

# Are Food Ingredient Social? An Empirical Investigation

Sandeep Khanna<sup>a,\*</sup>, Suman Kundu<sup>b</sup> and Chiranjoy Chattopadhyay<sup>c</sup>

<sup>a</sup> *Department of Computer Science & Engineering, Indian Institute of Technology Jodhpur, India*  
E-mail: [khanna.1@iitj.ac.in](mailto:khanna.1@iitj.ac.in); ORCID: <https://orcid.org/0000-0001-7272-6124>

<sup>b</sup> *Department of Computer Science & Engineering, Indian Institute of Technology Jodhpur, India*  
E-mail: [suman@iitj.ac.in](mailto:suman@iitj.ac.in); ORCID: <https://orcid.org/0000-0002-7856-4768>

<sup>c</sup> *School of Computing and Data Sciences, FLAME University, Pune, India*

E-mail: [chiranjoy.chattopadhyay@flame.edu.in](mailto:chiranjoy.chattopadhyay@flame.edu.in); ORCID: <https://orcid.org/0000-0002-3431-0483>

**Abstract.** Understanding the structural organization of ingredient relationships within cuisines can reveal fundamental patterns in culinary traditions and ingredient co-occurrence. In this paper, we constructed Ingredient Networks (InN) from two recipe ingredient datasets encompassing recipes from ten worldwide cuisines. We then performed an empirical investigation of these multi-cuisine Ingredient Networks to examine their structural characteristics. Our analysis demonstrates that the networks exhibit scale-free behavior, with their degree distributions following a power-law characterized by exponents ranging from  $\gamma = 1.96$  to  $\gamma = 2.38$ . This further aligns with statistical validation, where R-squared values range from 0.9965 to 0.9991, and p-values are extremely low ( $10^{-25}$  to  $10^{-30}$ ), reinforcing the robustness of the power-law fit. Additionally, the networks display ultra-small-world properties, as evidenced by their short network diameter of approximately 4. These structural measurements highlight striking similarities between ingredient networks and widely studied social networks, suggesting underlying patterns reflective of social-like dynamics. Furthermore, the communities formed within these ingredient networks show a strong correlation with the categorical grouping of recipes, providing insights into the evolution of culinary traditions and ingredient compatibility.

**Keywords:** Social Network Analysis, Ingredient Network, Power Law, Social Network metrics, Community structure

## 1. Introduction

A social network can be represented as a graph consisting of individuals or entities, their interconnections, and the modes of communication between them [35, 70]. Social network analysis (SNA) is a method used to examine and study the different aspects and characteristics of such networks [56, 88]. While SNA was developed with networks formed by human society, it has now been utilized in many different kinds of networks. SNA principles may equally relevant to understanding the structure and dynamics of ingredient networks, where ingredients and their relationships form the basis of the network.

Unlike random networks, which exhibit a homogeneous distribution of degrees, real-world networks such as ingredient networks often follow a scale-free structure. Scale-free networks are characterized by a power-law degree distribution, where a few highly connected nodes (hubs) coexist with many sparsely connected nodes [10]. This structure arises from growth and preferential attachment, where highly connected nodes are more likely to attract new connections. In the context of ingredient networks,

---

\*Corresponding author. E-mail: [khanna.1@iitj.ac.in](mailto:khanna.1@iitj.ac.in).

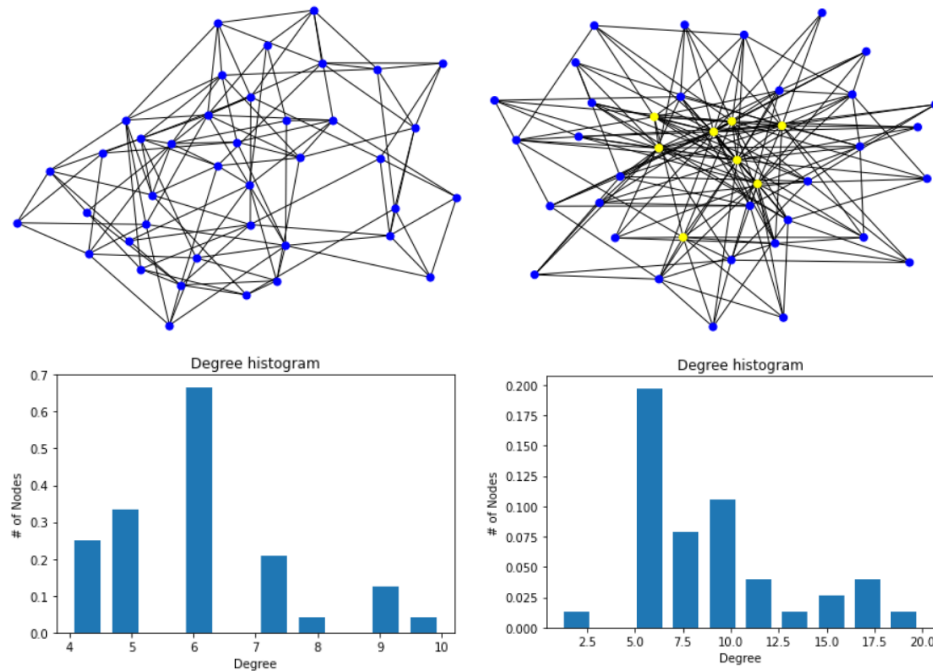


Fig. 1. Random Network vs Scale Free Network

this implies that certain ingredients (hubs) play a central role in connecting diverse components, while others remain peripheral.

Understanding the scale-free nature of ingredient networks is crucial for analyzing their structure and evolution. For example, the ingredients of the central region may represent foundational or widely used components, while the sparsely connected ingredients could indicate niche or specialized elements. In addition, the formation of communities within such networks can reveal patterns of ingredient usage and compatibility, offering insight into culinary practices and recipe design. Figure 1 illustrates the distinction between random and scale-free networks, highlighting the presence of hub nodes in the latter. In ingredient networks, these hubs are critical for maintaining connectivity and facilitating interactions between less connected ingredients. By leveraging SNA techniques, this study aims to uncover the underlying structure of ingredient networks, identify key ingredients, and explore their roles in shaping culinary systems.

In this study, we draw a conceptual parallel between ingredient networks and social networks by demonstrating that ingredient networks exhibit structural and organizational properties commonly observed in social systems. While traditional social networks model relationships between individuals or entities, ingredient networks capture the co-occurrence and collaborative relationships between ingredients in recipes. These relationships give rise to patterns such as scale-free degree distributions, community structures, and centrality hierarchies, which are hallmarks of social networks. The emergence of these patterns suggests that ingredient networks, like social networks, are governed by principles of preferential attachment, collaboration, and modular organization. By applying social network metrics—such as centrality measures, diameter, and community detection—we not only quantify these properties but also provide a novel perspective on the organizational dynamics of culinary systems. This analogy strengthens the case for interpreting ingredient interactions as a form of 'social' behavior, where

ingredients 'collaborate' in recipes to create complex flavor profiles, much like individuals collaborate in social systems to achieve collective outcomes.

This paper analyzes the Indian Dataset of Recipes and Ingredients (INDoRI), encompassing a total of 5187 recipes and Yummly dataset comprises of multi-cuisine details spanning more than 12 global cuisine.. Further, the dataset includes a graph-based representation of ingredient relationships, namely, ingredient network (InN). InN is formed by capturing ingredient relationships based on their co-occurrence within recipes. Ingredient Network (InN) is essentially a graph  $G(V, E, w)$  of ingredients [71], constructed by joining ingredients that appear together in different recipes. The vertices  $V$  represent the ingredients, edges  $E$  contain the connection between vertices, and the weight  $w$  represents the strength of the association. The association is stronger if ingredients appear multiple times in different recipes. For example, the ingredients 'onion' and 'tomato' have a strong association as they seem numerous times in many recipes (refer Fig. 3). The dataset and its characteristics were earlier published in [50]. Extended research involves studying the dataset property and comparison with other cuisine through detailed empirical analysis. This work is mostly emphasized in this paper.

Any study on a novel network starts with empirical analysis to know whether the network is generated randomly or through preferential attachment. In the same sense, it is interesting to know the characteristics of InN. The heterogeneous degree distribution in ingredient networks becomes evident as certain ingredients are essential in numerous recipes while others are only present in a few dishes. This observation highlights the varying degrees of ingredient usage, emphasizing each ingredient's diverse role within the network. However, *are these differences in degree significant enough to call this network a social network? Does the degree of this network a power law distribution?*. In this paper, we have addressed these questions through an extensive empirical analysis.

The organization of the remainder of this paper is as follows: Section 2 provides an overview of the related work on SNA and InN's. Section 3 explains the methodology, datasets, and metrics involved. Section 4 displays the SNA experimental results. Section 5 offers an in-depth discussion, while Section 6 explores the applications. Finally, Section 7 wraps up with conclusions and suggests future research avenues.

## 2. Related Work

The theory of complex networks characterizes social networks by a scale-free property [11]. This property allows it to pose specific properties different from the random network. Although the study of social network analysis started with the human relational network [63], researchers have investigated and explored various other networks and found that many are structurally similar to social networks [96]. The original attempt of Watts and Strogatz in their work on small-world networks [98] was to construct a network model with a small average path length as a random graph and a relatively large clustering coefficient as a regular lattice, which evolved to become a new network model as it stands today. On the other hand, the discovery of scale-free networks was based on the observation that the degree distributions of many real networks have a power-law form, albeit power-law distributions. Another significant recent discovery is that many large-scale complex networks are scale-free; that is, their connectivity distributions are in a power-law form independent of the network scale [10, 12]. Unlike an exponential network, a scale-free network is homogeneous: most nodes have very few link connections, yet few nodes have many connections.

Social Network Analysis (SNA) has been applied to various networks, with some of the notable examples being Twitter [20, 61, 107], Facebook [21, 107], human interaction networks [23], Internet [93],

WWW [3], E-mail [31], software [92], electronic circuits [43], language [19], movie actors [10, 98], math coauthorship [67], food web [62, 100], metabolism [45] where these networks were analyzed and proved that they all follow the small world pattern and are scale-free. In contrast, the analysis of the biomedical research collaboration network [14] shows that the collaboration networks are not scale-free but have small-world properties. Similarly, the Power Grid network is also not scale-free [73].

Ingredient networks have been widely studied to understand the relationships between food components, culinary traditions, and cultural influences. Researchers have investigated ingredient networks within different contexts, including identification of communities [89], recipe recommendation [71], and recipe recognition [60]. Like [41] offers a comprehensive review of network science applications in food studies, which contextualizes our work within the broader field. Similarly, the studies by [1] and [2] provide insights into flavor networks and food pairing principles, which complement our analysis of ingredient connectivity and co-occurrence patterns. Building upon this, [89] developed an ingredient network-based recipe recommendation system, showing that ingredient connectivity can be leveraged for personalized recipe suggestions. Their approach highlighted the role of network structures in food preferences, yet it primarily focused on algorithmic recommendations rather than structural properties of multi-cuisine networks. More recently, [83] utilized semantic knowledge graphs to model ingredient relationships, enabling dietary reasoning and nutrition-based food suggestions. Their work integrated ontology-based analysis, whereas our study focuses on empirical network structures in real-world recipe datasets. Additionally, [22] explored ingredient networks from a health perspective, applying natural language processing (NLP) and content analysis to assess the nutritional quality of recipes shared online. While their study focused on dietary health, our research examines ingredient networks from a structural and cultural perspective, revealing insights into how culinary traditions shape ingredient co-occurrence patterns. These prior works establish the significance of ingredient network analysis, and our study extends this research by providing a multi-cuisine comparative analysis with a focus on Indian and international culinary structures.

By integrating these references, our study is positioned as an extension and refinement of previous research, moving beyond flavor-based networks to a more structural and community-based understanding of ingredient networks across global cuisines. Unlike previous works that primarily focus on ingredient similarity and pairing principles, our study introduces a deeper network-based analysis incorporating metrics such as eigen centrality, clustering coefficients, and modularity to capture the hierarchical structure of culinary ingredient usage. Additionally, by analyzing ten global cuisines, we provide a comparative perspective that broadens the scope of existing research, offering novel insights into the social-like structures and small-world properties of ingredient networks.

Thus, while the existing literature lays a strong foundation for food network analysis, our work contributes by introducing a more comprehensive, structural, and empirical analysis of ingredient networks across multiple cuisines, offering new perspectives on how culinary traditions evolve through network-based interactions. However, none of these studies investigate whether InN shows randomness or scale-free characteristics.

### 3. Methodology

The overall pipeline is structured into several key stages: initially introducing the dataset, followed by cleaning it and constructing an Ingredient Network (InN). Subsequently, the focus shifts to analyzing

Table 1  
Comparison of INDoRI with other datasets

Dataset Name	#of Recipes	Ingredients	Category	Type	Cooking Instructions	Time to Prepare	Regional Information	Image Link	Recipe Link	Corresponding Graphical Data
Recipe Ingredients Dataset [106]	12000+	Yes	Yes	No	No	No	Yes	No	No	No
IndianFoodDataset [44]	6000	Yes	No	Yes	Yes	Yes	Yes	No	Yes	No
IndianFood101 [74]	255	Yes	No	Yes	No	Yes	Yes	No	No	No
INDoRI [50]	5187	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

various cuisines through social network metrics. Additionally, the investigation explores the communities within the InN, comparing these findings across different community detection algorithms. These steps are elaborated in the subsequent Sections.

### 3.1. INDoRI

Creating a comprehensive dataset of Indian cuisines possesses unique challenges. One of them is to compiling recipes that span diverse cultural landscape of India. Due to the same reason one may not find all the recipes from one single web portal. As there is no common data format available, each portal present data differently and the data are unstructured. Hence the second challenge is to extract meaningful information from it. We consider multiple recipe websites to address the first challenge. All the unstructured data therein are crawled using Python script.

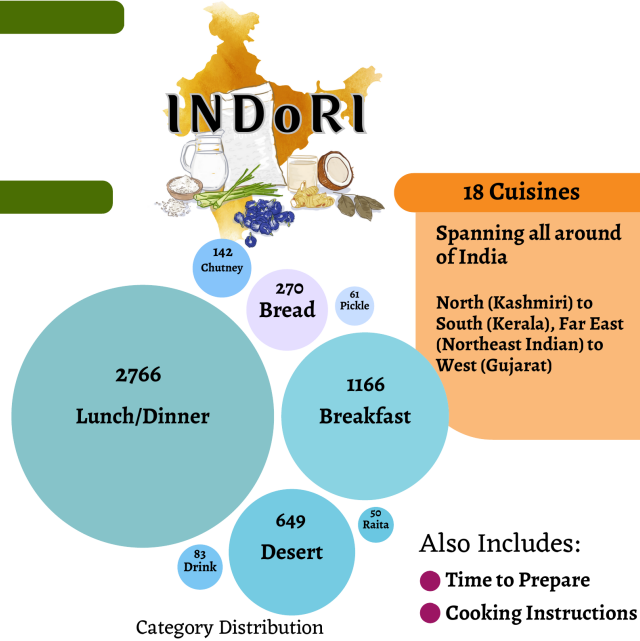
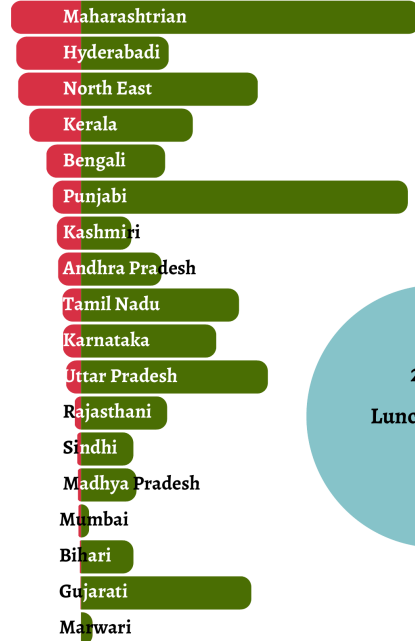
The distinctive feature of INDoRI compared to other food datasets is its collection of 5,187 recipes spanning 18 unique Indian cuisines. Furthermore, it includes additional details such as cuisine, category, preparation time, and cooking instructions. The recipes are organized into 8 specific categories. Excluding the 925 uncategorized recipes, the remainder are also divided among 18 distinct cuisines. Table 1 provides a comparison of INDoRI with existing Indian food-related datasets. Fig. 2 highlights the primary characteristics of INDoRI.

Detailed information about the INDoRI dataset, the cleaning process and ingredient stop words can be found in [50].

### 3.2. Dataset for Empirical Analysis

We utilized two datasets to create, analyze, and understand ingredient networks. The first dataset, INDoRI, is a comprehensive compilation of Indian recipes and their corresponding ingredients. In addition to the INDoRI dataset, we incorporated the recipe ingredient dataset sourced from Yummly [106]. The primary rationale for using the Yummly dataset was to enable a comparative analysis of ingredient properties across diverse global cuisines. While INDoRI provides a rich and unique representation of Indian cuisine, the inclusion of Yummly's dataset allows us to contextualize our findings within a broader global framework. This comparative approach is essential for understanding how ingredient usage patterns in Indian cuisine differ from or align with those in other cuisines. The Yummly dataset contains ingredient information from over 14 global cuisines, and we deliberately selected and filtered 9 cuisines based on their popularity and the quantity of available records. Combined with the Indian cuisine data from INDoRI, our study focuses on analyzing ingredient networks from a total of ten cuisines. Importantly, the INDoRI dataset remains central to our analysis, as it provides a detailed and culturally specific representation of Indian cuisine, which is not available in the Yummly dataset. By performing combined and separate analyses, we ensure that the unique characteristics and contributions of INDoRI are highlighted, while also leveraging the Yummly dataset to provide a global perspective.

## Cuisines Distribution (Non-vegetarian vs Vegetarian)



## 18 Cuisines

Spanning all around of India

North (Kashmiri) to South (Kerala), Far East (Northeast Indian) to West (Gujarat)

Also Includes:

Time to Prepare

Cooking Instructions

## 5187 Recipes

Sizeable compilation of recipes

Comprehensive asset for food-related computational tasks

## 1400+ Ingredients

Large corpus of ingredients



Fig. 2. Key Characteristics of INDoRI

Table 2  
Cuisine Wise Statistics

Cuisine	Total Number of Recipes
Indian	5187
Chinese	2673
Japanese	1423
SouthernUS	4320
French	2646
Italian	7838
Spanish	989
Mexican	6438
British	4320
Thai	1539

Ingredients contained extraneous details beyond ingredient names, requiring the removal of specific words like “cup,” “chopped,” and “boiled,” categorized as Ingredient Stop Words (ISW). This filtering process, based on [50], refines ingredient names for a cleaner dataset. For example, the phrase “half cup of sugar” is reduced to “sugar” through a sequence of actions. Initially, ISW are eliminated, leading to the removal of the words ‘half’ and ‘cup’. The word ‘of’ is also discarded as an NLP stop word, thereby facilitating a more precise Ingredient Network (InN). The detailed process is provided in the referenced source [50].



### 3.3. Ingredient Network (InN) Construction

We developed an ingredient network, represented as a graph  $G(V, E, w)$ , where  $V$  is a set of ingredients,  $E$  is the connections between ingredients and  $w : V \times V \rightarrow R$  of an edge signifies the number of association between ingredients in different recipes. Here each node corresponds to an ingredient and an edge is formed between two nodes if the corresponding ingredients appear together in the same recipe. The more they appear together in diverse recipes, the stronger the association. A visual representation of this network for Indian cuisine is illustrated in Fig. 3, where the thickness of an edge correlates to the strength of the association. Notably, the strongest link is between salt and oil, which co-occur in 1523 recipes, highlighted by one of the thickest edges in the graph.

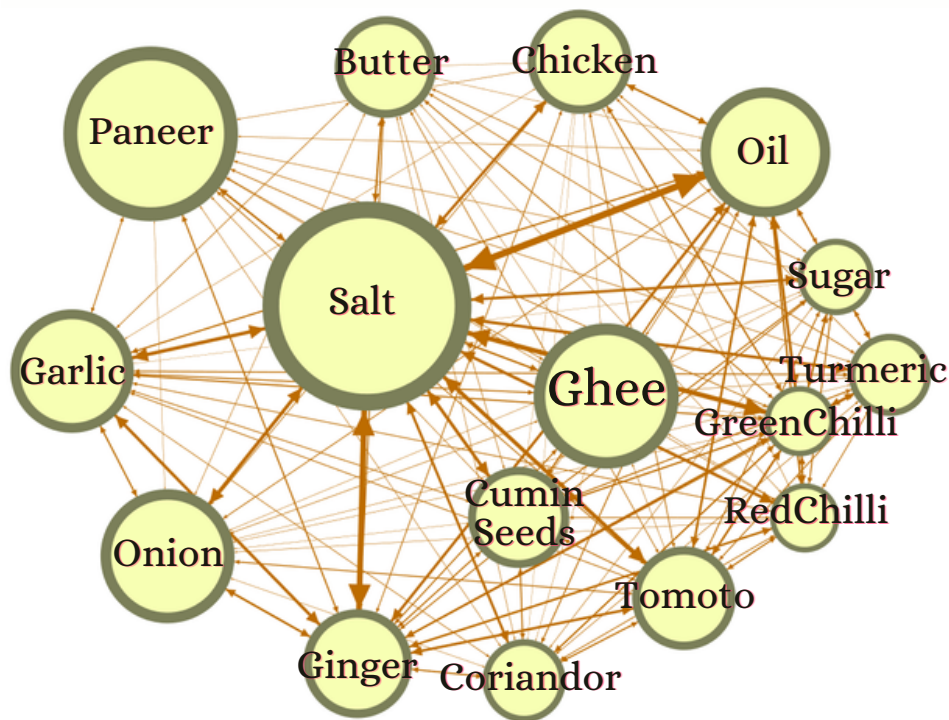


Fig. 3. Ingredient Network (InN) Representation.

### 3.4. Social Metrics Evaluation

In order to examine the dynamics and behaviour of InN, we utilised a range of social network measurements, which we classified into macro and micro metrics. Macro metrics, such as distance, diameter, and density, provide valuable information about the overall structure and connectivity of the network. On the other hand, micro metrics, such as node degree, clustering coefficient, closeness centrality, and eigencentrality, offer a detailed perspective by assessing the behaviour and roles of individual components in the network. Specifically, closeness centrality helps identify ingredients that are closely connected to others, suggesting their potential as central or bridging components in the network. Eigencentrality,

on the other hand, highlights ingredients that are connected to other highly connected ingredients, indicating their influence and importance in the network. These metrics collectively provide insights into the functional roles and relative significance of individual ingredients, enabling a deeper understanding of their usage patterns and interactions within the network. The results obtained from these metrics are explained in the Results Section, demonstrating how each statistic adds to a thorough comprehension of the InN features.

### 3.5. Community Structure Identification

Scale-free networks exhibits community structure. In the course of studying and analyzing InN, we also tried to identify communities within InN. We employed diverse community detection algorithms to partition the InN into multiple communities. Specifically, we implemented weighted versions of the Leiden [90], Louvain [15], and Weighted Association Based Community Detection (WABCD) [50] algorithms. It was observed that InN too exhibits community structure like other scale-free networks. Notably, the weighted Louvain and Leiden algorithms organized communities based on a weighted modularity score, while the WABCD method grouped ingredients by the strength of their connections. We used standard implementations of the Leiden, Louvain, and WABCD algorithms for weighted community detection, ensuring edge weights represented ingredient co-occurrence frequencies. The resolution parameter was set to 1.0 for both Leiden and Louvain to maintain consistent community granularity. WABCD followed weight-based adjustments as per [45]. These algorithms were chosen for their effectiveness: Leiden optimizes modularity for well-separated communities, Louvain provides fast hierarchical clustering, and WABCD is tailored for weighted networks, making it ideal for ingredient co-occurrence analysis.

The details of these community structures and their implications are further discussed in the Results and Analysis Section.

## 4. Results and Analysis

This section provides an in-depth analysis of the outcomes of different macroscopic and microscopic measurements of social networks, along with the observed trends within communities. The subsequent Discussion Section further explores and elaborates on the implications of these metrics. For our network analysis, we employed standard Python libraries, including NetworkX, to compute key network metrics such as degree distribution, shortest path distance, network diameter, density, clustering coefficient, closeness centrality, and eigenvector centrality. All calculations were performed using default settings. For instance, centrality measures were computed using the built-in functions of NetworkX and igraph, ensuring consistency and accuracy.

### 4.1. Degree Distribution of InN

The degree distribution of the Ingredient Network (InN) for 10 different cuisines adheres closely to a power law, as demonstrated by our analysis. We modeled this behavior using a power law distribution, detailed in Equation 1, and visually represented the fit with a red line in Fig. 4. Notably, the parameter  $\gamma$  varies among the cuisines, ranging from 1.96 in Italian Cuisine to 2.38 in Indian Cuisine, highlighting the distinct ingredient combinations characteristic of each cuisine. From the linear regression analysis performed on the log-transformed data across 10 global cuisine ingredient networks, we observed a



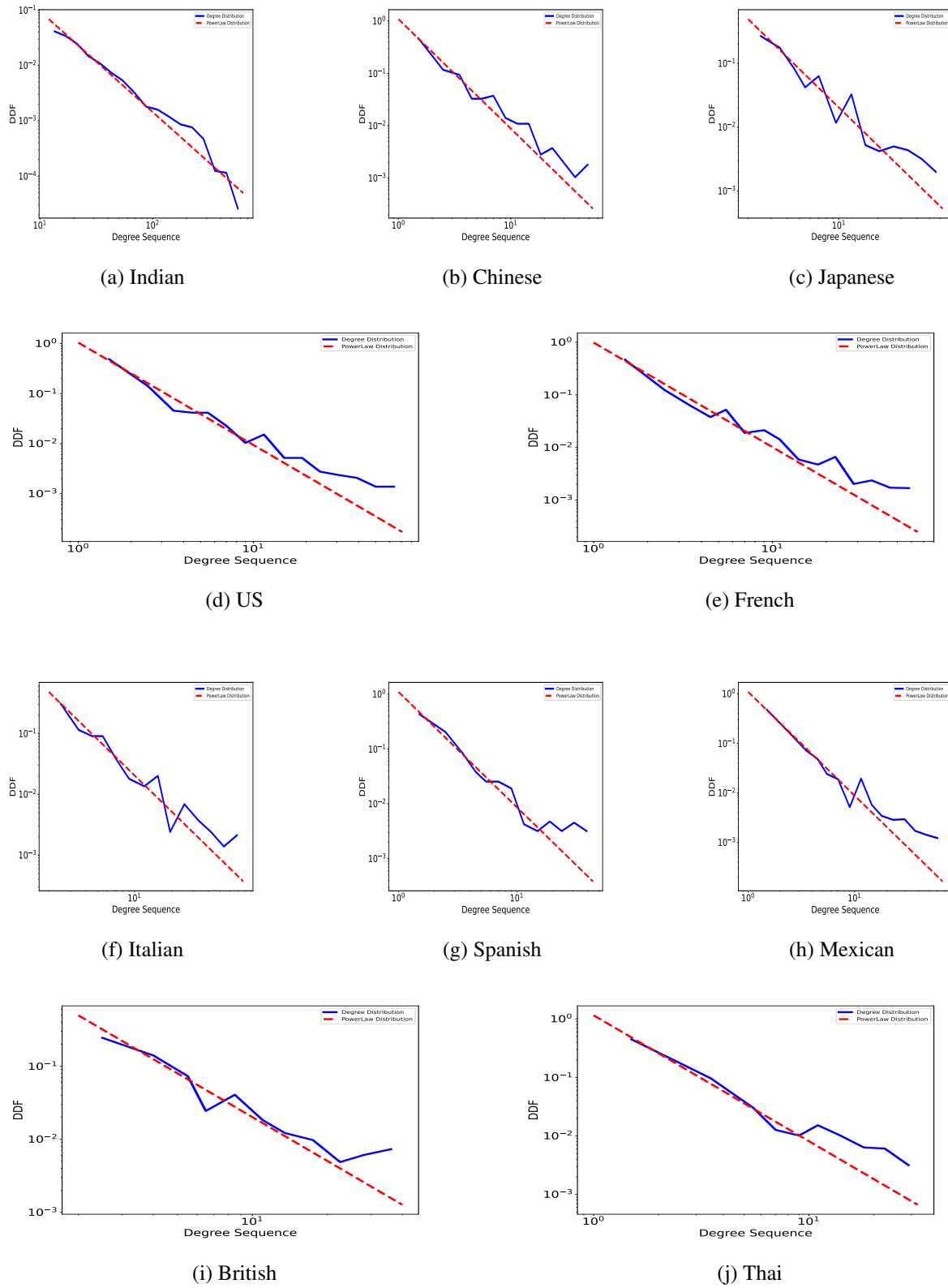


Fig. 4. Degree Distribution Comparison of 10 Cuisine's Ingredient Network along with fitted power law.

consistent range of values for the slope, intercept, R-squared, and p-values. Specifically, the slope ranged from -2.45 to -2.68, with an intercept between 0.18 and 0.22. The R-squared values remained high, ranging from 0.9965 to 0.9991, indicating an excellent fit between the log-transformed degree sequence and degree distribution. Furthermore, the p-values were extremely low (ranging from  $10^{-25}$  to  $10^{-30}$ ), providing strong statistical evidence that the slope of the regression line is significantly different from zero. These results consistently support the hypothesis that ingredient networks across diverse cuisines exhibit power-law behavior, reinforcing the robustness of our findings.

$$p_k \sim k^{-\gamma} \quad (1)$$

#### 4.2. Distance

In graph theory, the distance between two vertices is defined as the number of edges in the shortest path connecting them, often referred to as the hop distance. This concept is mathematically expressed as:

$$d(V_i, V_j) = \min \text{hopdistance}(V_i, V_j) \quad (2)$$

where  $d(V_i, V_j)$  represents the minimum number of hops between vertex  $V_i$  and vertex  $V_j$ . The distance between ingredients in the Ingredient Network (InN) reflects the degree of interconnection and cohesiveness within a cuisine. A shorter maximum distance, such as in Thai cuisine (2 steps), suggests a tightly integrated ingredient network, where most ingredients frequently co-occur, forming a highly cohesive flavor structure. In contrast, a larger maximum distance, as observed in Indian cuisine (4 steps), indicates a more diverse and modular network, where distinct ingredient groups exist, often used in specialized combinations within different regional or traditional dishes. The fact that most cuisines exhibit a maximum distance of 3 reinforces the small-world nature of culinary networks, where even seemingly unrelated ingredients can be connected through a few intermediate ingredients. This structure enables both traditional recipe formation and culinary innovations, as the short paths between ingredients facilitate cross-category ingredient substitutions and fusion cuisine development.

Our analysis of different cuisines revealed varying maximum distances, underscoring the ultra-small world nature of these networks. For example, the maximum distance in the InN of Indian cuisine is 4, indicating a wider variety of ingredient pairings, whereas in Thai cuisine the maximum distance is only 2, suggesting a tighter integration of ingredients. Other cuisines typically exhibited a maximum distance of 3, further demonstrating the close-knit structure of these culinary networks.

#### 4.3. Diameter

The diameter of a graph is determined by its maximum eccentricity, which is calculated as follows.

$$E(V_i) = \max_{V_j \in V \setminus V_i} (d(V_i, V_j)) \quad (3)$$

$$D(G) = \max_{V_i \in V} (E(V_i)) \quad (4)$$

In these equations,  $E(V_i)$  represents the eccentricity of vertex  $V_i$ , defined as the greatest distance from  $V_i$  to any other vertex  $V_j$ . The graph's diameter,  $D(G)$ , is then the largest eccentricity among all vertices in the graph. This metric reflects the farthest distance between any two nodes in the network. Notably, in the Ingredient Network (InN) for each cuisine InN analyzed, the diameter consistently measures 4. The diameter of the Ingredient Network (InN) provides insights into the reachability and connectivity of ingredients within a cuisine. The observation that the diameter remains consistently 4 across all cuisines suggests that ingredient networks are highly interconnected, meaning that even the most distantly related ingredients can be linked through a small number of intermediate connections. This reflects the efficient structure of culinary traditions, where core ingredients act as bridges, connecting diverse ingredient groups. For example, in Italian cuisine, ingredients like olive oil and garlic serve as key connectors across various dishes, while in Japanese cuisine, ingredients such as soy sauce and dashi link otherwise distinct flavor profiles. The small diameter highlights the flexibility of culinary systems, allowing ingredients to be combined in innovative ways while maintaining a cohesive flavor network. This structural characteristic underscores the adaptability of cuisines, where a limited number of widely used bridging ingredients facilitate the fusion and evolution of culinary traditions.

#### 4.4. Density

Density in a network measures the proportion of actual edges to the maximum possible edges between nodes, reflecting the network's connectivity. The denser a network, the more interconnected it is, with a high number of edges relative to the number of nodes. The density of the Ingredient Network (InN) varies across different cuisines, with node counts ranging from 784 in British cuisine to 1926 in Italian cuisine, and edge counts from 14479 to 62528, respectively. Density is calculated as follows:

$$\text{Density} = \frac{\text{Actual number of Edges}}{\text{Maximum possible edges}} \quad (5)$$

where the maximum possible edges for a graph with  $n$  nodes is  $\binom{n}{2}$ . The density of the Ingredient Network (InN) provides important insights into the interconnectivity of ingredients within different cuisines. A higher density suggests that a cuisine has a highly integrated ingredient network, where many ingredients frequently co-occur in diverse recipes, enabling greater flexibility in ingredient combinations. For instance, Thai cuisine, with the highest observed density (0.067), features a tightly interconnected ingredient network, reflecting the frequent pairing of core ingredients like fish sauce, lemongrass, chili, and coconut milk across various dishes. In contrast, Indian cuisine, which has a lower density (0.0296), exhibits a more modular structure, where ingredients like spices form distinct clusters, often used in specialized combinations rather than universally across all dishes. The observed variation in density across cuisines suggests that some culinary traditions rely on highly versatile, broadly connected ingredient sets, while others emphasize distinctive, clustered ingredient groupings, reinforcing the unique structural and cultural organization of different cuisines.

#### 4.5. Cluster Coefficient

The clustering coefficient is an essential metric in social network analysis as it indicates the degree to which nodes tend to cluster together, reflecting community formation and shared interests, as highlighted by Katzir et al. [49]. Applied to the Ingredient Network (InN), this coefficient can identify groups of

ingredients that are commonly used together in specific types of recipes. The formula for calculating the clustering coefficient for a node  $i$ , which has  $n_i$  neighbors, is:

$$C_i = \frac{2e_i}{n_i(n_i - 1)} \quad (6)$$

Here,  $C_i$  represents the clustering coefficient for node  $i$ , where  $e_i$  is the number of actual connections between the node's  $n_i$  neighbors.

The clustering coefficient in the Ingredient Network (InN) provides key insights into the structural organization of culinary ingredient relationships. A high clustering coefficient suggests that certain groups of ingredients are frequently used together, forming tightly-knit culinary clusters that define the flavor profiles of specific cuisines. For example, the high clustering coefficient observed in Mexican cuisine (0.8399) reflects the strong interconnectedness of ingredients like chilies, tomatoes, cilantro, and lime, which commonly co-occur across a variety of traditional dishes. Similarly, the French cuisine network (0.7986) indicates well-established ingredient pairings, such as butter, garlic, wine, and thyme, which frequently appear together in classical French cooking. The variation in clustering coefficients across cuisines suggests that some culinary traditions rely on highly modular ingredient networks, where groups of ingredients form cohesive communities that characterize distinct regional flavors. This highlights how different cuisines develop unique ingredient interaction patterns, reinforcing their cultural identity and traditional cooking methods.

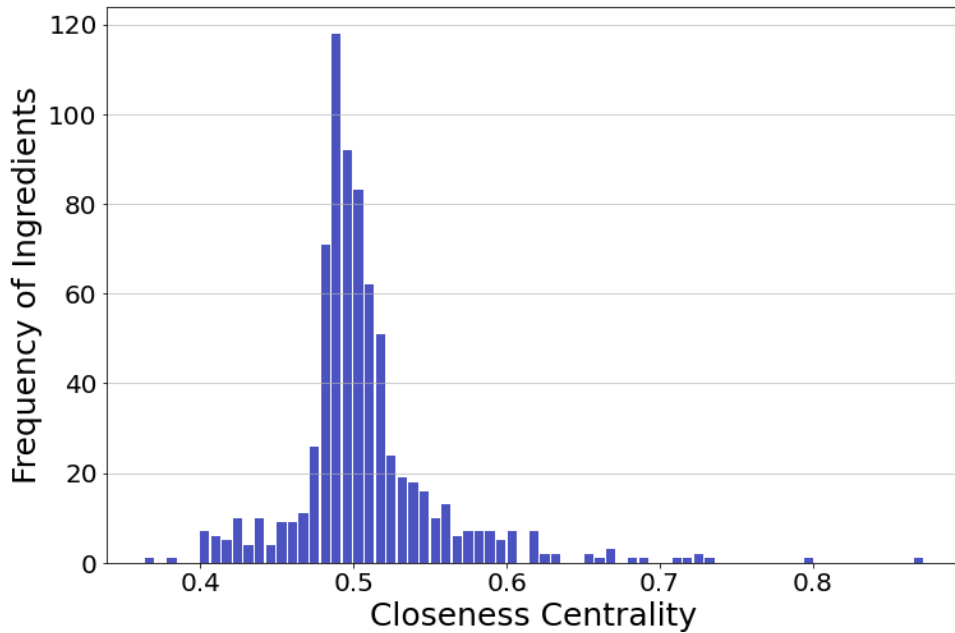


Fig. 5. Closeness Centrality Distribution for the Ingredient Network.

#### 4.6. Closeness Centrality

Closeness Centrality (CC) is a critical measure for understanding the proximity of vertices in a graph, which in the context of the Ingredient Network (InN), helps identify how readily accessible ingredients are to each other and reveals the flexibility in ingredient usage across recipes. CC is calculated using the formula:

$$C(V_i) = \sum_{V_j \in G} \frac{1}{d(V_i, V_j)} \quad (7)$$

A histogram illustrating the CC values for all ingredients is presented in Fig. 5, highlighting the range of closeness centrality, which typically spans from 0.4 to 0.6 for every cuisine in the InN. This range indicates the degree of closeness and connectivity among ingredients within the various cuisines.

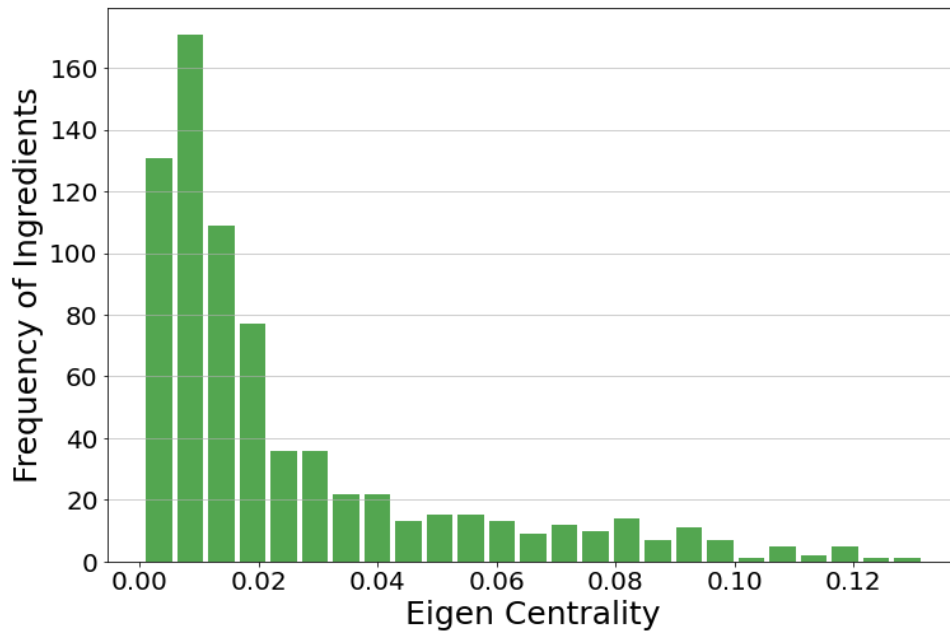


Fig. 6. Eigen Centrality Distribution for the Ingredient Network.

The Closeness Centrality (CC) of ingredients in the Ingredient Network (InN) provides valuable insights into the accessibility and functional versatility of ingredients within a cuisine. Ingredients with higher CC values tend to be widely used and easily combinable, allowing for greater flexibility in recipe formulation. These ingredients often serve as bridging elements, enabling diverse ingredient combinations and fostering innovation in cooking. For example, ingredients such as onion, salt, butter, and lemon frequently appear in multiple recipes across different cuisines, demonstrating high accessibility and broad culinary adaptability. Conversely, ingredients with lower CC values may be more specialized or region-specific, primarily appearing in select dishes or niche culinary traditions. The observed CC range (0.4 to 0.6) across cuisines suggests that most ingredients maintain a moderate level of accessibility, ensuring a balance between core staple ingredients and specialized components that contribute

to a cuisine's unique identity. This further underscores the structural efficiency of ingredient networks, where highly central ingredients enhance the cohesion and adaptability of a culinary tradition.

#### 4.7. Eigen Centrality

Eigen centrality (EC) is a key metric for determining the importance of nodes within a network, as it quantifies the influence of a node based on its connections to other highly significant nodes, as outlined by South et al. [87]. In the context of the Ingredient Network (InN), EC has been applied to assess the significance of each ingredient. It has been found that the EC values for most ingredients in every cuisine generally fall between 0.01 and 0.08. However, as depicted in Fig. 6, there are a few ingredients that stand out with notably higher EC values, underscoring their pivotal roles within their respective culinary networks.

In the context of culinary patterns, the EC of ingredients offers valuable insights into their functional roles within different cuisines. Ingredients with higher EC values tend to be those that frequently co-occur with other influential ingredients, often serving as essential flavor bases, binding agents, or key enhancers in traditional dishes. For example, in many global cuisines, garlic, onion, and olive oil exhibit high EC values, highlighting their widespread importance across multiple dishes and their strong connections to other significant ingredients. Conversely, ingredients with lower EC values may represent niche or region-specific components that are less interconnected but still contribute to the distinctiveness of a cuisine. The observed EC distribution across cuisines suggests that staple ingredients with high culinary adaptability and cross-linking properties tend to emerge as central nodes in ingredient networks, reinforcing their fundamental role in shaping flavor complexity and regional food identities.

#### 4.8. Community Structure

The Ingredient Network (InN) was analyzed using the Weighted Leiden (W-Leiden), Louvain (W-Louvain), and WABCD algorithms [50], to identify distinct culinary communities within various cuisines. The results from the W-Leiden algorithm revealed that the number of communities ranged from 3 in Italian cuisine to 8 in Japanese cuisine. In contrast, the W-Louvain algorithm detected between 4 communities in several cuisines such as Chinese, Japanese, Italian, and Thai, and up to 9 in British cuisine, with Southern US cuisine standing out with 17 communities. The WABCD algorithm showed a variation from 6 communities in Chinese cuisine to 9 in Southern US cuisine. These findings underscore the tendency of the InN to form community structures similar to other social networks.

Fig. 7 presents the community detection results within the Indian cuisine InN, illustrating the segmentation achieved by the WABCD, W-Leiden, and W-Louvain algorithms, which identified 7, 5, and 4 distinct communities, respectively. The figure clearly shows that the first community identified by each algorithm is characterized by tightly interconnected nodes, indicating strong cohesion. However, the density of nodes and the connectivity within communities tend to decrease in subsequent groups identified by each algorithm, reflecting a varying degree of association among different ingredient groups. These patterns suggest underlying structures in the data that mirror complex relationships within the culinary domain.

To uncover the inherent characteristics of each partition we have created multiple sub-graphs based on the category of recipes and compare them with the communities obtained from all three algorithms. The results were shown in Table 3. One may observe that with both weighted Leiden and Louvain algorithm, the second community exhibit connection with recipe category Desert whereas the rest tend



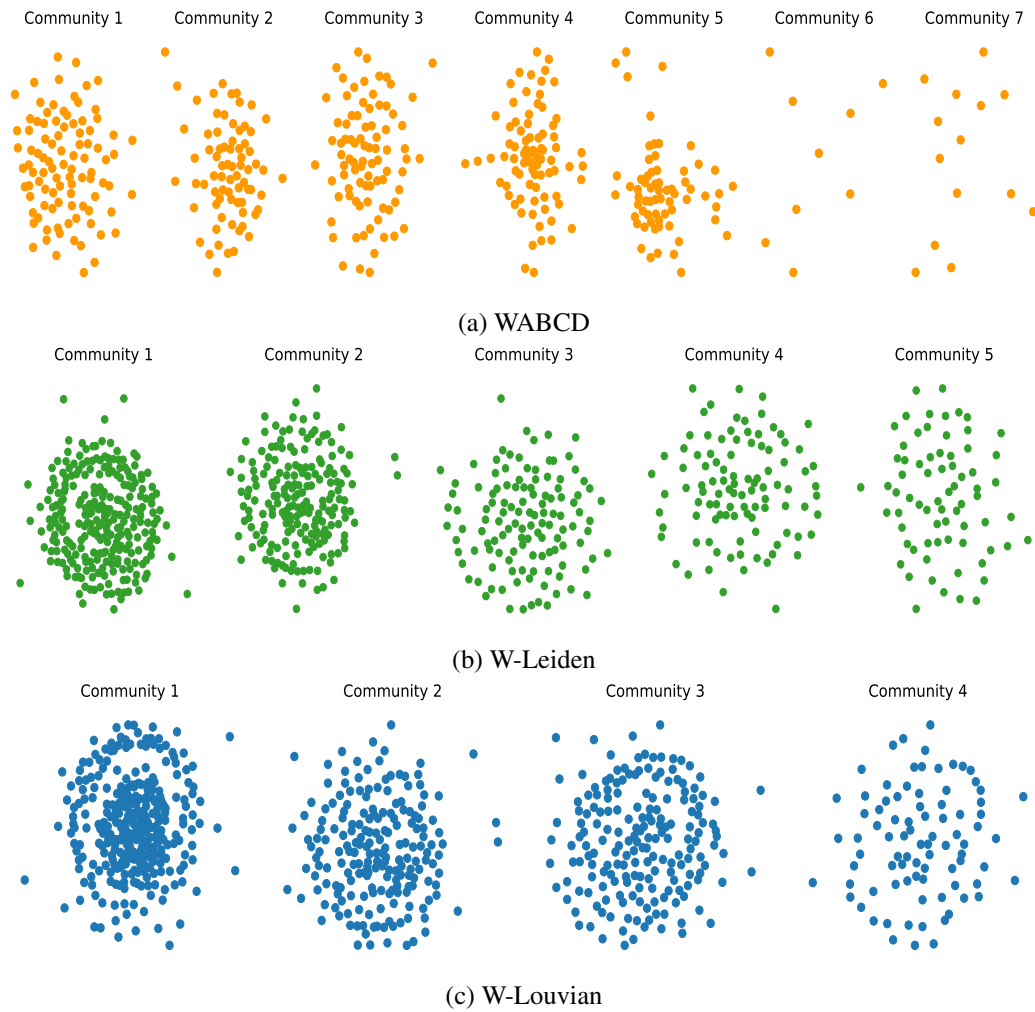


Fig. 7. Results from Different Community Detection Algorithms a) WABCD detects 7 communities b) W-Leiden detects 5 communities c) W-Louvian detects 4

to have more association with Lunch/Dinner category. Conversely, the WABCD identifies four prominent recipe categories: Bread, Lunch/Dinner, Drink, and Deserts.

## 5. Discussion

Results of the empirical study reveal that InN shows properties which resemble scale-free networks. For example, Fig. 4 demonstrates that InN follows a power-law degree distribution with exponents ranging from  $\gamma = 1.96$  to  $\gamma = 2.38$ . These values of  $\gamma$  ensure the ultra-small property [26], i.e., the average diameter of the network is minimal, indicating efficient connectivity. In the context of food networks, this range of  $\gamma$  reflects a scale-free structure, where a few ingredients (hubs) are central to many recipes, while most ingredients have limited connections. This heterogeneity is consistent with other real-world

Table 3  
Comparison of community detection algorithms

	Weighted Leiden	Weighted Louvain	WABCD
<b>C1</b>	Lunch/Dinner Recipes	Lunch/Dinner Recipes	Bread Recipes
<b>C2</b>	Desert Recipes	Desert Recipes	Bread Recipes
<b>C3</b>	Lunch/Dinner Recipes	Lunch/Dinner Recipes	Lunch/Dinner Recipes
<b>C4</b>	Lunch/Dinner Recipes	Lunch/Dinner Recipes	Drink Recipes
<b>C5</b>	Lunch/Dinner Recipes	-	Lunch/Dinner Recipes
<b>C6</b>	-	-	Dessert Recipes
<b>C7</b>	-	-	Lunch/Dinner Recipes

networks, such as social networks ( $\gamma = 2 - 3$ ), biological networks ( $\gamma = 2 - 2.5$ ) and technological networks ( $\gamma = 2.1 - 2.4$ ). The slightly lower gamma values in food networks may highlight the unique role of staple ingredients (e.g., salt, onions, or rice) that are ubiquitous across recipes, shaping the culinary structure of cuisines. On the contrary, a random network usually has  $\gamma \geq 3$  [9].

For reader's reference, the degree of separation comparison between real networks and InN for Indian cuisine is shown in Table 4. The table shows the average and maximum distances of the 5 real undirected networks and InN. The maximum distance in a network represents the longest shortest path between any two nodes, reflecting the degree of connectivity and integration within the network. In the context of ingredient networks (InN), the maximum distance varies across cuisines due to differences in culinary traditions, ingredient usage patterns, and recipe structures. Cuisines with smaller maximum distances (e.g., Italian or French) often rely on a core set of staple ingredients (e.g., olive oil, tomatoes, or butter) that are widely used across recipes, creating a highly interconnected network. This results in shorter paths between ingredients, as most ingredients are linked through these central hubs. Cuisines with larger maximum distances (e.g., Indian or Thai) may exhibit greater diversity in ingredient usage, with distinct regional or cultural variations in recipes. This can lead to less overlap between ingredient clusters, resulting in longer paths between certain ingredients. For instance, the use of specialized spices or herbs in specific dishes may create "bridges" that increase the maximum distance. The columns N, L, k, d and d\_max represent the number of nodes, links, average degree, average distance and maximum distance, respectively. Degrees of separation specify how many hops one must reach from any randomly chosen node to another. For InN, the maximum distance is 4, indicating that every node can be reached from any node through a maximum hop of 4. The average degree of separation is 3.12. This corresponds to the phenomenon of six degrees of separation [30, 59, 91], which says one node can be reachable through a maximum of 6 hops [66]. Although theoretically, this represents the diameter as 6, practically, the average distance stays at 6 or less. The actual diameter sometimes shows higher values because of outliers. For all pairs of Facebook users worldwide and within the United States, the average distance separation is only 3.90 [8]. However, there are networks like Power Grid where the average distance between nodes exceeds 10 degree of separation. Most real-world social networks adhere to the six degrees of separation principle. Such a network does not show power law distribution.

The nodes of InN adhere to an 80:20 ratio, meaning that 80 per cent of the nodes have a low degree and 20 per cent have a higher degree. For InN, most nodes (597) have degrees ranging from 0-100, and a smaller number of nodes (156) have degrees ranging from 200 - 500.

Table 4 also shows the degree fluctuations in real networks, indicating the average degree  $\langle k \rangle$  and the second moment  $\langle k^2 \rangle$  for 5 undirected real networks and InN. Except for the power grid, every network is scale-free as the estimated degree component  $\gamma$  offers a statistically significant fit. The confidence of the fit to the degree distribution is indicated by the star next to the given values.

Table 4  
Comparison of Real World Undirected Networks with Indian InN as to Degree separation and Fluctuations [66].

Network	$N$	$L$	$d$	$d_{max}$	$\langle k \rangle$	$\langle k^2 \rangle$	$\gamma$
Internet	192,244	609,066	6.98	26	6.34	240.10	3.42*
Power Grid	4,941	6,594	18.99	46	2.67	10.30	Exp.
Science Collaboration	23,133	93,437	5.35	15	8.08	178.20	3.35*
Actor Network	702,388	29,397,908	3.91	14	83.71	47,353.70	2.12*
Protein Interactions	2,018	2,930	5.61	14	2.90	32.30	2.89*-
<b>InN (Indian Cuisine)</b>	<b>1433</b>	<b>30,464</b>	<b>3.12</b>	<b>4</b>	<b>1.17</b>	<b>14.69</b>	<b>2.38*</b>

The Ingredient Networks (InN) for ten global cuisines shown in Table 5 exhibit distinct structural properties, reflecting the unique culinary traditions, cooking methods, and ingredient co-occurrence patterns in each cuisine. The table provides insights into key network metrics such as network size ( $N$ ), edge count ( $L$ ), power-law exponent ( $\gamma$ ), maximum degree ( $d_{max}$ ), diameter, density, clustering coefficient, centrality measures, and community structures. Below is a detailed analysis comparing these cuisines based on their network properties and their implications in culinary practices.

First, the network size and connectivity vary significantly across cuisines. Italian cuisine has the largest ingredient network with 1,926 nodes and 62,528 edges, indicating high ingredient diversity and frequent co-occurrence in recipes. Similarly, Indian, Chinese, and Mexican cuisines also have large ingredient networks, reflecting their rich culinary traditions. In contrast, British cuisine has the smallest network (784 nodes, 14,479 edges), suggesting a more limited ingredient repertoire and simpler recipe structures. The power-law exponent ( $\gamma$ ) ranges from 1.96 to 2.38, with Indian cuisine exhibiting the highest value (2.38), suggesting a strong hierarchical structure dominated by a few essential ingredients.

Second, the small-world properties of these networks demonstrate their ingredient connectivity and accessibility. The maximum shortest path length ( $d_{max}$ ) is lowest for Thai cuisine (2) and highest for Indian cuisine (4), indicating that Thai cuisine has a tightly connected core set of ingredients, while Indian cuisine exhibits a broader range of ingredient pairings. Despite variations in  $d_{max}$ , the network diameter remains consistently at 4 across all cuisines, reinforcing the small-world nature of ingredient relationships. The density of the networks is highest for Thai (0.07) and Spanish (0.05) cuisines, indicating strong ingredient co-occurrence, while cuisines like Indian, French, and Southern US (0.03) exhibit more modular structures with diverse ingredient groupings.

Third, clustering coefficients reveal the tendency of ingredients to form cohesive communities. The highest clustering is observed in Mexican (0.84) and Chinese (0.83) cuisines, reflecting their strong reliance on foundational ingredient combinations (e.g., chilies, corn, beans in Mexican; soy sauce, garlic, ginger in Chinese). In contrast, French cuisine has the lowest clustering coefficient (0.79), suggesting a more diverse range of ingredient combinations without strongly interconnected clusters. The number of communities detected varies widely, with Southern US cuisine displaying the highest number of communities (17 in W-Leiden), reflecting the influence of multiple regional culinary traditions.

Fourth, centrality measures highlight the influence and accessibility of key ingredients in each cuisine. Closeness centrality (0.40 - 0.60) remains relatively consistent across cuisines, indicating that ingredients are well-connected in most recipes. However, eigen centrality varies more significantly; for example, Japanese cuisine has the highest eigen centrality (0.01 - 0.08), suggesting a balanced network where multiple ingredients share influence, whereas Indian and Mexican cuisines have lower values (0.01 - 0.04), indicating dominance by a few key ingredients like spices, chilies, or staple grains.

Finally, these network properties provide valuable insights into cooking and culinary practices across different cultures. Highly clustered cuisines like Mexican, Thai, and Chinese rely on strong ingredient groups that frequently appear together, reflecting their use of staple sauces, spice bases, and well-defined flavor profiles. In contrast, less clustered cuisines like French, Indian, and British suggest a more segmented approach, where ingredients are used in distinct recipe contexts. The large, diverse ingredient networks in Italian, Indian, and Chinese cuisines reflect their culinary complexity and regional variations, while tightly connected networks like Thai and British indicate a more compact set of frequently used ingredients.

Table 5  
Social Metrics Summary of 10 Cuisine's Ingredient Network

Cuisine Wise Ingredient Network	$N$	$L$	$\gamma$	$d_{\max}$	Diam.	Dens.	Cluster Coefficient	Closeness Centrality	Eigen Centrality	Communities		
										W-Louvian	W-Leiden	WABCD
Indian	1433	30464	2.38	4	4	0.03	0.81	0.40 - 0.60	0.01 - 0.04	4	5	7
Chinese	1748	62062	2.12	3	4	0.04	0.83	0.40 - 0.60	0.01 - 0.06	6	4	6
Japanese	959	21552	2.06	3	4	0.05	0.81	0.40 - 0.60	0.01 - 0.08	8	4	7
SouthernUS	1606	42802	2.04	3	4	0.03	0.81	0.40 - 0.60	0.01 - 0.06	7	17	9
French	1432	35028	1.98	3	4	0.03	0.79	0.40 - 0.60	0.01 - 0.06	7	6	7
Italian	1926	62528	1.96	3	4	0.03	0.81	0.40 - 0.55	0.01 - 0.04	3	4	8
Spanish	836	16864	2.10	3	4	0.05	0.81	0.40 - 0.55	0.01 - 0.06	6	5	7
Mexican	1756	58972	2.09	3	4	0.04	0.84	0.40 - 0.55	0.01 - 0.04	4	5	8
British	784	14479	1.99	3	4	0.05	0.80	0.40 - 0.55	0.01 - 0.07	5	9	7
Thai	889	26531	2.15	2	4	0.07	0.83	0.45 - 0.55	0.01 - 0.06	4	4	7

Similarly The Fig. 8 presents a scatter plot comparing the standard deviation of node degrees with the mean degree ( $\langle k \rangle$ ) for different real-world networks, including Ingredient Network (InN), Power Grid, Protein Interactions, Internet, and Science Collaboration networks. Each point represents a specific network, and its position on the x-axis corresponds to  $\langle k \rangle$  (the average number of connections per node), while the y-axis represents the standard deviation of the degree distribution.

**Low  $\langle k \rangle$  and Low Standard Deviation for InN, Power Grid, and Protein Interactions:** The Ingredient Network (InN), Power Grid, and Protein Interactions networks have a low mean degree ( $\langle k \rangle \approx 3$ ) and low standard deviation ( $\approx 1.5-1.8$ ). This indicates that most nodes (ingredients, power stations, or proteins) have a similar number of connections, with fewer extreme hubs. In the culinary context of InN, this suggests that most ingredients in the network have a balanced number of connections with other ingredients, reflecting the structured nature of recipes.

**Higher  $\langle k \rangle$  and Standard Deviation for Internet and Science Collaboration Networks:** The Internet and Science Collaboration networks have much higher  $\langle k \rangle$  (between 6 and 8) and greater standard deviation (above 2.5). This means these networks have many highly connected hub nodes, which are significantly more connected than the average node. The Internet follows a scale-free structure where a few nodes (high-traffic websites) dominate connectivity. The Science Collaboration network shows that some researchers have an exceptionally high number of collaborations, creating a hub-like structure. The low standard deviation in InN suggests that most ingredients are relatively similar in terms of their usage frequency across recipes. Unlike the Internet, where some nodes (websites) dominate, there are no extreme hubs in InN, meaning that recipes are not overly dependent on a few highly connected ingredients. The low  $\langle k \rangle$  value (3) means each ingredient connects to a small number of other ingredients on average, highlighting distinct ingredient groupings common in culinary traditions. This aligns with the idea that cuisines have staple ingredient sets that frequently appear together rather than an arbitrary combination of all available ingredients. Since the InN does not have an extremely high variance in connectivity, it suggests that most ingredients can be substituted with similar alternatives, maintaining the network's

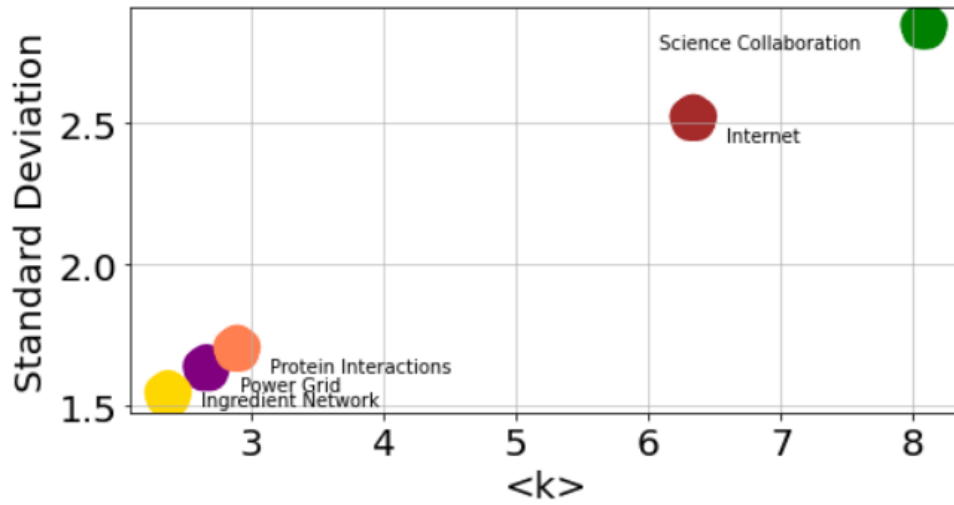


Fig. 8. An illustration of the comparison of standard deviation in Real World Networks with Ingredient Network of Indian cuisine

structure. This is particularly important in culinary practices where ingredients can be swapped based on regional availability, dietary restrictions, or personal preferences.

The structural differences observed in these global ingredient networks align with their respective culinary traditions, historical influences, and regional variations. Cuisines with high clustering and density reflect cohesive ingredient usage, while those with diverse community structures exhibit more varied and modular culinary patterns. These insights not only help in understanding the fundamental structure of cuisines but also offer potential applications in food science, recipe recommendation systems, and gastronomy research.

## 6. Applications

Understanding whether ingredient networks exhibit social properties is significant for both food computing and network science. By analyzing their structure and comparing them with real-world social systems, we gain valuable insight into how ingredients interact and cluster based on culinary traditions, regional preferences, and flavor compatibility. The small-world and scale-free properties of ingredient networks further reinforce their social nature, with profound implications beyond theoretical significance.

### 6.1. Based on Small-world Property

However, the small-world property, characterized by high clustering and short average path lengths, indicates that ingredient networks are highly interconnected and efficiently organized. This can be used in recipe recommendation systems, where understanding the proximity and relationships of ingredients enhances the accuracy and diversity of recommendations. Ingredients within the same community or cluster are likely to share similar flavor profiles, facilitating better substitutions of ingredients and suggestions of complementary pairing.

## 6.2. Based on Scale-free Network

Similarly, the scale-free nature of ingredient networks, with some highly connected hub ingredients, highlights the critical role of these hubs in culinary systems. These staple ingredients serve as foundational elements in many recipes, making them essential to predict culinary trends. Furthermore, the robustness of scale-free networks to random failures suggests that culinary systems can adapt to the removal of less central ingredients, informing strategies for ingredient substitution and dietary adaptation. Using these network properties, our work advances the field of food computing, offering a foundation for developing innovative applications such as flavor pairing prediction, trend analysis, and personalized recipe generation, new recipe generation. Furthermore, demonstrating social behavior in ingredient networks extends social network analysis techniques to food systems, opening new avenues for interdisciplinary research in computational gastronomy.

## 7. Conclusion and Future Work

In this paper, we conducted an empirical study to demonstrate that a food ingredient network (InN) exhibits structural properties similar to social networks, such as scale-free behavior, community structure, and centrality hierarchies. Through a thorough examination of ten diverse cuisines, we showed that InNs consistently follow a power-law distribution and display social network-like patterns, as evidenced by the computation of relevant metrics. These findings suggest that ingredient interactions in culinary systems are governed by principles similar to those observed in social networks, such as preferential attachment and modular organization.

While our study primarily focuses on the structural analysis of InNs, the observed properties provide a strong foundation for future research into culinary trends and user interactions. For example, the scale-free nature of InNs implies that certain ingredients act as hubs, playing a critical role in the composition of the recipe, which could inform the prediction of the popularity of the ingredients or the evolution of culinary practices. Similarly, the community structure of InNs highlights ingredient groupings that reflect cultural or flavor-based affinities, offering insight into how users might interact with or perceive different cuisines. However, these would require additional empirical studies, such as analyzing temporal data on recipe creation or user preferences, to validate their feasibility.

Our work establishes a robust framework for understanding the organizational dynamics of ingredient networks and opens new avenues to explore their applications in culinary science and user behavior analysis. Future research could build on these findings to develop predictive models or investigate the cultural and psychological factors driving ingredient co-occurrence in recipes.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## INDoRI

The dataset used in this study is available online at <https://figshare.com/s/12a1bce0210a7f031168>



## References

- [1] Y.-Y. Ahn, S.E. Ahnert, J.P. Bagrow and A.-L. Barabási, Flavor network and the principles of food pairing, *Scientific reports* **1**(1) (2011), 196.
- [2] S.E. Ahnert, Network analysis and data mining in food science: the emergence of computational gastronomy, *Flavour* **2** (2013), 1–3.
- [3] R. Albert, H. Jeong and A.-L. Barabási, Diameter of the world-wide web, *nature* **401**(6749) (1999), 130–131.
- [4] C.L. Apicella, F.W. Marlowe, J.H. Fowler and N.A. Christakis, Social networks and cooperation in hunter-gatherers, *Nature* **481**(7382) (2012), 497–501.
- [5] S. Aslan and M. Kaya, Link prediction methods in bipartite networks, in: *2017 International Conference on Computer Science and Engineering (UBMK)*, IEEE, 2017, pp. 1095–1099.
- [6] L. Atzori, A. Iera, G. Morabito and M. Nitti, The social internet of things (siot)—when social networks meet the internet of things: Concept, architecture and network characterization, *Computer networks* **56**(16) (2012), 3594–3608.
- [7] V. Azmanov, EAST INDIAN RECIPES, <https://eastindianrecipes.net> [accessed May 6, 2021].
- [8] L. Backstrom, P. Boldi, M. Rosa, J. Ugander and S. Vigna, Four degrees of separation, in: *Proceedings of the 4th Annual ACM Web Science Conference*, 2012, pp. 33–42.
- [9] A.-L. Barabási, Network science book, *Network Science* **625** (2014).
- [10] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *science* **286**(5439) (1999), 509–512.
- [11] A.-L. Barabási and E. Bonabeau, Scale-free networks, *Scientific american* **288**(5) (2003), 60–69.
- [12] A.-L. Barabási, R. Albert and H. Jeong, Mean-field theory for scale-free random networks, *Physica A: Statistical Mechanics and its Applications* **272**(1–2) (1999), 173–187.
- [13] J.A. Barnes, Class and committees in a Norwegian island parish, *Human relations* **7**(1) (1954), 39–58.
- [14] J. Bian, M. Xie, U. Topaloglu, T. Hudson, H. Eswaran and W. Hogan, Social network analysis of biomedical research collaboration networks in a CTSA institution, *Journal of biomedical informatics* **52** (2014), 130–140.
- [15] V.D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, Fast unfolding of communities in large networks, *Journal of statistical mechanics: theory and experiment* **2008**(10) (2008), P10008.
- [16] C. Böde, I.A. Kovács, M.S. Szalay, R. Palotai, T. Korcsmáros and P. Csermely, Network analysis of protein dynamics, *FEBS letters* **581**(15) (2007), 2776–2782.
- [17] A.D. Broido and A. Clauset, Scale-free networks are rare, *Nature communications* **10**(1) (2019), 1–10.
- [18] U. Can and B. Alatas, A new direction in social network analysis: Online social network analysis problems and applications, *Physica A: Statistical Mechanics and its Applications* **535** (2019), 122372.
- [19] R.F.I. Cancho and R.V. Solé, The small world of human language, *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**(1482) (2001), 2261–2265.
- [20] M. Cataldi, L. Di Caro and C. Schifanella, Emerging topic detection on twitter based on temporal and social terms evaluation, in: *Proceedings of the tenth international workshop on multimedia data mining*, 2010, pp. 1–10.
- [21] V. Chandola, A. Banerjee and V. Kumar, Anomaly detection: A survey, *ACM computing surveys (CSUR)* **41**(3) (2009), 1–58.
- [22] X. Cheng, S.-Y. Lin, K. Wang, Y.A. Hong, X. Zhao, D. Gress, J. Wojtusiak, L.J. Cheskin and H. Xue, Healthfulness assessment of recipes shared on Pinterest: natural language processing and content analysis, *Journal of medical Internet research* **23**(4) (2021), e25757.
- [23] Y. Cho, J. Hwang and D. Lee, Identification of effective opinion leaders in the diffusion of technological innovation: A social network approach, *Technological Forecasting and Social Change* **79**(1) (2012), 97–106.
- [24] J.H. Choi, G.A. Barnett and B.-S. CHON, Comparing world city networks: a network analysis of Internet backbone and air transport intercity linkages, *Global Networks* **6**(1) (2006), 81–99.
- [25] M. Ciba, Synchrony Measurement and Connectivity Estimation of Parallel Spike Trains from in vitro Neuronal Networks, PhD thesis, 2021. doi:10.25972/OPUS-22364.
- [26] R. Cohen and S. Havlin, Scale-free networks are ultrasmall, *Physical review letters* **90**(5) (2003), 058701.
- [27] M. Coscia, G. Rossetti, F. Giannotti and D. Pedreschi, Uncovering hierarchical and overlapping communities with a local-first approach, *ACM Transactions on Knowledge Discovery from Data (TKDD)* **9**(1) (2014), 1–27.
- [28] Dassana, DASSANA'S VEG RECIPES, <https://www.vegrecipesofindia.com> [accessed May 7, 2021].
- [29] E. De Gioannis, F. Bianchi and F. Squazzoni, Gender bias in the classroom: A network study on self and peer ability attribution, *Social Networks* **72** (2023), 44–51.
- [30] I. de Sola Pool and M. Kochen, Contacts and influence, *Social networks* **1**(1) (1978), 5–51.
- [31] H. Ebel, L.-I. Mielsch and S. Bornholdt, Scale-free topology of e-mail networks, *Physical review E* **66**(3) (2002), 035103.
- [32] D.A. Eisenberg, J. Park and T.P. Seager, Sociotechnical network analysis for power grid resilience in South Korea, *Complexity* **2017** (2017).

- [33] D.I.H. Farías, V. Patti and P. Rosso, Irony detection in twitter: The role of affective content, *ACM Transactions on Internet Technology (TOIT)* **16**(3) (2016), 1–24.
- [34] I.F. Forever, <https://nerdyfoodies.com/indian-spices-list-3291.html> [accessed May 5, 2021].
- [35] N.A. Ghani, S. Hamid, I.A.T. Hashem and E. Ahmed, Social media big data analytics: A survey, *Computers in Human Behavior* **101** (2019), 417–428.
- [36] H. Ghotra, <https://www.harighotra.co.uk/indian-recipes> [accessed May 5, 2021].
- [37] M. Gong, Q. Cai, M. Lijia and L. Jiao, *Big Network Analytics Based on Nonconvex Optimization*, 2016, pp. 345–373. ISBN 978-3-319-30263-8. doi:10.1007/978-3-319-30263-8-215.
- [38] M.F. Group., allrecipes, <https://www.allrecipes.com> [accessed May 7, 2021].
- [39] B. Hajian and T. White, Modelling influence in a social network: Metrics and evaluation, in: *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, IEEE, 2011, pp. 497–500.
- [40] C. He, X. Fei, Q. Cheng, H. Li, Z. Hu and Y. Tang, A survey of community detection in complex networks using nonnegative matrix factorization, *IEEE Transactions on Computational Social Systems* **9**(2) (2021), 440–457.
- [41] J.C. Herrera, The contribution of network science to the study of food recipes. A review paper, *Appetite* **159** (2021), 105048.
- [42] P.W. Holland, K.B. Laskey and S. Leinhardt, Stochastic blockmodels: First steps, *Social networks* **5**(2) (1983), 109–137.
- [43] R.F. i Cancho, C. Janssen and R.V. Solé, Topology of technology graphs: Small world patterns in electronic circuits, *Physical Review E* **64**(4) (2001), 046119.
- [44] K. Jain, 6000+Indian Food Recipes Dataset”, *Mendeley Data*, V1, doi: 10.17632/xsphgmmh7b.1 (2020).
- [45] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai and A.-L. Barabási, The large-scale organization of metabolic networks, *Nature* **407**(6804) (2000), 651–654.
- [46] S.Y. Jeong, Y.S. Koh and G. Dobbie, Phishing detection on Twitter streams, in: *Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2016 Workshops, BDM, MLSDA, PACC, WDMBF, Auckland, New Zealand, April 19, 2016, Revised Selected Papers 20*, Springer, 2016, pp. 141–153.
- [47] D. Jin, Z. Yu, P. Jiao, S. Pan, D. He, J. Wu, P. Yu and W. Zhang, A survey of community detection approaches: From statistical modeling to deep learning, *IEEE Transactions on Knowledge and Data Engineering* (2021).
- [48] S. Kapoor, <https://www.sanjeevkapoor.com/> [accessed May 6, 2021].
- [49] L. Katzir and S.J. Hardiman, Estimating clustering coefficients and size of social networks via random walk, *ACM Transactions on the Web (TWEB)* **9**(4) (2015), 1–20.
- [50] S. Khanna, C. Chattopadhyay and S. Kundu, INDoRI: Indian Dataset of Recipes and Ingredients and Its Ingredient Network, in: *International Conference on Complex Networks and Their Applications*, Springer, 2023, pp. 234–244.
- [51] D. Knoke and S. Yang, Network fundamentals, *Social network analysis* **154** (2008), 3–14.
- [52] S. Kolli and D. Khajeheian, How actors of social networks affect differently on the others? Addressing the critique of equal importance on actor-network theory by use of social network analysis, *Contemporary Applications of Actor Network Theory* (2020), 211–230.
- [53] L. Laporta, J. Afonso and I. Mesquita, The need for weighting indirect connections between game variables: Social Network Analysis and eigenvector centrality applied to high-level men’s volleyball, *International Journal of Performance Analysis in Sport* **18**(6) (2018), 1067–1077.
- [54] D.D. Lee and H.S. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature* **401**(6755) (1999), 788–791.
- [55] R. Liu, S. Feng, R. Shi and W. Guo, Weighted graph clustering for community detection of large social networks, *Procedia Computer Science* **31** (2014), 85–94.
- [56] Y. Liu, A. Liu, X. Liu and X. Huang, A statistical approach to participant selection in location-based social networks for offline event marketing, *Information Sciences* **480** (2019), 90–108.
- [57] D. Lusseau, The emergent properties of a dolphin social network, *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**(suppl\_2) (2003), S186–S188.
- [58] P. Meena, M. Pawar and A. Pandey, A survey on community detection algorithm and its applications, *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* **12**(6) (2021), 4807–4815.
- [59] S. Milgram, The small world problem, *Psychology today* **2**(1) (1967), 60–67.
- [60] W. Min, L. Liu, Z. Luo and S. Jiang, Ingredient-guided cascaded multi-attention network for food recognition, in: *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 1331–1339.
- [61] S. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu and C. Cherry, Semeval-2016 task 6: Detecting stance in tweets, in: *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, 2016, pp. 31–41.
- [62] J.M. Montoya and R.V. Solé, Small world patterns in food webs, *Journal of theoretical biology* **214**(3) (2002), 405–412.
- [63] J.L. Moreno, Foundations of sociometry: An introduction, *Sociometry* (1941), 15–35.
- [64] D.A. Munoz, J.P. Queuepil and P. Fraser, Assessing collaboration networks in educational research: A co-authorship-based social network analysis approach, *International Journal of Educational Management* **30**(3) (2016), 416–436.

- [65] A. Nasiruzzaman, H. Pota and M. Mahmud, Application of centrality measures of complex network framework in power grid, in: *IECON 2011-37th Annual Conference of the IEEE Industrial Electronics Society*, IEEE, 2011, pp. 4660–4665.
- [66] M.E. Newman, A.-L.E. Barabási and D.J. Watts, *The structure and dynamics of networks.*, 2006.
- [67] M.E. Newman, Scientific collaboration networks. I. Network construction and fundamental results, *Physical review E* **64**(1) (2001), 016131.
- [68] M.E. Newman, Fast algorithm for detecting community structure in networks, *Physical review E* **69**(6) (2004), 066133.
- [69] M.E. Newman, Finding community structure in networks using the eigenvectors of matrices, *Physical review E* **74**(3) (2006), 036104.
- [70] M.E. Newman, D.J. Watts and S.H. Strogatz, Random graph models of social networks, *Proceedings of the national academy of sciences* **99**(suppl\_1) (2002), 2566–2572.
- [71] U. Nyati, S. Rawat, D. Gupta, N. Aggrawal and A. Arora, Characterize ingredient network for recipe suggestion, *International Journal of Information Technology* **13** (2021), 2323–2330.
- [72] E. Otte and R. Rousseau, Social network analysis: a powerful strategy, also for the information sciences, *Journal of information Science* **28**(6) (2002), 441–453.
- [73] G.A. Pagani and M. Aiello, The power grid as a complex network: a survey, *Physica A: Statistical Mechanics and its Applications* **392**(11) (2013), 2688–2700.
- [74] N. Prabhavalkar, Indian Food 101, <https://www.kaggle.com/datasets/nehaprabhavalkar/indian-food-101>.
- [75] S. Qu, Y. Du, M. Zhu, G. Yuan, J. Wang, Y. Zhang and X. Duan, Dynamic Community Detection Based on Evolutionary DeepWalk, *Applied Sciences* **12**(22) (2022), 11464.
- [76] U.N. Raghavan, R. Albert and S. Kumara, Near linear time algorithm to detect community structures in large-scale networks, *Physical review E* **76**(3) (2007), 036106.
- [77] R.A. Rossi and N.K. Ahmed, Role discovery in networks, *IEEE Transactions on Knowledge and Data Engineering* **27**(4) (2014), 1112–1131.
- [78] M. Rosvall and C.T. Bergstrom, Maps of random walks on complex networks reveal community structure, *Proceedings of the national academy of sciences* **105**(4) (2008), 1118–1123.
- [79] N. Roy, K. Pushkarna, R. Raj, M. Kaur and S. Mishra, TIMES FOOD, <https://recipes.timesofindia.com> [accessed May 10, 2021].
- [80] A. Said, R.A. Abbasi, O. Maqbool, A. Daud and N.R. Aljohani, CC-GA: A clustering coefficient based genetic algorithm for detecting communities in social networks, *Applied Soft Computing* **63** (2018), 59–70.
- [81] S. Samanta, V.K. Dubey and B. Sarkar, Measure of influences in social networks, *Applied Soft Computing* **99** (2021), 106858.
- [82] K. Selvarajah, P.M. Zadeh, M. Kargar and Z. Kobti, Identifying a team of experts in social networks using a cultural algorithm, *Procedia Computer Science* **151** (2019), 477–484.
- [83] S.S. Shirai, O. Seneviratne, M.E. Gordon, C.-H. Chen and D.L. McGuinness, Identifying ingredient substitutions using a knowledge graph of food, *Frontiers in Artificial Intelligence* **3** (2021), 621766.
- [84] S. Shreekanth, SWASTHI'S RECIPES, <https://www.indianhealthyrecipes.com> [accessed May 8, 2021].
- [85] T.A. Snijders and K. Nowicki, Estimation and prediction for stochastic blockmodels for graphs with latent block structure, *Journal of classification* **14**(1) (1997), 75–100.
- [86] S. Sosa, C. Sueur and I. Puga-Gonzalez, Network measures in animal social network analysis: Their strengths, limits, interpretations and uses (2020).
- [87] T. South, M. Roughan and L. Mitchell, Popularity and centrality in Spotify networks: critical transitions in eigenvector centrality, *Journal of Complex Networks* **8**(6) (2020), cnaa050.
- [88] S. Tabassum, F.S. Pereira, S. Fernandes and J. Gama, Social network analysis: An overview, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **8**(5) (2018), e1256.
- [89] C.-Y. Teng, Y.-R. Lin and L.A. Adamic, Recipe recommendation using ingredient networks, in: *Proceedings of the 4th annual ACM web science conference*, 2012, pp. 298–307.
- [90] V.A. Traag, L. Waltman and N.J. Van Eck, From Louvain to Leiden: guaranteeing well-connected communities, *Scientific reports* **9**(1) (2019), 1–12.
- [91] J. Travers and S. Milgram, An experimental study of the small world problem, in: *Social networks*, 1977, pp. 179–197.
- [92] S. Valverde, R.F. Cancho and R.V. Sole, Scale-free networks from optimal design, *Europhysics Letters* **60**(4) (2002), 512.
- [93] A. Vazquez, R. Pastor-Satorras and A. Vespignani, Internet topology at the router and autonomous system level, *arXiv preprint cond-mat/0206084* (2002).
- [94] L.M. Vicsek, G. Király and H. Konya, Networks in the social sciences: Comparing actor-network theory and social network analysis, *Corvinus Journal of Sociology and Social Policy* **7**(2) (2016), 77–102.
- [95] P. Wagenseller, F. Wang and W. Wu, Size matters: A comparative analysis of community detection algorithms, *IEEE Transactions on Computational Social Systems* **5**(4) (2018), 951–960.

- [96] F.-Y. Wang, Computational social systems in a new period: A fast transition into the third axial age, *IEEE Transactions on Computational Social Systems* **4**(3) (2017), 52–53.
- [97] S. Wasserman, *Advances in social network analysis: Research in the social and behavioral sciences*, Sage, 1994.
- [98] D.J. Watts and S.H. Strogatz, Collective dynamics of ‘small-world’ networks, *nature* **393**(6684) (1998), 440–442.
- [99] B. Wei and Y. Deng, A cluster-growing dimension of complex networks: From the view of node closeness centrality, *Physica A: Statistical Mechanics and its Applications* **522** (2019), 80–87.
- [100] R.J. Williams, E.L. Berlow, J.A. Dunne, A.-L. Barabási and N.D. Martinez, Two degrees of separation in complex food webs, *Proceedings of the National Academy of Sciences* **99**(20) (2002), 12913–12916.
- [101] J. Wu, T. Vallenius, K. Ovaska, J. Westermarck, T.P. Mäkelä and S. Hautaniemi, Integrated network analysis platform for protein-protein interactions, *Nature methods* **6**(1) (2009), 75–77.
- [102] J. Xia, M.J. Benner and R.E. Hancock, NetworkAnalyst-integrative approaches for protein–protein interaction network analysis and visual exploration, *Nucleic acids research* **42**(W1) (2014), W167–W174.
- [103] D. Yang, E. Yurtsever, V. Renganathan, K.A. Redmill and Ü. Özgüner, A vision-based social distancing and critical density detection system for COVID-19, *Sensors* **21**(13) (2021), 4608.
- [104] C.-J. Yen, C.-H. Tu, E.E. Ozkeskin, H. Harati and L. Sujo-Montes, Social network interaction and self-regulated learning skills: Community development in online discussions, *American Journal of Distance Education* (2022), 1–18.
- [105] X. Ying and X. Wu, On randomness measures for social networks, in: *Proceedings of the 2009 SIAM International Conference on Data Mining*, SIAM, 2009, pp. 709–720.
- [106] Yummly, Recipe Ingredients Dataset, 2024, Accessed: 2024-06-21. <https://www.kaggle.com/datasets/kaggle/recipe-ingredients-dataset>.
- [107] Q. Zhang, S. Zhang, J. Dong, J. Xiong and X. Cheng, Automatic detection of rumor on social network, in: *Natural Language Processing and Chinese Computing: 4th CCF Conference, NLPCC 2015, Nanchang, China, October 9-13, 2015, Proceedings 4*, Springer, 2015, pp. 113–122.