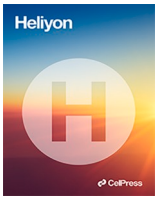
Heliyon 10 (2024) e40873

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

A rapid approach based on empirical formulas and graph traversal algorithms to automatic generation of fragrance formulas

Jiaxiao Cai [a], Suxing Tuo [a], Yanchun Li [a], Hongbing Lu [a], Yizi Wu [a], You Zou [b], Zhen Ma [a], Yuqi Cui [a], Bo Kong [a,*], Kejun Zhong [a,c,**]

[a] Technology Center of China Tobacco Hunan Industrial Co., Ltd., Changsha, 410007, China
[b] Information and Network Center of Central South University, Changsha, 410083, China
[c] China Tobacco Hunan Industrial Co., Ltd., Changsha, 410007, China

ARTICLE INFO

Keywords:
Fragrance formula
Automatic formula creation
Traversal algorithm
Olfactory profiles

ABSTRACT

This article proposes a novel approach for improving the efficiency of fragrance designing and the accuracy of automatic fragrance formula creation based on empirical fragrance formulas and graph traversal algorithms. By effectively extracting the composition information and further analyzing the combination of fragrance materials in 210 fragrance formulas, a relational network model was constructed in the form of a graph to illustrate the relationship between the ingredients used in the formulas. Additionally, a fragrance ingredients information database of 344 common ingredients was constructed and used as a reference for perfumers when setting algorithmic constraints based on their experience. Finally, an automatic fragrance formula creation algorithm was established by constructing a relational network subgraph and finding fragrance formula solutions with the help of depth-first search algorithm that satisfies the constraint conditions and combining with appropriate statistical strategy that could determine the usage of each component in the new fragrance formula. By testing the algorithm with the goal of creating a floral fragrance, the resulting formula well fulfilled our expectations and had practical application value.

1. Introduction

Fragrances are integral to the marketability of various consumer products ranging from perfumes to detergents and personal care products [1]. Their significance stems from the fact that fragrances serve as a distinguishing factor for these products, influencing consumers' purchasing and repurchasing decisions [2,3]. However, fragrance design has traditionally been a time-consuming and labor-intensive process that relies on a perfumer's experience and creativity [4]. And fragrance formulas have long been considered the cornerstone of the fragrance industry.

With the rapid development of scientific technology, fragrance research has benefited greatly from the widespread application of chemical component separation, analytical techniques, chemometrics multivariate analysis techniques, and artificial intelligence and computer technology. These advancements have provided a systematic way to optimize the fragrance design process [5–8]. At the same time, the flavor and fragrance industry has made significant progress, with many mature aroma components that have undergone

* Corresponding author.
** Corresponding author. China Tobacco Hunan Industrial Co., Ltd., Changsha, 410007, China.
E-mail addresses: kongb0706@hngytobacco.com (B. Kong), zhongkejun@yeah.net (K. Zhong).

https://doi.org/10.1016/j.heliyon.2024.e40873
Received 17 July 2023; Received in revised form 11 November 2024; Accepted 2 December 2024
Available online 3 December 2024

long-term experimentation and optimization, making them valuable data resources [7,9,10]. In this context, the utilization of product engineering in fragrance design provides a valuable means to incorporate scientific knowledge into an area that has traditionally relied solely on empiricism and experimentation [11–13].

As our understanding of fragrance design has advanced, it has become increasingly clear that empirical fragrance formulas developed by fragrance specialists can be represented as a data structure resembling a knowledge graph, where nodes represent fragrance raw materials and edges reflect their relationships. This graph-like structure offers a useful means to analyze and optimize fragrance formulas through the application of computational methods. Graph traversal algorithms can facilitate the identification of intrinsic connections among various materials in empirical fragrance formulas, enabling systematic and insightful analysis of the empirical formula data [14–16]. This, in turn, provides a foundation for the automatic generation of novel fragrance formulas through the application of computational tools. Among various graph algorithms, the Depth First Search (DFS) algorithm is considered a fundamental graph traversal algorithm, renowned for its efficacy in solving graph traversal problems [17,18]. Furthermore, DFS serves as a building block for other graph theory algorithms, including graph connectivity problems, topological sorting, and bipartite graph matching, among others [19–21]. Its pivotal role in addressing graph-related problems and broader implications in the field ensure its ongoing importance.

In order to improve the efficiency of fragrance design, this paper aims to comprehensively investigate the characteristics of flavor materials and the corresponding empirical fragrance formulas. With the aid of computers, empirical fragrance formulas can be represented as a relationship graph model, and the DFS algorithm leveraged to transform the fragrance goal into a graph traversal constraint. This approach enables the construction of a relationship subgraph that satisfies the imposed constraint, thereby facilitating the generation of new fragrance formulas. Additionally, suitable statistical methods can be employed to determine the optimal dosage of each fragrance component in the newly generated formula, resulting in an intelligent and automatic fragrance design process.

2. Materials and methods

2.1. Description of the study area and study period

The experiment was conducted in Changsha, Hunan Province, China. Geographically, it lies in the central-south part of China at approximately 28°12′ N latitude and 112°59′ E longitude. The region experiences an average annual rainfall ranging from 1300 to 1500 mm, with a primary rainy season from April to June and a secondary rainy period from August to September. Changsha's elevation is about 38 m above sea level, and it has an average annual temperature of approximately 18 °C. This study, conducted from January 2022 to July 2023, encompassed multiple seasons and regional climate variations. This timeframe allowed for a thorough observation of seasonal environmental fluctuations and their impact on the research objectives.

2.2. Materials

All fragrance ingredients and solvents used in the fragrances design process were food grade and supplied by Aladdin Reagent (Shanghai) Co., Ltd., China. The equipment used included an electronic analytical balance (sensitivity 0.0001g, Mettler Toledo X3603SX, Switzerland).

2.3. Olfactory evaluations

In the sensory evaluation, 18 expert panelists (12 women and 6 men, all non-smokers), aged 20–55 years, were employed. These evaluations took place across two consecutive days in a controlled environment with stable temperature and consistent lighting conditions. The evaluation sessions were scheduled at the same time each day for all samples and stimuli, ensuring minimal variation in conditions. Each of the 18 sensory attributes, which represent distinct olfactory families, was assessed for every sample. These attributes included hay, resinous, fresh, fruity, woody, roasted, spicy, green, herbal, floral, chocolatey, beany, milky, sweet, balsamic, smoky, winy, and sour. Panelists rated the intensity of each attribute on a 0–9 scale, with 1-point increments. The specific criteria for these ratings were detailed in Table 1 of the study. To maintain objectivity, a blinded assessment approach was adopted, where each sample was identified only by a unique 3-digit code. The sequence of sample evaluation was randomized. For statistical robustness, outlier screening was conducted using JMP software. The ratings for each attribute were then averaged using arithmetic mean to derive the final scores for each sample [1].

Table 1
Scoring criteria of olfactory evaluation.

Sensory intensity	Score
None	0
Weak	1~2
Slightly weak	3~4
Moderate	5
Slightly strong	6~7
Strong	8~9

2.4. Depth-first search algorithm

The depth-first search (DFS) algorithm is a recursive algorithm that follows a path as far as possible, and backtracks when a dead end is reached, i.e., when all the neighbors of a given node have already been visited. By doing so, DFS visits all the nodes in the graph in a depth-first order, which can be useful for many different types of problems [17,18]. The DFS algorithm has been extensively studied and applied in various domains, such as computer science, mathematics, physics, biology, and social sciences [22]. Many researchers have explored different aspects of the DFS algorithm, such as its time and space complexity, correctness and completeness properties, practical implementations, and extensions and variations [18]. The DFS algorithm can be implemented using either a recursive approach or an iterative approach with an explicit stack [23]. In both cases, the algorithm consists of four main steps.

1. Initialization: The algorithm begins by selecting a starting node and initializing a stack or a recursive call with this node. In addition, a Boolean array or set is used to keep track of the visited nodes.
2. Recursive Procedure: In the recursive approach, the algorithm recursively visits each neighbor of the current node until there are no unvisited nodes left. The recursive procedure involves marking a node as visited and then recursively visiting all its unvisited neighbors. In the iterative approach, the algorithm uses an explicit stack to maintain a list of the nodes to be visited. As each node is visited, its unvisited neighbors are added to the stack.
3. Marking of Visited Nodes: After visiting a node, the algorithm marks it as visited, usually using a Boolean array or set.
4. Backtracking: If the current node has no unvisited neighbors, the algorithm backtracks to the previous node and continues the search process from there until all nodes have been visited.

2.5. Construction of fragrance ingredients information database (FIID)

The core of this article focuses on the fragrance formulation algorithm, which incorporates both qualitative and quantitative data of fragrance ingredients, as well as empirical fragrance formulas. The qualitative and quantitative data of fragrance ingredients used in intelligent fragrance creation include their names, material types (synthetic or natural), solubility, vapor pressure at 25 °C, olfactory profiles, odor threshold, and type of notes (top, middle, or base) [10,24–26]. By collecting and compiling these descriptive indicators for 344 fragrance ingredients, including 158 natural and 186 synthetic ingredients commonly used within the food industry and our own company, a Fragrance Ingredients Information Database (FIID) was ultimately established for calling upon the digital fragrance formulation algorithm. FIID serves mainly as a point of reference for perfumers to establish constraints for subsequent digital fragrance formulation algorithms, as well as to provide targeted directions for perfumers based on the evaluation results of digital fragrance formulas. As an example, Table 2 illustrates the basic information of linalool.

2.6. Construction of empirical formulas database

We collected 210 empirical fragrance formulas based on our company's existing products. These were adjusted and verified by perfumers and stored as empirical formulas in a fragrance database, with the main odors and their intensity scores, ranked in the top three, as the labels. This database was designed to establish a relationship diagram of ingredients (Table 3). An example of the formula composition for empirical formula 1# is shown in Table 4.

2.7. Relationship diagram of ingredients used in empirical formulas

Denoted $PF = \{X_1, X_2, X_i \ldots, X_n\}$ be the set of empirical fragrance formulas, where X_i represents the i-th empirical formula and n is the total number of empirical formulas. Each empirical formula in PF belongs to a certain type of fragrance, such as empirical formula 3# with the main odors of floral and herbal. Each empirical formula X_i is composed of multiple fragrance ingredients and their

Table 2
Descriptive indicators of linalool.

Number	Item	Details
1	Material name	Linalool
2	Material type	Synthesized ingredient
3	Solubility	Soluble in ethanol and 1,2-propanediol
4	Vapor pressure	0.17 mmHg (25 °C)
5	Olfactory profiles	
6	Odor threshold	6 ppb[a]
7	Type of note	Middle note

[a] Parts(mL) of compound per billion(10^8) parts(mL) of water [27].

Table 3
Empirical fragrance formulas used in fragrance database.

Formula number	Main odors	Formula number	Main odors	Formula number	Main odors
1#	Sweet, fruity, milky	71#	Hay, sweet, fresh	141#	Fruity, sweet, sour
2#	Spicy, herbal, green	72#	Floral, herbal, green	142#	Fruity, sweet, sour
3#	Fruity, sweet, winy	73#	Spicy, sweet, floral	143#	Fruity, sweet, green
4#	Fruity, winy, sweet	74#	Spicy, herbal, green	144#	Floral, green, fresh
5#	Herbal, green, sweet	75#	Woody, sweet, herbal	145#	Hay, sweet, fresh
6#	Roasted, sweet, smoky	76#	Herbal, woody, sweet	146#	Floral, sweet, fruity
7#	Spicy, herbal, sweet	77#	Floral, sweet, resinous	147#	Fruity, winy, sweet
8#	Sour, fruity, milky	78#	Spicy, sweet, green	148#	Green, fruity, sweet
9#	Sreen, hay, fresh	79#	Fruity, winy, green	149#	Floral, sweet, fruity
10#	Sour, milky, fruity	80#	Hay, sweet, roasted	150#	Sweet, fruity, floral
11#	Green, sweet, herbal	81#	Sweet, floral, fruity	151#	Milky, sweet, resinous
12#	Winy, fruity, sour	82#	Beany, milky, hay	152#	Roasted, milky, sweet
13#	Woody, sweet, resinous	83#	Floral, green, sweet	153#	Woody, smoky, herbal
14#	Floral, sweet, spicy	84#	Green, sweet, floral	154#	Resinous, sour, fruity
15#	Sweet, balsamic, spicy	85#	Milky, sweet, beany	155#	Floral, sweet, sour
16#	Milky, sweet, fruity	86#	Herbal, spicy, sweet	156#	Sweet, woody, hay
17#	Floral, woody, resinous	87#	Balsamic, sweet, resinous	157#	Hay, green, sweet
18#	Herbal, green, sweet	88#	Sweet, balsamic, fruity	158#	Woody, floral, sweet
19#	Roasted, chocolatey, milky	89#	Green, sweet, fresh	159#	Sweet, hay, balsamic
20#	Hay, herbal, sweet	90#	Spicy, sweet, herbal	160#	Spicy, herbal, sweet
21#	Herbal, green, sweet	91#	Herbal, floral, green	161#	Herbal, green, spicy
22#	Winy, floral, fruity	92#	Hay, woody, sweet	162#	Spicy, sweet, herbal
23#	Sweet, hay, herbal	93#	Hay, roasted, sour	163#	Herbal, spicy, floral
24#	Hay, sweet, balsamic	94#	Floral, spicy, sweet	164#	Sweet, floral, woody
25#	Floral, herbal, green	95#	Spicy, herbal, green	165#	Spicy, sweet, woody
26#	Herbal, floral, hay	96#	Spicy, sweet, herbal	166#	Roasted, smoky, sweet
27#	Fruity, floral, sweet	97#	Floral, fruity, sweet	167#	Spicy, sweet, sour
28#	Fruity, sweet, smoky	98#	Sweet, balsamic, resinous	168#	Woody, sweet, floral
29#	Winy, fruity, sweet	99#	Balsamic, resinous, sweet	169#	Fruity, sweet, green
30#	Fruity, winy, sweet	100#	Woody, herbal, green	170#	Floral, woody, fruity
31#	Smoky, herbal, roasted	101#	Balsamic, sweet, smoky	171#	Sweet, spicy, balsamic
32#	Milky, roasted, sour	102#	Hay, sweet, roasted	172#	Winy, fruity, sweet
33#	Beany, sweet, milky	103#	Chocolatey, sweet, roasted	173#	Hay, sweet, roasted
34#	Spicy, sweet, floral	104#	Fruity, sweet, sour	174#	Smoky, sweet, roasted
35#	Sour, fruity, milky	105#	Spicy, herbal, sweet	175#	Fruity, floral, sweet
36#	Floral, green, fresh	106#	Green, sweet, fruity	176#	Floral, sweet, fresh
37#	Herbal, sweet, sour	107#	Herbal, sweet, woody	177#	Hay, roasted, sweet
38#	Herbal, sweet, hay	108#	Fruity, sweet, winy	178#	Spicy, sweet, herbal
39#	Fruity, sweet, sour	109#	Sweet, fruity, spicy	179#	Sweet, woody, hay
40#	Fruity, fresh, sweet	110#	Sweet, floral, fruity	180#	Winy, milky, sweet
41#	Spicy, sweet, herbal	111#	Green, fresh, sweet	181#	Floral, sweet, fresh
42#	Herbal, sweet, hay	112#	Fruity, winy, sweet	182#	Fruity, winy, sweet
43#	Fruity, sweet, floral	113#	Spicy, sweet, herbal	183#	Roasted, smoky, sweet
44#	Spicy, sweet, woody	114#	Sweet, fruity, milky	184#	Green, resinous, herbal
45#	Hay, roasted, sweet	115#	Fruity, sweet, floral	185#	Floral, sweet, green
46#	Hay, sour, sweet	116#	Spicy, sweet, woody	186#	Chocolatey, beany, roasted
47#	Woody, herbal, floral	117#	Woody, smoky, spicy	187#	Fruity, sweet, floral
48#	Sweet, floral, fresh	118#	Spicy, herbal, sweet	188#	Herbal, sweet, hay
49#	Sweet, fruity, roasted	119#	Smoky, chocolatey, roasted	189#	Herbal, sweet, green
50#	Milky, sweet, beany	120#	Floral, sweet, green	190#	Sweet, resinous, beany
51#	Herbal, smoky, sweet	121#	Balsamic, sweet, resinous	191#	Smoky, herbal, woody
52#	Floral, sweet, fruity	122#	Roasted, chocolatey, milky	192#	Fresh, herbal, sweet
53#	Sweet, chocolatey, winy	123#	Hay, balsamic, sweet	193#	Spicy, sweet, green
54#	Herbal, woody, sweet	124#	Green, fruity, floral	194#	Floral, woody, sweet
55#	Floral, fruity, sweet	125#	Milky, sweet, chocolatey	195#	Spicy, sweet, herbal
56#	Spicy, sweet, woody	126#	Fruity, sweet, beany	196#	Hay, sweet, herbal
57#	Chocolatey, woody, sweet	127#	Spicy, herbal, sweet	197#	Herbal, woody, floral
58#	Chocolatey, smoky, balsamic	128#	Fruity, sweet, fresh	198#	Hay, sweet, roasted
59#	Milky, sweet, fruity	129#	Fruity, sweet, green	199#	Fruity, sweet, herbal
60#	Fruity, sweet, sour	130#	Spicy, sweet, floral	200#	Herbal, woody, green
61#	Milky, fruity, sweet	131#	Floral, sweet, fresh	201#	Floral, fruity, sweet
62#	Roasted, sweet, chocolatey	132#	Green, sweet, floral	202#	Green, herbal, resinous
63#	Floral, green, fresh	133#	Herbal, woody, spicy	203#	Herbal, green, sweet
64#	Fruity, sweet, fresh	134#	Resinous, green, herbal	204#	Floral, sweet, fresh
65#	Balsamic, sweet, resinous	135#	Fruity, sweet, green	205#	Roasted, smoky, sweet
66#	Herbal, green, sweet	136#	Beany, milky, sweet	206#	Herbal, woody, green
67#	Herbal, sweet, woody	137#	Herbal, woody, green	207#	Sweet, balsamic, woody
68#	Milky, sweet, beany	138#	Sweet, fruity, balsamic	208#	Green, floral, woody
69#	Floral, sweet, fresh	139#	Fruity, sweet, herbal	209#	Woody, resinous, sweet
70#	Green, floral, herbal	140#	Fruity, sweet, resinous	210#	Milky, sweet, beany

Table 4
Formula composition of empirical formula 1# (with the main odors of sweet, fruity and milky).

Number	Fragrance ingredient	Amount (wt. %)	Material type	Type of note
1	Acetic acid	0.05	Synthesized	Top
2	Ethyl acetate	0.02	Synthesized	Top
3	Butyric acid	0.20	Synthesized	Top
4	Ethyl butyrate	0.60	Synthesized	Top
5	Ethyl 2-methylbutyrate	0.07	Synthesized	Top
6	Cis-3-hexen-1-ol	0.04	Synthesized	Top
7	Cis-3-hexen-1-yl acetate	0.02	Synthesized	Top
8	Ethyl maltol	0.30	Synthesized	Middle
9	Geraniol	0.01	Synthesized	Middle
10	β-damascone	0.08	Synthesized	Middle
11	α-ionone	0.05	Synthesized	Middle
12	β-ionone	0.01	Synthesized	Middle
13	4-(4-hydroxyphenyl)-2-butanone	1.50	Synthesized	Middle
14	Date tincture	0.30	Natural	Middle
15	Vanillin	2.00	Synthesized	Base
16	Ethyl vanillin	1.50	Synthesized	Base
17	1,2-propanediol	93.25	Solvent	–
Total		100.00	–	–

corresponding amount information, i.e., $X_i = \{(C1_i, P1_i), (C2_i, P2_i), \ldots, (Ct_i, Pt_i)\}$, where Ct_i is the fragrance ingredient and Pt_i is the corresponding amount (Table 5).

Obtain the set of fragrance ingredients by taking the union of all ingredients included in the empirical formula sheets in *PF*. Define C as the resulting set in Equation (1):

$$C = \bigcup_{i=1}^{i=n} Ct_i = \{C1, C2, \ldots, Cm\} \tag{1}$$

Where m is the total number of fragrance ingredients. For each ingredient in C, count its occurrence in all empirical formulas to obtain $n_c (1 \le n_c \le n)$. Sort the fragrance ingredients in C in descending order based on n_c, and let the sorted set be $\overline{C} = \{\overline{C1}, \overline{C2}, \ldots, \overline{Cm}\}$. Therefore, X_i can be represented by $\overline{C}$ as shown in Equation (2):

$$X_i = \{P1_i, P2_i, \ldots, Pm_i\} \tag{2}$$

If X_i does not include a certain ingredient, its amount is 0.

After this process, all empirical formulas (*PF*) can be represented as an m-by-n matrix, as shown in Equation (3):

$$PF = \left(P_{j,i}\right)_{m \times n} \tag{3}$$

Where the j-th row and i-th column of the matrix represent the amount of the j-th ingredient in the i-th formula.

Finally, construct a square adjacency matrix $Q = \left(q_{i,j}\right)_{m \times m}$ to represent the relationships between fragrance ingredients. For all n formulas, if the i-th fragrance ingredient appears in the same formula as the j-th fragrance ingredient k times, then $Q_{i,j} = k$, where k represents the weight of the edge connecting the two fragrance ingredients. Q is symmetric, so its diagonal elements that represent each ingredient itself can be set to 0.

2.8. Construction of an automatic fragrance formula creation algorithm (AFFC)

New fragrance formulas are generated by traversing the relationship diagram generated in section 2.7 using a depth-first search (DFS) algorithm. Prior to traversal, common constraint information must be established to guide the traversal process. This constraint information includes.

Table 5
Information of X_i in empirical fragrance formulas.[a].

Number	Fragrance ingredient	Amount (wt. %)
1	$C1_i$	$P1_i$
2	$C2_i$	$P2_i$
…	…	…
t	Ct_i	Pt_i

[a] Exclude the solvents and their corresponding amounts from the formulas artificially.

i. Fragrance goal: defining the main odors desired for the new fragrance formula in order to provide the algorithm with an empirical formula containing the corresponding main fragrance style.
ii. Mandatory ingredients for the new formula: primarily based on user preferences or selecting one or more ingredients that match the fragrance goal, set them as the starting or passing point for the graph traversal.
iii. Rejected ingredients for the new formula: typically based on user preferences or usage experience, define ingredients that cannot be used in the new formula. To ensure graph connectivity, results containing the rejected ingredient(s) should be removed from the result set of the traversal. If only one result contains the rejected ingredient(s), the ingredient(s) should be removed directly from the formulation.
iv. Quantity of ingredients in the new formula: setting the number of fragrance ingredients included in the generated formula.
v. Other constraint information: such as the proportion of top note, middle note, and base note, the proportion of synthesized and natural ingredients, and other relevant ingredient information can be established as additional constraints during the traversal process to adjust the traversal path.

Using the DFS algorithm to implement the traversal process, the specific method is as follows.

i. A query is made to the empirical formulas database to retrieve all empirical formulations that contain the desired main odors for the new fragrance.
ii. Starting from the mandatory ingredient A, select the edge with the highest weight among the edges connected to A as the forward direction;
iii. If B is the edge connected to A with the highest weight, select the edge with the highest weight among the edges connected to B, except for A, as the forward direction.
iv. Set B as the new starting point and proceed according to the above steps;
v. If there is only one edge connected to B, which is from A, backtrack to A and select the edge connected to A with the second highest weight as the forward direction while retaining B in the path and repeat the process.
vi. Once all mandatory ingredients have been covered by the path, let L be the number of points traversed minus the number of rejected ingredients. Suppose the number of components in the preset formula is R. If $L = R$, stop the traversal. If $L < R$, continue the traversal until $L = R$. If $L > R$, select the point with the smallest degree (number of connected edges) in the traversed path for deletion and repeat until $L = R$.
vii. Remove the rejected ingredients from the traversed path and output the remaining portion as the new formula ingredients.

Based on prior research and the structure of empirical formulas, we have developed a practical method associated with the dosage amount of raw materials in empirical formulas for predicting the usage amount of raw materials in fragrance formulas.

We obtain the corresponding formula ingredients information, denoted as T, based on the new formula composition generated above. We then calculate the ratio of the intersection size (i.e., the number of common ingredients between T and each empirical formula X_i) to the union size (i.e., number of ingredients in both T and X_i), denoted as W_i in Equation (4):

$$W_i = \frac{|T \cap X_i|}{|T \cup X_i|} \tag{4}$$

The usage amount p of ingredient c in T is then given by Equation (5):

$$p = \frac{\sum_{i=1}^{n} W_i \times Pc_i}{\sum_{i=1}^{n} W_i \times |sgn(Pc_i)|} \tag{5}$$

Here, n is the number of empirical formulas, Pc_i is the usage amount of ingredient c in the corresponding empirical formula, and $|sgn(Pc_i)|$ is the sign function. When $Pc_i = 0$, $|sgn(Pc_i)|$ is 0, otherwise, it is 1.

3. Results and discussion

3.1. Intelligent creation of fragrance formula

In this study, the goal of fragrance creation was focused on floral fragrances. A total of 52 empirical formulas with floral odor as main odors were selected from the empirical formulas database, involving 285 types of fragrance raw materials. These empirical formulas were organized into a 285 × 52 matrix, and a diagram illustrating the relationships of ingredients used in these empirical formulas was constructed using the methodology described in section 2.7 (Fig. 1) Each point on Fig. 1 represents a fragrance ingredient, with the size of the point indicating the number of uses of this ingredient in the empirical formula. In other words, the more frequently a certain fragrance ingredient appears in the formula, the larger the point is. The lines between the points denote that these two fragrance ingredients have been used together in at least one empirical formula.

To create a fragrance formula utilizing computer intelligence, the following constraint conditions were implemented based on traditional flower fragrance formulas and product requirements.

1) The fragrance goal was set to "floral fragrance."
2) The formula was limited to a maximum of "20" fragrance ingredients.
3) The mandatory ingredients "geraniol" and "benzyl acetate" were included to enhance floral odors.
4) No rejected materials or other constraint conditions were specified.

Using the AFFC algorithm described in section 2.8 with the above constraint conditions, a semi-finished fragrance formula T_0# containing only the selected ingredients was generated by the computer (Table 6).

Based on the semi-finished fragrance formula T_0#, the W_i values between formula T_0# and the 52 empirical formulas used were automatically calculated by the computer according to Equation (4) in *section 2.8*, and the results are presented in Table 7.

According to Equation (5) in section 2.8, each component in formula T_0# was automatically calculated in sequence to determine the required percentage usage amount. Any remaining percentage less than 100 % was compensated with a solvent. The resulting primary fragrance formula X_1# was then obtained from the computer-generated formula that met our pre-defined fragrance targeting goals (Table 8).

3.2. Optimization of the primary fragrance formula generated by AFFC algorithm

A small sample of the formula X_1# was prepared according to the specifications in Table 8 and allowed to mix and settle appropriately. Subsequently, the olfactory profile of the sample was evaluated using the method outlined in Section 2.3.

The results of the evaluation (Fig. 2) indicated that the formula X_1# had predominant notes of sweet (4.10 points), spicy (3.88 points), and floral (3.66 points), accompanied by green (2.45 points), roasted (2.05 points), and fresh (1.86 points) aromas. Heavy use of ingredients with spicy and roasted notes resulted in poor overall harmony and comfort, which deviated from the intended goal of creating a floral fragrance. Therefore, it is necessary to adjust the algorithm constraints to optimize the formula based on the characteristics of the raw materials used.

Based on the olfactory profiles of fragrance formula X_1# and the characteristics of its materials as references according to FIID, the spicy cinnamon oil and roasted notes of 2,3,5-trimethylpyrazine were manually removed as constraint conditions for the AFFC algorithm. The new traversal constraint conditions for the algorithm were adjusted as follows.

1) The fragrance goal was set to "floral fragrance."
2) The formula was limited to a maximum of "20" fragrance ingredients.
3) Mandatory ingredients set for new formula were "geraniol" and "benzyl acetate".
4) Reject the use of "cinnamon oil" and "2,3,5-trimethylpyrazine" in the new formula.

After running the aforementioned AFFC algorithm, the optimized complete fragrance formula X_2# was obtained (Table 9).

Likewise, a small sample of the formula X_2# was prepared according to Table 9 and made it mix and settle appropriately, and the olfactory profiles of this sample was evaluated according to the method mentioned in section 2.3.

As depicted in Fig. 3, the olfactory sensory evaluation revealed that formula X_2# was primarily composed of floral (4.58 points), sweet (4.22 points), and fresh (3.66 points) notes, complemented by green (3.02 points), fruity (2.58 points), and wine (1.65 points) notes. The overall olfactory performance of this formula exhibited excellent consistency and stability, and it also had a comfortable and harmonious scent with long-lasting effects. The achieved results well aligned with the objective of creating a floral fragrance with practical value.

In summary, we introduced an advanced algorithm designed to analyze the interrelationships among ingredients in empirical fragrance formulas in this study. Leveraging statistical analysis, this algorithm strategically selected key components for devising

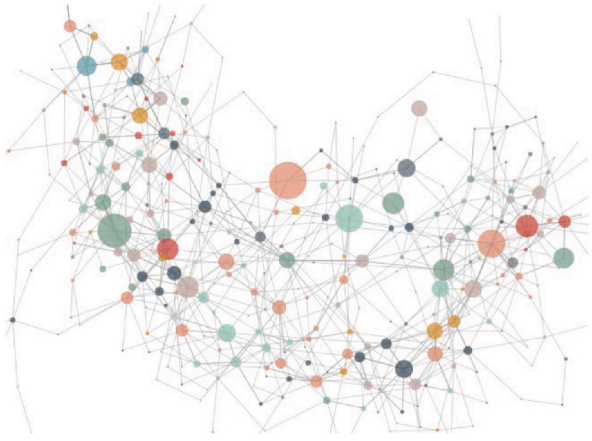


Fig. 1. Relationship diagram of empirical fragrance formulas.

Table 6
Fragrance components of semi-finished formula T_0# generated by AFFC algorithm.

Number	Fragrance ingredient	Material type	Type of note
1	Benzyl acetate	Synthesized	Top
2	Phenethyl acetate	Synthesized	Top
3	Phenethyl alcohol	Synthesized	Top
4	Linalool	Synthesized	Top
5	Orange flower absolute	Natural	Top
6	D,l-menthol	Synthesized	Top
7	Nerolidol	Synthesized	Middle
8	Anisyl alcohol	Synthesized	Middle
9	Geraniol	Synthesized	Middle
10	Menthone	Synthesized	Middle
11	β-damascone	Synthesized	Middle
12	Ethyl maltol	Synthesized	Middle
13	2,3,5-trimethyl pyrazine	Synthesized	Middle
14	β-ionone	Synthesized	Middle
15	Citronellal	Synthesized	Middle
16	Cassia bark oil	Natural	Middle
17	Lavender oil	Natural	Middle
18	Jasmine absolute	Natural	Base
19	Treemoss absolute	Natural	Base
20	Balsam peru	Natural	Base

Table 7
Values of W_i between formula T_0# and the 52 empirical formulas used.

Number of i	W value	Number of i	W value
1	0.1316	27	0.1862
2	0.1250	28	0.0426
3	0.2895	29	0.1500
4	0.0789	30	0.0189
5	0.1351	31	0.3912
6	0.2564	32	0.0910
7	0.2093	33	0.0278
8	0.2791	34	0.2162
9	0.2432	35	0.2195
10	0.1081	36	0.1842
11	0.2250	37	0.2368
12	0.0620	38	0.1676
13	0.0732	39	0.0271
14	0.0544	40	0.1514
15	0.0563	41	0.0859
16	0.0688	42	0.0786
17	0.0865	43	0.2452
18	0.3926	44	0.0812
19	0.0722	45	0.0468
20	0.0714	46	0.0725
21	0.3514	47	0.2764
22	0.0956	48	0.1096
23	0.0978	49	0.1358
24	0.2522	50	0.3580
25	0.1081	51	0.2021
26	0.0244	52	0.0232

innovative formulas. The algorithm's efficacy was significantly influenced by the quality of the empirical fragrance formulas it processed. For instance, an ingredient like 'A', if ubiquitous in empirical formulas, is consistently included in new formulations. Conversely, an ingredient such as 'B', absent from the empirical datasets, is similarly excluded from the new formulations. This presents a notable challenge in the algorithm's functionality. Currently, addressing these imbalances necessitates manual intervention, involving strategic inclusion or exclusion of certain ingredients in the new formulas, to align them with predefined research objectives.

4. Conclusion

This article presents a novel approach to improving the efficiency and accuracy of fragrance design using empirical fragrance formulas and graph traversal algorithms. We constructed a database of 344 common fragrance ingredients that allowed perfumers to set algorithmic constraints based on their experience and the characteristics of these materials. By analyzing the composition of 210 empirical fragrance formulas and constructing a relational network model in graph form, we illustrated the relationships between the

Table 8
Complete intelligent fragrance formula X_1# generated by AFFC algorithm.

Number	Fragrance ingredient	Amount(wt. %)	Material type	Type of note
1	Benzyl acetate	0.39	Synthesized	Top
2	Phenethyl acetate	0.88	Synthesized	Top
3	Phenethyl alcohol	2.12	Synthesized	Top
4	Linalool	0.15	Synthesized	Top
5	Orange flower absolute	0.26	Natural	Top
6	D,l-menthol	0.47	Synthesized	Top
7	Nerolidol	0.45	Synthesized	Middle
8	Anisyl alcohol	1.76	Synthesized	Middle
9	Geraniol	1.11	Synthesized	Middle
10	Menthone	0.47	Synthesized	Middle
11	β-damascone	0.52	Synthesized	Middle
12	Ethyl maltol	0.51	Synthesized	Middle
13	2,3,5-trimethylpyrazine	0.19	Synthesized	Middle
14	β-ionone	0.49	Synthesized	Middle
15	Citronellal	0.04	Synthesized	Middle
16	Cassia bark oil	0.72	Natural	Middle
17	Lavender oil	0.05	Natural	Middle
18	Jasmine absolute	0.13	Natural	Base
19	Treemoss absolute	0.97	Natural	Base
20	Balsam peru	0.18	Natural	Base
21	Solvent	88.14	Solvent	–
Total		100.00	–	–

Fig. 2. The olfactory profiles of fragrance formula X_1#.

ingredients used in these formulas. Finally, we developed an automatic fragrance formula creation (AFFC) algorithm that constructs a subgraph of the relational network and finds fragrance formula solutions based on the depth-first search algorithm, satisfying the constraint conditions, and combining the appropriate statistical strategy to determine the use of each component in the new fragrance formula. The algorithm introduced in our study innovatively utilized empirical fragrance formulas to determine various components and their respective quantities. Through empirical validation, it demonstrated satisfactory performance. However, this algorithm exhibited a significant dependency on empirical formula data. Any deviations in the empirical formulas or the inclusion of personal preferences in raw material selection markedly affected the algorithm's outcomes, leading to a decrease in its innovative capacity. Most notably, the algorithm was unable to accommodate ingredients not previously encountered in the empirical formula data. The results of testing the AFFC algorithm for creating a floral fragrance produced a formula with primary fragrance notes of floral, sweet, and fresh, exhibiting good consistency and stability which meets perfumers' requirements. Overall, this study provides a novel and convenient fragrance development approach for fragrance developers, which addresses the diversity deficiency problem caused by individual preferences and improves the efficiency of traditional fragrance development.

Table 9
Optimized intelligent fragrance formula X_2# generated by AFFC algorithm.

Number	Fragrance ingredient	Amount (wt. %)	Material type	Type of note
1	Benzyl acetate	0.26	Synthesized	Top
2	Phenethyl acetate	1.21	Synthesized	Top
3	Phenethyl alcohol	1.56	Synthesized	Top
4	Linalool	0.13	Synthesized	Top
5	Linalyl acetate	0.12	Synthesized	Top
6	D,l-menthol	0.63	Synthesized	Top
7	β-damascenone	0.25	Synthesized	Middle
8	Nerolidol	0.23	Synthesized	Middle
9	Anisyl alcohol	1.19	Synthesized	Middle
10	Geraniol	0.12	Synthesized	Middle
11	γ-valerolactone	0.48	Synthesized	Middle
12	Maltol	0.63	Synthesized	Middle
13	Styralyl acetate	0.05	Synthesized	Middle
14	β-ionone	0.26	Synthesized	Middle
15	Geranium oil	0.30	Natural	Middle
16	Spearmint oil	0.14	Natural	Middle
17	Aglaia odorata flower oil	0.06	Natural	Middle
18	Jasmine absolute	0.17	Natural	Base
19	Treemoss absolute	0.58	Natural	Base
20	Balsam peru	0.36	Natural	Base
Solvent		91.27	Solvent	
Total		100.00	–	

Fig. 3. The olfactory profiles of fragrance formula X_2#.

CRediT authorship contribution statement

Jiaxiao Cai: Writing – original draft, Methodology, Data curation. **Suxing Tuo:** Validation, Methodology, Formal analysis, Data curation. **Yanchun Li:** Formal analysis, Data curation. **Hongbing Lu:** Validation, Formal analysis, Data curation. **Yizi Wu:** Validation, Supervision, Software, Methodology. **You Zou:** Visualization, Validation, Supervision, Software. **Zhen Ma:** Resources, Investigation. **Yuqi Cui:** Methodology, Investigation, Data curation. **Bo Kong:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Kejun Zhong:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Data availability statement

Data will be made available on request.

Ethics declarations

Review and/or approval by an ethics committee was not needed for this study because the study used anonymized, published data.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors extend their appreciation to the State Tobacco Monopoly Administration of China for funding this research work through the project no. 110202202002 and no. 110202001001(YX-01).

Nomenclature

PF	the set of empirical fragrance formulas
X_i	empirical formula
n	total number of empirical formulas
Ct_i	fragrance ingredient
Pt_i	amount of fragrance ingredient
m	total number of fragrance ingredients
n_c	fragrance ingredients that occurrence in all empirical formulas
C	fragrance ingredient set
$\overline{C}$	sorted fragrance ingredient set
Q	a square adjacency matrix
k	weight of the edge connecting the two fragrance ingredients
L	number of points traversed minus number of rejected ingredients
R	number of components in preset formula
T	fragrance ingredients used in auto-created formula
W_i	ratio of the intersection size to the union size
p	usage amount of fragrance ingredients in auto-created formula

References

[1] W. Zhan, F. Doro, M.A. Teixeira, A rapid approach to optimize the design of fragrances for fabric care products, Flavour Fragrance J. 35 (2) (2019) 167–173.
[2] H.N. Schifferstein, B.F. Howell, Using color-odor correspondences for fragrance packaging design, Food Qual. Prefer. 46 (2015) 17–25.
[3] M.A. Teixeira, O. Rodriguez, V.G. Mata, A.E. Rodrigues, Perfumery quaternary diagrams for engineering perfumes, AIChE J. 55 (8) (2009) 2171–2185.
[4] V.G. Mata, P.B. Gomes, A.E. Rodrigues, Engineering perfumes, AIChE J. 51 (10) (2005) 2834–2852.
[5] A.C. Figueiredo, J.G. Barroso, L.G. Pedro, J.J.C. Scheffer, Factors affecting secondary metabolite production in plants: volatile components and essential oils, Flavour Fragrance J. 23 (4) (2008) 213–226.
[6] W. Jennings, S. Takayuki, Qualitative analysis of flavor and fragrance volatiles by glass capillary gas chromatography, Journal of Chemical Education 58 (12) (1981) A381.
[7] C.S. Letizia, J. Cocchiara, G.A. Wellington, C. Funk, A.M. Api, Monographs on fragrance raw materials, Food Chem. Toxicol. 38 (2000). S1-+.
[8] M.A. Teixeira, O. Rodriguez, A.E. Rodrigues, Prediction model for the odor intensity of fragrance mixtures: a valuable tool for perfumed product design, Ind. Eng. Chem. Res. 52 (2) (2013) 963–971.
[9] L. Buck, R. Axel, A novel multigene family may encode odorant receptors: a molecular basis for odor recognition, Cell 65 (1) (1991) 175–187.
[10] J. Wei, Product Engineering: Molecular Structure and Properties, Oxford University Press, 2007.
[11] L. Shang, C.J. Liu, Y. Tomiura, K. Hayashi, Machine-learning-based olfactometer: prediction of odor perception from physicochemical features of odorant molecules, Anal. Chem. 89 (22) (2017) 11999–12005.
[12] M.A. Teixeira, L. Barrault, O. Rodriguez, C.C. Carvalho, A.E. Rodrigues, Perfumery radar 2.0: a step toward fragrance design and classification, Ind. Eng. Chem. Res. 53 (21) (2014) 8890–8912.
[13] L. Turin, A method for the calculation of odor character from molecular structure, J. Theor. Biol. 216 (3) (2002) 367–385.
[14] H.S. Meda, A.K. Sen, A. Bagchi, Detecting data flow errors in workflows: a systematic graph traversal approach. 17th Annual Workshop on Information Technolgies & Systems (WITS) Paper, 2007.
[15] J.J. Holdsworth, Graph traversal and graph transformation, Theor. Comput. Sci. 321 (2–3) (2004) 215–231.
[16] R. Fleischer, G. Trippen, Experimental studies of graph traversal algorithms, in: K. Jansen, M. Margraf, M. Mastrolli, J.D.P. Rolim (Eds.), Experimental and Efficient Algorithms, 2003, pp. 120–133. Proceedings.
[17] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, Introduction to Algorithms, MIT press, 2022.
[18] R. Tarjan, Depth-first search and linear graph algorithms, SIAM J. Comput. 1 (2) (1972) 146–160.
[19] R. Rahim, D. Abdullah, J. Simarmata, A. Pranolo, A.S. Ahmar, R. Hidayat, D. Napitupulu, H. Nurdiyanto, B. Febriadi, Z. Zamzami, Iop, block architecture problem with depth first search solution and its application. Joint Workshop of KO2PI/2nd International Conference on Mathematics, Science, Technology, Education, and Their Applications (ICMSTEA), 2016. Makassar, INDONESIA.
[20] J. Xiong, Z. Li, L. Fan, D. Liu, Improved depth first algorithm and its application in information retrieval. 2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), IEEE, 2010, pp. 69–73.
[21] H.N.B. Moseley, N. Riaz, J.M. Aramini, T. Szyperski, G.T. Montelione, A generalized approach to automated NMR peak list editing: application to reduced dimensionality triple resonance spectra, J. Magn. Reson. 170 (2) (2004) 263–277.
[22] M. Angele Hamel, Combinatorial optimization: theory and algorithms (5th ed.), Comput. Rev. 53 (12) (2012) 717–718.
[23] J.E. Hopcroft, R. Motwani, J.D. Ullman, Introduction to automata theory, languages, and computation, Acm Sigact News 32 (1) (2001) 60–65.
[24] M.A. Teixeira, O. Rodriguez, P. Gomes, V. Mata, A. Rodrigues, Perfume Engineering: Design, Performance and Classification, Butterworth-Heinemann, 2012.

[25] C. Fisher, T.R. Scott, Food Flavours: Biology and Chemistry, Royal Society of chemistry, 2007.
[26] C.S. Sell, The Chemistry of Fragrances: from Perfumer to Consumer, Royal Society of Chemistry, 2006.
[27] L. Van Gemert, Compilation of Odor Threshold Values in Air and Water, RID, CIVO-TNO, Zeist, The Netherlands, 1977.