FlavorDiffusion: Modeling Food-Chemical Interactions with Diffusion

Joomidang AI Research

Junpyo Seo* Dongwan Kim*

jpseo99@joomidang.com ryankim@joomidang.com

Jaewook Jeong Ingyu Park Junho Min

jwjeong@joomidang.com s52091851@joomidang.com minjh0113@joomidang.com

Abstract

The study of food pairing has evolved beyond subjective expertise with the advent of machine learning. This paper presents FlavorDiffusion, a novel framework leveraging diffusion models to predict food-chemical interactions and ingredient pairings without relying on chromatography. By integrating graph-based embeddings [Perozzi et al., 2014], diffusion processes [Ho et al., 2020, Song et al., 2021, Sun and Yang, 2023], and chemical property encoding [Azambuja et al., 2023], FlavorDiffusion addresses data imbalances and enhances clustering quality. Using a heterogeneous graph derived from datasets like Recipe1M [Marín et al., 2019] and FlavorDB, our model demonstrates superior performance in reconstructing ingredient-ingredient relationships. The addition of a Chemical Structure Prediction (CSP) layer further refines the embedding space, achieving state-of-the-art NMI scores and enabling meaningful discovery of novel ingredient combinations. The proposed framework represents a significant step forward in computational gastronomy, offering scalable, interpretable, and chemically informed solutions for food science. The source code and dataset used in this study are publicly available at https://github.com/Giventicket/FlavorDiffusion.

1 Introduction

Food pairing has traditionally relied on the intuition and experience of chefs, yet scientific analysis and optimization of food combinations remain underexplored. Recent research has leveraged data-driven approaches to model the relationships between food ingredients and chemical compounds to predict novel food pairings.

Several computational approaches have been developed to model food pairings and ingredient relationships. Kitchenette [Park et al., 2021], for instance, applies Siamese neural networks to predict and recommend ingredient pairings based on a large annotated dataset. However, it suffers from key limitations, such as a lack of chemical interpretability and heavy reliance on labeled data, making it less generalizable across different cuisines and novel food combinations.

One of the key advancements in this domain is FlavorGraph [Park et al., 2021], a large-scale food-chemical deep neural network model comprising 6,653 ingredient nodes and 1,645 compound nodes. This graph captures two primary relationships: (1) ingredient-ingredient relations, representing co-occurrence patterns in recipes, and (2) ingredient-compound relations, indicating chemical composition links. These relationships are constructed using datasets such as Recipe1M [Marín et al., 2019], FlavorDB, and HyperFoods. FlavorGraph incorporates food-chemical associations into a neural network by leveraging the metapath2vec [Dong et al., 2017] algorithm, which embeds ingredient-compound relationships in a word2vec-like manner. Expanding on this approach, WineGraph [Gawrysiak et al., 2023] extends the framework by integrating wine-related datasets to define optimal food-wine pairings.

Despite progress in computational food science, major challenges remain. Chromatography-based methods, while precise, are costly and limit the acquisition of large-scale chemical interaction data. FlavorGraph effectively captures ingredient-compound relationships using metapath-based embeddings, but its reliance on random-walk sampling makes it difficult to incorporate edge weights and spatial information within the graph structure. These limitations hinder the full exploitation of food-chemical associations, leading to suboptimal ingredient relationship modeling. To address these challenges, we introduce FlavorDiffusion, a Diffusion Model-based framework that refines the representation of food-chemical interactions and ele-

*Co-first authors.

vates the quality of food pairing predictions.

Contributions

- We propose a graph-based diffusion modeling approach that leverages DIFUSCO [Sun and Yang, 2023] to capture richer and more structured representations of food-chemical interactions.
- We introduce a balanced subgraph sampling strategy to address data imbalance issues, ensuring fair representation across different ingredient-chemical associations.
- Our experimental results demonstrate improvements in Normalized Pointwise Mutual Information (NPMI) scores for node embeddings, facilitating more effective chemical inference.
- We establish a foundation for predicting chromatography results for non-hub chemicals, extending the applicability of our model beyond frequently occurring compounds.
- Our approach enables pairing inference using chemical properties, providing structured and interpretable recommendations for novel ingredient combinations.

2 Dataset

Our study builds upon FlavorGraph [Park et al., 2021] by utilizing the same large-scale datasets to construct a robust food-chemical network. These datasets provide a structured representation of ingredient relationships and chemical interactions. In the following sections, we summarize the key characteristics of these datasets and outline the preprocessing steps applied to ensure data consistency and usability in our framework.

Type	Source	Nodes	Edges
I-I	Recipe1M	6,653	111,355
I-FC	FlavorDB	1,561	35,440
I-DC	HyperFoods	84	386
Total	-	8,298	147,181

Table 1: Summary of the heterogeneous food-compound graph. I-I represents ingredient ingredient co-occurrence from Recipe1M, I-FC denotes ingredient-flavor compound associations from FlavorDB, and I-DC refers to ingredient-drug compound relations

2.1 Data Sources

This study utilizes the same datasets as FlavorGraph [Park et al., 2021] to construct a structured food-chemical network.

Recipe1M [Marín et al., 2019] contains 65,284 recipes with ingredient lists and cooking instructions, capturing ingredient co-occurrence patterns in real-world culinary practices.

FlavorDB compiles chemical composition data from multiple sources, including *FooDB*, *Flavornet*, and *BitterDB*. It originally includes 2,254 flavor compounds linked to 936 food ingredients, but only 400 commonly used ingredients were selected to align with Recipe1M, resulting in 1,561 flavor compound nodes and 164,531 ingredient-flavor compound edges.

HyperFoods maps drug compounds to food ingredients using machine learning based on food-gene interactions. From the original 206 food ingredients, 104 were selected, yielding 84 drug compound nodes and 386 ingredient-drug compound edges.

2.2 Data Processing

To construct a structured representation of food-chemical relationships, we build upon FlavorGraph [Park et al., 2021], a heterogeneous graph that integrates both culinary and chemical associations. The graph construction process follows a structured approach. First, an ingredient-ingredient graph is built by extracting co-occurrence patterns from Recipe1M [Marín et al., 2019], where edges between ingredients are established based on their Normalized Pointwise Mutual Information (NPMI) scores. Only statistically significant ingredient pairs appearing together in a substantial number of recipes are retained, resulting in a total of 111,355 edges. Second, an ingredient-chemical graph is formed by linking ingredients to their corresponding chemical compounds using FlavorDB and HyperFoods, leading to 35,440 edges between food ingredients and known chemical compounds. The final graph structure comprises 6,653 ingredient nodes and 1,645 compound nodes, forming a heterogeneous graph that encodes both culinary co-occurrence relationships and chemical interactions.

2.3 Chemical Property Encoding

To ensure chemically informed ingredient representations, each compound is characterized using CACTVS chemical fingerprints, which are encoded

as 881-dimensional binary vectors. These vectors represent molecular descriptors such as molecular weight, functional groups, and substructure patterns, using a binary encoding scheme where each bit indicates the presence or absence of a specific chemical substructure.

3 Related Work

3.1 FlavorGraph

FlavorGraph [Park et al., 2021] is a heterogeneous graph $G = (V, E)$ integrating ingredient co-occurrence and molecular profiling to model food-chemical interactions. By leveraging metapath-based learning [Dong et al., 2017], it enables systematic ingredient discovery and predictive food pairing through shared molecular properties.

3.1.1 Metapath2Vec

To learn chemically meaningful embeddings, we employ **Metapath2Vec**, which captures high-order relations via structured random walks. Ingredients are classified into hub ingredients (H), which directly connect to chemical compounds, and non-hub ingredients (N), which lack direct chemical links and rely on hub ingredients to acquire chemical insights.

The metapath sampling strategy follows:

$$N \to H \to C \to H \to N$$

where C represents chemical compounds. This structured propagation ensures that non-hub ingredients inherit chemical relevance, enhancing embedding robustness and interpretability.

3.1.2 Architecture

The network, parameterized by θ, takes node pairs (i, j) as input and outputs an edge score $s_\theta(i, j)$, normalized across all embeddings:

$$s_\theta(i, j) = \sigma(\mathbf{u}_i^T \mathbf{u}_j)$$

where $\mathbf{u}_i$ and $\mathbf{u}_j$ are the learned embeddings for nodes i and j, ensuring consistency across culinary co-occurrence and chemical similarity.

3.1.3 Loss Function

Embeddings are optimized using Skip-Gram with Negative Sampling (SGNS):

$$J_\theta = \sum_{(i,j)\in D} \log \sigma(\mathbf{u}_i^T \mathbf{u}_j) + \sum_{(i,j')\in D'} \log \sigma(-\mathbf{u}_i^T \mathbf{u}_{j'})$$

where D and D' are positive and negative sample pairs. To enforce chemical relevance, an additional **Chemical Structure Prediction (CSP)** loss is introduced:

$$L_{\text{CSP},\theta} = \sum_{d=1}^{D} y_d \log f_{\theta,d}(i) + (1-y_d) \log(1-f_{\theta,d}(i))$$

where $f_{\theta,d}(i)$ predicts the presence of the d-th molecular substructure y_d, refining embeddings with molecular fingerprints.

3.2 DIFUSCO

Graph-based diffusion models have recently emerged as powerful frameworks for solving combinatorial optimization problems by leveraging probabilistic generative processes. In our work, we leverage the fundamental principles of graph-based diffusion models, particularly the Gaussian diffusion framework, to reconstruct structured graph representations. By incorporating diffusion-driven embeddings into our heterogeneous network, we enhance the predictive accuracy of food-chemical interactions while maintaining interpretability. This approach allows for the seamless integration of molecular-level insights into ingredient pairing research, further advancing computational gastronomy.

4 Proposition: FlavorDiffusion

4.1 Sub-Graph Sampling

FlavorDiffusion is built upon the DIFUSCO Gaussian noise-based diffusion model, extending its capabilities to structured food-chemical graphs. The core objective is to train a model capable of reconstructing subgraphs sampled from the full heterogeneous graph $G = (V, E)$ while leveraging node attributes as guidance.

The full graph consists of a diverse set of nodes V, including hub ingredients, non-hub ingredients, flavor compounds, and drug compounds, with edges E encoding the strength of their relationships as continuous values in $[0, 1]$. We define a dataset of subgraphs, where each sample contains m nodes selected from G. These subgraphs are denoted as:

$$\mathcal{D}_m = \{G_i = (V_i, E_i)\}_{i=1}^{N},$$

where each subgraph G_i has $|V_i| = m$ nodes and an adjacency matrix E_i of size $m \times m$, representing pairwise edge scores. The dataset is partitioned into training (N_t) and validation (N_v) subsets.

4.2 Forward Diffusion Process

For a single data point $G_i = (V_i, E_i)$ sampled from the dataset, we define the diffusion process over its edge set E_i. By convention, we denote the corrupted version of E_i at timestep t as x_t, aligning with standard diffusion formalisms. The node representations, encompassing all vertex features, are denoted as $\mathbf{Emb}$.

The forward diffusion process follows a Markovian Gaussian noise injection, progressively perturbing the edges x_t while preserving node representations:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I),$$

where β_t is a predefined noise variance at timestep t. Given an initial clean edge matrix $x_0 = E_i$, we can analytically express the direct corruption of x_0 at any timestep t as:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I),$$

where $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$ represents the cumulative noise effect over time. This formulation allows direct sampling of x_t from x_0, bypassing iterative updates.

In this framework, the edge structure is progressively degraded into Gaussian noise, while node representations $\mathbf{Emb}$ remain unchanged, ensuring that denoising relies on learned node attributes.

4.3 Reverse Denoising Process

The reverse process seeks to recover x_0 from the fully corrupted state x_T, learning to remove noise in a stepwise manner. The key assumption is that the forward process follows a Gaussian transition, enabling an analytically derived reverse process.

Given the Markovian nature of the diffusion process, we define the true posterior:

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I),$$

where the posterior mean and variance are derived as:

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t,$$

$$\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t.$$

Since x_0 is unknown, we train a model $p_\theta(x_0|x_t)$ to approximate it. Substituting the predicted x_0, the learned reverse process is modeled as:

$$p_\theta(x_{t-1}|x_t, \mathbf{Emb}) = \mathcal{N}\big(x_{t-1}; \mu_\theta(x_t, t, \mathbf{Emb}), \Sigma_\theta(x_t, t)\big),$$

where μ_θ is the learned estimate for $\tilde{\mu}_t(x_t, x_0)$, and the variance term is fixed as $\Sigma_\theta(x_t, t) = \tilde{\beta}_t I$, avoiding the need for explicit learning. The function μ_θ is now conditioned on the node representations ($\mathbf{Emb}$) of the two vertices forming the edge.

Using the DDPM convention, we parameterize μ_θ as:

$$\mu_\theta(x_t, t, \mathbf{Emb}) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t, \mathbf{Emb})\right),$$

where $\epsilon_\theta(x_t, t, \mathbf{Emb})$ is the learned noise estimate, which is now explicitly conditioned on the representations of the two nodes forming the edge. The node representations provide additional context for denoising by leveraging node-specific features.

4.4 Optimization via Variational Lower Bound

To train the reverse model, we maximize the variational lower bound (ELBO), decomposed as:

$$\mathcal{L}_{\text{ELBO}} = E_q\Bigg[\log p_\theta(x_0|x_1, \mathbf{Emb}) - \sum_{t=1}^{T} D_{\text{KL}}\big(q(x_{t-1}|x_t, x_0)\|p_\theta(x_{t-1}|x_t, \mathbf{Emb})\big)\Bigg].$$

Here, T represents the total number of diffusion steps, defining the depth of the forward and reverse process. The KL divergence encourages the learned transitions to match the true posterior. Since $q(x_t|x_0)$ is Gaussian, minimizing D_{KL} is equivalent to predicting the noise component ϵ added during diffusion. Thus, the training objective simplifies to:

$$\mathcal{L}_{\text{recon}} = E_{t,x_0,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t, t, \mathbf{Emb})\|^2\right].$$

This loss ensures that ϵ_θ effectively estimates the noise introduced in the forward process while incorporating node representations. By iteratively refining the denoising function, FlavorDiffusion reconstructs the original ingredient-ingredient graph from noisy subgraphs, leveraging both the structural edge information and node attributes to enhance predictive modeling for food pairing analysis.

4.5 Inference

Graph reconstruction follows Denoising Diffusion Implicit Models (DDIM) for efficient and deterministic sampling. Unlike DDPM, DDIM removes noise via a non-Markovian update, accelerating inference.

Starting from $x_T \sim \mathcal{N}(0, I)$, the reverse process iterates:

$$x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{x}_0 + \sqrt{1-\bar{\alpha}_{t-1}} \cdot \epsilon_\theta(x_t, t, \mathbf{Emb}),$$

where the predicted clean graph is:

$$\hat{x}_0 = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t, t, \mathbf{Emb})}{\sqrt{\bar{\alpha}_t}}.$$

Iterating from T to 0, the model refines x_t to recover ingredient-ingredient relationships. DDIM ensures fast, stable, and chemically meaningful reconstructions.

4.6 Model Architecture

The noise prediction network $\epsilon_\theta(x_t, t, \mathbf{V})$ employs an anisotropic GNN to iteratively refine node and edge embeddings. Let $h_i^\ell \in \mathbf{R}^d$ and $e_{ij}^\ell \in \mathbf{R}^{d_e}$ denote the node and edge features at layer ℓ, respectively. The refinement process updates both edge and node embeddings through the following operations:

Edge Refinement The initial edge embeddings e_{ij}^0 are set as the corresponding values from the noisy edge representation x_t. At each layer ℓ, the intermediate edge embeddings $\hat{e}_{ij}^\ell$ are updated as:

$$\hat{e}_{ij}^\ell = P^\ell e_{ij}^\ell + Q^\ell h_i^\ell + R^\ell h_j^\ell,$$

where $P^\ell, Q^\ell, R^\ell \in \mathbf{R}^{d_e \times d_e}$ are learnable parameters. The refined edge embedding $e_{ij}^{\ell+1}$ is then computed as:

$$e_{ij}^{\ell+1} = e_{ij}^\ell + \text{MLP}_e\big(\text{BN}(\hat{e}_{ij}^\ell)\big) + \text{MLP}_t(t),$$

where MLP_e is a 2-layer perceptron and MLP_t embeds the diffusion timestep t using sinusoidal features.

Node Refinement The node embeddings h_i^ℓ are refined by aggregating information from neighboring nodes and their associated edges. The update rule for $h_i^{\ell+1}$ is given by:

$$h_i^{\ell+1} = h_i^\ell + \alpha \cdot \text{BN}\Big(U^\ell h_i^\ell + \sum_{j \in \mathcal{N}(i)} \sigma(\hat{e}_{ij}^\ell) \odot V^\ell h_j^\ell\Big),$$

where $U^\ell, V^\ell \in \mathbf{R}^{d \times d}$ are learnable parameter matrices, σ is the sigmoid activation function used for edge gating, $\odot$ denotes the Hadamard (element-wise) product, $\mathcal{N}(i)$ represents the set of neighbors for node i, and α is the ReLU activation applied after aggregation.

Final Prediction After L GNN layers, the final refined edge embeddings $E^{(L)} \in \mathbf{R}^{N \times N \times d_e}$ are passed through a ReLU activation and a multi-layer perceptron (MLP) to predict the noise:

$$\epsilon_\theta(x_t, t, \mathbf{V}) = \text{MLP}\big(\text{ReLU}(E^{(L)})\big).$$

This formulation ensures that both node and edge embeddings are iteratively refined to capture local and global graph structure, enabling robust denoising and reconstruction of ingredient-ingredient relationships.

5 Experimental Results

The evaluation consists of two primary experiments: (1) reproducing the NMI-based clustering performance evaluation originally conducted in FlavorGraph, and (2) assessing the generalization ability of our proposed Flavor Diffusion framework by testing on subgraphs of different sizes.

Subgraphs of size 25, 50, 100, and 200 nodes were sampled while maintaining an equal proportion of hub and non-hub ingredients. The number of subgraphs used for training and testing at each scale is shown in Table 2.

Table 2: Subgraph Composition for Training and Testing

Nodes per Subgraph	Train Set Size	Test Set Size
25	256,000	256
50	128,000	128
100	64,000	64
200	32,000	32

Generalization Ability To assess the generalization ability of the proposed framework, models trained on one subgraph size were tested on all sizes to observe performance across different scales. The results in Table 3 indicate that models trained on 25-node subgraphs generalize poorly to larger graphs, with an MSE of 0.025078 when tested on 100-node subgraphs. In contrast, the 100-node trained model demonstrates the most stable generalization across different test sizes, showing minimal MSE variation. The 200-node trained model, while excelling on large graphs with an MSE of 0.003692, exhibits difficulties in adapting to smaller structures, with a high error of 0.059557 when tested on 25-node subgraphs.

Table 3: Generalization Performance: Validation MSE Loss

Train Size	Test (25)	Test (50)	Test (100)	Test (200)
25	0.004589	0.010965	0.025078	0.019477
50	0.025235	0.005884	0.004420	0.004123
100	**0.003964**	**0.003678**	0.004232	0.003953
200	0.059557	0.007837	**0.003992**	**0.003692**

These results highlight that subgraph size significantly impacts both intra-subgraph clustering and cross-subgraph generalization performance. The Flavor Diffusion (100 nodes) model provides the best balance between clustering accuracy and scalability, demonstrating the ability to generalize well across varying ingredient graph structures. On the other hand, training on extremely small subgraphs limits generalization, while models trained on large subgraphs struggle when applied to smaller ingredient sets. These findings suggest that a mid-sized subgraph training approach (e.g., 100 nodes) is optimal for robust ingredient representation learning.

NMI-based Evaluation To construct the clustering test dataset, nine representative food categories were defined: *Bakery/Dessert/Snack, Beverage Alcoholic, Cereal/Crop/Bean, Dairy, Fruit, Meat/Animal Product, Plant/Vegetable, Seafood, and Others*. From these, 416 chemical hub ingredients with strong connections were selected to ensure diverse and well-defined clustering labels, enabling fair comparisons across models commonly used in related studies.

The NMI-based evaluation results in Table 4 demonstrate the clustering quality of different models. Among the non-CSP variants, the Flavor Diffusion (50 nodes) model achieves the highest NMI score of 0.3236, surpassing the baseline FlavorGraph model without CSP. The best overall performance is observed in the Flavor Diffusion_CSP (200 nodes) model, which achieves an NMI score of 0.3410, indicating that the CSP layer significantly improves the learned ingredient embeddings. Smaller subgraphs, such as the 25-node configuration, show the greatest improvement when using CSP (0.2970 vs. 0.2167), suggesting that the chemical structure prediction enhances clustering, particularly in more limited ingredient sets.

Table 4: Performance Comparison Using NMI Metric. *CSP shorts for chemical structure prediction.

Model	NMI Mean	NMI Std
FlavorGraph [Park et al., 2021]	0.2995	0.0403
FlavorGraph_CSP [Park et al., 2021]	0.3102	0.0407
Flavor Diffusion (25 nodes)	0.2167	0.0319
Flavor Diffusion (50 nodes)	**0.3236**	**0.0134**
Flavor Diffusion (100 nodes)	0.3170	0.0207
Flavor Diffusion (200 nodes)	0.2935	0.0300
Flavor Diffusion_CSP (25 nodes)	0.2970	0.0144
Flavor Diffusion_CSP (50 nodes)	0.2862	0.0152
Flavor Diffusion_CSP (100 nodes)	0.3169	0.0257
Flavor Diffusion_CSP (200 nodes)	**0.3410**	**0.0150**

6 Discussion

The visualization results highlight the impact of the proposed Flavor Diffusion framework on embedding quality, particularly with the CSP (Chemical Structure Prediction) layer, as shown in Figures 1 and 2.

Dynamic Reconstruction for Novel Insights The iterative reconstruction process visualized in Figure 1 showcases the Flavor Diffusion framework's ability to refine ingredient-ingredient relationships progressively. Starting from random initialization (Step 0), the edge scores evolve over diffusion steps, ultimately converging towards the ground truth structure by Step 10. The color intensity of the edges reflects their normalized scores, with higher values indicating stronger relationships. This gradual alignment with the ground truth demonstrates the model's capacity to encode meaningful relational patterns in a structured manner.

Embedding Space Analysis Figure 2 compares embedding spaces across model configurations. The baseline embeddings (left) show poor separation, forming diffuse clusters dominated by non-hub ingredients.

Flavor Diffusion (200 nodes) without CSP (center) improves clustering by grouping chemical compounds and hub ingredients, though some overlap

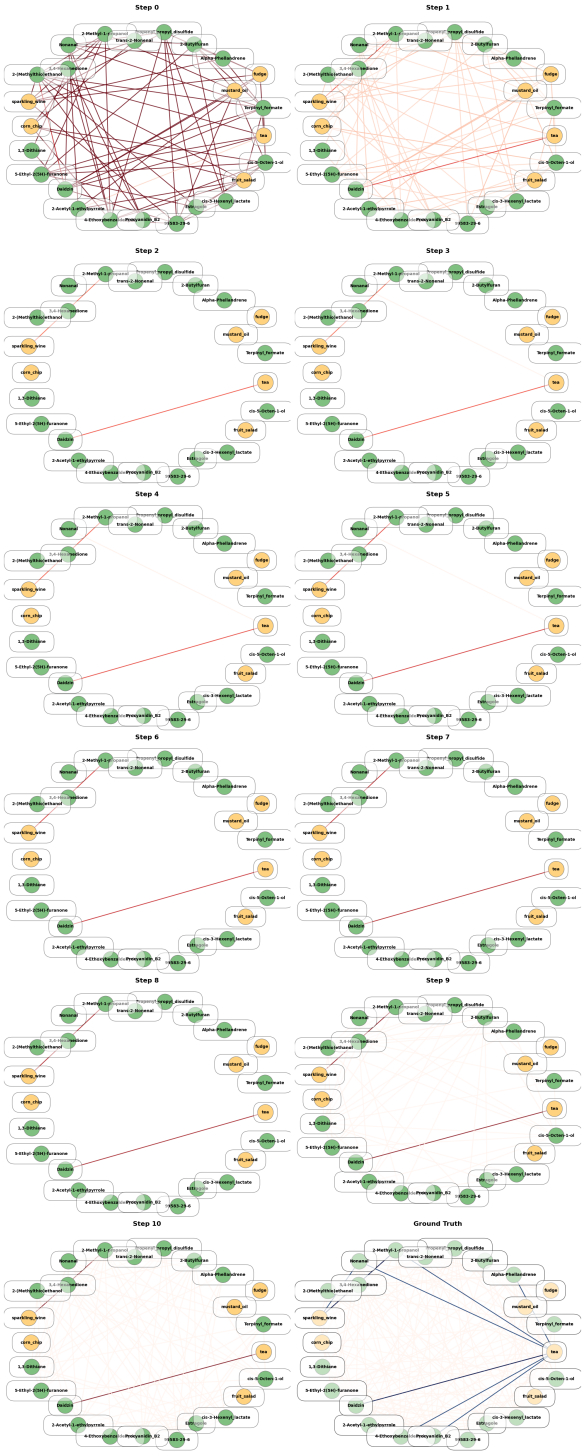

Figure 1: Progression of edge scores over diffusion steps for a 25-node subgraph. The color intensity represents edge scores normalized between 0 and 1. The reconstructed graph increasingly aligns with the ground truth structure.

remains. Adding the CSP layer (right) further refines the structure, yielding anisotropic clusters that better capture ingredient-compound relationships.

Potential for Ingredient Innovation To evaluate the predictive capacity of Flavor Diffusion, we randomly sampled 100 nodes and computed the mean edge score over 100,000 inferred edges. This large-scale evaluation ensures that the model captures both established and novel ingredient relationships, supporting its ability to reconstruct known pairings while suggesting unexplored flavor synergies.

Table 5: Top 5 High-Confidence Ingredient Pairings

Ingredient 1	Ingredient 2	Mean Score	Std Dev
Red Chili Powder	Turmeric Powder	0.7114	0.0882
Coriander Powder	Turmeric Powder	0.6057	0.0827
Asafoetida Powder	Turmeric Powder	0.5930	0.0846
Garam Masala Powder	Turmeric Powder	0.5178	0.1055
Cumin Powder	Turmeric Powder	0.4663	0.1525

These pairings align with traditional spice blends, frequently observed in Indian and Southeast Asian cuisine. Their strong co-occurrence validates Flavor Diffusion's ability to model established ingredient relationships. Beyond known pairings, the model also proposes conceptually novel combinations, potentially inspiring new culinary applications.

Table 6: Top 5 Creative Ingredient Pairings Suggested by Flavor Diffusion

Ingredient 1	Ingredient 2	Mean Score	Std Dev
Soy Sauce	Vanilla Extract	0.0006	0.0001
Garlic Paste	Dark Chocolate	0.0005	0.0001
Cumin Powder	Coffee Beans	0.0004	0.0002
Green Cardamom	Parmesan Cheese	0.0003	0.0002
Olive Oil	Black Tea	0.0004	0.0001

These unconventional combinations introduce potential for umami-sweet fusion (Soy Sauce, Vanilla Extract), savory-bitter contrast (Garlic Paste, Dark Chocolate), and aromatic synergies (Cumin Powder, Coffee Beans and Green Cardamom, Parmesan Cheese). Such findings demonstrate that Flavor Diffusion extends beyond known ingredient interactions, offering a data-driven approach for novel flavor discovery and AI-assisted recipe development.

Alignment with Culinary and Chemical Properties The reconstructed graphs closely align with ground truth structures, demonstrating the model's fidelity in capturing both culinary and chemical relationships. As diffusion progresses, the model

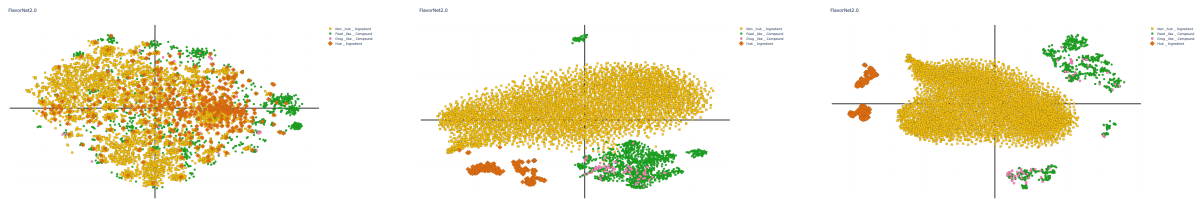

Figure 2: Embedding space comparison under different configurations, where each color represents a different category: **Yellow** (Non-hub Ingredient), **Green** (Food-like Compound), **Pink** (Drug-like Compound), and **Orange** (Hub Ingredient). (Left) Baseline embeddings show poor separation between ingredients and compounds. (Center) Flavor Diffusion (200 nodes) without CSP achieves improved clustering of chemical compounds and hub ingredients. (Right) Flavor Diffusion (200 nodes) with CSP results in well-defined clusters, leveraging chemical fingerprints to enhance separation.

effectively balances local (ingredient-level) and global (chemical-based) interactions, enhancing clustering quality and enabling meaningful extensions of ingredient networks.

7 Conclusion

This study introduced FlavorDiffusion, a diffusion model-based framework for predicting ingredient pairings and chemical interactions. The model's capability to integrate chemical fingerprints and optimize graph embeddings resulted in improved clustering quality and predictive accuracy. Experiments revealed that the inclusion of the CSP layer significantly enhanced the representation of food-chemical relationships, achieving the highest NMI scores across various configurations. The progressive nature of the diffusion process further demonstrated the model's ability to generalize, enabling the inference of novel ingredient combinations. By aligning culinary and chemical properties, FlavorDiffusion offers a robust tool for advancing food pairing discovery, with applications in flavor design and computational gastronomy. Future work will aim to expand dataset coverage, integrate multi-modal data, and explore new graph-sampling techniques to further enrich food science research.

References

Garima Gupta and Rahul Katarya, "A Computational Approach Towards Food-Wine Recommendations", *Expert Systems With Applications*, vol. 238, p. 121766, 2024. doi:10.1016/j.eswa.2023.121766.

Javier Marín, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba, "Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. doi:10.1109/TPAMI.2019.2908359.

Zuzanna Gawrysiak, Agata Żywot, and Agnieszka Ławrynowicz, "WineGraph: A Graph Representation For Food-Wine Pairing", *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Rogério Xavier de Azambuja, A. Jorge Morais, and Vítor Filipe, "X-Wines: A Wine Dataset for Recommender Systems and Machine Learning", *Big Data Cogn. Comput.*, vol. 7, no. 20, 2023. doi:10.3390/bdcc7010020.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena, "DeepWalk: Online Learning of Social Representations", *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14)*, 2014. doi:10.1145/2623330.2623732.

Zhiqing Sun and Yiming Yang, "DIFUSCO: Graph-based Diffusion Solvers for Combinatorial Optimization", *Neural Information Processing Systems (NeurIPS)*, 2023. doi:10.48550/arXiv.2302.08224.

Jiaming Song, Chenlin Meng, and Stefano Ermon, "Denoising Diffusion Implicit Models", *International Conference on Learning Representations (ICLR)*, 2021. doi:10.48550/arXiv.2010.02502.

Jonathan Ho, Ajay Jain, and Pieter Abbeel, "Denoising Diffusion Probabilistic Models", *Neural Information Processing Systems (NeurIPS)*, 2020. doi:10.48550/arXiv.2006.11239.

Yuxiao Dong, Nitesh V. Chawla, and Ananthram Swami, "metapath2vec: Scalable Representation Learning for Heterogeneous Networks", *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17)*, 2017. doi:10.1145/3097983.3098036.

Donghyeon Park, Keonwoo Kim, Seoyoon Kim, Michael Spranger, and Jaewoo Kang, "FlavorGraph: a large-scale food-chemical graph for generating food representations and recommending food pairings", *Scientific Reports*, vol. 11, p. 931, 2021. doi:10.1038/s41598-020-79422-8.