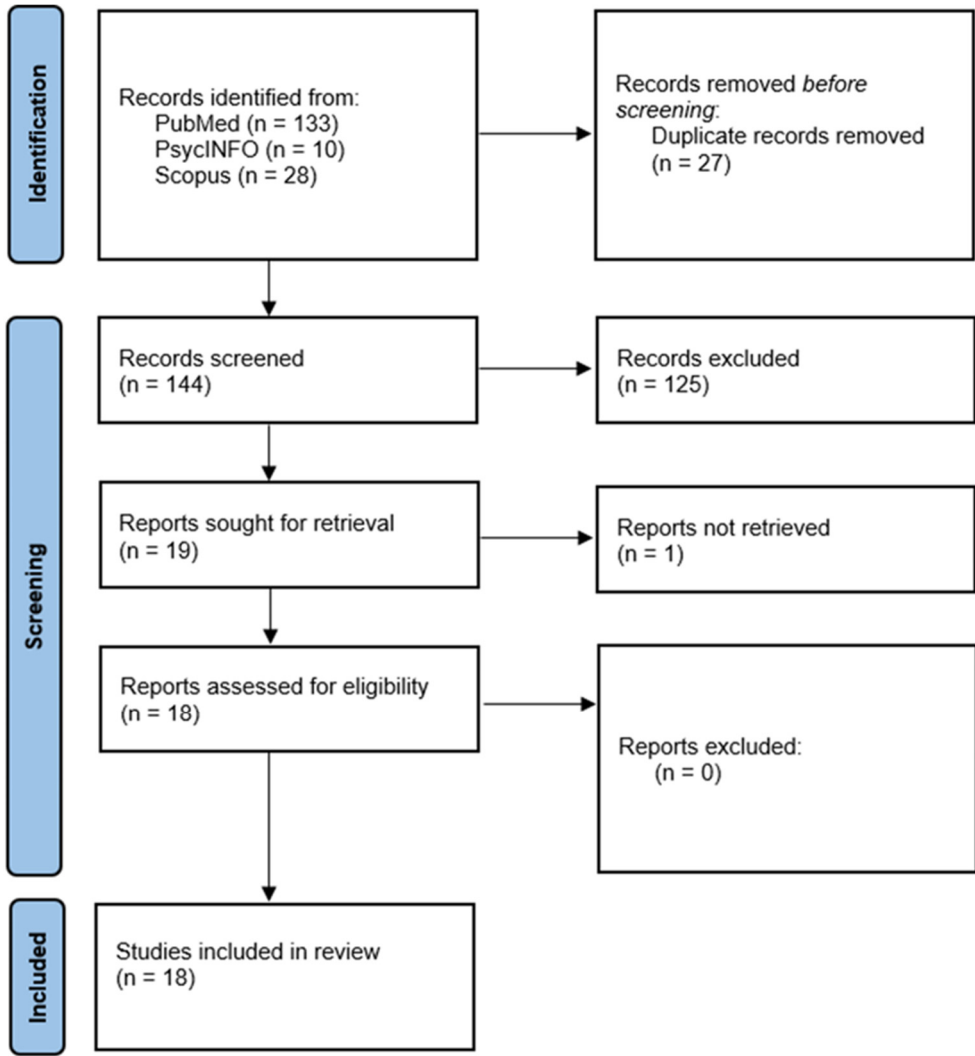


Figure 1. PRISMA flow diagram of literature screening and selection process.

3.2. Study Characteristics of Included Network Studies

The general characteristics of the network studies included in this review are summarised in Table 3. All studies were published between 2018 and 2024. Most (79%) studies used FFQs to collect their dietary data, while two studies used 24-hour dietary recall, and one study used a Mediterranean adequacy questionnaire.

Nine studies analysed data from large, pre-existing cohorts: four from the Cancer Screening cohort in South Korea, three from the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort, one from the Lifelines cohort, and one from the 3C study. The remaining nine studies used smaller, non-cohort-based samples. Sample sizes varied from 230 participants to 74 132 participants.

Regarding network approaches, eleven studies used Gaussian graphical models, with one of these confirming their results with a semiparametric extension – semiparametric Gaussian copular graphical model (SGCGM). Two studies used only SGCGM. Three studies used mutual information matrices. Two studies used mixed graphical models.

A wide range of health outcomes were analysed, highlighting the versatility of network approaches in dietary research. Cancer was the most frequently studied outcome, with five studies varying on various cancer types, including gastric and breast cancers. Other physical health outcomes analysed were incident prediabetes, metabolic syndrome, obesity, anhedonia, non-alcoholic fatty liver disease, and diet quality during pregnancy. Neurological disorders were also explored, with one study on dementia and one on multiple sclerosis. Additionally, one study compared dietary

patterns between men and women, and another examined differences between meal-specific and habitual dietary networks.

3.3. Adherence to Methodological Best Practices

The remainder of this results section provides a thematic analysis of the included studies, evaluating their adherence to several key methodological best practices for network analysis (Table 4).

3.3.1. Justifications for Using Network Models

The included studies justified their use of network analysis in several ways. The most common rationale was to counter the known limitations of traditional methods, a justification made by nine of the reviewed studies [45–53]. For instance, authors noted that network analysis was chosen because Principal Component Analysis (PCA) can explain only a small proportion of the variability in food intake [52] and fails to demonstrate pairwise correlations between food groups [50].

Another common theme was the use of network analysis to complement existing research and gain additional insights into complex data [33,54,55]. One study framed its use as complementary without explicitly mentioning the limitations of traditional methods [56]. Two studies specifically focused on using network analysis to overcome the limitations of diet scores, allowing for an assessment of diet as a pattern rather than a sum of single food items [57,58].

Furthermore, three studies used network analysis directly alongside traditional dietary pattern analysis methods [59–61]. This was often done to compare the dietary patterns identified by each approach, for example by using network analysis to derive dietary networks and then using PCA and reduced rank regression (RRR) for comparison [59], or by comparing new network-derived patterns to PCA patterns from the same cohort in an earlier study [60].

3.3.2. Study Design and Causal Inference

The vast majority of the reviewed studies utilised a cross-sectional design. Acknowledging the limitations of this approach, most of these studies (13 out of 18) did not attempt to make inferences about causality [33,46,47,50–55,57,58,60,61]. Among these, three studies explicitly stated that the use of cross-sectional data was the reason for this caution [48,49,56]. One paper acknowledged the cross-sectional design as a limitation but did not specify why [50]. In contrast, only one paper attempted to make inferences about causality from its cross-sectional data [59].

3.3.3. Network Estimation and Regularisation

The reviewed studies employed several techniques to estimate their networks and control for spurious connections, with approaches varying by the chosen network model.

Among the studies using Gaussian Graphical Models (GGMs), LASSO regularisation was the most common approach, used in thirteen of the fourteen GGM-based studies [45–47,54,57,59] [49–52,55,60,61]. Only one GGM study did not employ a regularisation method [58]. However, the transparency in reporting the specific LASSO tuning parameter was inconsistent; just seven of these thirteen studies provided this detail [46,47,54,55,57,59,61], while one study explored network structures across a range of different tuning parameters [45].

For the three studies that used mutual information networks, all applied thresholding to reduce network density [33,53,56], and two of these also used permutation testing to retain only statistically significant connections [33,56].

Regarding the stability and novelty of the findings, two studies explicitly refrained from drawing strong inferences from their results [56,60], while two others, noting they were the first in their specific populations, did make inferences from their findings [50,52].

Table 4. Adherence of studies to the guiding principles of using network analysis for dietary pattern analysis.

Author (Year)	Justification for Using Network Models	Study Design and Causal Inference	Network Estimation and Regularisation	Use of Centrality Metrics	Handling of Non-Normal Data
Slurink et al. (2023) [61]	Used network approach to aid interpretation of regression models by accounting for interrelatedness of risk factor. Holistic approach to aid traditional reductionism methods.	Does not attempt to make inferences about causality.	Notes that the weaker connections for lifestyle risk factors and food groups being compared to clinical markers may reflect a greater extent of measurement uncertainty. LASSO regularisation method used, tuning parameter 0.5.	Uses centrality metrics without discussing limitations. Uses strength as it has “shown the greatest stability of centrality indices.”	N/A
Schwedhelm et al. (2021) [52]	Discusses limitations of PCA (only explains small proportion of variability of food intake) and why GGM is a better alternative (reveals patterns of food group combinations specific to each meal).	Does not attempt to make inferences about causality.	Makes inferences despite being the first study to examine associations between foods consumed within meals during pregnancy. Used graphical LASSO method.	Does not use centrality metrics.	Addresses this by using a semiparametric extension of GGM and excluding food groups consumed in fewer than 5% of meals to avoid over-representation of the relationship between episodically consumed foods eating together on a few occasions.
Felicetti et al. (2022) [56]	Network analysis to see complex relations hidden in eating behaviour. Complementary to other research on dietary habits.	Acknowledged it used cross-sectional data which prevents conclusions being drawn about direct causality.	Does not attempt to draw inferences for clinical research.	Does not use centrality metrics.	N/A
Samieri et al. (2020) [33]	Network methods to provide complementary information to other approaches to gain additional insights into food-disease associations and patterns involved in reducing dementia risk.	Does not attempt to make inferences about causality.	Compared the strengths of associations (edge-weights) between the two networks.	Uses centrality metrics without discussing its limitations	N/A

Jayedi et al. (2021) [50]	Discusses limitations of PCA (does not demonstrate pairwise correlations between food groups). GGM to show how food groups are consumed in relation to one another.	Addresses that cross-sectional design is a limitation but does not say why.	Makes inferences despite being first study to investigate association between GGM networks and general and abdominal adiposity in adults. Used graphical LASSO.	Evaluates food groups belonging to more than one community for centrality to determine importance of a food group based on a number of communities it belongs to. Does not discuss limitation of centrality metrics.	Mentions that GGMs assume a multivariate normal distribution for underlying data. Does not say this is a limitation or attempt to fix it.
Iqbal et al. (2016) [55]	Limitations of existing methods of dietary pattern analysis warrant investigation of complementary approaches.	Does not attempt to make inferences about causality.	Used network stability analysis (repeated bootstrapping 80% of original sample with replacement) – found that identified networks were stable in current population. Used graphical LASSO, tuning parameter 0.25.	Does not use centrality metrics.	Addresses this limitation. Log-transformed all the data and confirmed results of GGMs with SGCGMs.
Schwedhelm et al. (2018) [51]	When using traditional methods, understanding of how dietary patterns arise from food intake is limited.	Does not attempt to make inferences about causality.	Similarities between previous GGMs using participants data from EPIC-Potsdam cohort which used FFQ instead of 24-hour recalls. Used LASSO using cross-validation.	Used centrality metrics to assist interpretation. Did not discuss limitations.	Addresses this limitation. Uses SGCGMs instead of GGMs.
Gunathilake et al. (2022) [59]	Used GGM to derive communities of dietary networks and evaluate contributions of dietary networks to development of cancer. Used PCA and RRR to compare the identified DPs.	Attempts to make inferences about causality despite using between-person cross-sectional data.	Used LASSO regularisation. Optimal λ values for total study population and matched subgroup were 0.32 and 0.34.	Uses centrality metrics without discussing its limitations.	Log-transformed the weight intake to improve the normal distribution.
Iqbal et al. (2019) [60]	Use GGM to investigate relationship between dietary patterns and risk of chronic diseases.	Does not attempt to make inferences about causality.	Does not attempt to draw inferences. Details of regularization were not reported in this paper and referred to a previous publication - graphical LASSO.	Uses centrality metrics without discussing its limitations.	Applied GGM to log-transformed intakes of food groups.

Reconstructed PCA patterns to compare the approaches.					
Hoang et al. (2021a) [47]	Discusses limitations of PCA and RRR, strength of one is a limitation of the other. Used GGM to resolve this issue.	Identified that the “cross-sectional design was not strong enough to identify the temporal and causal relationships between dietary intake and cancer development.”	Makes inferences. LASSO regularisation used, optimum values of 0.48 for total study population, 0.52 for male subgroup, 0.46 for female subgroup.	Uses centrality metrics without discussing its limitations.	Weight intake values of the food groups were log-transformed to improve normality distribution.
Jahanmiri et al. (2022) [49]	PCA and cluster analysis are reduction techniques. GGM a “commanding method” for DPA.	Acknowledged it used cross-sectional data which prevents conclusions being drawn about a cause-and-effect relationship.	LASSO regularisation used.	Uses centrality metrics without discussing its limitations.	Discusses that the data needs to be “Gaussian-distributed” which is not possible for all variables. Does nothing to counter limitation.
Aguirre-Quezada and Aranda-Ramírez (2024) [45]	Previous studies have limitations in analysis. GGMs provide a comprehensive and easy to understand overview of relationships between variables.	Does not attempt to make inferences about causality.	Graphical LASSO used. Does not choose a regularisation parameter, instead they repeatedly create networks for the different parameters.	Uses centrality metrics without discussing its limitations.	Acknowledges that data should follow a Gaussian distribution which is not met by all variables. Does nothing to counter limitation.
Hoang et al. (2021b) [48]	Conventional approaches have limitations in explaining complex relationships. Network analysis to provide insights into interactions among all variables and explore how a single variable is impacted by multiple factors.	Acknowledged that it used cross-sectional design which may not have allowed for a full investigation of a causal relationship.	LASSO regularisation with extended Bayesian information criteria selection applied and set at 0.5. Network accuracy assessed by bootstrapping 80% of original sample with a replacement.	Uses centrality metrics without discussing its limitations.	N/A
Landaeta-Díaz et al. (2023) [58]	Previous studies used diet scores which has a limited scope as focuses on representing a conceptual diet but does	Does not attempt to make inferences about causality.	Does not use any regularisation techniques.	Uses centrality metrics without discussing its limitations.	Does not acknowledge any limitations of using GGM. Does nothing to fix the limitations.

	not show how food relates to each other. GGM represents underlying structure of food groups.				
Xia et al. (2020) [53]	Limitations of traditional statistical methods. Network science can help to discover the potential role of food groups in overall dietary pattern, providing new insight into complexity and non-linearity of dietary patterns.	Does not attempt to make inferences about causality.	Plots were limited to edges with inferred weight $\geq .30$ for better interpretability.	Does not use centrality metrics.	N/A
Fereidani et al. (2021) [46]	Limitations of existing methods. GGM patterns show how foods are consumed in different combinations.	Does not attempt to make inferences about causality.	Used graphical LASSO. λ value of 0.3 chosen and applied for all analyses.	Does not use standard named centrality metrics (e.g., betweenness, closeness). Defines central food groups as those with a high correlation with 4 or more other food groups.	Dietary data log-transformed to improve normality.
Gunathilake et al. (2020) [54]	Complementary strategy for investigating diet and disease relationships.	Does not attempt to make inferences about causality.	Used graphical LASSO. Optimum λ value of 0.38.	Used strength as centrality metric, did not discuss limitations.	Discusses that data needs to follow a Gaussian distribution so log transformation should be applied.
Gunathilake et al. (2021) [57]	Wanted to assess dietary intake as a pattern rather than a sum of single food items.	Does not attempt to make inferences about causality.	Used graphical LASSO. Optimum λ value of 0.37.	Used strength as centrality metric, did not discuss limitations.	Log-transformed the dietary intake values.

3.3.4. Use and Interpretation of Centrality Metrics

The application of centrality metrics to identify important foods or nutrients was inconsistent across the reviewed literature. Five of the eighteen studies avoided using centrality metrics in their analysis [46,52,53,55,56]. In contrast, the majority of studies (13 out of 18) did employ centrality metrics, and none of these studies acknowledged or discussed the potential limitations of this approach in the context of dietary network analysis [33,47–51,54,57–61]. For instance, one study examined node centrality using strength, betweenness, and closeness, ultimately opting to use the strength metric for its final analysis [61].

3.3.5. Handling of Non-Normal Data

The studies employing Gaussian Graphical Models (GGMs) used three distinct strategies to address the assumption of normally distributed data. The most robust approach was to use a non-parametric extension of the GGM; two studies used the semiparametric Gaussian copular graphical model (SGCGM) exclusively [51], and a third study used SGCGM to confirm the results of their primary GGM analysis [55].

The most common strategy was data transformation, with six studies applying a log-transformation to their data to improve normality [46,47,54,57,59,60].

Finally, three of the GGM studies did not apply any correction for non-normal data. Of these, two acknowledged the issue as a limitation but did not address it [49,50], while one study did not acknowledge the limitation at all [58].

4. Discussion

This scoping review provides a comprehensive overview of the emerging field of dietary network analysis. The 18 identified studies, all published since 2018, demonstrate the versatility of these methods across a wide range of populations and health outcomes, from cancer to neurological disorders. Our findings show that researchers are actively employing network analysis to overcome the recognised limitations of traditional methods and to gain more nuanced insights into complex food co-consumption patterns. However, our thematic analysis also reveals significant methodological heterogeneity and several recurrent challenges across the literature. These include a heavy reliance on cross-sectional data which limits causal inference, inconsistencies in network estimation and reporting, and the widespread, often uncritical, application of analytical techniques - such as centrality metrics and simple data transformations - that may be inappropriate for complex dietary datasets. These identified challenges highlight the need for a more standardised approach. Therefore, the remainder of this discussion will use these findings to establish a set of guiding principles designed to enhance the rigour and reliability of future research in this promising field.

4.1. Guiding Principles for Future Research

One objective of this review was to create guiding principles for the use of the network approach in dietary pattern analysis. This was achieved by using the critiques of using network analysis to analyse multivariate data [62] and applying them to dietary pattern analysis. A summary of these guiding principles can be found in Figure 2.

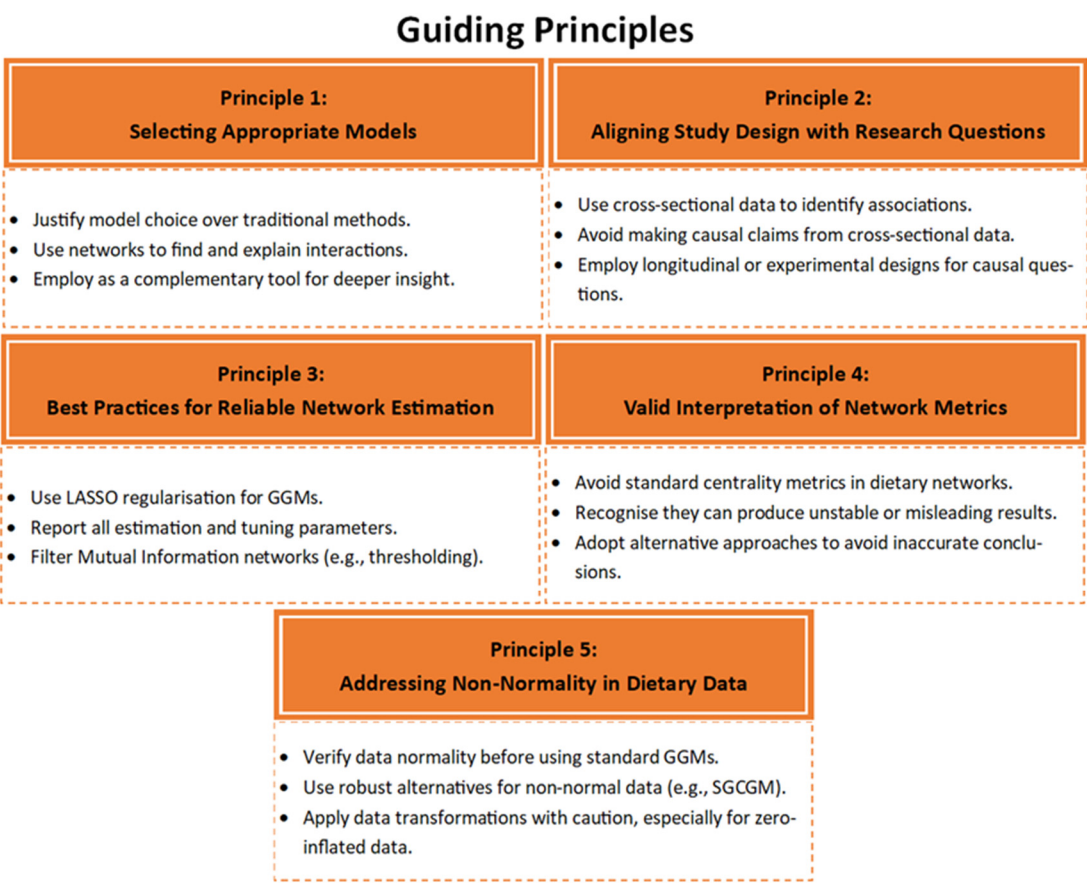


Figure 2. Five guiding principles for research using network approaches to dietary pattern analysis.

Principle 1: Selecting Appropriate Models

Researchers should only use network analysis when it addresses specific limitations of traditional multivariate methods or provides complementary insights that align with the research question [62]. Using network analysis to counter the limitations of widely used dietary pattern analysis methods and to complement the data obtained by these methods is effective in building on the existing knowledge to uncover interactions that have previously been concealed. It is critical, however, to ensure that network analysis is the most appropriate method for the research question at hand.

Principle 2: Aligning Study Designs with Research Questions

Researchers need to avoid making causal inferences from between-person cross-sectional data and ensure that the study design aligns with the research objective. When using between-person cross-sectional data, it limits the inferences that can be made about causation, as correlations among variables do not indicate causality [62]. It is important to refrain from making any inferences about causation, as it is not possible to use between-person statistical associations to draw conclusions about how food groups are related for a given individual [62]. Making inappropriate causal inferences can lead to misleading conclusions about relationships between food groups and health outcomes. For causality-focused research questions, longitudinal data may be better suited.

Principle 3: Best Practices for Reliable Network Estimation

The robustness of estimated networks can be questionable, as networks may appear similar in their global characteristics while their detailed characteristics vary substantially [62,63]. Therefore, inferences should not be drawn before results have been rigorously replicated [62]. To create more reliable and interpretable networks, specific best practices should be followed.

When using GGMs, robust methods such as LASSO regularisation should be used to minimise spurious connections, as this technique limits the number of spurious edges to obtain more interpretable networks [64]. It is also important for researchers to provide details of their tuning parameter choices to ensure reproducibility, as this parameter controls the level of sparsity in the network [64]. Similarly, when using mutual information networks, techniques like thresholding (retaining edges above a certain value) and permutation testing (recalculating mutual information on randomised data to keep only significant connections) should be used to create a more robust network with fewer spurious edges [65,66].

Principle 4: Valid Interpretation of Network Metrics

Researchers should avoid relying on centrality metrics (e.g., degree, closeness, betweenness) for dietary networks. Most network centrality metrics were developed for bounded networks, where the nodes are clearly defined and fixed [62]. Dietary pattern analysis, however, often involves unbounded networks, which are dynamic and do not have a fixed number of nodes or connections. When centrality metrics are applied to unbounded networks, their interpretability is compromised, making them unstable and unreliable [62]. This can lead to inaccurate conclusions. For example, one study concluded that “a low centrality predictability of low-fat milk intake in the networks may indicate that significant associations in the regression analysis could be due to intake levels coinciding with influential risk factors” [61]. However, given the limitations of centrality metrics in unbounded networks, this conclusion may not be valid. While the choice of the ‘strength’ metric in that study was based on its relative stability, this does not negate the broader issue that these metrics can produce misleading information [67]. Therefore, researchers should either avoid centrality metrics entirely or adopt alternative approaches that account for the unbounded nature of dietary networks.

Principle 5: Addressing Non-Normality in Dietary Data

When using GGMs, researchers must address the fact that dietary data is rarely normally distributed, which violates a core assumption of the model [68]. Dietary data is often skewed and zero-inflated (i.e., contains many “no consumption” reports), which presents significant challenges. While transformations such as log-transformation can mitigate some skewness, they are not a robust solution for zero-inflated data [69]. Adding a small constant to avoid errors with zero values can distort the data and complicate the interpretation of the results [70]. More appropriate approaches should be employed, such as using the semiparametric Gaussian copular graphical model (SGCGM), which is a non-parametric extension designed for skewed data. Alternatively, opting for a different method entirely, such as a mutual information network or mixed graphical model, may be a better choice when dealing with non-normally distributed data.

4.2. Strengths and Limitations of This Review

A key strength of this review is its systematic and rigorous methodology. We conducted a comprehensive search across multiple relevant databases and followed the PRISMA-ScR guidelines for scoping reviews. The inclusion of two independent reviewers for the study selection process minimises the risk of bias and enhances the reliability of our findings. Furthermore, this review moves beyond a simple summary of the literature. By critically evaluating the included studies against an established framework of best practices and generating a novel set of guiding principles, this work provides a tangible roadmap for future researchers in this emerging field.

Despite these strengths, several limitations should be acknowledged. First, as a scoping review, our aim was to map the field rather than conduct a formal quality assessment or risk-of-bias analysis for each individual study; therefore, the quality of the primary studies was not formally appraised. Second, our review may be subject to publication bias, as studies with null or non-significant findings may be underrepresented in the published literature. Additionally, our search was limited to English-language publications, however, only one relevant study was excluded because no English translation was available. Finally, dietary network analysis is a rapidly evolving field. While our search was comprehensive up to its cut-off date, new studies will have been published in the interim.

4.3. Conclusion

In conclusion, network analysis offers a powerful and promising alternative to traditional methods for dietary pattern research, allowing for a more holistic understanding of food synergies and co-consumption. However, our scoping review has revealed that the application of these methods is currently marked by significant methodological heterogeneity and a frequent lack of adherence to best practices, particularly concerning study design, network estimation, and the interpretation of metrics. The guiding principles established in this review provide a foundational framework to address these challenges. By adopting these principles, future research can harness the full potential of network analysis to generate more robust, reliable, and interpretable findings, ultimately advancing our understanding of the complex relationship between diet and health.

Author Contributions: Conceptualization, HY, RT; Methodology, HY, RT; Formal Analysis, HY, RT; Investigation, HY, RT; Resources, HY; Data Curation, RT; Writing – Original Draft, HY, RT; Writing – Review & Editing, HY, RT, JM, AC; Supervision, HY, AC; Project Administration, HY. All authors read and approved the final article for submission.

Funding: This research received no external funding

Acknowledgments: We would like to thank Swansea University for supporting this research

Conflicts of Interest: The authors declare no conflicts of interest

Abbreviations

The following abbreviations are used in this manuscript:

DASH	Dietary Approach to Stop Hypertension
PCA	Principal component analysis
GGM	Gaussian Graphical Model
MGM	Mixed Graphical Model
MI	Mutual Information
BN	Bayesian Networks
EPIC	European Prospective Investigation into Cancer and Nutrition
SGCGM	Semiparametric Gaussian copular graphical model
RRR	Reduced rank regression
MeDi	Mediterranean diet
PwMS	People with multiple sclerosis
HC	Healthy controls

References

1. Martínez-González, M.A., et al., Mediterranean diet and the incidence of cardiovascular disease: a Spanish cohort. *Nutrition, Metabolism and Cardiovascular Diseases*, 2011. 21(4): p. 237-244.
2. Becerra-Tomás, N., et al., Mediterranean diet, cardiovascular disease and mortality in diabetes: A systematic review and meta-analysis of prospective cohort studies and randomized clinical trials. *Critical reviews in food science and nutrition*, 2020. 60(7): p. 1207-1227.
3. Petersson, S.D. and E. Philippou, Mediterranean diet, cognitive function, and dementia: a systematic review of the evidence. *Advances in Nutrition*, 2016. 7(5): p. 889-904.
4. Sezaki, A., et al., Association between the Mediterranean diet score and healthy life expectancy: a global comparative study. *The journal of nutrition, health & aging*, 2022. 26(6): p. 621-627.
5. Peng, W., et al., Major dietary patterns and their relationship to obesity among urbanized adult Tibetan pastoralists. *Asia Pacific Journal of Clinical Nutrition*, 2019. 28(3): p. 507-519.
6. 6Rakhra, V., et al., Obesity and the western diet: How we got here. *Missouri medicine*, 2020. 117(6): p. 536.
7. 7Fabiani, R., et al., A western dietary pattern increases prostate cancer risk: a systematic review and meta-analysis. *Nutrients*, 2016. 8(10): p. 626.
8. 8Young, H.A., et al., Multi-nutrient interventions and cognitive ageing: are we barking up the right tree? *Nutr Res Rev*, 2023. 36(2): p. 471-483.
9. Bánáti, D., et al., Defining a vitamin A5/X specific deficiency–vitamin A5/X as a critical dietary factor for mental health. *International Journal for Vitamin and Nutrition Research*, 2024.
10. Zhao, J., et al., A review of statistical methods for dietary pattern analysis. *Nutr J*, 2021. 20(1): p. 37.
11. Naska, A., A. Lagiou, and P. Lagiou, Dietary assessment methods in epidemiological research: current state of the art and future prospects. *F1000Research*, 2017. 6.
12. Schulz, C.-A., K. Oluwagbemigun, and U. Nöthlings, Advances in dietary pattern analysis in nutritional epidemiology. *European journal of nutrition*, 2021: p. 1-16.
13. Lin, Y.C., et al., The Protective Effect of Garlic Essential Oil in Carnitine-Induced Cardiovascular Disease apoE-/-Mice Model. *Current Developments in Nutrition*, 2020. 4: p. nzaa062_029.
14. Appel, L.J., et al., A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. *N Engl J Med*, 1997. 336(16): p. 1117-24.
15. Miller, E.R., T.P. Erlinger, and L.J. Appel, The effects of macronutrients on blood pressure and lipids: an overview of the DASH and OmniHeart trials. *Current Cardiovascular Risk Reports*, 2007. 1(1): p. 46-51.
16. Morgenstern, J.D., et al., Perspective: Big Data and Machine Learning Could Help Advance Nutritional Epidemiology. *Adv Nutr*, 2021. 12(3): p. 621-631.
17. Hu, F.B., Dietary pattern analysis: a new direction in nutritional epidemiology. *Current opinion in lipidology*, 2002. 13(1): p. 3-9.
18. O’Leary, D., et al., Negative Affect, Affect Regulation, and Food Choice: A Value-Based Decision-Making Analysis. *Social Psychological and Personality Science*, 2023. 14(3): p. 295-304.

19. Newby, P. and K.L. Tucker, Empirically derived eating patterns using factor or cluster analysis: a review. *Nutrition reviews*, 2004. 62(5): p. 177-203.
20. D'Alessandro, A. and G. De Pergola, The Mediterranean Diet: Its definition and evaluation of a priori dietary indexes in primary cardiovascular prevention. *International journal of food sciences and nutrition*, 2018. 69(6): p. 647-659.
21. Heidari, H., et al., Association of priori-defined DASH dietary pattern with metabolic health status among Iranian adolescents with overweight and obesity. *Scientific Reports*, 2024. 14(1): p. 4993.
22. Jacques, P.F. and K.L. Tucker, Are dietary patterns useful for understanding the role of diet in chronic disease? 12. *The American journal of clinical nutrition*, 2001. 73(1): p. 1-2.
23. Panagiotakos, D., α -Priori versus α -posterior methods in dietary pattern analysis: a review in nutrition epidemiology. *Nutrition bulletin*, 2008. 33(4): p. 311-315.
24. Bodnar, L.M., et al., Machine learning as a strategy to account for dietary synergy: an illustration based on dietary intake and adverse pregnancy outcomes. *Am J Clin Nutr*, 2020. 111(6): p. 1235-1243.
25. Benton, D. and H.A. Young, Early exposure to sugar sweetened beverages or fruit juice differentially influences adult adiposity. *European Journal of Clinical Nutrition*, 2024: p. 1-6.
26. Jacobs, D.R., L.C. Tapsell, and N.J. Temple, Food synergy: the key to balancing the nutrition research effort. *Public Health Reviews*, 2011. 33: p. 507-529.
27. Hevey, D., Network analysis: a brief overview and tutorial. *Health psychology and behavioral medicine*, 2018. 6(1): p. 301-328.
28. Zhang, C.Q. and J. Huang, Examining the network dynamics of daily movement and dietary behaviors among college students: A diary study. *Appl Psychol Health Well Being*, 2024.
29. Brown, I.J., et al., Sugar-sweetened beverage, sugar intake of individuals, and their blood pressure: international study of macro/micronutrients and blood pressure. *Hypertension*, 2011. 57(4): p. 695-701.
30. Altenbuchinger, M., et al., Gaussian and Mixed Graphical Models as (multi-) omics data analysis tools. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 2020. 1863(6): p. 194418.
31. Yang, E., et al. Mixed graphical models via exponential families. in *Artificial intelligence and statistics*. 2014. PMLR.
32. Reshef, D.N., et al., Detecting novel associations in large data sets. *science*, 2011. 334(6062): p. 1518-1524.
33. Samieri, C., et al., Using network science tools to identify novel diet patterns in prodromal dementia. *Neurology*, 2020. 94(19): p. e2014-e2025.
34. Conrady, S. and L. Jouffe, Introduction to bayesian networks & bayesialab. *Bayesia SAS*, 2013.
35. Sadler, M.J., H. McNulty, and S. Gibson, Sugar-fat seesaw: a systematic review of the evidence. *Critical reviews in food science and nutrition*, 2015. 55(3): p. 338-356.
36. Regazzoni, C., V. Murino, and G. Vernazza, Distributed propagation of a-priori constraints in a Bayesian network of Markov random fields. *IEE Proceedings I (Communications, Speech and Vision)*, 1993. 140(1): p. 46-55.
37. Niedermayer, D., An introduction to Bayesian networks and their contemporary applications, in *Innovations in Bayesian networks: Theory and applications*. 2008, Springer. p. 117-130.
38. Casteigts, A., et al., Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems*, 2012. 27(5): p. 387-408.
39. Holme, P. and J. Saramäki, Temporal networks. *Physics reports*, 2012. 519(3): p. 97-125.
40. Bretto, A., Hypergraph theory. An introduction. *Mathematical Engineering*. Cham: Springer, 2013. 1.
41. Hayat, M.K., et al., Heterogeneous hypergraph embedding for node classification in dynamic networks. *IEEE Transactions on Artificial Intelligence*, 2024.
42. Marinazzo, D., et al., An information-theoretic approach to hypergraph psychometrics. *arXiv preprint arXiv:2205.01035*, 2022.
43. Kivelä, M., et al., Multilayer networks. *Journal of complex networks*, 2014. 2(3): p. 203-271.
44. Chen, M., et al., [Children's dietary patterns and dietary networks in five regions in China]. *Wei Sheng Yan Jiu*, 2024. 53(2): p. 195-208.

45. Aguirre-Quezada, M.A. and M.P. Aranda-Ramírez, Irruption of Network Analysis to Explain Dietary, Psychological and Nutritional Patterns and Metabolic Health Status in Metabolically Healthy and Unhealthy Overweight and Obese University Students: Ecuadorian Case. *Nutrients*, 2024. 16(17): p. 2924.
46. Fereidani, S.S., et al., Gaussian Graphical Models Identified Food Intake Networks among Iranian Women with and without Breast Cancer: A Case-Control Study. *Nutr Cancer*, 2021. 73(10): p. 1890-1897.
47. Hoang, T., J. Lee, and J. Kim, Differences in dietary patterns identified by the Gaussian graphical model in Korean adults with and without a self-reported cancer diagnosis. *Journal of the Academy of Nutrition and Dietetics*, 2021a. 121(8): p. 1484-1496. e3.
48. Hoang, T., J. Lee, and J. Kim, Network Analysis of Demographics, Dietary Intake, and Comorbidity Interactions. *Nutrients*, 2021b. 13(10).
49. Jahanmiri, R., et al., Saturated fats network identified using Gaussian graphical models is associated with metabolic syndrome in a sample of Iranian adults. *Diabetology & Metabolic Syndrome*, 2022. 14(1): p. 123.
50. Jayedi, A., et al., Dietary networks identified by Gaussian graphical model and general and abdominal obesity in adults. *Nutrition Journal*, 2021. 20: p. 1-12.
51. Schwedhelm, C., et al., Meal and habitual dietary networks identified through semiparametric Gaussian copula graphical models in a German adult population. *PLoS One*, 2018. 13(8): p. e0202936.
52. Schwedhelm, C., et al., Using food network analysis to understand meal patterns in pregnant women with high and low diet quality. *International Journal of Behavioral Nutrition and Physical Activity*, 2021. 18: p. 1-13.
53. Xia, Y., et al., Complex Dietary Topologies in Non-alcoholic Fatty Liver Disease: A Network Science Analysis. *Front Nutr*, 2020. 7: p. 579086.
54. Gunathilake, M., et al., Identification of Dietary Pattern Networks Associated with Gastric Cancer Using Gaussian Graphical Models: A Case-Control Study. *Cancers (Basel)*, 2020. 12(4).
55. Iqbal, K., et al., Gaussian graphical models identify networks of dietary intake in a German adult population. *The Journal of nutrition*, 2016. 146(3): p. 646-652.
56. Felicetti, F., et al., Eating hubs in multiple sclerosis: exploring the relationship between Mediterranean diet and disability status in Italy. *Frontiers in Nutrition*, 2022. 9: p. 882426.
57. Gunathilake, M., et al., Effect of the Interaction between Dietary Patterns and the Gastric Microbiome on the Risk of Gastric Cancer. *Nutrients*, 2021. 13(8).
58. Landaeta-Díaz, L., S. Durán-Agüero, and G. González-Medina, Exploring food intake networks and anhedonia symptoms in a Chilean Adults sample. *Appetite*, 2023. 190: p. 107042.
59. Gunathilake, M., et al., Association between dietary intake networks identified through a Gaussian graphical model and the risk of cancer: a prospective cohort study. *European Journal of Nutrition*, 2022. 61(8): p. 3943-3960.
60. Iqbal, K., et al., Gaussian graphical models identified food intake networks and risk of type 2 diabetes, CVD, and cancer in the EPIC-Potsdam study. *European journal of nutrition*, 2019. 58: p. 1673-1686.
61. Slurink, I.A., et al., Dairy consumption and incident prediabetes: prospective associations and network models in the large population-based Lifelines Study. *The American journal of clinical nutrition*, 2023. 118(6): p. 1077-1090.
62. Neal, Z.P., et al., Critiques of network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, 2022. 2(1): p. 90.
63. Forbes, M.K., et al., Quantifying the reliability and replicability of psychopathology network characteristics. *Multivariate behavioral research*, 2021. 56(2): p. 224-242.
64. Epskamp, S. and E.I. Fried, A tutorial on regularized partial correlation networks. *Psychological methods*, 2018. 23(4): p. 617.
65. Margolin, A.A., et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. in *BMC bioinformatics*. 2006. Springer.
66. Mokhtari, E.B. and B.J. Ridenhour, Filtering asvs/otus via mutual information-based microbiome network analysis. *BMC bioinformatics*, 2022. 23(1): p. 380.

67. Hallquist, M.N., A.G. Wright, and P.C. Molenaar, Problems with centrality measures in psychopathology symptom networks: Why network psychometrics cannot escape psychometric theory. *Multivariate behavioral research*, 2021. 56(2): p. 199-223.
68. Epskamp, S., et al., The Gaussian graphical model in cross-sectional and time-series data. *Multivariate behavioral research*, 2018. 53(4): p. 453-480.
69. Yang, S., A comparison of different methods of zero-inflated data analysis and its application in health surveys. 2014: University of Rhode Island.
70. Thelwall, M. and P. Wilson, Regression for citation data: An evaluation of different methods. *Journal of Informetrics*, 2014. 8(4): p. 963-971.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.