

Article

Prediction of Tea Varieties' "Suitable for People" Relationship: Based on the InteractE-SE+GCN Model

Qiang Huang^{1,*}, Zongyuan Wu^{1,†}, Mantao Wang¹, Youzhi Tao¹, Yinghao He¹ and Francesco Marinello² 

¹ College of Information Engineering, Sichuan Agricultural University, Ya'an 625000, China; zywu@stu.sicau.edu.cn (Z.W.)

² Department of Land, Environment, Agriculture and Forestry, University of Padua, 35020 Legnaro, Italy; francesco.marinello@unipd.it

* Correspondence: 13499@sicau.edu.cn; Tel.: +86-151-8435-7781

† These authors contributed equally to this work.

Abstract: This study proposes an improved link prediction model for predicting the "suitable for people" relationship within the knowledge graph of tea. The relationships between various types of tea and suitable target groups have yet to be fully explored, and the existing InteractE model still does not adequately capture a portion of the complex information around the interactions between entities and relationships. In this study, we integrate SENet into the feature layer of the InteractE model to enhance the capturing of helpful information in the feature channels. Additionally, the GCN layer is employed as the encoder, and the SENet-integrated InteractE model is used as the decoder to further capture the neighbour node information in the knowledge graph. Furthermore, our proposed improved model demonstrates significant improvements compared to several standard models, including the original model from public datasets (WN18RR, Kinship). Finally, we construct a tea dataset comprising 6698 records, including 330 types of tea and 29 relationship types. We predict the "suitable for people" relationship in the tea dataset through transfer learning. When comparing our model with the original model, we observed an improvement of 1.4% in H@10 for the WN18RR dataset, a 7.6% improvement in H@1 for the Kinship dataset, and a 5.2% improvement in MRR. Regarding the tea dataset, we achieved a 4.1% increase in H@3 and a 2.5% increase in H@10. This study will help to fully exploit the value potential of tea varieties and provide a reference for studies assessing healthy tea drinking.

Keywords: tea; suitable for people; knowledge graph; link prediction; knowledge graph completion; transfer learning; deep learning



Citation: Huang, Q.; Wu, Z.; Wang, M.; Tao, Y.; He, Y.; Marinello, F.

Prediction of Tea Varieties' "Suitable for People" Relationship: Based on the InteractE-SE+GCN Model.

Agriculture **2023**, *13*, 1732. <https://doi.org/10.3390/agriculture13091732>

Academic Editor: Roberto Alves Braga Junior

Received: 30 June 2023

Revised: 17 August 2023

Accepted: 30 August 2023

Published: 31 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tea is a beverage traditionally considered to have health-promoting properties [1]. Tang et al. [2] showed that tea also has preventive effects on inflammation, cancer, and obesity, providing a valuable reference for further research on tea components' health-related functions and mechanisms of action. Sae-Tan et al. [3] explored the effects of tea consumption on weight loss and the prevention of metabolic syndrome (MetS) in animals and confirmed that green tea consumption had a fat-reducing or weight-reducing effect. The discovery of tea's preventive and therapeutic effects was, however, accompanied by cases of injury or illness induced by inappropriate tea consumption, such as the case of 79 patients who developed liver disease as a result of consuming green tea extracts [4]. Exploring the suitability of various types of tea for human health, taking into account the properties of each tea variety, is essential.

According to Pan et al. [5], based on the degree of fermentation, tea tree growing conditions, cultivation methods, and the tea-making process, tea is generally classified into six major categories: green tea, green (oolong) tea, white tea, yellow tea, black tea, and dark black tea. It can thus be concluded that factors such as the nature, genus, suitable

type, and value efficacy of tea are the leading indicators through which to classify tea as suitable for people. Mahdavi-Roshan et al. [6] studied the effects of black and green tea beverage intake on hypertension through a search strategy. Yan et al. [7] analysed winter tea's value efficacy and social benefits by testing its composition and contents, aiming to improve these features. Both of these studies utilised statistical analysis methods, and this single form of analysis may have led to the poor comparability of their data. Lee et al. [8] studied a recommendation service of blended tea in conjunction with a food recommendation system and verified the feasibility of blended tea recommendations after final consumer acceptance tests. In recent years, with the advancement and application of artificial intelligence technology, deep learning has been incorporated into tea quality research. Chen et al. [9] used image information and environmental parameters (EPs) to construct convolutional neural networks and gated recursive units (GRUs) with which to predict the moisture content and product quality of Pu-erh tea during the sun-drying process, which has guiding significance for tea research combined with deep learning. However, the objectives of these studies were still mainly focused on external characteristics. They considered the components of tea leaves without a comprehensive analysis of the six major tea types or a sufficient focus on the relationship between each tea type and the people who consume it.

Google proposed the knowledge graph (KG) in 2012, and Hogan et al. [10] provided a comprehensive introduction to the knowledge graph. At the same time, many knowledge graph methods have sound application effects in various agricultural fields, such as agricultural knowledge services and pest diagnoses [11]. Chen et al. [12] proposed an Agricultural KG (AgriKG) for effectively integrating fragmented information generated using multiple applications for agricultural entity retrieval and agricultural knowledge Q&A.

Link prediction, which uses existing relationships to infer new connections and thus build a complete knowledge graph, is a fundamental task in knowledge graph completion. Rossi et al. [13] classified link prediction models into three families: tensor decomposition, geometric, and deep learning models. On the other hand, Wang et al. [14] showed that neural network-based models demonstrate superior performance in knowledge graph link prediction tasks compared to other traditional methods. Advanced neural network structures are capable of generating expressive feature embeddings; for example, the ConvE model [15] addresses the problem that previous models for knowledge graph link prediction tasks had—the fact that they are primarily shallow and have a weaker ability to learn features than deep multilayer models. It utilises a two-dimensional convolutional neural network to extract features from the stitching matrix of entities and relations. After linear transformation to conduct matrix multiplication with the entity matrix, the graph obtained effectively improves the link prediction performance of the model. The ParamE model [16] treats the neural network parameters as the embedding of relations and, using head entities as input values and tail entities as output values, trains different networks for different relations. The InteractE [17] model is still lacking in capturing detailed interaction information. However, the InteractE model also achieved good results in the knowledge graph link prediction task with feature replacement on the information matrix, “Chequer” reshaping operations, and cyclic convolution operations. Models based on ordinary neural networks have achieved better results in knowledge graph link prediction tasks; however, the ability to capture interaction information still needs to be improved upon in order for these models to be comparable to that of graph neural networks, which can fully account for the neighbourhood information of knowledge graph entities and capture a more informative representative embedding between entities. R-GCN [18] applied graph convolutional networks to knowledge graph link prediction tasks, using the representation of neighbouring nodes as the representation of the current node, based on the idea of “information propagation”, which takes into account the multi-relational data within the knowledge graph and constructs an end-to-end encoder–decoder model. The KBGAT model [19] proposes using graph attention for relation prediction, using neighbouring nodes to represent the current node, further facilitating the flow of information and achieving better results on

knowledge graph link prediction tasks. In addition, the ComplexGCN model [20] is a new extension of standard graph convolutional networks (GCNs) in complex space; it combines the symbolic power of complex geometry with GCNs to improve the quality of the representation of KG components. CompGCN [21] achieved excellent results in the link prediction task, using multiple entity–relationship combination operations in the knowledge graph embedding technique and fully demonstrating the effectiveness of combining graph neural networks with ordinary neural networks in the link prediction task. The link prediction models illustrate a progressive trend within CNN-based architectures, transitioning from the initial ConvE model to the more proficient InteractE model. Similarly, in the realm of GCN-based architectures, the evolution has advanced from the initial adoption of the R-GCN model for link prediction to culminate in the refined and more effective CompGCN model through the continuous efforts of researchers. However, despite the enhancement in expressive capacity achieved with using the InteractE model through the “Chequer” reshaping operation, such a technique disrupts the spatial information contained within the original entity and relationship embeddings. Furthermore, it is essential to note that the InteractE model, as part of the CNN framework, lacks the ability to effectively process and consolidate relational information across various hierarchical levels, which is inherent in the CompGCN model.

Link prediction methods are prominent in numerous fields. In particular, McCoy et al. [22] developed an end-to-end machine learning pipeline through which to train and serve link prediction models, using link prediction methods to predict missing links in the biomedical literature for drug discovery, which can effectively suggest repurposed drugs for emergent diseases. Huo et al. [23] proposed a personalisation-based social influence link prediction approach with which to predict link relationships between users by modelling personalised influences in their social networks. Nasiri et al. [24] introduced a link prediction method based on familiar neighbours and various centrality metrics (including degree, k-core, closeness, betweenness, Eigenvector, and PageRank) to forecast new links in a multiplex network. By leveraging existing health conditions, Shabaz et al. [25] employed multiple link prediction techniques to anticipate future diseases. Nasiri et al. [26] proposed a feature selection-based random walk approach for link prediction between proteins, enhancing the discovery of their interactions. These studies show that link prediction algorithms have been extensively investigated in various domains yet still need to be explored with regard to tea. Furthermore, the aforementioned studies [5–9] mainly focused on the substance composition of specific tea types, and studies on the associated relationships between tea and people can still be fully explored.

SENet [27] represents a network architecture designed for classification tasks through which to enhance convolutional neural networks’ feature expression capability. Transfer learning [28] can effectively use existing knowledge and data resources to elevate model performance, generalisation capacity, and training efficiency. In response to the original InteractE model’s challenge, where the “Chequer” reshaping operation disrupts the spatial information encoded within the original entity and relationship embeddings, we integrated SENet into the InteractE model framework. Addressing the inherent limitations of the original InteractE model in handling relational information across different hierarchical levels, we leveraged the GCN layers from the CompGCN model as encoders to further capture diversified relational information across varying levels. Finally, we constructed a tea dataset (ID_Tea) and used the improved model to predict the relationships between types of tea and the variable “suitable for people”. The main contributions of this paper are as follows:

- We improved the initial InteractE model by combining it with SENet (the improved model is called InteractE-SE) and incorporated SENet after the feature layer of the InteractE model to enhance the capture of helpful information in the feature channel.
- We combined the above model with GCN to improve the InteractE model so that the GCN layer in the CompGCN model is used as an encoder and the SENet-incorporated InteractE model is used as a decoder (the improved model is called IntGCN), which

strengthens the model's ability to extract complex interaction information between entities and relationships. After several experiments, the improved model significantly improved the prediction metrics on public datasets (WN18RR, Kinship).

- We constructed a dataset containing 6698 records, including 330 types of tea and 29 types of relationships. Combining the improved model (IntGCN) with migration learning, we comprehensively used the knowledge and patterns the improved model learned in WN18RR to predict the “suitable for people” relationships in the tea dataset and complete the tea knowledge graph using the prediction results. This study thereby helps to explore the value potential of tea varieties and provides some references for tea research.

2. Materials and Methods

2.1. Research Process

Figure 1 presents this study's workflow, which included improving the InteractE link prediction model, testing the improved model on public datasets (Step 1), and constructing a tea knowledge graph (Step 2). In Step 2, the knowledge graph dataset was organised to form the ID_Tea dataset. We used the IntGCN model to predict the “suitable for people” relationships for different teas in the ID_Tea dataset, combined with transfer learning (Step 3). Specifically, the improvements made to the InteractE link prediction model involve using the GCN layer as the encoder and integrating SENet into the InteractE model as the decoder. The final model was named IntGCN. After training, we can obtain the pre-trained version of the IntGCN model under the WN18RR dataset. For Step 1, we validated the performance of the IntGCN model on public datasets (WN18RR, Kinship). Establishing and constructing the tea knowledge graph was mainly achieved through a “bottom-up” approach [29], and the constructed knowledge graph is stored in the Neo4j database [30]. Predicting the suitable target groups for tea involves utilising the transfer learning method. First, we trained the pre-trained model obtained by training the IntGCN model on the public dataset WN18RR and fine-tuning it for link prediction on the ID_Tea dataset. Then, we used the IntGCN model to predict the “suitable for people” relationship for the ID_Tea dataset. Through multiple iterations of model training, we obtained the suitable target groups for each tea category based on the scores obtained by evaluating the test triplets corresponding to that category.

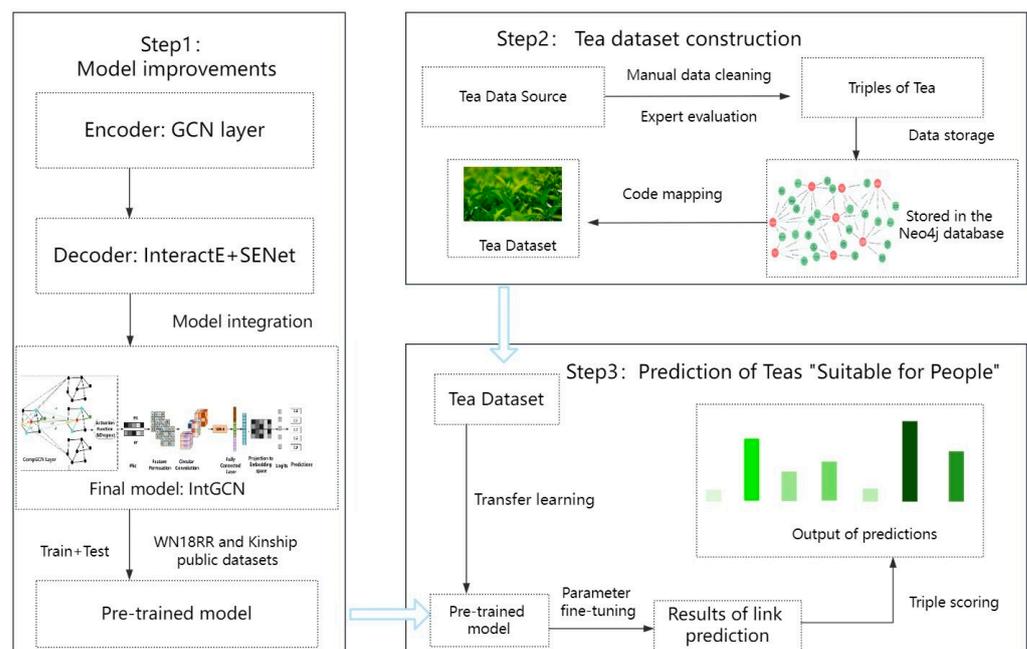


Figure 1. Overall flow chart of this study's workflow.

2.2. Model Design

This study proposes an end-to-end structural model that progressively refines the initial InteractE model. Initially, while the “Chequer” reshaping operation enhances the expressive capacity of the initial InteractE model, it compromises the spatial information embedded within entity and relationship embeddings. To address this concern, we incorporated SENet following the convolutional layers of the InteractE model. This integration selectively extracts valuable feature channel information while suppressing redundant information. The model resulting from this fusion is named InteractE-SE and is depicted in Figure 2. Furthermore, InteractE-SE inherently follows a CNN-based structural paradigm, limiting its ability to fully capture the intricate interaction information between entities and relationships. We employed the graph convolutional layer from the CompGCN model as the encoder with which to overcome this limitation. This choice facilitates the acquisition of enriched embedded vectors representing combinations of entities and relationships. With InteractE-SE serving as the decoder, we constructed an end-to-end IntGCN model. This comprehensive model was subsequently utilised in link prediction experiments. The model structure diagram of IntGCN is shown in Figure 3.

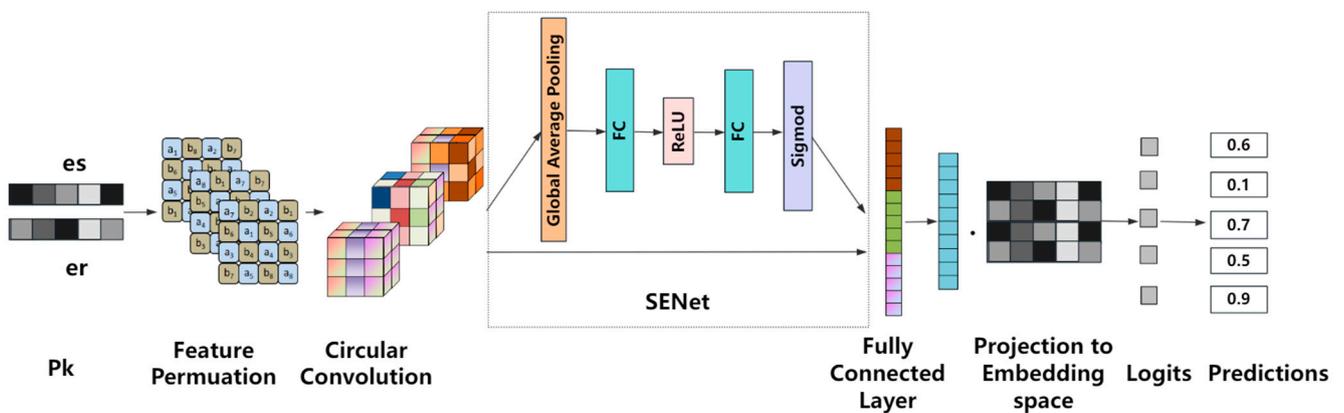


Figure 2. Structure of the InteractE-SE model.

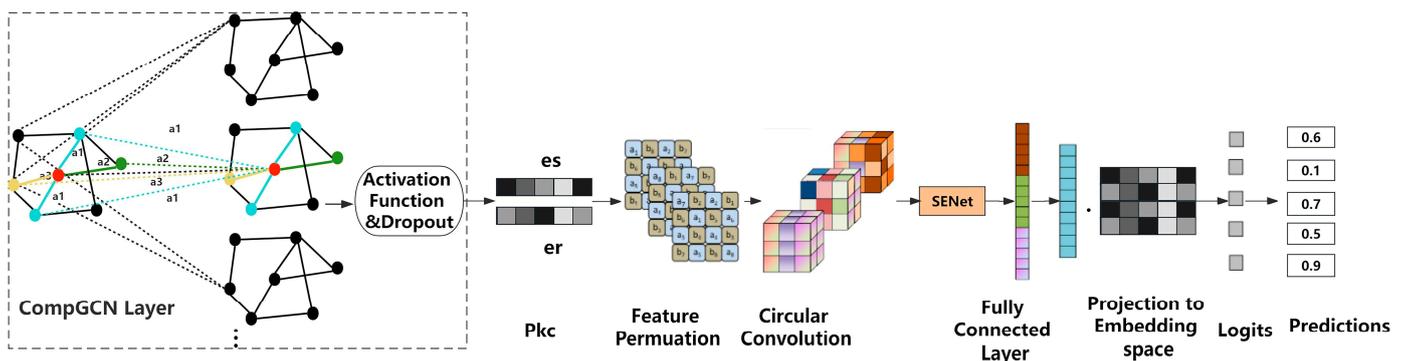


Figure 3. Structure of the IntGCN model.

2.2.1. GCN Layer

The CompGCN model demonstrates superior performance in link prediction tasks, primarily due to its unique graph convolutional layer (henceforth referred to as the GCN layer). Compared to other graph convolution models, it offers several advantages. Firstly, the GCN layer explicitly models the relationships between entities by utilising composition operations to interact with entity and relationship embeddings, and this enables the model to capture complex interactions between entities and relationships more effectively, thereby improving the accuracy of link prediction. Secondly, the GCN layer supports multiple composition operations, providing greater flexibility. This flexibility allows the model to

select suitable composition operations based on the characteristics of different tasks and datasets, further enhancing the capability of representation learning. Such flexibility helps the model better adapt to different knowledge graph data types.

For instance, for the tea dataset (see Section 2.3 for details), the specific workflow of the GCN layer is as follows: The initial step involves the representation of entities and relationships. Each entity (representing a particular type of tea or its attributes) and relationship is represented using vectors. These vectors can either be randomly initialised or pre-trained embedding vectors. Subsequently, diverse relationship aggregation operations are conducted to achieve relational updates. GCN layers cater to different relationships, combining them in specific ways that involve concatenation, weighted summation, matrix operations, and more. These approaches are employed to capture the interactive dynamics between various relationships. Such relationship aggregations can be performed across different layers. Within each layer, the model amalgamates different relationships to acquire enriched representations of relationships. After relationship aggregation, the updating of tea entity representations occurs. Then, building upon the relationship aggregation, tea entity updates are realised by combining the entity's representation with the updated relationship representation, typically through concatenation. The final phase involves the iterative stacking of GCN layers to extract more advanced features, thereby enhancing the model's expressive capabilities. Across each layer, the model executes relationship aggregation and entity updating for richer information on entities and relationships. It becomes evident that the computed tea entity and relationship embeddings through the GCN layers for the tea dataset encapsulate significantly enriched interactive information. This research integrates the GCN layers as the encoders of the ultimate model (IntGCN).

2.2.2. InteractE-SE

Building upon the InteractE model, this study integrates SENet into the InteractE framework. The InteractE model captures high-order interactions between entities and relationships by introducing feature permutation operations within entity and relationship embeddings. This feature permutation operation enables the InteractE model to outperform many other link prediction models with regard to representation learning capability. Additionally, the InteractE model possesses richer feature representations. Using cyclic convolutional operations, the InteractE model learns interaction features between entities and relationships from various perspectives. This multi-perspective learning feature contributes to improved accuracy in link prediction.

Notably, the InteractE model, in capturing high-order interactions and rich feature representations in entity and relationship embeddings, exhibits good generalisation ability across different tasks and datasets. However, the "Chequer" reshaping operation enhances the expressive capacity of the initial InteractE model, and this compromises the spatial information embedded within entity and relationship embeddings. To further enhance the feature extraction capabilities of the InteractE model, we integrated SENet into the feature map extracted using the InteractE model. SENet is a deep learning-based convolutional neural network model that adaptively weights channel features, effectively enhancing the network's expression of crucial features. Through learning appropriate weights, SENet automatically focuses on the most helpful channel features for the classification task, thereby improving model performance. Due to its ability to adaptively weight channel features, SENet exhibits robustness. Even in the presence of interference or noise within the input data, SENet can correctly identify critical features. The SENet-integrated model is named InteractE-SE, and the model architecture is depicted in Figure 2 below.

The scoring function for the InteractE-SE model is as follows:

$$g(\text{vec}(\text{se}(f(\phi(P_k) \star \omega)))W) \cdot e_o \quad (1)$$

where P_k refers to the randomly initialised embedding, $\phi(P_k)$ refers to the feature replacement of the embedding stacked with P_k (Step 1), \star refers to the circular convolution through which to obtain the feature map (Step 2), and se indicates the execution of SENet on the

obtained features (Step 3), the projection to the multidimensional space after a fully connected layer (Step 4), and matching with the embedding of the candidate target in the inner product layer (Step 5).

The InteractE-SE model, as a decoder, retains the valuable features within the channel information and suppresses the other useless features through SENet. It can thus be determined that InteractE-SE is an excellent CNN-based decoder model.

2.2.3. IntGCN

Due to the early random initialisation of entity and relationship embeddings in InteractE-SE, it remains a CNN-based structural model that does not fully integrate the graph structure information from the knowledge graph. The GCN layer can capture neighbourhood information effectively, while the InteractE-SE model captures high-order interactions between entities and relationships through convolutional operations and cyclic feature matrices. Combining the GCN layer with the InteractE-SE model allows us to leverage their respective advantages, facilitating the learning of more prosperous and expressive representations of entities and relationships. On the one hand, the GCN layer focuses on local structural information between entities, while the InteractE-SE model emphasises the capture of high-order interactions between entities and relationships. By combining the two, we can simultaneously consider high-order interactions and local structural information, thereby improving link prediction accuracy.

Furthermore, the GCN layer and the InteractE-SE model each capture different graph structure features. Integrating them can enhance the model's generalisation ability when dealing with various tasks and datasets, resulting in improved stability and robustness in different scenarios. Therefore, for this study, we decided to fuse the GCN layer as the encoder and InteractE-SE as the decoder. The model structure of IntGCN is illustrated in Figure 3 below.

The scoring function for the IntGCN model is as follows:

$$g(\text{vec}(\text{se}(f(\phi(P_{kc}) \star \omega))W)) \cdot e_o \quad (2)$$

Similarly to the InteractE-SE model, here, P_{kc} refers to an initial embedding that has information about the structure of the graph, $\phi(P_{kc})$ refers to the feature replacement of the embedding stacked with P_{kc} (Step 1), \star refers to the circular convolution through which to obtain the feature map (Step 2), and se indicates the execution of SENet on the obtained features (Step 3), the projection to the multidimensional space after a fully connected layer (Step 4), and matching with the embedding of the candidate target in the inner product layer (Step 5).

Compared to the original InteractE-SE model, the main difference in the IntGCN model is that its initial entity and relationship embeddings already incorporate more graph structure information, which enriches the following interaction information. To better train the model parameters, we employed binary cross-entropy loss with label smoothing as the loss function in this study, as shown in Equation (3):

$$\mathcal{L}(p, t) = -\frac{1}{N} \sum_i (t_i \cdot \log(p_i) + (1 - t_i) \cdot \log(1 - p_i)) \quad (3)$$

where p denotes the score on the fact triple, and t is the smoothed label.

As seen above, IntGCN belongs to an end-to-end model that enhances the accessibility of the graph structure information on top of InteractE-SE.

2.3. Constructing the Tea Knowledge Graph

To address the challenges of varying developmental stages in the agricultural productive service industry across different regions, limited information flow among service supply and demand entities, difficulties in resource allocation for large-scale service operations, subjective measurement of service quality, and crucial aspects of agricultural

production management, this study adopts a combined approach of literature research and investigations within representative regions and industries. Taking tea production in typical counties (districts) in China as a case study, we employed a “bottom-up” method through which we could construct a tea knowledge graph. The primary data sources include platforms such as “China Tea Net” and “Baidu Health”. These prominent platforms encompass an extensive repository of tea-related knowledge and health content. Renowned for their professional and objective content, these platforms wield substantial influence within China. Data extraction predominantly relies on manual curation, serving to mitigate data redundancy and errors to a considerable extent. Given the relatively limited dataset, an additional layer of refinement involves manual disambiguation, facilitated by multiple domain experts within the tea field. The culmination of these efforts results in the formation of a manageable collection of tea triples amenable for storage. The main tea entities include six significant types: green tea, black tea, oolong tea, white tea, yellow tea, and dark tea. The various attributes and relationships among these tea entities encompass tea properties, suitable tea processing methods, value and benefits, characteristics, and places of origin. The ‘bottom-up’ approach to constructing the knowledge graph is shown in Figure 4.

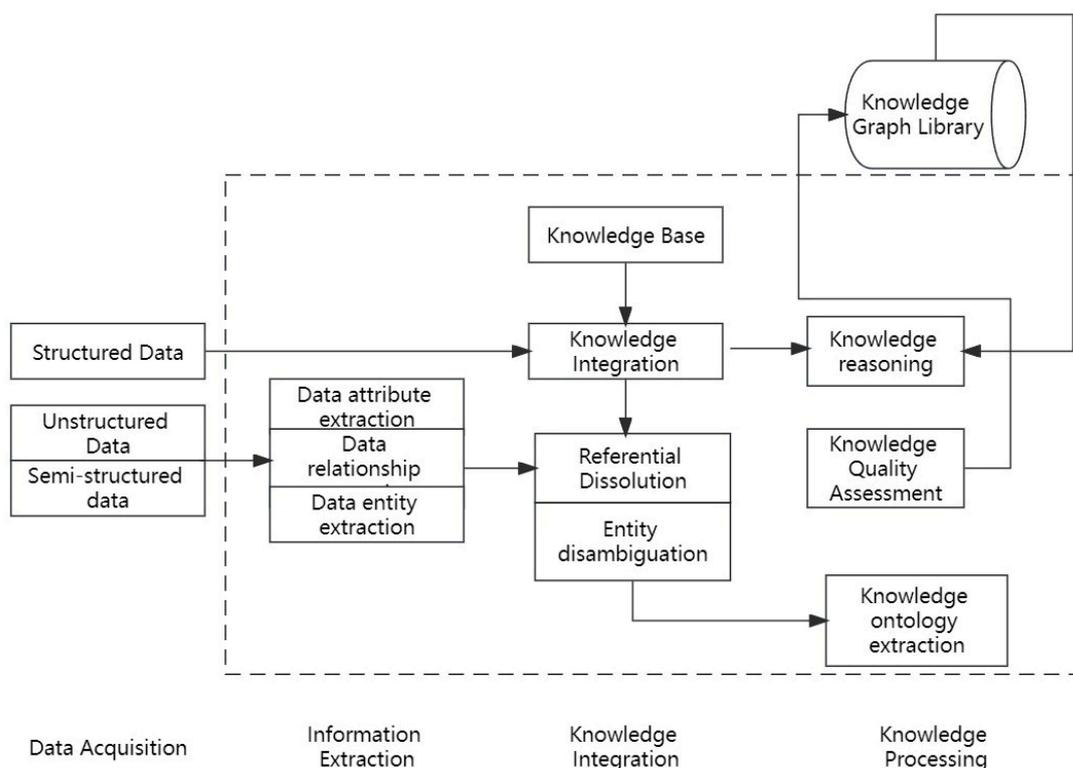


Figure 4. “Bottom-up” approach to tea knowledge graph construction.

After identification by several tea experts, the resulting tea knowledge graph included 6698 records, comprising 330 types of tea divided into six main categories—green tea, green (oolong) tea, white tea, yellow tea, black tea, and dark black tea—with a total of 1064 entities and 29 relationships. This study combines the characteristics and value efficacy of each type of tea. The dataset consists of 12 categories of the more common populations summarised on the “Baidu Health” platform as obese people, frequent smokers, people experiencing feelings of heat or dryness, people experiencing feelings of coldness, people with constipation, people with poor stomach and intestinal health, people who are easily fatigued, people who often drink alcohol, people with greasy diets, people who suffer from the three highs (hypertension, hyperglycaemia, hyperlipidaemia), people with poor immunity, and people who often use computers. Examples are included in Table 1.

Table 1. Sample data for selected properties or relations for the six tea categories.

Entity 1 (Pictures and Tea Names)	Relations (Properties)	Entity 2 or Property Values
 Maolv	Suitable for tea Tea quality Suitable for people Value Effectiveness Propagation method Germination time Characteristics	Green tea Cool Obese people, people experiencing heat/dryness Cooling, slows ageing, weight loss Asexual Early life Leaves with lots of fuzz
 Foxiang No. 4	Suitable for tea Tea quality Suitable for people Value Effectiveness Characteristics Place of origin	(Dark) black tea Hot People who often drink alcohol Slows ageing, promotes digestion, diuretic, relieves fatigue Leaves with lots of fuzz, high yield Yunnan Province
 Jianghua Bitter Tea	Suitable for tea Tea quality Place of origin Propagation method Germination time Characteristics Value Effectiveness Suitable for people	Black tea Hot Jianghua Yao Autonomous County, Hunan Province Asexual Mid-life High yield, leaves with lots of fuzz Promotes digestion, diuretic, relieves fatigue People with constipation
 Almond Tea	Suitable for tea Tea quality Place of origin Propagation method Germination time Characteristics Value Effectiveness Suitable for people	Green (oolong) tea Neutral Jianghua Yao Autonomous County, Hunan Province Asexual Late-life High yield, leaves with less fuzz Slimming and fat loss, slows ageing People who are easily fatigued
 Fuding Great White Tea	Suitable for tea Tea quality Place of origin Propagation method Germination time Characteristics Value Effectiveness Suitable for people	White tea Cool Dutou Town, Fuding City, Fujian Province Asexual Early birth High yield, cold resistant Antidiarrhoeal, germicidal People with poor immunity
 Junshanyinzen	Suitable for tea Tea quality Place of origin Category Characteristics Value Effectiveness Suitable for people	Yellow tea Cool Dongting Lake, Yueyang, Hunan Province Yellow tea Resembles silver needles Cooling, relieves fatigue People who often use computers

Images obtained from the “Baidu Baike” platform (China’s popular knowledge-sharing platform; link: <https://baike.baidu.com>, accessed on 20 June 2023).

Additionally, in order to clearly and visually display our constructed knowledge graphs, we have selected some of the stored knowledge graphs for visualisation, as shown in Figure 5.

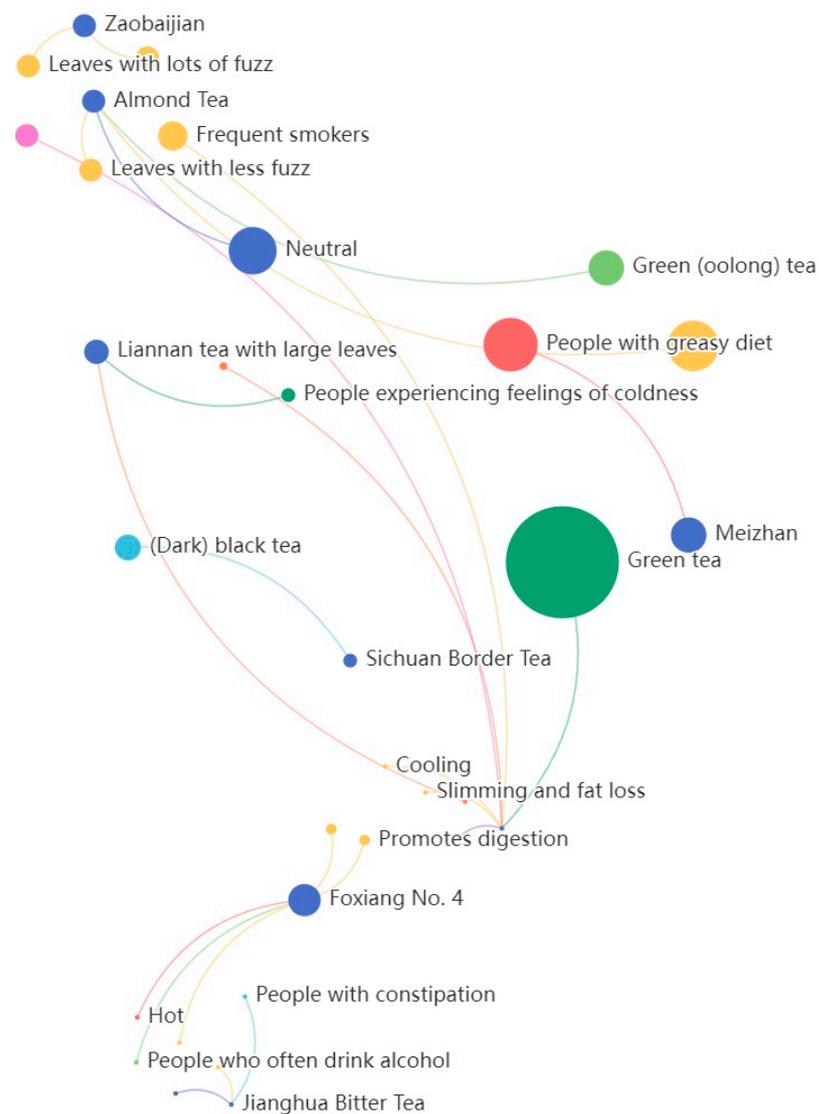


Figure 5. Visualisation of the tea knowledge graph.

2.4. Experimental Method Design

2.4.1. Dataset and Evaluation Metrics

This study first evaluated the performance of the IntGCN model on the publicly available datasets WN18RR [31] and Kinship. Subsequently, the model was applied to the tea dataset (ID_Tea) to predict the “suitable for people” relationships. Experiments were conducted on three datasets in total. The tea dataset was randomly divided into training, validation, and test sets in an 8:1:1 ratio. A summary statistics of the datasets is presented in Table 2. The IntGCN model was employed to predict suitable populations for tea. This study employed several standard link prediction task metrics through which to evaluate the model’s performance, namely H@k, MR, and MRR. H@k refers to the average proportion of triplets with a rank less than k in the link prediction task. It is used to signify how many of the top k predicted results are correct. Typically, different values of k can be used, such as H@1, H@3, or H@10. MR (mean rank) is used to evaluate the average rank of the predicted relationship within the entire candidate set. For each test sample, the rank of the correct relationship in the predictions is identified, and the average rank across all samples is computed. MRR (mean reciprocal ranking) is an indicator for assessing the ranking capability of the link prediction model, focusing on the rank of the correct relationship in predictions. MRR is used to calculate the reciprocal rank of the correct relationship for each test sample and then compute the average reciprocal rank across all

samples. The MRR value lies between 0 and 1. A lower MR value, indicating that the model can accurately place the correct relationships at the forefront, is preferred. Conversely, higher values of H@k and MRR are desirable, as these reflect the improved performance of the link prediction model. The formulae for H@k, MR, and MRR are presented below:

$$H@k = \frac{1}{|N|} \sum_{i=1}^{|N|} \mathbb{I}(\text{rank}_i \leq k) \quad (4)$$

$$MR = \frac{1}{|N|} \sum_{i=1}^{|N|} \text{rank}_i \quad (5)$$

$$MRR = \frac{1}{|N|} \sum_{i=1}^{|N|} \frac{1}{\text{rank}_i} \quad (6)$$

where N denotes the set of test triples, and $|N|$ denotes the number of test triples. rank_i refers to the link prediction ranking of the i -th triple, while \mathbb{I} denotes the INDICATOR function (which has a value of 1 if $\text{rank}_i \leq k$ is accurate, and a value of 0 otherwise).

Table 2. Statistics for the three datasets.

Dataset	Entities	Relations	Train	Validation	Test
WN18RR	40943	11	86835	3034	3134
Kinship	104	25	8544	1068	1074
ID_Tea	1064	29	5368	665	665

2.4.2. Training Environment and Parameter Settings

For all experiments, training was performed using the Adam optimiser [32], and the parameters were initialised using Xavier initialisation [33]. The model training environment utilised had an RTX 2080 Ti graphics card, Pytorch 1.6.0 framework, and Python 3.8. Combining the InteractE model parameters with those in Table 3, we used a grid search and trained 500 rounds to find the best set of hyperparameters for this study, and the optimal hyperparameters for the utilised model were selected based on the MRR of the validation set; for the WN18RR and ID_Tea datasets, the optimal parameters were as follows: lr = 0.001, batch = 256, k = 11, q = 8, d = 1; for the Kinship dataset, the optimal parameters were as follows: lr = 0.005, batch = 256, k = 11, q = 8, d = 0.95.

Table 3. Details of the hyperparameters.

Hyperparameter	Values
Learning rate (lr)	{0.0001, 0.001, 0.005}
Batch size (batch)	{128, 256}
Convolutional kernel size (k)	{3, 5, 7, 9, 11}
Dimensional reduction setting for SENet (q)	{4,8,16}
Learning rate decay (d)	{1,0.95}

2.4.3. Transfer Learning

Applying transfer learning in link prediction tasks can improve model performance, generalisation ability, and training efficiency through leveraging existing knowledge and data resources. In Section 2, we constructed a knowledge graph for various types of tea and their corresponding target populations. To fully realise the prediction of the “suitable population” relationship in tea, we employed transfer learning methods in order to transfer the knowledge learnt from the link prediction task on the WN18RR dataset to the ID_Tea dataset.

The IntGCN model uses transfer learning to apply the features learnt from the link prediction task on the WN18RR dataset to the link prediction task on the ID_Tea dataset. By

utilising the well-trained weights of the IntGCN model for the WN18RR dataset, we further enhanced the prediction of suitable populations for different tea variations. Although the WN18RR and ID_Tea datasets belong to different domains, they share similarities and commonalities. Since the ultimate goal of both tasks is link prediction, transfer learning can effectively leverage these similarities and commonalities, thereby improving the model's generalisation ability and performance compared to not using transfer learning methods.

3. Results

3.1. Evaluation of Public Datasets

In this segment of the study, we focus on the performance comparison of the proposed IntGCN model with the original InteractE model and some standard base models on public datasets (WN18RR, Kinship), as well as the ablation evaluation of the IntGCN model.

3.1.1. Comparison of Link Prediction Performance

One purpose of this study was to evaluate the performance and plausibility of the proposed model by comparing the results with several existing methods developed in recent years. The baseline mainly includes non-neural network models such as TransE [34], DistMult [35], and ComplEx [36] and neural network models such as R-GCN, ConvTransE [37], SACN [37], ConvE, CompGCN, and InteractE.

As can be observed from Table 4, the IntGCN and InteractE + SE models significantly outperformed the InteractE model in all indicators for the Kinship and WN18RR datasets. Therefore, our proposed algorithms have some superiority and can be used as predictive models for tea suitability population prediction.

Table 4. Comparison of the link prediction performance of the improved InteractE models with several recent models for the Kinship and WN18RR datasets.

Model	Kinship				WN18RR			
	MRR	MR	H@10	H@1	MRR	MR	H@10	H@1
TransE	0.309	6.8	0.841	0.009	0.226	3384	0.501	-
DistMult	0.516	5.26	0.867	0.367	0.430	5110	0.490	0.390
ComplEx	0.823	2.48	0.971	0.733	0.440	5216	0.510	0.410
R-GCN	0.109	25.92	0.239	0.030	-	-	-	-
KBGAN	0.165	-	0.347	-	0.214	-	0.472	-
ConvTransE	0.824	2.53	0.972	0.734	0.460	-	0.520	0.430
SACN	0.759	3.25	0.951	0.643	0.470	-	0.540	0.430
ConvE	0.833	2.03	0.981	0.738	0.430	4187	0.520	0.400
CompGCN	0.840	2.10	0.982	0.753	0.469	3307	0.536	0.434
InteractE	0.806	2.32	0.974	0.706	0.463	5202	0.528	0.430
Interact-SE	0.810	2.31	0.974	0.716	0.467	4900	0.530	0.436
IntGCN	0.858	1.93	0.983	0.782	0.474	3533	0.542	0.438

The bold-formatted values and those denoted as—represent the best and missing scores, respectively.

As shown in Figure 6, we conducted a visual analysis of the IntGCN and InteractE-SE models, and all metrics of IntGCN relative to InteractE-SE on both public datasets improved relative to the original InteractE model. In particular, we observed an improvement of 1.4% in H@10 for the WN18RR dataset, 7.6% in H@1 for the Kinship dataset, and a 5.2% improvement in MRR.

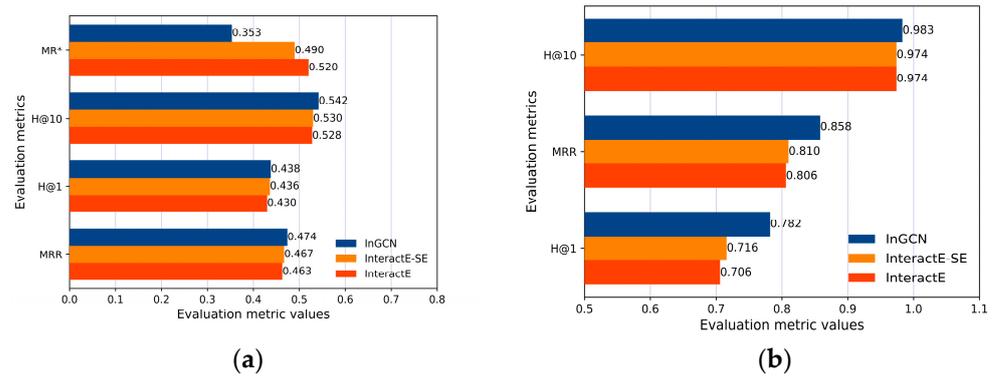


Figure 6. IntGCN comparative analysis on public datasets (MR* indicates MR/10000). (a) Comparison of results for the WN18RR dataset. (b) Comparison of results for the Kinship dataset.

3.1.2. Ablation Evaluation

This study introduced the GCN layer and SENet for InteractE separately (see Section 2.2 for details). The changes in the metrics before and after adding the GCN layer and SENet to the InteractE model for the WN18RR and Kinship datasets, respectively, are shown in Tables 5 and 6.

Table 5. Comparison of the link prediction performance of InteractE with and without GCN layers or SENet for the WN18RR dataset.

Model	GCN Layer	SENet	WN18RR			
			MRR	MR	H@10	H@1
InteractE	No	No	0.463	5202	0.528	0.430
	No	Yes	0.467	4900	0.530	0.436
	Yes	No	0.472	3266	0.540	0.437
	Yes	Yes	0.474	3533	0.542	0.438

The bold-formatted values represent the best scores.

Table 6. Comparison of the link prediction performance of InteractE with and without GCN layers or SENet for the Kinship dataset.

Model	GCN Layer	SENet	Kinship			
			MRR	MR	H@10	H@1
InteractE	No	No	0.806	2.32	0.974	0.706
	No	Yes	0.810	2.31	0.974	0.716
	Yes	No	0.844	2.06	0.982	0.757
	Yes	Yes	0.858	1.93	0.983	0.782

The bold-formatted values represent the best scores.

The results presented in Tables 5 and 6 show that, for public datasets (WN18RR, Kinship), the IntGCN model can significantly improve the link prediction performance of the original model by fusing the SENet and GCN layers on top of the original InteractE model.

3.2. Evaluation of ID_Tea Dataset

In this section of the study, we focus on the performance comparison of the proposed transfer learning-enhanced IntGCN model with the original InteractE model and some standard neural network models for the ID_Tea dataset, as well as the ablation evaluation of the IntGCN model.

3.2.1. Comparison of Link Prediction Performance

For this section, we applied the IntGCN model to the ID_Tea dataset for link prediction. We compared the experimental results using the metrics MRR and H@k to evaluate

the performance of the IntGCN model. Table 7 shows the comparison between the link prediction performance of the IntGCN model and those of some recent state-of-the-art neural network models. In Table 7, 'noTransfer' indicates that no transfer learning method was used.

Table 7. Comparison of the results of IntGCN with those of common neural network models for the ID_Tea dataset.

Model	MRR (%)	H@1 (%)	H@3 (%)	H@10 (%)
ConvE	56.4	47.6	61.4	72.9
CompGCN	60.2	53.4	63.3	72.7
InteractE	59.6	52.9	61.7	72.5
InteractE-SE	60.0	53.8	62.2	72.5
IntGCN(noTransfer)	61.3	54.3	64.9	74.2
IntGCN	61.6	54.4	65.8	75.0

From Table 7, it can be observed that the IntGCN model, with the application of transfer learning, demonstrates better link prediction performance than other methods for the ID_Tea dataset.

3.2.2. Ablation Evaluation

This study introduced the GCN layer and SENet for InteractE separately (see Section 2.2 for details). The changes in the metrics before and after adding the GCN layer and SENet to the InteractE model for the ID_Tea dataset are shown in Table 8.

Table 8. Comparison of the link prediction performance of InteractE with and without GCN layers or SENet for the ID_Tea dataset.

Model	GCN Layer	SENet	MRR (%)	H@1 (%)	H@3 (%)	H@10 (%)
InteractE	No	No	59.6	52.9	61.7	72.5
	No	Yes	60.0	53.8	62.2	72.5
	Yes	No	61.0	54.2	64.2	74.0
	Yes	Yes	61.3	54.3	64.9	74.2

The results from Table 8 show better link prediction performance for the ID_Tea dataset using the IntGCN model after integrating SENet and the GCN Layer.

3.2.3. Relationship Prediction and Knowledge Graph Completion

The tea knowledge graph is utilised to rank all entities related to tea in order to predict the suitable populations for different types of tea. By predicting the missing relationship in a given triplet (tea, relation, people), the suitability of tea for specific populations can be determined. The experimental results listed in Table 7 indicate that, among the top 10 predicted target entities, the average probability of the target entity appearing is 75%.

Finally, 200 prediction triples, which were randomly combined from each type of tea and each type of suitable population and were independent of the available data, were selected for prediction in this study. The scores of each triple were calculated using Equation (2) in Section 2.2.3, and the scores of these 200 prediction triples are shown in Figure 7.

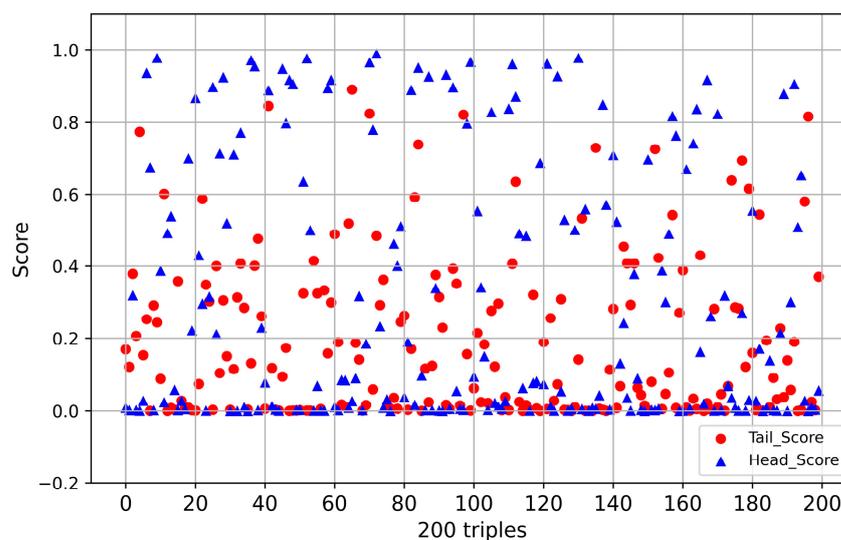


Figure 7. Scores of the 200 target triples.

Where Head_Score and Tail_Score refer to the scores of the fact triple when predicting the head entity based on the relation and the tail entity, and the scores of the fact triple when predicting the tail entity based on the head entity and the relation, respectively, taking into consideration the feature that there are more tea types than suitable types of people in the tea knowledge graph. By comparing the “Head_Score” and “Tail_Score” in a realistic situation to determine which entity pair is more likely to constitute the correct triad, this study uses the “Head_Score” as the final score for the triad of the tea complementary knowledge graph. After combining several experiments with reality, we set the score threshold to 0.9. The target triad with a score greater than or equal to 0.9 was taken as the new triplet, which was subsequently added to the original tea knowledge graph for completion, and at the same time, we also obtained the “suitable for people” relationship. Some of the link prediction triples and their scores are shown in Table 9.

Table 9. Prediction triples and their scores.

Prediction Triples	Scores
(Zaobaijian, suitable for people, obese people)	0.952
(Foxiang No. 4, suitable for people, people who suffer from three highs)	0.977
(Jianghua Bitter Tea, suitable for people, people experiencing feelings of coldness)	0.965
(Almond Tea, suitable for people, people with greasy diet)	0.941
(Foshou, suitable for people, people experiencing feelings of heat and dryness)	0.709
...	...

As can be seen by comparing Figure 8a,b, the link prediction experiments of the IntGCN model, followed by the scoring function through which the predicted triple scores are obtained, play a crucial role in determining the “suitable for people” relationship of the tea knowledge graph to be predicted. This process assists in the decision-making process for research studies aiming to assess tea’s impact on human health, make tea recommendations, and realise the accurate matching of supply and demand for social tea production services.

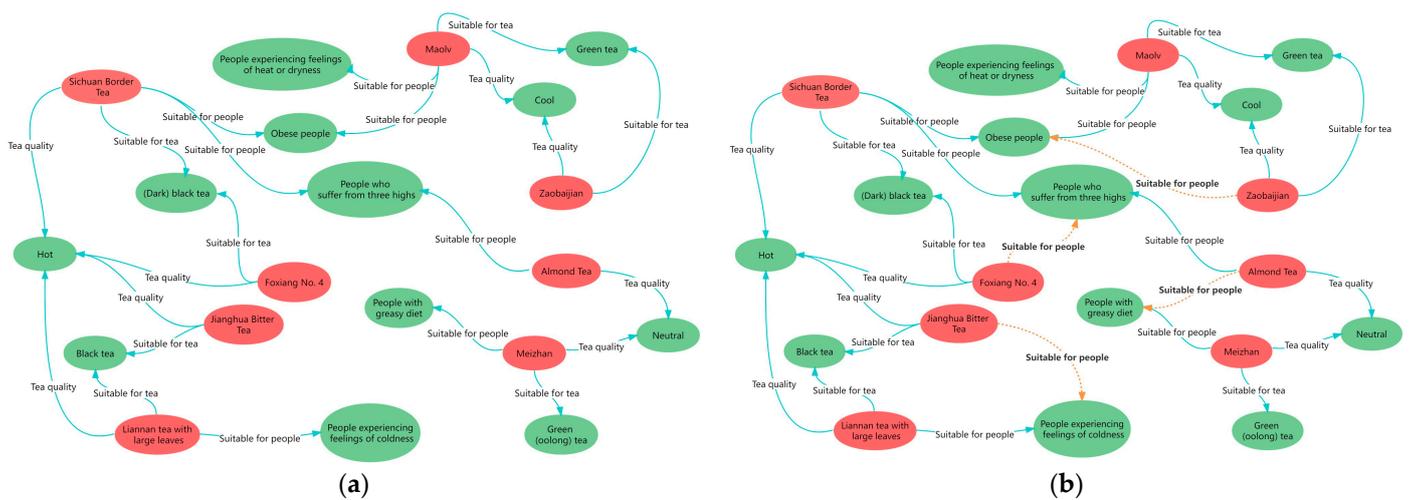


Figure 8. Updated tea knowledge graph after link prediction. (a) Initial partial tea knowledge graph. (b) Selected tea knowledge graphs updated after link prediction.

4. Discussion

In this section, the experimental results obtained in this study are discussed and compared with the results of other studies. Some of the excellent work on tea includes the following publications: P Olha et al. [38] studied the positive antioxidant and anti-diabetic effects of tea polyphenols on tea drinkers, P Chen et al. [39] conducted a study to analyse the cultivation suitability of tea trees, J Ye et al. [40] studied the effects of applying different amounts of organic fertilisers on tea yield and quality, and XQ Zheng et al. [41] studied the effects of tea plant chemical composition. Unlike previous work, this study used a link prediction algorithm to link multiple types of tea and tea-suitable people, combining some of the characteristics of tea and using an improved link prediction algorithm for the predictive analysis of tea-suitable people.

The original InteractE model lacked the ability to capture graph structural information. To address this, we improved the InteractE model by incorporating SENet and GCN layers, resulting in the IntGCN model (detailed in Section 2.2). We conducted link prediction experiments on public datasets (WN18RR and Kinship) using the IntGCN model and observed significant improvements in MRR and H@k. Specifically, H@10 increased by 1.4% for the WN18RR dataset and by 7.6% for the Kinship dataset, while MRR improved by 5.2% (see Section 3.1). These results convincingly demonstrate the superior performance of the proposed IntGCN model.

In addition, more research is needed on the relationships among the six major types of tea and their suitability for different populations, presenting an opportunity for further exploration. In this study, we constructed a tea dataset (ID_Tea) comprising 6698 records, 330 tea types, and 29 relationship types, focusing on green tea, green (oolong) tea, white tea, yellow tea, black tea, and dark black tea. The improved model introduced in Section 2.2 was applied to predict the “suitable for people” relationship in the ID_Tea dataset. Notably, the performance improvements varied when the GCN layer and SENet were individually applied to the initial InteractE model (see Sections 3.2.1 and 3.2.2).

When the training dataset is small or lacks diversity, transfer learning can leverage the abundant data from the source task for pre-training and then transfer the learned knowledge to the target task, thereby enhancing performance for the target task. In this study, we utilised the feature information learnt by the IntGCN model from the WN18RR dataset for link prediction experiments on the ID_Tea dataset. The experimental results revealed improvements in all prediction metrics when using transfer learning compared to training from scratch. The results prove the effectiveness of introducing transfer learning to the IntGCN model in improving its generalisation ability, accelerating the learning process for the target task, and saving computational resources. Lastly, we supplemented the

knowledge graph by incorporating triplets with scores equal to or above a threshold (see Section 3.2.3).

5. Conclusions

In the first step of this study, we introduced an improved version of the InteractE model, the IntGCN model. It is important to note that the IntGCN model further incorporates graph structural information through the GCN layer and preserves more useful feature information through SENet. For the task of link prediction, the IntGCN model demonstrates better performance in predicting links. Additionally, we constructed a knowledge graph for tea by combining the IntGCN model with transfer learning and applying it to predict suitable populations for different types of tea. Finally, we completed the tea knowledge graph based on the prediction results. Experimental results indicate the feasibility of applying this model, combined with a knowledge graph, to predict the suitable populations for different types of tea. This approach provides valuable insights for applying knowledge graphs in the tea field and helps us to fully explore the value potential of tea varieties.

In the future, we will aim to collect a wider variety of textual data on tea varieties based on the existing data. Future research will enable a more granular categorisation of suitable populations for different types of tea, facilitating a more comprehensive investigation that combines link prediction methods with tea research. Furthermore, we seek to expand the application of link prediction techniques to various other domains in agriculture, including the following:

- Crop–soil adaptability prediction: By constructing knowledge graphs for crops and soils and leveraging link prediction algorithms, we can forecast the adaptability relationships between different crops and soils. This would aid farmers in selecting the most suitable crops for cultivation and optimising soil management strategies.
- Agricultural product quality assessment: By constructing knowledge graphs for agricultural products, link prediction algorithms can forecast these products' quality characteristics and relevant attributes. For instance, they could predict fruit ripeness or the nutritional values of agricultural products, thereby assisting farmers and consumers in making informed decisions.
- Agricultural disease prediction: By constructing a knowledge graph that connects crops, diseases, and environmental conditions, it is possible to utilise link prediction algorithms to predict the probability of crops being affected by specific diseases. This approach can assist farmers in taking timely preventive measures and reducing damage to their crops caused by diseases. A well-designed and adequately implemented agricultural disease prediction system could significantly impact crop yields and the agricultural industry.
- Optimisation of agricultural supply chains: By constructing knowledge graphs for agricultural supply chains, link prediction algorithms can predict partner relationships, resource allocation, and the feasibility of transactions at various stages. This would optimise the agricultural supply chain's operational efficiency and profit distribution.

Furthermore, the application of link prediction algorithms can extend beyond agriculture to other domains, such as predicting the likelihood of medical accidents or forecasting failures in vehicles. These applications demonstrate the versatility of link prediction techniques and their potential impacts across various practical fields.

Author Contributions: Conceptualisation, Q.H., Z.W. and M.W.; data curation, Y.T. and Z.W.; formal analysis, M.W. and Q.H.; funding acquisition, Q.H.; investigation, Z.W. and Q.H.; methodology, Q.H. and Y.H.; project administration, M.W. and Q.H.; resources, Q.H. and M.W.; software, Z.W.; supervision, Y.H. and F.M.; validation, Y.H. and Z.W.; visualisation, Y.T. and Z.W.; writing—original draft, Z.W. and Q.H.; writing—review and editing, Q.H., Z.W. and F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Research on the Application of Spatio-temporal Big Data Analysis in Agricultural Production Services (Sichuan Provincial Finance Independent Innovation Special Project, grant number 2022ZZCX034).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The tea dataset constructed for this study is available at “https://github.com/5zongyuan/ID_Tea (accessed on 20 June 2023)” and can be shared upon request.

Acknowledgments: Thanks to Minglin He, Siqi Liu, and Luyu Shuai for their advice throughout the research process.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shen, W.; Xiao, Y.; Ying, X.; Li, S.; Zhai, Y.; Shang, X.; Li, F.; Wang, X.; He, F.; Lin, J. Correction: Tea consumption and cognitive impairment: A cross-sectional study among Chinese elderly. *PLoS ONE* **2015**, *10*, e0140739. [CrossRef] [PubMed]
2. Tang, G.-Y.; Meng, X.; Gan, R.-Y.; Zhao, C.-N.; Liu, Q.; Feng, Y.-B.; Li, S.; Wei, X.-L.; Atanasov, A.G.; Corke, H. Health functions and related molecular mechanisms of tea components: An update review. *Int. J. Mol. Sci.* **2019**, *20*, 6196. [CrossRef]
3. Sae-Tan, S.; Grove, K.A.; Lambert, J.D. Weight control and prevention of metabolic syndrome by green tea. *Pharmacol. Res.* **2011**, *64*, 146–154. [CrossRef] [PubMed]
4. Schönthal, A.H. Adverse effects of concentrated green tea extracts. *Mol. Nutr. Food Res.* **2011**, *55*, 874–885. [CrossRef] [PubMed]
5. Pan, S.-Y.; Nie, Q.; Tai, H.-C.; Song, X.-L.; Tong, Y.-F.; Zhang, L.-J.-F.; Wu, X.-W.; Lin, Z.-H.; Zhang, Y.-Y.; Ye, D.-Y. Tea and tea drinking: China’s outstanding contributions to the mankind. *Chin. Med.* **2022**, *17*, 27. [CrossRef]
6. Mahdavi-Roshan, M.; Salari, A.; Ghorbani, Z.; Ashouri, A. The effects of regular consumption of green or black tea beverage on blood pressure in those with elevated blood pressure or hypertension: A systematic review and meta-analysis. *Complement. Ther. Med.* **2020**, *51*, 102430. [CrossRef]
7. Yan, W.; Ge, Z. Research on Winter Tea Application and Promotion Value. *Mod. Econ.* **2020**, *11*, 817–828. [CrossRef]
8. Lee, J.; Kang, S. Consumer-driven usability test of mobile application for tea recommendation service. *Appl. Sci.* **2019**, *9*, 3961. [CrossRef]
9. Chen, C.; Zhang, W.; Shan, Z.; Zhang, C.; Dong, T.; Feng, Z.; Wang, C. Moisture contents and product quality prediction of Pu-erh tea in sun-drying process with image information and environmental parameters. *Food Sci. Nutr.* **2022**, *10*, 1021–1038. [CrossRef] [PubMed]
10. Hogan, A.; Blomqvist, E.; Cochez, M.; d’Amato, C.; Melo, G.d.; Gutierrez, C.; Kirrane, S.; Gayo, J.E.L.; Navigli, R.; Neumaier, S. Knowledge graphs. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 71. [CrossRef]
11. Xiaoxue, L.; Xuesong, B.; Longhe, W.; Bingyuan, R.; Shuhan, L.; Lin, L. Review and trend analysis of knowledge graphs for crop pest and diseases. *IEEE Access* **2019**, *7*, 62251–62264. [CrossRef]
12. Chen, Y.; Kuang, J.; Cheng, D.; Zheng, J.; Gao, M.; Zhou, A. AgriKG: An agricultural knowledge graph and its applications. In Proceedings of the Database Systems for Advanced Applications: DASFAA 2019 International Workshops: BDMS, BDQM, and GDMA, Chiang Mai, Thailand, 22–25 April 2019; pp. 533–537. [CrossRef]
13. Rossi, A.; Barbosa, D.; Firmani, D.; Matinata, A.; Merialdo, P. Knowledge graph embedding for link prediction: A comparative analysis. *ACM Trans. Knowl. Discov. Data (TKDD)* **2021**, *15*, 1–49. [CrossRef]
14. Wang, M.; Qiu, L.; Wang, X. A survey on knowledge graph embeddings for link prediction. *Symmetry* **2021**, *13*, 485. [CrossRef]
15. Dettmers, T.; Minervini, P.; Stenetorp, P.; Riedel, S. Convolutional 2D knowledge graph embeddings. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018. [CrossRef]
16. Che, F.; Zhang, D.; Tao, J.; Niu, M.; Zhao, B. Parame: Regarding neural network parameters as relation embeddings for knowledge graph completion. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 2774–2781. [CrossRef]
17. Vashishth, S.; Sanyal, S.; Nitin, V.; Agrawal, N.; Talukdar, P. Interact: Improving convolution-based knowledge graph embeddings by increasing feature interactions. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 3009–3016. [CrossRef]
18. Schlichtkrull, M.; Kipf, T.N.; Bloem, P.; Van Den Berg, R.; Titov, I.; Welling, M. Modeling relational data with graph convolutional networks. In Proceedings of the Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Greece, 3–7 June 2018; pp. 593–607. [CrossRef]
19. Nathani, D.; Chauhan, J.; Sharma, C.; Kaul, M. Learning attention-based embeddings for relation prediction in knowledge graphs. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019. [CrossRef]
20. Zeb, A.; Saif, S.; Chen, J.; Haq, A.U.; Gong, Z.; Zhang, D. Complex graph convolutional network for link prediction in knowledge graphs. *Expert Syst. Appl.* **2022**, *200*, 116796. [CrossRef]
21. Vashishth, S.; Sanyal, S.; Nitin, V.; Talukdar, P. Composition-based multi-relational graph convolutional networks. *arXiv* **2019**, arXiv:1911.0308.

22. McCoy, K.; Gudapati, S.; He, L.; Horlander, E.; Kartchner, D.; Kulkarni, S.; Mehra, N.; Prakash, J.; Thenot, H.; Vanga, S.V. Biomedical text link prediction for drug discovery: A case study with COVID-19. *Pharmaceutics* **2021**, *13*, 794. [[CrossRef](#)] [[PubMed](#)]
23. Huo, Z.; Huang, X.; Hu, X. Link prediction with personalized social influence. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018. [[CrossRef](#)]
24. Nasiri, E.; Berahmand, K.; Samei, Z.; Li, Y. Impact of centrality measures on the common neighbors in link prediction for multiplex networks. *Big Data* **2022**, *10*, 138–150. [[CrossRef](#)] [[PubMed](#)]
25. Shabaz, M.; Garg, U. Predicting future diseases based on existing health status using link prediction. *World J. Eng.* **2022**, *19*, 29–32. [[CrossRef](#)]
26. Nasiri, E.; Berahmand, K.; Rostami, M.; Dabiri, M. A novel link prediction algorithm for protein-protein interaction networks by attributed graph embedding. *Comput. Biol. Med.* **2021**, *137*, 104772. [[CrossRef](#)]
27. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
28. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. *Proc. IEEE* **2020**, *109*, 43–76. [[CrossRef](#)]
29. Ma, X. Knowledge graph construction and application in geosciences: A review. *Comput. Geosci.* **2022**, *161*, 105082. [[CrossRef](#)]
30. Cheng, S.; Wang, T.; Guo, X.; Wang, Y. Knowledge Graph construction of Thangka icon characters based on Neo4j. In Proceedings of the 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), Sanya, China, 4–6 December 2020; pp. 218–221. [[CrossRef](#)]
31. Lin, X.V.; Socher, R.; Xiong, C. Multi-hop knowledge graph reasoning with reward shaping. *arXiv* **2018**, arXiv:1808.10568.
32. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
33. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
34. Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; Yakhnenko, O. Translating embeddings for modeling multi-relational data. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 2787–2795.
35. Yang, B.; Yih, W.-t.; He, X.; Gao, J.; Deng, L. Embedding entities and relations for learning and inference in knowledge bases. *arXiv* **2014**, arXiv:1412.6575.
36. Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; Bouchard, G. Complex embeddings for simple link prediction. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 20–22 June 2016; pp. 2071–2080.
37. Shang, C.; Tang, Y.; Huang, J.; Bi, J.; He, X.; Zhou, B. End-to-end structure-aware convolutional networks for knowledge base completion. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 3060–3067. [[CrossRef](#)]
38. Olcha, P.; Winiarska-Mieczan, A.; Kwiecień, M.; Nowakowski, Ł.; Miturski, A.; Semczuk, A.; Kiczorowska, B.; Gałczyński, K. Antioxidative, anti-inflammatory, anti-obesogenic, and antidiabetic properties of tea polyphenols—The positive impact of regular tea consumption as an element of prophylaxis and pharmacotherapy support in endometrial cancer. *Int. J. Mol. Sci.* **2022**, *23*, 6703. [[CrossRef](#)] [[PubMed](#)]
39. Chen, P.; Li, C.; Chen, S.; Li, Z.; Zhang, H.; Zhao, C. Tea Cultivation Suitability Evaluation and Driving Force Analysis Based on AHP and Geodetector Results: A Case Study of Yingde in Guangdong, China. *Remote Sens.* **2022**, *14*, 2412. [[CrossRef](#)]
40. Ye, J.; Wang, Y.; Kang, J.; Chen, Y.; Hong, L.; Li, M.; Jia, Y.; Wang, Y.; Jia, X.; Wu, Z. Effects of Long-Term Use of Organic Fertilizer with Different Dosages on Soil Improvement, Nitrogen Transformation, Tea Yield and Quality in Acidified Tea Plantations. *Plants* **2023**, *12*, 122. [[CrossRef](#)] [[PubMed](#)]
41. Zheng, X.-Q.; Dong, S.-L.; Li, Z.-Y.; Lu, J.-L.; Ye, J.-H.; Tao, S.-K.; Hu, Y.-P.; Liang, Y.-R. Variation of Major Chemical Composition in Seed-Propagated Population of Wild Cocoa Tea Plant *Camellia ptilophylla* Chang. *Foods* **2023**, *12*, 123. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.