Are Food Ingredient Social? An Empirical Investigation

Sandeep Khanna [a,*], Suman Kundu [b] and Chiranjoy Chattopadhyay [c]

[a] *Department of Computer Science & Engineering, Indian Institute of Technology Jodhpur, India*
E-mail: khanna.1@iitj.ac.in; ORCID: https://orcid.org/0000-0001-7272-6124
[b] *Department of Computer Science & Engineering, Indian Institute of Technology Jodhpur, India*
E-mail: suman@iitj.ac.in; ORCID: https://orcid.org/0000-0002-7856-4768
[c] *School of Computing and Data Sciences, FLAME University, Pune, India*
E-mail: chiranjoy.chattopadhyay@flame.edu.in; ORCID: https://orcid.org/0000-0002-3431-0483

Abstract. This paper introduces INDoRI (Indian Dataset of Recipes and Ingredients) dataset comprising of rich collection of recipes and ingredients from 18 cuisine within the Indian subcontinent. Additionally we examined two recipe ingredient datasets comprising of ten worldwide cuisines and constructed its Ingredient Network (InN). Further, an empirical investigation was performed on these multi-cuisine InN's, uncovering resemblances to social networks. The research reveals that the distribution of InN follows a power-law distribution, with an alpha exponent ranging from $\gamma = 1.96$ to $\gamma = 2.38$. This suggests that InN exhibits an ultra-small world characteristic, further supported by a network diameter of 4. These measurements indicate that the InN, similar to numerous social networks, exhibits scale-free characteristics indicative of social behaviour patterns.

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 [32, 65]. Social network analysis (SNA) is a method used to examine and study the different aspects and characteristics of social networks [51, 81]. SNA helps in understanding the interdependence of social entities, characterizing their activities and their impact on the network.

However, social networks do not have the homogeneous distribution of degrees that random networks possess. A random network is constructed by placing an edge between a pair of vertices with probability p, and each node's expected number of connections is equal. On the contrary, social networks are real-world networks with a few highly connected nodes and many less connected nodes. These networks show power-law degree distribution and are called scale-free. A scale-free network recognizes that network structure and its evolution are inextricably linked [8], i.e., the power law distribution results from growth and preferential attachment (PA). PA says that the more connected a node is, the more likely it is to receive new links. One essential characteristic of scale-free networks is their propensity to form communities.

*Corresponding author. E-mail: khanna.1@iitj.ac.in.

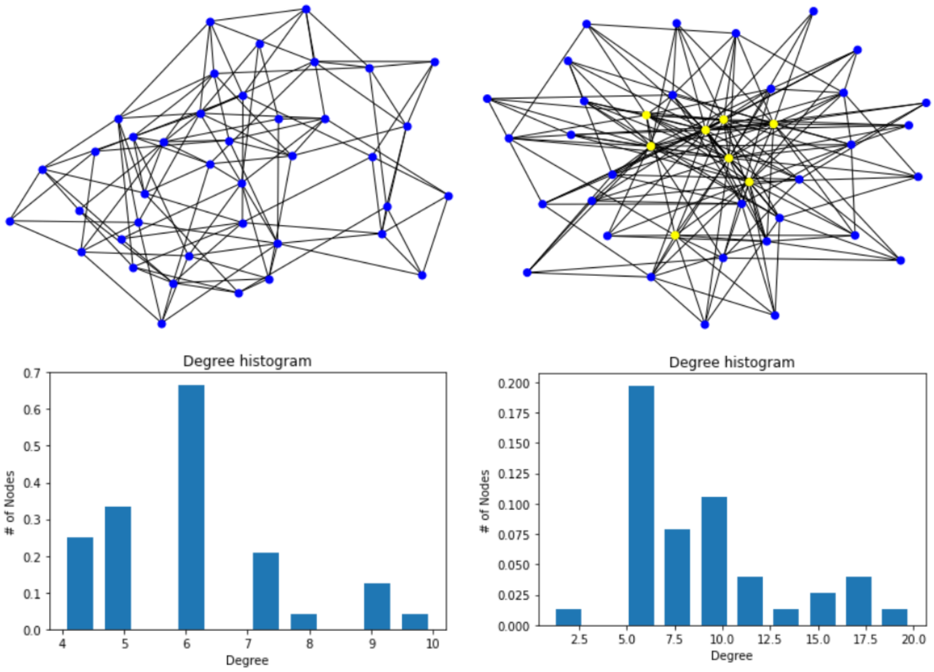


Fig. 1. Random Network vs Scale Free Network

Fig 1 shows a random network and a scale-free network side by side. The degree distribution for random networks follows Poison distribution, whereas scale-free networks offer power law. Another property of scale-free networks is that they contain hub nodes. In Fig. 1, the hub nodes are the nodes that have more connections than the average and are shown in yellow. As a result, many nodes are sparsely connected, whereas hubs can have many links. Another essential feature of a scale-free network is that the clustering coefficient drops with node degree, with progressively lower-degree nodes forming the regions between the core and the periphery.

This paper introduces the Indian Dataset of Recipes and Ingredients (INDoRI), encompassing a total of 5187 recipes. Recipes were extracted and gathered from multiple online platforms [5, 25, 31, 35, 43, 77]. These recipes span a variety of Indian cuisines, reflecting the rich cultural diversity across regions such as Punjabi, Bengali, and Gujarati. INDoRI stands as a structured repository of recipes and their corresponding ingredients. 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 [66], 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 [45]. 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.

2. Related Work

The theory of complex networks characterizes social networks by a scale-free property [9]. 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 [58], researchers have investigated and explored various other networks and found that many are structurally similar to social networks [89]. The original attempt of Watts and Strogatz in their work on small-world networks [91] 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 [8, 10]. 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 [18, 56, 100], Facebook [19, 100], human interaction networks [20], Internet [86], WWW [1], E-mail [28], software [85], electronic circuits [39], language [17], movie actors [8, 91], math coauthorship [62], food web [57, 93], metabolism [40] 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 [12] shows that the collaboration networks are not scale-free but have small-world properties. Similarly, the Power Grid network is also not scale-free [68]. Researchers have investigated ingredient networks within different contexts, including identification of communities [82], recipe recommendation [66], and recipe recognition [55]. 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 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

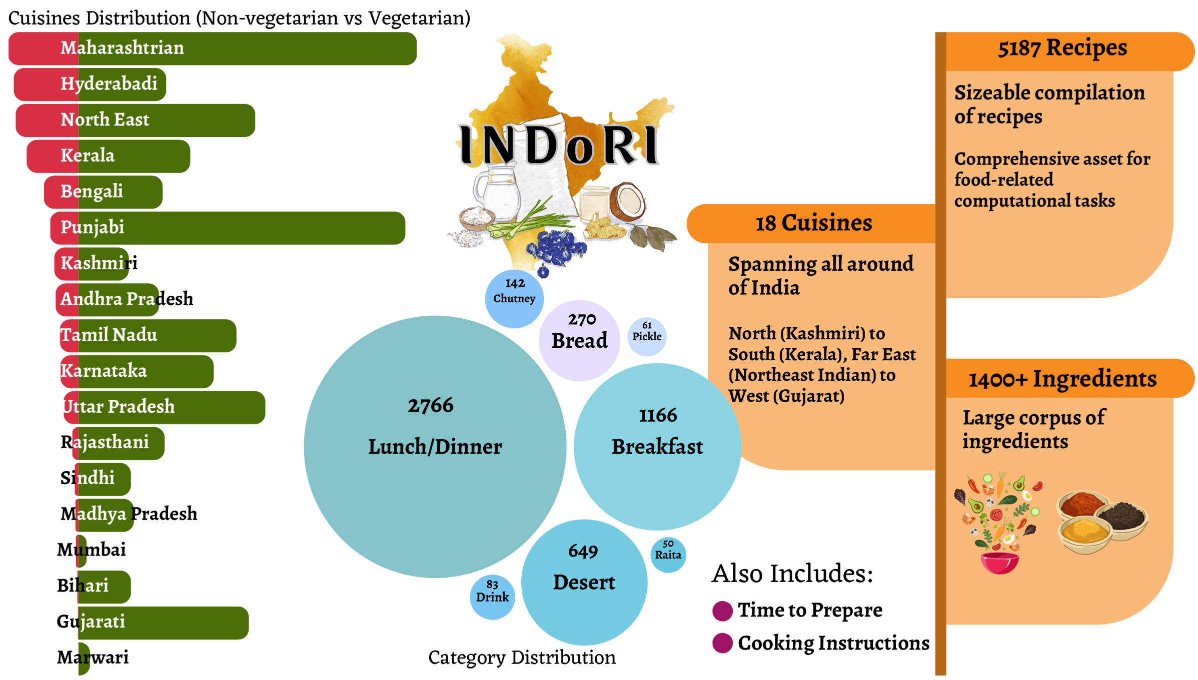


Fig. 2. Key Characteristics of INDoRI

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.

INDoRI encompasses 5,187 recipes representing 18 diverse Indian cuisines. Additionally it encompasses additional attributes such as cuisine, category, and preparation time. All recipes are classified into 8 different types. Apart from 925 unclassified recipes rest are also categorized into 18 different cuisines Fig. 2 shows the key characteristics of INDoRI.

Detailed information about the INDoRI dataset, the cleaning process and ingredient stop words can be found in [45].

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 [99]. This dataset contains ingredient information from over 14 global cuisines. For our analysis, we deliberately selected and filtered nine cuisines based on their popularity and the quantity of available records. In summary, our study focuses on analysing ingredient networks from 10 cuisines. Table 1, you can observe the distribution of recipes for each cuisine. Notably, Italian and Mexican cuisines have the highest number of recipes, followed closely by Indian cuisine.

Table 1
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

3.3. Data Cleaning

The datasets we used contain extraneous details beyond just ingredient names, necessitating the removal of certain words like “kg," “gms," “cup," “tbls," “pinch," “chopped," “boiled," “sliced," and “split" to refine our data. These words, categorized as Ingredient Stop Words (ISW), are sourced from [45] and include a total of 527 such terms. For instance, the phrase “half cup of sugar" includes unnecessary descriptors; by applying ISW filters, we simplify it to just “sugar." This process of removing ISWs ensures that our Ingredient Network (InN) is constructed with only the essential ingredient names, leading to a cleaner and more accurate dataset. For reference, examples of such words can be found in the referenced Table 2.

Table 2
Sample Ingredient Stop Words.

drops	sunflower	kept	florets	adding
handful	chopped	colostrum	shah	consistency
karhai	take	strips	sabut	handful
ones	cup	magaz	tightly	pass
sprig	strong	must	to	boiling
kootu	quick	purpose	seedless	pieces
substitute	more	pour	ounces	pinches
mix	quality	inches	approx	required
Kg	grain	drop	outer	sprigs
fat	crumbled	out	lengthwise	each

3.4. 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.

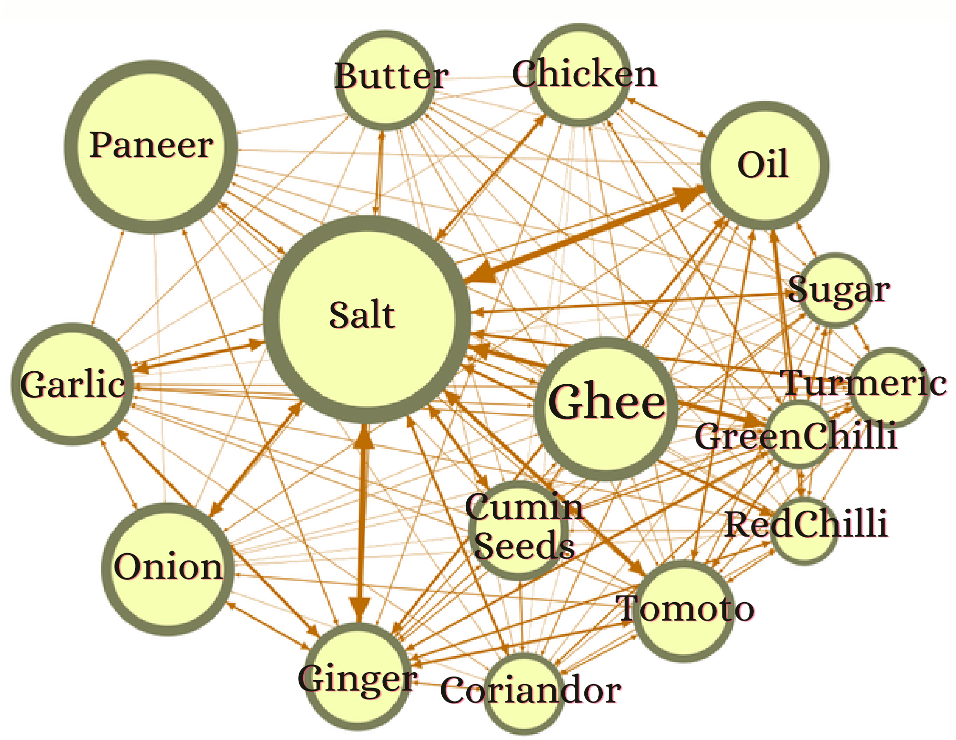


Fig. 3. Ingredient Network (InN) Representation.

3.5. 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 insights of the network. On the other hand micro metrics, such as node degree, cluster coefficient, proximity centrality, and eigencentrality, provide a detailed perspective by assessing the behaviour of individual components in 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.

3.6. 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 [83], Louvain [13], and Weighted Association Based Community Detection (WABCD) [45] 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. The details of these community structures and their implications are further discussed in the Results Section.

4. Results

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.

4.1. Degree Distribution of InN

The degree distribution of the Ingredient Network (InN) for ten 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 and Fig. 5. Notably, the parameter γ 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.

$$p_k \sim k^{-\gamma} \tag{1}$$

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:

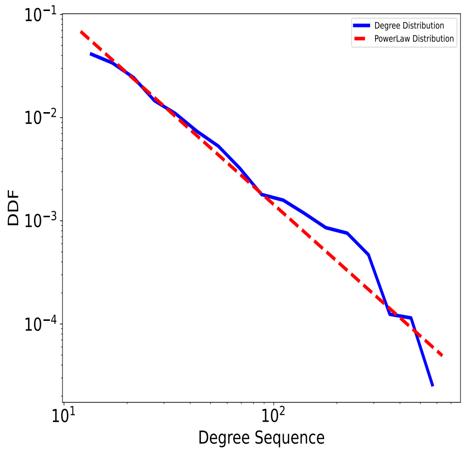
$$d(V_i, V_j) = \min hopdistance(V_i, V_j) \tag{2}$$

where $d(V_i, V_j)$ represents the minimum number of hops between vertex V_i and vertex V_j. In the context of the Ingredient Network (InN), this distance can reveal how closely two ingredients are related in terms of their use in recipes. A shorter distance between two ingredients suggests they frequently appear together in recipes, indicating a strong culinary link.

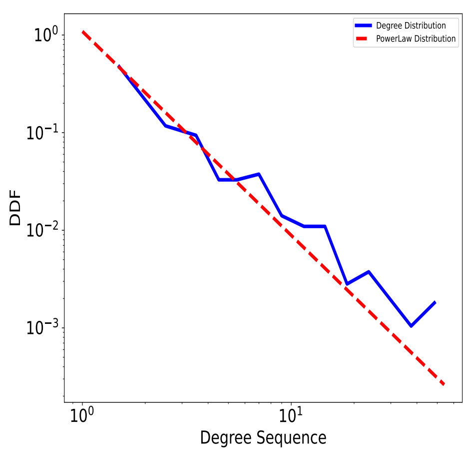
Our analysis of different cuisines revealed varying maximum distances, underscoring the ultra-small-world nature of these networks. For instance, the maximum distance in the InN of Indian cuisine is 4, indicating a broader variety of ingredient pairings, whereas in Thai cuisine, the maximum distance 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

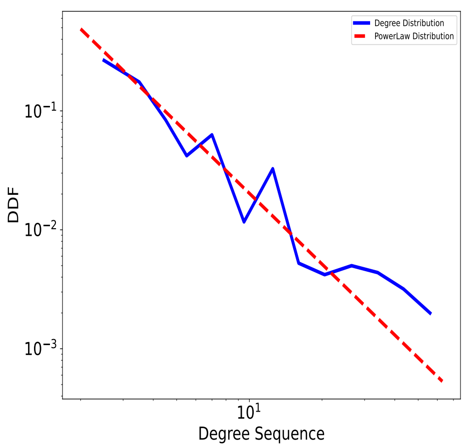
The diameter of a graph is determined by its maximum eccentricity, which is calculated as follows:



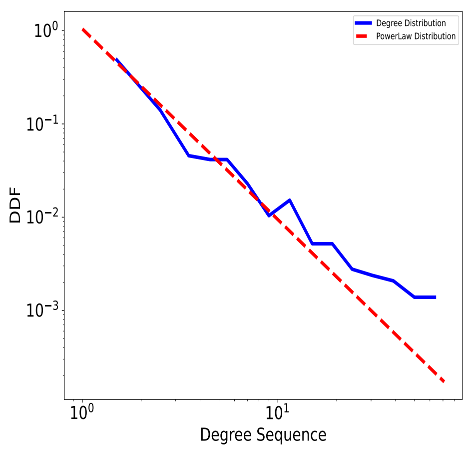
(a) Indian



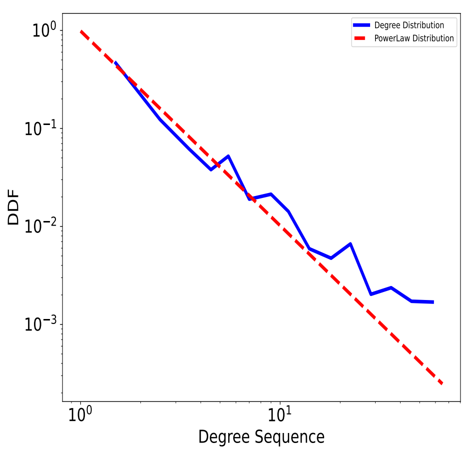
(b) Chinese



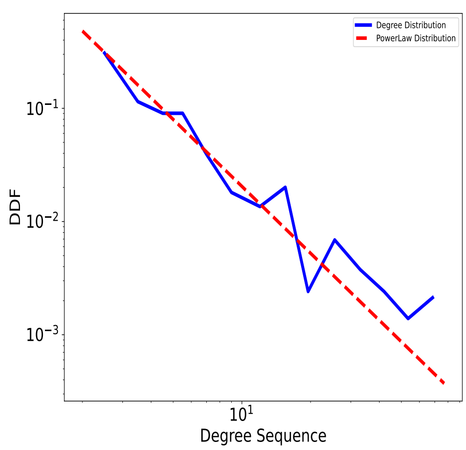
(c) Japanese



(d) US

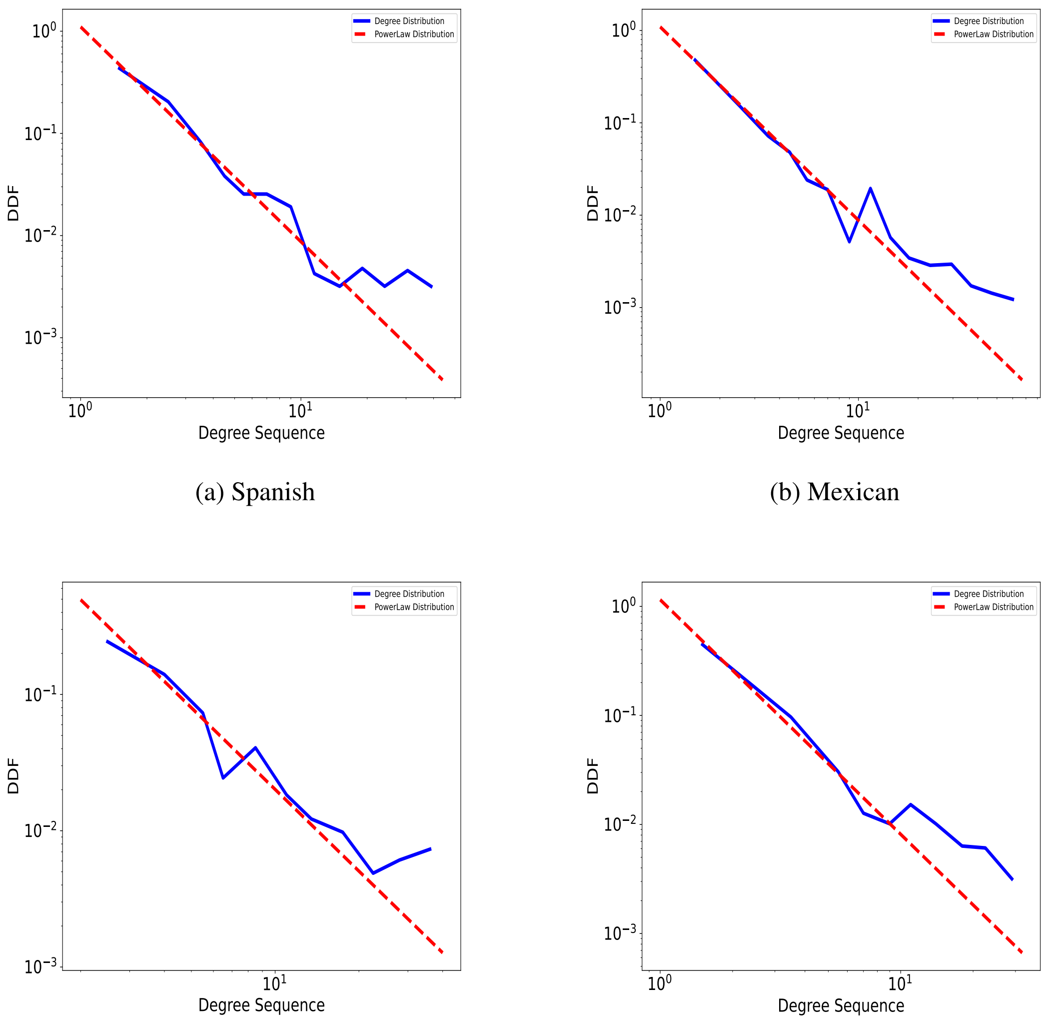


(e) French



(f) Italian

Fig. 4. Degree Distribution Comparison of 6 Cuisine's Ingredient Network along with fitted power law.



(a) Spanish

(b) Mexican

(c) British

(d) Thai

Fig. 5. Degree Distribution Comparison of 4 Cuisine's Ingredient Network along with fitted power law.

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

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

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 analyzed, the diameter consistently measures

4. This uniformity across different cuisines suggests that, regardless of the specific ingredients involved, any ingredient can be connected to any other through no more than four steps, illustrating the closely interconnected nature of culinary ingredients within these networks.

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:

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

where the maximum possible edges for a graph with n nodes is $\binom{n}{2}$. In our analysis, the density values for InN across cuisines ranged from 0.0296 for Indian cuisine to 0.067 for Thai cuisine. These values indicate that the ingredient relationships in the InN for each cuisine are relatively sparse, as densities close to 0 suggest a network with few connections relative to the number of possible connections, whereas densities near 1 indicate a highly interconnected network.

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. [44]. 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:

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

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 value of C_i is essentially the proportion of these connections relative to the total possible connections among the neighbors. In our analysis of the InN, the average clustering coefficients vary significantly across cuisines, from 0.7986 in French cuisine to 0.8399 in Mexican cuisine, indicating a strong tendency for ingredients within these cuisines to be interconnected within recipes.

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:

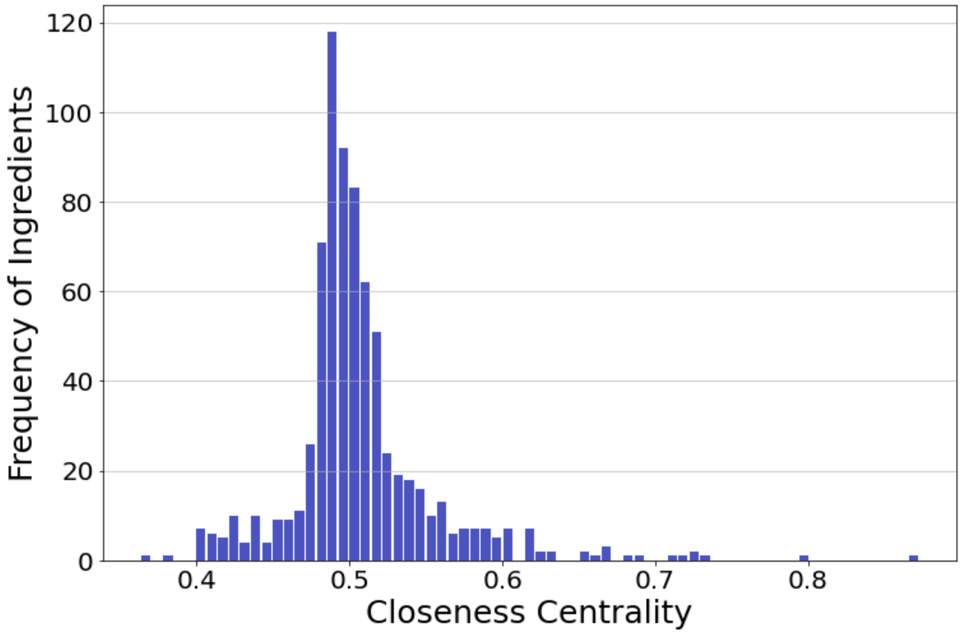


Fig. 6. Closeness Centrality Distribution for the Ingredient Network.

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

A histogram illustrating the CC values for all ingredients is presented in Fig. 6, 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.

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. [80]. 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. 7, there are a few ingredients that stand out with notably higher EC values, underscoring their pivotal roles within their respective culinary networks.

4.8. *Community Structure*

The Ingredient Network (InN) was analyzed using the Weighted Leiden (W-Leiden), Louvain (W-Louvain), and WABCD algorithms [45], 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

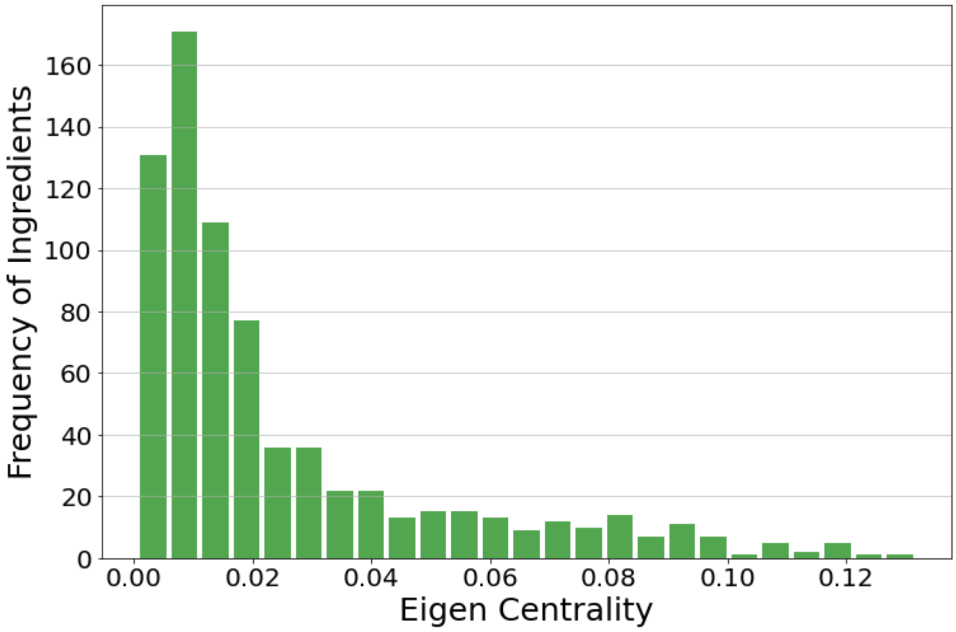


Fig. 7. Eigen Centrality Distribution for the Ingredient Network.

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. 8 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.

5. Discussion

Results of the empirical study reveal that InN shows properties which resemble scale-free networks. For example, Fig. 4 and Fig. 5 demonstrates that InN follows a power law degree distribution with exponent $\gamma = 1.96$ to $\gamma = 2.38$. This value of γ ensures ultra-small property [23], i.e., the average diameter is minimal. The same is found in our experiment; the diameter value is 4. On the contrary, a random network usually has $\gamma \geqslant 3$ [7]. For reader's reference, the degree of separation comparison between real networks and InN for Indian cuisine is shown in Table 3. The table shows the average and maximum distances of the 5 real undirected networks and InN. The columns N, L, k, d and d_max represent the number of nodes, links, average degree, average distance and maximum distance, respectively. In other words, d_max is the diameter. Degrees of separation specify how many hops one must reach from any

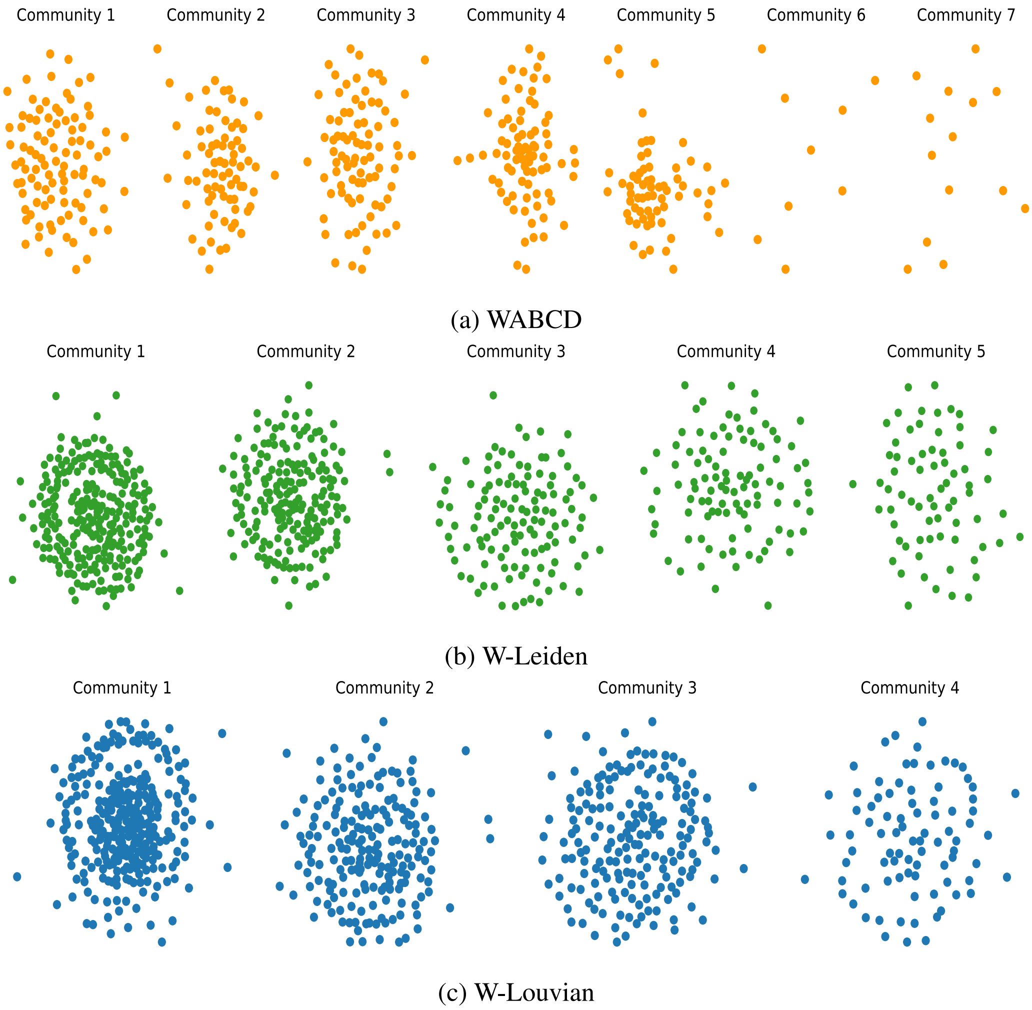


Fig. 8. Results from Different Community Detection Algorithms a) WABCD detects 7 communities b) W-Leiden detects 5 communities c) W-Louvian detects 4

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 [27, 54, 84], which says one node can be reachable through a maximum of 6 hops [61]. 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 [6]. 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 3

Comparison of Real World Undirected Networks with Ingredient Network as to Degree separation and Fluctuations [61].

Network	N	L	d	d_max	$\langle k \rangle$	$\langle k2 \rangle$	γ
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*-
InN (Indian Cuisine)	**1433**	**30,464**	**3.12**	**4**	**1.17**	**14.69**	**2.38***

Table 3 also shows the degree fluctuations in real networks, indicating the average degree $\langle k \rangle$ and the second moment $\langle k2 \rangle$ for 5 undirected real networks and InN. Except for the power grid, every network is scale-free as the estimated degree component γ 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 compares ingredient networks for ten cuisines. It is noteworthy that the InN exhibits similar characteristics for most cuisines. These networks' distance and diameter metrics are consistently around 3 and 4, respectively, indicating an ultra-small world property. Additionally, the degree distribution metric varies from 1.96 (Italian cuisine) to 2.38 (Indian cuisine), indicating a scale-free network property. This suggests that some ingredients have many connections in these networks, while most have fewer connections, contributing to a scale-free structure.

Furthermore, the density metric, which reflects the level of connectivity within the InNs, ranges from 0.033 (for Italian and Southern US cuisines) to 0.067 (for Thai cuisine). This metric further supports the scale-free nature of these networks, as it indicates that specific nodes have a high degree of connections while the majority have fewer connections.

A particularly intriguing observation is that these InNs tend to form communities, with the number of communities ranging from 3 to 9. These communities were identified through weighted Louvain, weighted Leiden, and WABCD. The only exception to this pattern is the InN for Southern US cuisine, where weighted Leiden identifies more communities, precisely 17.

Table 4

Social Metrics Summary of 10 Cuisine's Ingredient Network

Cuisine Wise Ingredient Network	$\lvert V \rvert$	$\lvert E \rvert$	γ	**Dist.**	**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

Fig 4 shows the degree distribution of InN for ten cuisines. It is evident that the degree distribution for each InN closely adheres to a power law, with the degree distribution curves nearly aligning along the central diagonal line.

For a random network, the standard deviation follows $SD = \langle k \rangle\ 1/2$ as shown in Fig. 9, which depicts SD for 4 out of 5 reference networks calculated using Table 3 plus InN. The actor network has a very

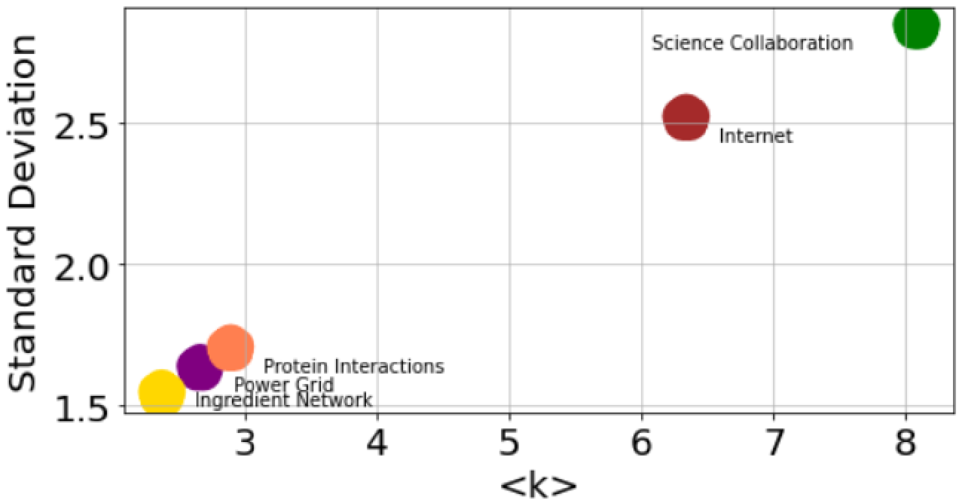


Fig. 9. An illustration of the comparison of standard deviation in Real World Networks with Ingredient Network.

large $\langle k \rangle$ and SD; hence it is omitted for clarity. Each network's SD is more significant than predicted for a random network of the same $\langle k \rangle$. The power grid is the only exception, which is not scale-free. With its small standard deviation, InN is effectively scale-free.

The InN was also tested on other social network metrics and was discovered to behave similarly to most real-world social networks. Both macroscopic and microscopic metrics were calculated on InN. Distance metric shown in Section 4.2 is computed to identify the reachability of one node to the other. Most nodes can be visited from any node in a maximum hop of 3, whereas some require 4 hops. Most real-world social networks follow this property. The density metric is calculated for the InN shown in Section 4.4, and it has been found that InN has sparse connections, similar to most real-world social networks. The average clustering coefficient of InN is calculated, shown in Section 4.5, to be 0.8176, indicating the formation of a cluster where maximum nodes tend to form a more extensive cluster. The majority of real-world social networks follow this property. Another metric, closeness centrality, shown in Section 4.6 is also calculated for InN, which verifies that the new addition of nodes to the graph is more likely to connect with the existing popular ones. Eigen centrality shown in Section 4.7 for InN indicates that few nodes had high connections, with most popular ones resembling the same behaviour as a social network. These calculated graph metrics also added weightage to the conclusion that InN mimics social behaviour and property. A summary of the computed social metrics on InN is shown in Table 4.

6. Conclusion and Future Work

In this paper, we conducted an empirical study to demonstrate that a network of food ingredients (InN), like any other social network, manifests social behaviour. We carried out a thorough examination and the computation of pertinent metrics. The study revealed that InNs follow a power law, a defining feature of a scale-free network. We have discussed various social network metrics and community detection results on InN. We have conducted experiments on ten types of cuisines and evaluation metrics across different cuisines that conform to social network-like behaviour.

In food computing, future research can analyse user interactions within cuisine networks to understand cultural nuances and culinary evolution. Future work paves the way for developing advanced algorithms and models for predicting culinary trends and fostering cross-cultural culinary exchanges within these networks. Additionally, our findings serve as a stepping stone for prospective research endeavours that focus on the ever-evolving impact of social networks on the dynamic landscape of food computing.

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] R. Albert, H. Jeong and A.-L. Barabási, Diameter of the world-wide web, *nature* **401**(6749) (1999), 130–131.

[2] 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.

[3] 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.

[4] 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.

[5] V. Azmanov, EAST INDIAN RECIPES, https://eastindianrecipes.net [accessed May 6, 2021]..

[6] 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.

[7] A.-L. Barabási, Network science book, *Network Science* **625** (2014).

[8] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *science* **286**(5439) (1999), 509–512.

[9] A.-L. Barabási and E. Bonabeau, Scale-free networks, *Scientific american* **288**(5) (2003), 60–69.

[10] 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.

[11] J.A. Barnes, Class and committees in a Norwegian island parish, *Human relations* **7**(1) (1954), 39–58.

[12] 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.

[13] 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.

[14] 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.

[15] A.D. Broido and A. Clauset, Scale-free networks are rare, *Nature communications* **10**(1) (2019), 1–10.

[16] 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.

[17] 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.

[18] 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.

[19] V. Chandola, A. Banerjee and V. Kumar, Anomaly detection: A survey, *ACM computing surveys (CSUR)* **41**(3) (2009), 1–58.

[20] 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.

[21] 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.
[22] M. Ciba, Synchrony Measurement and Connectivity Estimation of Parallel Spike Trains from in vitro Neuronal Networks, PhD thesis, 2021. doi:10.25972/OPUS-22364.
[23] R. Cohen and S. Havlin, Scale-free networks are ultrasmall, *Physical review letters* **90**(5) (2003), 058701.
[24] 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.
[25] Dassana, DASSANA'S VEG RECIPES, https://www.vegrecipesofindia.com [accessed May 7, 2021].
[26] 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.
[27] I. de Sola Pool and M. Kochen, Contacts and influence, *Social networks* **1**(1) (1978), 5–51.
[28] H. Ebel, L.-I. Mielsch and S. Bornholdt, Scale-free topology of e-mail networks, *Physical review E* **66**(3) (2002), 035103.
[29] D.A. Eisenberg, J. Park and T.P. Seager, Sociotechnical network analysis for power grid resilience in South Korea, *Complexity* **2017** (2017).
[30] 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.
[31] I.F. Forever, https://nerdyfoodies.com/indian-spices-list-3291.html [accessed May 5, 2021]..
[32] 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.
[33] H. Ghotra, https://www.harighotra.co.uk/indian-recipes [accessed May 5, 2021]..
[34] 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:$10.1007/978-3-319-30265-2_15$.
[35] M.F. Group., allrecipes, https://www.allrecipes.com [accessed May 7, 2021].
[36] 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.
[37] 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.
[38] P.W. Holland, K.B. Laskey and S. Leinhardt, Stochastic blockmodels: First steps, *Social networks* **5**(2) (1983), 109–137.
[39] 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.
[40] 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.
[41] 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.
[42] 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).
[43] S. Kapoor, https://www.sanjeevkapoor.com/ [accessed May 6, 2021]..
[44] 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.
[45] 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.
[46] D. Knoke and S. Yang, Network fundamentals, *Social network analysis* **154** (2008), 3–14.
[47] 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.
[48] 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.
[49] D.D. Lee and H.S. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature* **401**(6755) (1999), 788–791.
[50] 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.
[51] 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.

[52] 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.
[53] 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.
[54] S. Milgram, The small world problem, *Psychology today* **2**(1) (1967), 60–67.
[55] 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.
[56] 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.
[57] J.M. Montoya and R.V. Solé, Small world patterns in food webs, *Journal of theoretical biology* **214**(3) (2002), 405–412.
[58] J.L. Moreno, Foundations of sociometry: An introduction, *Sociometry* (1941), 15–35.
[59] D.A. Munoz, J.P. Queupil 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.
[60] 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.
[61] M.E. Newman, A.-L.E. Barabási and D.J. Watts, *The structure and dynamics of networks.*, 2006.
[62] M.E. Newman, Scientific collaboration networks. I. Network construction and fundamental results, *Physical review E* **64**(1) (2001), 016131.
[63] M.E. Newman, Fast algorithm for detecting community structure in networks, *Physical review E* **69**(6) (2004), 066133.
[64] M.E. Newman, Finding community structure in networks using the eigenvectors of matrices, *Physical review E* **74**(3) (2006), 036104.
[65] 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.
[66] 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.
[67] 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.
[68] 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.
[69] 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.
[70] 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.
[71] R.A. Rossi and N.K. Ahmed, Role discovery in networks, *IEEE Transactions on Knowledge and Data Engineering* **27**(4) (2014), 1112–1131.
[72] 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.
[73] N. Roy, K. Pushkarna, R. Raj, M. Kaur and S. Mishra, TIMES FOOD, https://recipes.timesofindia.com [accessed May 10, 2021].
[74] 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.
[75] S. Samanta, V.K. Dubey and B. Sarkar, Measure of influences in social networks, *Applied Soft Computing* **99** (2021), 106858.
[76] 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.
[77] S. Shreekanth, SWASTHI'S RECIPES, https://www.indianhealthyrecipes.com [accessed May 8, 2021].
[78] 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.
[79] S. Sosa, C. Sueur and I. Puga-Gonzalez, Network measures in animal social network analysis: Their strengths, limits, interpretations and uses (2020).
[80] 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.
[81] 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.
[82] 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.
[83] 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.

[84] J. Travers and S. Milgram, An experimental study of the small world problem, in: *Social networks*, 1977, pp. 179–197.
[85] S. Valverde, R.F. Cancho and R.V. Sole, Scale-free networks from optimal design, *Europhysics Letters* **60**(4) (2002), 512.
[86] A. Vazquez, R. Pastor-Satorras and A. Vespignani, Internet topology at the router and autonomous system level, *arXiv preprint cond-mat/0206084* (2002).
[87] 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.
[88] 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.
[89] 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.
[90] S. Wasserman, *Advances in social network analysis: Research in the social and behavioral sciences*, Sage, 1994.
[91] D.J. Watts and S.H. Strogatz, Collective dynamics of 'small-world'networks, *nature* **393**(6684) (1998), 440–442.
[92] 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.
[93] 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.
[94] 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.
[95] 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.
[96] 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.
[97] 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.
[98] 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.
[99] Yummly, Recipe Ingredients Dataset, 2024, Accessed: 2024-06-21. https://www.kaggle.com/datasets/kaggle/recipe-ingredients-dataset.
[100] 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.