

Responses to Reviewers' Comments for Manuscript 859-1839

Are Food Ingredient Social? An Empirical Investigation

Addressed Comments for Publication to

DATA SCIENCE

Dear Editor,

Please find enclosed the revised version of our previous submission entitled “Are Food Ingredient Social? An Empirical Investigation” with the old manuscript number “859-1839”. We would like to thank you and the reviewers for the valuable comments which helped improving the quality of our manuscript. In this updated version, we have carefully addressed the reviewers’ comments. A summary of main modifications and a detailed point-by-point response to the comments from Reviewers 1 to 3 (following the reviewers’ order in the decision letter) are given below.

Sincerely,

Note: To enhance the legibility of this response letter, all the editor’s and reviewers’ comments are typeset in boxes. The respective parts in the revised manuscript are highlighted in blue to indicate changes.

Authors' Response to the Editor

General Comments. I wish to inform you that the acceptance or rejection of your manuscript is still UNDECIDED (we don't use "major revisions" or "minor revisions").

Response: We appreciate your handling of the review process.

Comment 1

The contribution relative to the previously published dataset INDoRI should be clarified, along with claim that the datasets have a graph based representation (See Reviewer 1 comments)

Response:

We appreciate the meta-reviewer's comments. In the revised manuscript, we modified the abstract as well as changed the narrative structure of the introduction to clarify our contribution related to the INDoRI dataset. We have also expanded the discussion on data curation and quality control to ensure transparency.

Additionally, we have clearly established the graph-based representation of the dataset, explaining how ingredient networks were constructed and analyzed. Furthermore, we have addressed all the concerns raised by Reviewer 1, specifically focusing on the dataset structure, methodology, and its positioning within the existing literature.

Comment 2

The Lack of Interpretation and Justification should be addressed - in particular with respect to Social Behaviour and Community Structure.

Response:

In the revised manuscript, we have significantly enhanced the interpretation and justification of our findings, particularly in relation to social behavior and community structure.

We have expanded the discussion on how ingredient networks exhibit characteristics similar to social networks, explaining the implications of clustering, connectivity, and modularity in the context of culinary traditions.

Additionally, we have provided a more detailed analysis of the community structure, drawing connections between ingredient groupings and cultural, regional, and functional aspects of cooking. [The same is reflected in the Results & Analysis and Discussion Section on page 8 and 15 of the revised manuscript]

Comment 3

The link to ‘social’ networks should be underpinned by additional theoretical grounding or the claim should be relaxed.

Response:

We appreciate the meta-reviewer’s comment regarding the theoretical grounding of the link to social networks. In the revised manuscript, we have strengthened our argument by providing additional statistical validation of the scale-free nature of ingredient networks. From the linear regression analysis performed on the log-transformed data across 10 global cuisine ingredient networks, we observed a consistent range of values for the slope (-2.45 to -2.68), intercept (0.18 to 0.22), and high R-squared values (0.9965 to 0.9991), indicating an excellent fit between the log-transformed degree sequence and degree distribution. Furthermore, the extremely low p-values (10^{-25} to 10^{-30}) provide strong statistical evidence supporting the robustness of our findings.

These results reinforce the argument that ingredient networks exhibit power-law behavior, a fundamental property observed in many real-world social networks. By integrating these statistical validations, we have strengthened the theoretical grounding of our claims while ensuring that our conclusions remain well-supported and justified. [The same is reflected in the Degree Distribution of InN subsection of Result & Discussion Section on page 8 of the revised manuscript]

Comment 4

There a number of gaps in the related work and the contribution of the paper needs updated accordingly

Response:

In the revised manuscript, we have incorporated additional literature, including the references suggested by the reviewers, to better contextualize our study within the existing body of research. This collection includes key works on ingredient networks, food pairing principles, and computational gastronomy, which help bridge the gap between previous studies and our contributions.

By integrating these references, we have strengthened the positioning of our work, clearly outlining how our study extends beyond existing research. This ensures that our contributions are well-differentiated and highlights the novelty of our findings in the domain of ingredient network analysis. [The same is reflected in the Related Work Section on page 4 of the revised manuscript]

Comment 5

The overall technical rigour of the manuscript needs significant improvement and comparisons across cuisines types should be more systematic and consistent

Response:

In the previous version of the manuscript, we have analyzed 10 global cuisines. In the revised version, we have enhanced the systematic comparison of ingredient networks across different cuisine types by ensuring consistency in the evaluation metrics and analysis framework.

We have refined the result and discussion sections, providing clearer justifications for our analytical choices and ensuring uniformity in metric reporting (e.g., degree distribution, clustering coefficient, eigen centrality, and community structure). Additionally, we have strengthened the statistical validation of our findings, reinforcing the robustness of our cross-cuisine comparisons. These improvements collectively enhance the rigor and clarity of our study. [The same is reflected in the Results & Analysis and Discussion Section on page 8 and 15 of the revised manuscript]

Authors' Response to Reviewer 1

Comment 1

The authors state “Further, the dataset includes a graph-based representation of ingredient relationships, namely, ingredient network (InN).” However the dataset provided does not contain a graph based representation and requires substantial cleaning/processing to extract ingredients.

Response: The authors are thankful to the reviewer for pointing this out.

We have added the graphical dataset to the existing repository. <https://figshare.com/s/12a1bce0210a7f031168>

Comment 2

The authors state “The dataset and its characteristics were earlier published in [45].” If this is the case, the stated contribution is in direct conflict with the abstract “This paper introduces INDoRI (Indian Dataset of Recipes and Ingredients)” – if the dataset has already been presented and published elsewhere the authors should modify their contribution. If this is an updated version of the original the authors should state that fact to avoid confusion

Response:

We would like to clarify that the dataset referenced in [45] is indeed the same dataset; however, the current paper introduces significant updates, enhancements, and additional analyses that were not part of the earlier publication. However, as the reviewer pointed out, some parts of the abstract and introduction did not reflect the contribution appropriately. We have revised the abstract and relevant sections of the manuscript to explicitly state the contributions of the present paper. We hope this clarification addresses the reviewer’s concern. [Please refer to the Abstract and Introduction Section on page 1, 2 and 3 to review the changes in the revised manuscript]

Comment 3

In Table 1 the authors detail the distribution of different cuisines used to create the InN, this data does not appear to be available at the figshare link provided in the document, which solely contains the INDoRI dataset. The authors state “The dataset used in this study is available online at <https://figshare.com/s/12a1bce0210a7f031168>” but only the INDoRI dataset is available at this address.

Response:

We sincerely appreciate the reviewer’s observation regarding the availability of the data used to create the graphical ingredient network. We have included the graphical dataset in the original repository. <https://figshare.com/s/12a1bce0210a7f031168>

Comment 4

The authors state “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 cuisine” Perhaps this is a typo in the first sentence, as Table 1 contains 10 cuisines

Response: We thank the reviewer for their observation.

We acknowledge that the phrasing in the original manuscript may have caused some confusion. To clarify, our analysis incorporates two datasets: (1) the INDoRI dataset, from which we use one cuisine, and (2) the Yummly dataset, from which we selected nine cuisines based on their popularity and the quantity of available records. This results in a total of ten cuisines analyzed in our study. We have revised the relevant paragraph in the manuscript to make this distinction clearer and avoid any ambiguity. [Please refer to the INDoRI Section on page 5 to review the changes in the revised manuscript]

Comment 5

I am confused as to why there is such a comprehensive breakdown across cuisines in S4, however table 3 here focuses solely on the Indian specific subgraph. As this data is already present it would improve the quality of the paper to merge Tables 3 and 4 together to provide a single comprehensive breakdown. - Similarly differing notation is used across the two tables e.g. “N” in Table 3 and “|V|” in Table 4 to refer to the same concepts (the number of nodes)

Response: We thank the reviewer for pointing this out.

The inclusion of Table 3 (in the revised manuscript Table 4) was intentional, as it provides a comparison of at least one ingredient network (in our case we have included Indian) with other real world social networks to highlight structural similarities. In contrast, Table 4 (in the revised manuscript Table 5) was specifically included to demonstrate the social properties across all ten cuisines. However, we acknowledge the inconsistency in notation and have standardized it across both tables for clarity. [Please refer to the Table 4 and Table 5 on page 17 and 18 to review the changes in the revised manuscript]

Comment 6

Is Figure 9 referring to the entire InN or specifically the Indian cuisine subgraph?

Response:

Figure 9 (in the revised manuscript Figure 8) specifically refers to the Indian cuisine subgraph of the Ingredient Network (InN). To avoid any confusion, we have updated the figure caption to explicitly state that the visualization represents the Indian cuisine subgraph of the InN. The revised caption now reads: “An illustration of the comparison of standard deviation in Real World Networks with Ingredient Network of Indian cuisine”. We hope this clarification resolves any ambiguity and improves the clarity of the figure. [Please refer to the Figure 8 on page 19 to review the changes in the revised manuscript]

Authors' Response to Reviewer 2

Comment 1

Lack of Interpretation and Justification: The paper presents a lot of quantitative data but fails to provide sufficient interpretation or justification.

Response:

We appreciate the reviewer's critical feedback. In the revised manuscript, we have added interpretation and justification of our analysis with culinary patterns of cooking. The revised manuscript now includes these additions in both the Results and Discussion sections, providing a clearer explanation of our findings and their significance. We believe this enhances the overall clarity and impact of our study. [Please refer to the Result & Analysis and Discussion Section on page 8 and 15 to review the changes in the revised manuscript]

Comment 2

Gamma Range: The range of gamma values (1.96-2.38) is mentioned, but its significance in the context of food is not explained. How does this compare to other networks?

Response:

Thank you for your valuable feedback. The range of gamma values indicates that the ingredient networks (InN) exhibit similar scale-free properties to other real-world networks, including social, biological, and technological networks. In the context of food, this suggests that ingredient networks are highly heterogeneous, with a few ingredients (hubs) being central to many recipes, while most ingredients have limited connections. This aligns with the observation that certain staple ingredients (e.g., salt, onions, or rice) are ubiquitous across recipes, while others are more specialized. [Please refer to Degree Distribution subsection of the Results and Analysis Section on page 8 to review the changes in the revised manuscript]

Comment 3

Maximum Distance: While different maximum distances across cuisines are noted, no culinary explanations are offered.

Response:

In the revised manuscript, we have included culinary explanations for the observed differences in maximum distances. [Please refer to Discussion Section on page 24 to review the changes in the revised manuscript]

Comment 4

Micro Metrics: The choice of micro metrics is not justified, and their results are not interpreted in a meaningful way. What do the closeness and eigen centrality values imply about ingredient usage?

Response:

We have revised our analysis to provide a clearer justification for the selected micro metrics and their implications for understanding ingredient usage in the network. Specifically, we have elaborated on how closeness centrality and eigen centrality values offer insights into the influence and accessibility of individual ingredients within the network. Closeness centrality highlights ingredients that are closely connected to others, suggesting their potential role as central or bridging components, while eigen centrality identifies ingredients that are connected to other highly connected ingredients, indicating their influence in the network. These metrics collectively help us understand the relative importance and functional roles of ingredients in the network. [Please refer to the Result & Analysis and Discussion Section on page 8 and 15 to review the changes in the revised manuscript]

Comment 5

Community Structure: The paper mentions the number of communities found by different algorithms but fails to provide examples of these communities or explain their culinary significance. This is a crucial flaw. Simply finding clusters does not prove a meaningful "community structure" analogous to social networks. How do these communities relate to known culinary practices or ingredient pairings? The paper does not address the crucial difference between simple co-occurrence clustering and true community structure.

Response:

Thank you for your valuable suggestions. In response, we have added a table that maps the communities of identified ingredients to relevant culinary practices. These communities align with recipe categories such as breakfast, lunch / dinner, and beverages, illustrating how ingredients tend to cluster according to meal preparation times. This addition clarifies the culinary significance of the detected communities and differentiates true community structures from simple co-occurrence clustering. [Please refer to Table 3 and Community Structure subsection of Results & Analysis Section on page 16 and 14 respectively, to review the changes in the revised manuscript]

Comment 6

Table Interpretation: Both Table 3 and Table 4 are presented with minimal analysis. For example, the discussion of Table 4 merely restates the numbers without explaining the observed differences between cuisines. Why does Southern US cuisine have so many more communities according to Leiden?

Response:

We have added a more detailed discussion and analysis of Tables 3 and 4, explaining the observed differences between cuisines. Specifically, we have examined why certain cuisines, such as Southern US, exhibit more communities under the Leiden algorithm, linking these patterns to their diverse ingredient usage and culinary traditions. Please note the purpose of designing INDoRI was to perform food computing related task as well as to understand the structure of the graphical network, therefore inherently the category of recipes are labeled within the dataset like Breakfast, lunch etc and our

experiments revealed that these are nothing but communities for Indian InN. However, other cuisine networks such as US Southern cuisine are not designed for such analysis and hence lack the related labeling. Therefore we could not empirically verify the meaning of the communities for other cuisines. [Please refer to the Discussion Section from page 16-18 to review the changes in the revised manuscript]

Comment 7

Weak Justification for ‘Social Behaviour’: The paper claims that InNs exhibit ‘social behaviour’ but does not adequately justify this claim. Simply showing that InNs have properties like power-law distributions and community structure is not enough. Many non-social networks also exhibit these properties. The analogy to social networks feels forced and lacks a strong theoretical foundation. The paper does not distinguish its findings from simple frequency-based co-occurrence or topic modelling approaches for data clustering.

Response:

We appreciate the thoughtful feedback from the reviewer. We agree that many non-human or non-animal networks show power-law distribution. However, we would like to highlight that the research community of social network analysis identifies all such scale-free networks as analogous to social networks, often referring to them as such. The same ideas are presented in the book on Network Science by Albert-Barabasi (<http://www.networksciencebook.com>) and many other textbooks on social network analysis. In fact, the current paper explores the same question: does the InN exhibit properties similar to those found in social and its analogous networks? None of the previous studies tried to verify it with the ingredient networks. In the manuscript, we have strengthened our justification by performing an in-depth empirical analysis in 10 global cuisine ingredient networks. In particular, we show through a comparison study how ingredient networks show important traits of social behavior, including preferential attachment, community formation based on modularity, and complex assortativity patterns. Further, our revised manuscript now explicitly highlights how the observed patterns go beyond simple frequency-based co-occurrence or topic modeling approaches.

Comment 8

However, the current manuscript has a significant weakness in establishing the motivation and significance of this study. The authors do not adequately explain why it is important to understand whether ingredient networks exhibit ‘social behaviour’. What are the potential applications or benefits of this knowledge? How does it advance the field of data science or, more specifically, food computing?

Response:

We appreciate your thoughtful feedback. While we recognize that it is important to think about potential applications, empirical understanding of data and component therein plays equal importance in scientific explorations. Our study primarily focuses on the structural analysis of InNs. The observed properties provide a 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. Both of these may lead into alternate ingredient predictions or even innovative recipe generations. However, these applications remain speculative at this stage and would require additional studies, such as analyzing temporal data on recipe creation or user preferences, to validate their feasibility.

We have incorporated an application section in our revised manuscript, which elaborates on the potential advantages in the realm of food computing. [Please refer to the Applications Section on page 19 to review the changes in the revised manuscript]

Comment 9

The paper needs to clearly articulate the potential impact of this research. Some possible areas of impact, which the authors could explore and develop, include:

- Recipe Recommendation: Understanding ingredient relationships could lead to more effective recipe recommendation systems that go beyond simple co-occurrence and consider more complex culinary patterns.
- Culinary Trend Prediction: Analyzing changes in ingredient networks over time could help predict emerging culinary trends and identify new ingredient combinations.
- Cross-Cultural Culinary Analysis: Comparing ingredient networks across different cuisines could reveal cultural influences and similarities or differences in culinary practices.
- Food Product Development: Insights from ingredient networks could be used to inform the development of new food products and flavour combinations.
- Nutritional Analysis: Combining ingredient network analysis with nutritional data could provide insights into the nutritional properties of different cuisines and identify potential nutritional deficiencies or imbalances.

Response:

Thank you for your valuable feedback. In response, we have expanded the discussion on the potential impact of our research by incorporating a dedicated applications section in the revised manuscript. This section explores key areas such as recipe recommendation, culinary trend prediction, cross-cultural culinary analysis, food product development, and nutritional analysis. By leveraging ingredient network properties, we highlight how our findings contribute to advancements in food computing and culinary science. These additions provide a clearer articulation of the broader implications of our work. [Please refer to the Applications Section on page 19 to review the changes in the revised manuscript]

Comment 10

However, the paper's connection to the existing literature on ingredient network analysis is significantly weaker. While some relevant papers are cited, the authors fail to adequately contextualize their work within this specific subfield. This is a major shortcoming, as it leaves the reader wondering how this work builds upon or differs from previous research on ingredient networks. Specifically, the paper should more thoroughly discuss and relate its findings to the following works:

- 1) Ahn, Y. Y., Ahnert, S. E., Bagrow, J. P., & Barabási, A. L. (2011). Flavor network and the principles of food pairing. *Scientific reports*, 1(1), 196.:
- 2) Teng, C. Y., Lin, Y. R., & Adamic, L. A. (2012, June). Recipe recommendation using ingredient networks. In *Proceedings of the 4th annual ACM web science conference* (pp. 298-307).:
- 3) Shirai, S. S., Seneviratne, O., Gordon, M. E., Chen, C. H., & McGuinness, D. L. (2021).
- 4) Cheng, X., Lin, S. Y., Wang, K., Hong, Y. A., Zhao, X., Gress, D., ... & Xue, H. (2021).
- 5) Healthfulness assessment of recipes shared on Pinterest: natural language processing and content analysis. *Journal of Medical Internet Research*, 23(4), e25757:

By not adequately addressing these related works on ingredient networks, the reviewed paper fails to establish its place within the existing research landscape. It's crucial for the authors to explain how their work contributes new insights beyond what has already been established in the field. While the paper provides a reasonable background on general SNA concepts, it lacks a sufficient connection to the specific literature on ingredient network analysis. The authors need to significantly expand this section by discussing and comparing their work to existing research on ingredient networks, particularly the papers mentioned above.

Response:

We sincerely appreciate the reviewer's feedback regarding the need to strengthen the connection between our work and existing literature on ingredient network analysis. In the revised manuscript, we have significantly expanded our discussion of related works and explicitly positioned our study within this research landscape. We have expanded the related work section to incorporate a detailed discussion of recent studies on ingredient networks. and also clarified how our work builds upon and differs from prior research

by explicitly comparing methodologies, findings, and contributions. [Please refer to the Related Work Section on page 4 to review the changes in the revised manuscript]

Comment 11

INDoRI Dataset: The creation and presentation of the INDoRI dataset, a collection of 5187 Indian recipes covering 18 diverse Indian cuisines, can be considered a substantial contribution. If this dataset is indeed comprehensive and well-curated, it could be a valuable resource for future research in food computing and culinary analysis. The authors should emphasize the unique characteristics of this dataset and how it compares to existing recipe datasets. For example: 1) What is the size and scope of other publicly available Indian recipe datasets (if any)? 2) Does INDoRI include metadata or attributes beyond ingredients (e.g., preparation time, nutritional information, regional origin)? 3) How was the data collection and cleaning process performed to ensure data quality? If the INDoRI dataset offers unique advantages or fills a gap in existing resources, this would significantly strengthen the paper's novelty. It is commendable that the authors have provided a link to the dataset. ◦ Combined Analysis: The analysis performed on the aggregated dataset, which includes INDoRI and a dataset from Yummly covering international cuisines, could also be considered a novel aspect. Comparing ingredient network properties across different cuisines (Indian vs. international) could reveal interesting cultural and culinary insights. However, the current analysis is quite superficial and fails to fully exploit this potential. To enhance the novelty of this combined analysis, the authors should: 1) Provide a more detailed comparison of the network properties across different cuisines. What are the key differences and similarities? What culinary factors might explain these differences? 2) Explore how the network structure reflects cultural differences in ingredient usage and culinary traditions. 3) Consider more advanced comparative analysis techniques to identify statistically significant differences between cuisines.

Response:

We sincerely appreciate the reviewer's insightful comments and suggestions regarding the INDoRI dataset and the comparative analysis between Indian and international cuisines. In the revised manuscript, we have thoroughly addressed each of the concerns by adding

a detailed explanation. [Please refer to the Result & Analysis Section on page 8 to review the changes in the revised manuscript]

Comment 12

Lack of New Methods: The paper does not introduce any new methods for network analysis or ingredient network construction. It relies on standard SNA metrics and community detection algorithms. This limits the methodological novelty of the work.

Response:

While our work utilizes standard Social Network Analysis (SNA) metrics, we provide novel insights by analyzing ten different ingredient networks and demonstrating that they follow a power-law distribution and exhibit social properties. In addition, we apply a community detection approach to analyze ingredient communities across networks, offering a deeper understanding of ingredient interactions. These contributions add methodological value beyond standard network analysis techniques. To the best of our knowledge, ours is the first work to analyze the social behavior in Food Ingredient network.

Comment 13

Superficial Analysis: The analysis performed on the combined dataset is currently too superficial to be considered a significant contribution. Simply calculating and comparing basic network metrics is not enough. The authors need to provide a deeper analysis and interpretation of the results to extract meaningful insights.

Response:

Thank you for your feedback. In response, we have performed additional statistical analysis on the log-transformed data across 10 global cuisine ingredient networks, we observed a 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. [Please refer to the Degree Distribution of InN subsection of Result & Analysis and Discussion Section on page 8 and 15 to review the changes in the revised manuscript]

Comment 14

Missing Comparison to Existing Ingredient Network Research: As already mentioned, the paper does not adequately compare its findings to existing research on ingredient networks. This makes it difficult to assess the true novelty of the work.

Response:

Thank you for your feedback. We acknowledge the importance of comparing our findings with existing research on ingredient networks. In response, we have incorporated a discussion on relevant literature, highlighting previous studies on ingredient networks and positioning our work in relation to them. This addition clarifies the novelty of our approach and provides a comprehensive comparison with existing research. [Please refer to the Related Work Section on page 4 to review the changes in the revised manuscript]

Comment 15

Network Analysis Methods: The paper mentions using standard network metrics (degree distribution, distance, diameter, density, clustering coefficient, closeness centrality, and eigen centrality), but provides little detail on how these metrics were calculated. Were standard libraries used (e.g., NetworkX in Python)? Were any specific parameters or settings used? This lack of detail hinders reproducibility.

Response:

In the revised manuscript, we have explicitly stated that we used standard Python libraries, including NetworkX, igraph, for computing network metrics and community.

Additionally, we have provided details on the specific parameters and settings used to ensure reproducibility. We believe these additions enhance the clarity and transparency of our methodology. [Please refer to the Results and Analysis Section on page 8 to review the changes in the revised manuscript]

Comment 16

Community Detection Algorithms: The paper mentions using weighted versions of Leiden, Louvain, and WABCD algorithms, but provides insufficient information about their implementation. 1) Were standard implementations used, or were they modified in any way?

2) What parameters were used for each algorithm?

3) The paper mentions that WABCD is described in [45], but this paper should be self-contained.

4) Critically, the paper does not adequately justify the choice of these specific algorithms. Why were these algorithms chosen over other community detection methods? What are the strengths and weaknesses of each algorithm in the context of ingredient networks?

Response:

Thank you for your detailed feedback. In response, we have expanded the discussion on community detection by including an additional table and a dedicated paragraph in the revised manuscript. This enhancement provides more meaningful insights related to the results of the community detection algorithms, explaining how different methods influence the detected ingredient communities and their culinary significance.

Additionally, we address your specific queries as follows:

1) Implementation of Algorithms: We used standard implementations of the Leiden, Louvain, and WABCD algorithms, available in well-established network analysis libraries. However, for weighted community detection, we ensured that the algorithms considered edge weights, which represent ingredient co-occurrence frequencies.

2) Parameters Used: Leiden Algorithm: The resolution parameter was set to 1.0 (default), ensuring a balance between detecting smaller and larger communities.

Louvain Algorithm: The resolution parameter was also set to 1.0 to maintain consistency in community granularity across methods.

WABCD Algorithm: We followed the parameter settings described in [45], incorporating weight-based adjustments to detect ingredient communities more effectively. These parameters are detailed in the revised manuscript.

3) Self-Containment of WABCD Explanation: We acknowledge that relying on [45] for the WABCD description might make the paper less self-contained. In response, we have added a brief explanation of the WABCD algorithm, summarizing its key working principles and how it applies to ingredient networks.

4) Justification for Algorithm Choice: We selected these three algorithms due to their effectiveness in weighted network analysis and their ability to detect meaningful community structures.

Leiden Algorithm: Chosen for its optimization of modularity and ability to detect well-separated communities efficiently.

Louvain Algorithm: Selected as it is widely used in community detection and provides fast, hierarchical clustering.

WABCD Algorithm: Included because it is specifically designed for weighted networks, making it well-suited for ingredient networks where co-occurrence frequency is critical.

The revised manuscript now includes a discussion of the strengths and weaknesses of each algorithm in the context of ingredient networks, explaining why these methods were chosen over others. [Please refer to the Community Structure Identification subsection of the Methodology Section on page 8 to review the changes in the revised manuscript]

Comment 17

Lack of Comparison to Other Clustering Methods: The paper does not discuss how its network-based community detection compares to other clustering methods, such as topic modeling. This is a crucial omission. Topic modeling is a common technique for analyzing text data, including recipe descriptions, and can also be used to identify groups of related ingredients. The authors could: 1) Discuss the similarities and differences between network-based community detection and topic modeling in the context of ingredient analysis. 2) Explain why they chose a network-based approach over topic modeling or other clustering methods. 3) Ideally, they should perform a comparison between network-based communities and topic-based clusters to demonstrate the advantages (if any) of their approach.

Response:

Thank you for your suggestions. The choice between a network-based approach and topic modeling for analysis of ingredients network depends on the goal and nature of the data. In our work, we have performed the analysis on InN's keeping in mind applications such as recipe recommendation, alternative ingredient sharing, and ingredient replacement. Our choice of network-based approach directly captures the structural relationships between ingredients, enabling more precise and actionable insights. Unlike topic modeling, which focuses on semantic clustering of ingredients based on textual co-occurrence, network analysis leverages explicit connections between ingredients, such as shared flavor compounds, co-occurrence in recipes, or functional roles in dishes. This allows for identifying substitutions based on structural compatibility (e.g., flavor similarity or bridging) and predicting novel pairings through graph-based principles like food-pairing and food-bridging. Furthermore, network-based methods are well suited for handling large-scale datasets with sparse or dynamic interactions, ensuring scalability and adaptability to diverse culinary contexts. By modeling the interconnection of ingredients, this approach provides a robust framework to optimize recipes while preserving flavor profiles and nutritional balance. However, as per the suggestion of the reviewers, one could carry out a separate analysis of the application of the topic modelling on InNs, for a different set of applications. Currently, such an analysis is beyond the scope of this paper and will be explored in the future.

Comment 18

Statistical Analysis: The paper makes claims about power-law distributions but doesn't provide any statistical measures of fit (e.g., R-squared values, p-values). This makes it impossible to assess the statistical significance of these claims. Similarly, the paper doesn't provide any statistical comparisons between different cuisines.

Response:

We appreciate the feedback of the reviewer and have now included statistical measures to support our claims about power-law distributions. From the linear regression analysis performed on the log-transformed data across 10 global cuisine ingredient networks, we observed a consistent range of values for 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 the 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 in diverse cuisines exhibit power-law behavior, reinforcing the robustness of our findings. We have updated the revised manuscript to reflect these statistical analyses. The same has been included in the revised manuscript. [Please refer to the Degree Distribution of InN subsection of Result and Analysis Section on page 8 and 10 to review the changes in the revised manuscript]

Comment 19

Interpretation of Results Even if the methods were adequately described, the interpretation of the results is weak. The paper presents numbers without providing sufficient culinary context or explanation.

Response:

Thank you for your insightful feedback. We acknowledge the need for a stronger interpretation of our results. In response, we have added detailed explanations for each result in relation to the Ingredient Network (InN), providing the necessary culinary context. These additions clarify the significance of our findings and enhance the overall understanding of the results. [Please refer to the Result and Analysis Section on page 8 to review the changes in the revised manuscript]

Comment 20

Contradictory statements on INDoRI Dataset: There's a clear contradiction on page 2. Lines 30-32 suggest that the paper introduces the INDoRI dataset, while lines 41-42 state that it was presented in a previous publication [45]. This needs to be clarified. If INDoRI was previously published, this paper should clearly state that it is using or extending the previously published dataset, not introducing it. This contradiction undermines the perceived novelty of the work.

Response:

Thank you for your pointing out the error. It was an honest mistake. To clarify, our work analyzes and utilizes the INDoRI dataset rather than introducing it. We apologize for the inconsistency and have updated the revised manuscript to explicitly state that we are using a previously published dataset, ensuring clarity and avoiding any misrepresentation of novelty. [Please refer to the Abstract and Introduction Section on page 1 and 2 to review the changes in the revised manuscript]

Comment 21

Redundant data cleaning section: Given that the INDoRI dataset and its cleaning process are deemed described in [45], the detailed explanation of the cleaning process in Section 3.3 is largely redundant. The authors should simply briefly summarize the cleaning steps and refer the reader to [45] for more details.

Response:

We appreciate your input; we have removed Section 3.3 and included a brief paragraph outlining the cleaning steps in the revised manuscript. The reader is directed to consult the referenced work for further details. [Please refer to Empirical Analysis subsection of Methodology Section on page 6 to review the changes in the revised manuscript]

Comment 22

Missing context for Stop Words: While Table 2 shows ingredient stop words (ISW), it would be much more helpful to provide examples of these stop words in context before cleaning. For example, instead of just listing "chopped," the authors could provide an example like "1 cup of chopped onions" and then show how it is reduced to "onions" after cleaning. This would make the purpose and effect of ISW filtering much clearer.

Response:

Thank you for your valuable feedback. As per your previous suggestions, we have removed Table 1. Additionally, we have incorporated more contextual examples to illustrate the ISW filtering process. These examples demonstrate how ingredient stop words like “chopped” in phrases such as “1 cup of chopped onions” are cleaned to retain only the essential ingredient name, such as “onions.” This enhancement provides clearer insight into the purpose and effect of ISW filtering. [Please refer to the Dataset for Empirical Analysis subsection of Methodology Section on page 6 to review the changes in the revised manuscript]

Comment 23

The phrase "examples of such words can be found in referenced Table 2" should be simplified to "examples of such words can be found in Table 2."

Response:

Thank you for your valuable feedback. We have made the change in the revised manuscript, simplifying the phrase as suggested.

Comment 24

Inconsistent Number Formatting: The inconsistent use of numerical (e.g., 9) and textual (e.g., nine) representations of numbers should be corrected. The authors should consistently use numerical representations for numbers greater than ten and follow a consistent style guide.

Response:

We have carefully reviewed the entire document and corrected the formatting to ensure uniformity. As per the reviewer's suggestion, we have consistently used numerical representations (e.g., 9) for numbers greater than ten and followed a uniform style guide throughout the manuscript.

Comment 25

Inconsistent cuisine count: The discrepancy between mentioning 9 cuisines in line 42 and 10 cuisines in line 43 on page 4 needs to be resolved. The authors should clearly state that they used 9 cuisines from Yummly in addition to the Indian cuisine from INDoRI, resulting in a total of 10 cuisines.

Response:

We thank the reviewer for pointing out the inconsistency in the cuisine count mentioned in the manuscript. We acknowledge the discrepancy between the mention of 9 cuisines and 10 cuisines on page 4. To address this, we have revised the text to clearly state that we used 9 cuisines from the Yummly dataset in addition to the Indian cuisine from the INDoRI dataset, resulting in a total of 10 cuisines analyzed in the study. This clarification ensures accuracy and eliminates any confusion for the reader. The revised manuscript now reflects this correction. [Please refer to the Dataset for Empirical Analysis subsection of Methodology Section on page 5 to review the changes in the revised manuscript]

Comment 26

Impact of using external data: The authors mentioned that they used 9 cuisines from Yummly dataset which I beleieve, overshadows the contribution of INDoRI – their main contribution supposedly. The authors need to address this directly. For instance, they might find it beneficial to:

- 1) Clearly explain the rationale for using the Yummly dataset.
- 2) Emphasize the unique contributions of INDoRI, even within the combined analysis.
- 3) Consider performing separate analyses on INDoRI to highlight its specific characteristics.

Response:

We appreciate the reviewer's concern. To address this, we have revised the manuscript to clearly explain the rationale for incorporating the Yummly dataset, emphasize the unique contributions of INDoRI, and highlight its distinct characteristics. [Please refer to the Dataset for Empirical Analysis subsection of Methodology Section on page 5 to review the changes in the revised manuscript]

Comment 27

In page 3, line 41-45, the authors wrote –"One of them is to compilling recipes that span diverse cultural. . . ." The authors should consider changing "Compiling" → compile

Response:

We have addressed this issue in the updated manuscript.

Comment 28

Discussion : The authors provide different evaluation metrics without adequately discussing their implications. Each metric should be explained in the context of ingredient networks and the results should be interpreted accordingly.

Response:

We appreciate your recommendations. As per your advice, we have integrated explanations for each metric concerning the ingredient network into both the Result Section and the Discussion Section. [Please refer to the Result and Discussion Section on page 8 and 15 to review the changes in the revised manuscript]

Comment 29

Irrelevant Introduction: The introduction's lengthy discussion of general social network concepts without a clear connection to the specific work on ingredient networks is a concern for me. The introduction should be focused on motivating the study of ingredient networks and establishing the research question. The general background on social networks should be significantly shortened and integrated more seamlessly into the context of ingredient analysis.

Response:

We thank the reviewer for their valuable feedback regarding the introduction. To address this concern, we have revised the introduction to focus more directly on motivating the study of ingredient networks and establishing the research question. The general background on social networks has been significantly shortened and integrated more seamlessly into the context of ingredient analysis. This revision ensures that the introduction provides a clearer and more relevant foundation for the study. [Please refer to the Introduction Section on page 2 to review the changes in the revised manuscript]

Authors' Response to Reviewer 3

Comment 1

The link to “social networks”. The paper refers to ‘social network metrics,’ but these are well established graph-theoretic measures widely used across diverse network types. While many networks naturally exhibit scale-free and community structures, this does not inherently imply ‘social’ behavior. Drawing a stronger connection to ‘social’ would require additional theoretical grounding. The link to ‘social’ appears to be tenuous.

Response: We sincerely thank the reviewer for their insightful comment regarding the connection between ingredient networks and social networks.

Our intention was not to claim that ingredient networks are social networks in the traditional sense, but rather to highlight that they exhibit patterns and properties commonly observed in social networks, such as scale-free behavior and community structure. These properties suggest that ingredient networks share underlying organizational principles with social networks, even if the entities and interactions differ.

To address the reviewer’s concern, we have revised the manuscript to provide a stronger theoretical grounding for the “social” analogy. Specifically, we have added a paragraph in the introduction that elaborates on the conceptual parallels between ingredient networks and social networks. We emphasize that the “social” behavior in ingredient networks arises from the collaborative and co-occurrence nature of ingredients in recipes, which mirrors the interactions and relationships in social networks. This addition strengthens the justification for using social network metrics and provides a clearer rationale for the observed patterns. [Please refer to the Introduction Section on page 2 to review the changes in the revised manuscript]

Comment 2

measures are applied to individual cuisine types in the paper in some cases (granular) presumably for the purpose of comparison (Fig 4 & 5 for Degree of Distribution), but other measures are seemingly applied across all cuisines (e.g. Fig 6 - Closeness Centrality and Fig 7 - Eigen Centrality). If this paper is focused on differences in these measures across different cuisine types, the way these differences are shown should be more consistent and systematic i.e. why is Fig 4 & 5 split over two separate figures? Could these not all be meaningfully combined into a single figure for clarity and to aid in comparison given they have the same y-axes and x-axes scales?

Response:

Thank you for your insightful comment. In the revised version of the paper, we have combined Figures 4 and 5 into a single figure. This unified presentation ensures consistency in how differences across cuisine types are visualized, particularly since both measures share the same y-axis and x-axis scales. [Please refer to the Application Section on page 5 to review the changes in the revised manuscript]

Comment 3

How complete/correct is this dataset that has been created using existing datasets (e.g. 95,25,31,35,43,77), and how does it compare to similar datasets? It's not clear how these differences in measures (e.g. number of communities) across cuisines might be biased by the data selection i.e. how do you know for one type of cuisine recipes (sourced from a particular dataset) is actually representative? These points about the data should be addressed so that a reader is aware of potential limitations in the analysis and comparisons that may stem from the dataset i.e. the paper does not detail how data sources might bias the results such as repeated recipes, etc.

Response:

We thank the reviewer for the suggestions.

Based on the suggestions we have included Table 1 which shows the comparisons across similar datasets and highlights necessary details in the subsequent paragraph. [Please refer to the Table 1 on page 5 to review the changes in the revised manuscript]

Comment 4

(from a cursory search) there seems to be relevant literature to the analysis being conducted that is not referred to. The references overall seem to miss important related work in this space. Besides including these references, how does existing literature change how the contribution in the paper is positioned? e.g. Herrera, Juan CS. "The contribution of network science to the study of food recipes. A review paper." *Appetite* 159 (2021): 105048.

Ahn, Yong-Yeol, Sebastian E. Ahnert, James P. Bagrow, and Albert-László Barabási. "Flavor network and the principles of food pairing." *Scientific reports* 1, no. 1 (2011): 196.

Ahnert, Sebastian E. "Network analysis and data mining in food science: the emergence of computational gastronomy." *Flavour* 2 (2013): 1-3.

Response:

We appreciate the reviewer's suggestion and have incorporated the recommended references into the revised manuscript. The inclusion of these works strengthens the contextual foundation of our study and ensures that our analysis is well aligned with existing research in the field of computational gastronomy and network science in food studies.

[Please refer to the Related Work Section on page 4 to review the changes in the revised manuscript]

Comment 5

Applying WABCD, Weighted-Louvain, and Weighted-Leiden is interesting, however, the paper could explain more how communities map onto meaningful culinary groupings.

Response:

We sincerely thank the reviewer for the valuable feedback and for highlighting the importance of explaining how the detected communities map onto meaningful culinary groupings. In response to this suggestion, we have provide a more detailed explanation of how the communities identified by WABCD, Weighted-Louvain, and Weighted-Leiden algorithms align with category of the recipe. [Please refer to the Community Structure Section on page 14 to review the changes in the revised manuscript]

Comment 6

The conclusion talks about predicting culinary trends and understanding user interactions, which is interesting but speculative. The paper needs empirical or theoretical grounding for how current findings would this.

Response:

We thank the reviewer for their valuable feedback regarding the speculative nature of the conclusion in our paper. We have revised the conclusion to provide a more focused and evidence-based discussion. The revised conclusion now emphasizes the implications of our findings for understanding the structural properties of ingredient networks and how these properties can serve as a foundation for future research on culinary trends and user interactions, rather than making speculative claims. [Please refer to the Conclusion Section on page 20 to review the changes in the revised manuscript]

Comment 7

Although the network analysis indicates small world / scale free traits, it doesn't clearly clarify why these aspects matter from a practical point of view e.g. how do they truly benefit food computing beyond just drawing parallels to social networks?

Response:

We appreciate the reviewer's comment. We have revised the manuscript to explicitly discuss how these network properties can benefit applications in food computing, such as recipe recommendation systems, flavor pairing prediction, and culinary trend analysis. [Please refer to the Application Section on page 19 to review the changes in the revised manuscript]