ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Computational Linguistic Interpretation and Modeling of Dress Metaphors in Ming Dynasty Novel Texts

Menghe Tian[1] *and Xiangyang Bian*[1,*]

[1] Collage of Fashion and Design, Donghua University, Shanghai, 200050, China

SUMMARY: *Although metaphor has long been recognized as a rhetorical pattern, it is in fact more of a cognitive device and a way of thinking that plays an indispensable role in enriching the form and connotation of language. The study extracts the characteristics of characters' costumes in the novel text by utilizing the CRF method, and applies the Jieba participle and deactivation lexicon to clean the text with data. TF-IDF algorithm, graph theory and social network analysis, LDA topic modeling and sentiment analysis are used to construct a multi-level data mining method. Taking the Ming Dynasty novel Jin Ping Mei as the research object, high-frequency feature words are extracted from the text of the novel, the relationship network of the core characters of the novel is analyzed, and the LDA theme model is constructed to comprehensively excavate the metaphorical themes of the novel's costumes, and four themes are found: the costumes of the noble women of the powerful families, the daily costumes of the women of the city, the costumes of the male government officials and businessmen, and the costumes of the festivals and special occasions, so as to achieve the visualization of the costume metaphors and the depth of interpretation. We realized the visualization and in-depth interpretation of clothing metaphors.*

KEYWORDS: *LDA theme; social network analysis; visualization analysis; theme mining; Ming Dynasty novels*

1 Introduction

Metaphor is a very important form of expression that has a wide range of applications in literature, art and design. Metaphor is literally understood as a rhetorical device. In relation to the clothing metaphor, it can be understood as the emotional expression of the design of clothing to play a certain role, that is, for the clothing itself, without other external design of the clothing itself, leaving aside the fabric modification, structural division and the use of color, etc., it is a simple monolithic type of the clothing world itself, which is called the “clothing ontology” [1-4]. For clothing metaphor, two points are particularly important, one is the aesthetics of clothing design; the other is the meaning of clothing metaphor. A deep understanding of the meaning, stance, creativity and concept of metaphorical design can enable designers to better devote themselves to the design field. Through the study of traditional clothing metaphors, the specific meaning and implication of the metaphors can be better understood, so that the national culture can be continuously improved [5].

The Ming Dynasty, as the most perfect dynasty of Han Chinese public opinion clothing system, can be studied in depth to understand in detail the cultural and aesthetic characteristics of Chinese national history in terms of clothing [6, 7]. The important function of the Ming

*xiangyanbian@163.com
https://doi.org/10.65102/is2026238
Published: 30 April, 2026

Dynasty costumes was to maintain the hierarchical social order. As early as in the Qin and Han Dynasties, China used crowns and ornaments to indicate official rank [8]. As for the Ming Dynasty, the crown clothes of the officials were divided into three kinds of official clothes, regular clothes, court dress, and all of them had the provisions of the material of the clothes, the background pattern, the flowers, the size, etc., and the top of the hat, the cap beads, the beam crowns, the corded brocade, the Ge Dai, the wutu boards to indicate the grade level, and to follow the way of the military officials embroidered beasts, the civil officials embroidered birds, were strictly regulated [9, 10]. Consort, the fate of women's clothing is no exception, from one product to nine products, the official rank is different, the difference between noble and lowly, clear at a glance. This extremely standardized Ming dress system, its social identity labeling function can be seen. Therefore, the Ming dress metaphor covers identity, culture, morality and other aspects. As one of the main forms of Ming Dynasty cultural carriers, Ming Dynasty novels have certain identity, cultural, and moral significance embedded in the dress costumes of the characters [11]. Therefore, novels have also gradually become the main body of research on dress metaphors of different dynasties, providing reference materials for the study of Ming Dynasty dress.

In this paper, the famous novel “Jin Ping Mei” of the Ming Dynasty is selected as the research object, and the text data integration and processing of the novel text is carried out, the CRF method is used to extract the characters' costumes in the novel text, and the Jieba participle technology is applied to clean the text data. On the basis of TF-IDF algorithm, graph theory and network analysis, and LDA theme model, the relationship between the characters' costumes and the social environment of the Ming Dynasty in the novel text is visualized and analyzed, and R software is used to establish the LDA theme model and extract the theme keywords.

2 Acquisition and Processing of Multidimensional Text Data for Ming Dynasty Novels

2.1 Natural language processing techniques

2.1.1 Named entity identification

Named Entity Recognition NER word analysis of text extracts entities such as names of people, places, times, proper nouns, etc., and is also able to recognize other types of entities. In chapter book novels, named entity technology is a very effective method of text analysis, often used to reveal information related to specific characters in the text. The entity extraction can directly extract the characters and other identifiers in the text of chapter book novels, and it can also be used to analyze the important characters and their relationships. Common models include: Hidden Markov Model, Support Vector Machine, Conditional Random Field, etc. The Conditional Random Field open source model CRF is used.

CRF is a sequence labeling algorithm commonly used for labeling and analyzing sequence data, and belongs to a type of random field model. In this case, the undirected graph $G=(V,E)$, given the observations as random variables X gets the conditional probability distribution Y. Here upper case X is used and Y denotes the sequence. Assume that the input sequence $X=(x_1,x_2,\ldots,x_n)$ and the output target sequence $Y=(y_1,y_2,\ldots,y_n)$. where V denotes the set of vertices and E denotes the set of edges, and a conditional random field is formed by random variables X and Y.

2.1.2 Text pre-processing methods

(1) Text Separation. The first step in text preprocessing is usually word separation, where the separation operation consists of separating words with commas or quotation marks and treating punctuation marks as individual words. Modern word separation algorithms usually use standard corpora as samples, and among the word separation algorithms, the Jieba library is commonly used for Chinese word separation.

(2) Word shape normalization. Word form normalization is a text preprocessing technique that aims to convert the different forms of a word into the same basic form for better text analysis and processing. In natural language processing, a word may have multiple forms, such as verb tenses, plural forms of nouns, etc. These deformations, although semantically similar, will be regarded as different words in text analysis, affecting the performance of the model.

(3) Deletion of stop words. Deleting deactivated words is a common step in text processing, which refers to removing words that are considered “useless” from the text to reduce the noise in the text and make the text clearer, more compact and meaningful.

(4) Word Vector. Word vectors are a technique for converting words in natural language into vector representations. Usually generated using word embedding algorithms, each word is represented as a fixed-length vector in word vectors, and the similarity between words is determined by calculating the distance between word vectors.

2.2 Acquisition and Preprocessing of Ming Dynasty Novels Dataset

2.2.1 Introduction and Acquisition of Fiction Datasets

The Ming Dynasty Text Dataset is a collection of all the text content of Jin Yong's martial arts novels, including classic works such as "The Romance of the Three Kingdoms", "Heroes of the Marshes", "Journey to the West", "Jin Ping Mei", and "Creation of the Gods". The dataset exists in the form of text files, with each file corresponding to one novel. The files contain information such as chapter titles and the main text content of the novels. It includes old editions, new editions, and revised editions, among which the revised editions have undergone more reasonable deletions and modifications.

The encyclopedia data is first crawled using crawling techniques to extract key data from the web content. During the experiment, the urllib module in Python is mainly used to initiate Ajax page requests to crawl the data.Ajax is a technology used to create dynamic user interfaces, commonly used in web applications. By sending requests to the server and getting responses in an asynchronous manner, it makes the user interface smoother and the user experience better.

2.2.2 Ming Dynasty Novels Dataset Extraction

In order to improve the accuracy of the analyzed data and analysis results, the data source needs to be preprocessed before analyzing the data:

(1) Segment the text using Jieba Segmentation, after which a deactivation word list is used to filter out symbols and words that are meaningless for the next step of text semantic processing.

(2) Extract the key information in the text data, and form a preliminary corpus by using the CRF method and other methods for the rapid identification of characters, places, and times in the text.

(3) Through the construction of character co-occurrence matrix, firstly, the names of people identified from the text are made into a character dictionary, and the number of times different characters appear is counted; then, the names of people appearing at the same time are looked up in the text passages, and the correlation statistics are carried out; finally, the number of times the obtained characters co-occur is stored in the co-occurrence matrix.

(4) For the same character, organize the set of utterances and cut them into chronological order by using word splitting technique. Then, the character trajectory information, including latitude and longitude information and event information, is sequentially organized according to the time sequence.

(5) Compare the contents of the corpus extracted from different text datasets, and filter and integrate the data for processing. Relationships and related information extracted from biographical texts and online texts are summarized and classified according to character ID to improve the dataset.

2.2.3 Construction of a corpus of character costumes

In order to achieve better character analysis, character corpus construction is needed. First of all, the novel “Jin Ping Mei”, which has more complicated characters' costumes, is crawled from the website as the data source, and the text format is converted. Secondly, data preprocessing of the novel data, the novel text is stored in chapters, and the content of the novel, as well as the characters' costumes and patterns are extracted to construct the character corpus. Finally, stuttering is used to perform operations such as word division and stop word filtering.

3 Interpretation of Dress Metaphors in Ming Dynasty Novels Based on NLP Technology

3.1 Ming Dynasty Novel Costume Feature Extraction

3.1.1 Eigen term frequency TF

In the TF-IDF algorithm, the feature item frequency is to achieve the normalization of the comparison of documents of different lengths in the same environment, and the feature items are phrases and words, as well as processed semantic conceptual feature units. It indicates the number of times a text phrase feature item appears in a certain text data, the more times the phrase feature item appears in the text, the higher the feature item frequency, and accordingly the more relevant it is to the topic of the text, which also indicates that the frequency information is one of the important indexes of text classification. Of course, in certain language contexts, the feature items of many words do not have practical value. Through these word items, it is impossible to perform text feature annotation for the text, so they need to be excluded. For example, in Chinese, the words“you” and “get”, and in English, the words "is", "am" and "are", etc., are meaningless function words.

The theoretical formula for word frequency (TF) is as follows, where $n_{i,j}$ is the number of occurrences of the word in the document d_j, and the denominator is the total number of occurrences of all words in the document d_j.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{1}$$

3.1.2 Inverse Text Frequency IDF

Inverse text frequency is an important index to quantify the distribution of feature terms in the text. Quantitative feature terms are also a kind of feature assignment, and feature assignment is divided into “equal weight assignment” and “non-equal weight assignment”, the former means that all the feature terms in the training set have the same degree of importance, which means

that these feature terms will not have any practical impact on the classification results. The latter is that different weights represent the different importance of feature values, which needs to be processed differently, and “IDF” is the most representative “non-uniform weight assignment”.

The inverse text frequency examines the recognition ability of the words in the whole, and measures the weight of the words in the text set, the larger the weight, the stronger the representation of the words, indicating a better ability to distinguish between categories. The theoretical formula for inverse text frequency is as follows: where D represents the total number of documents, n_t is the number of texts containing the word t in the text corpus, and 1 needs to be added to the denominator in order to avoid the denominator being zero.

$$IDF = \log \frac{D}{n_t + 1} \tag{2}$$

3.2 Ming Dynasty Novel Dress Network

3.2.1 Graph theory

Graph theory is an important mathematical tool for carrying out research in the field of humanities and social sciences as well as in the field of natural sciences. The graph in graph theory is not a geometric figure in the general sense, it is a method of describing any system containing binary relations, which is the main object of graph theory research.

A graph in graph theory is a graph formed by edges formed by connecting a number of given points, with points representing things, two points connecting to form edges, and edges representing the relationship between two things.Mathematically, a graph is a bipartite group graph $G = (V(G), E(G))$ consisting of vertices and edges, where V represents the set of all vertices and E represents the set of all edges, and graphs constructed from these points and edges have: Directed and Undirected.

3.2.2 Social network analysis

The essential meaning of social network refers to the online social network using the Internet as a bridge, which is an expansion of people's real social activities. In large networks with complex character relationships, identifying key nodes in the network, i.e., real key characters, based on factors such as the deterrence and influence of the characters, is a major feature of social network research. An exhaustive study of the social networks mapped by these key nodes (core characters) will help to uncover important information, which in turn will play a pivotal role in the study of the overall situation.

Four typical approaches to social network information dissemination are: centralized broadcast dissemination, (multi)center penetration dissemination, cross-regional chain dissemination, and hybrid (a mixture of the above three dissemination approaches).

3.2.3 Graph theory and social network analysis methodology applications

From the perspective of graph theory analysis, there are currently four evaluation metrics for measuring key nodes, namely: point degree centrality, proximity centrality, mediator centrality, and eigenvector centrality. Point degree centrality refers to how many nodes a particular node is connected to, emphasizing the individual value of the point, which is mathematically represented as follows: $d(x_i, y_j) = 1$, the nodes at the ends of the x and y lines.

$$DEC(x_i) = \sum_{x_i \neq y_j} d(x_i, y_j) \tag{3}$$

Proximity centrality mainly measures the inverse of the sum of the distances from each node to the other nodes, indicating both the closeness of the node to the other nodes and its value to the other nodes, which is mathematically represented as follows:

$$CLC(x_i) = \frac{1}{\sum_{x_i \neq y_j} d\left(x_i, y_j\right)} \tag{4}$$

Intermediate centrality is an important measure of the control and influence of the node, it refers to a point in the network with a point of the shortest path to the total number of short-circuit path line ratio of this point, the node's ability to act as a mediator (bridge), intermediation, the higher the intermediation, indicating that the node is more important, and its mathematical representation is:

$$BTC(x_i) = \sum_{i<j} \frac{g_{ij}}{a_i} \tag{5}$$

i denotes distinct nodes, $i > j$, g_{ij} denotes the number of shortest distances between node i and node j, and g_{ij} denotes the sum of the number of shortest distances passing through node i.

Eigenvector centrality mainly indicates the importance of neighboring nodes because a few influential nodes are more important than nodes with larger centrality values, therefore, as an extension of pointwise centrality, the importance of a node is determined by calculating the eigenvector values of the adjacency matrix of neighboring nodes. A particular node i eigenvector is $AX = \lambda X$, where A is the eigenvalue of the eigenvector. The four metrics of point centrality, proximity centrality, intermediate centrality and eigenvector centrality are used to perform a comprehensive measure of importance in the network of key stages and identify the most influential key nodes.

In the dataset, the important information corresponding to the important nodes is identified using graph theory and social network analysis, while the edges connected to nodes in the large social network represent the linkage relationship between the author's costumes and the characters in the Ming Dynasty novels.

3.3 Modeling the Metaphorical Theme of Clothing in Ming Dynasty Novels

The LDA model involves a great deal of mathematical knowledge such as Gamma function, binomial distribution, multinomial distribution, Beta distribution, Dirichlet distribution, conjugate prior and Bayesian framework, Gibbs sampling and so on.

(1) Gamma function: it is essentially a generalization of the factorial function to the real numbers, and the factorial formula when it is an integer and its generalization formula when it is a real number are Eq. (6), respectively:

$$\begin{cases} \Gamma(n) = (n-1)!, \text{ Integer} \\ \Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx, \text{ Real number} \end{cases} \tag{6}$$

(2) Binomial distribution: generally has only two values, is the frequency of the number of times the event A occurs in k experiments of two outcomes of the desired outcome. The outcome of each experiment is independent of the last, if B appears with probability $P(B)$ and $P(A)+P(B)=1$.

$$P(A)=\begin{bmatrix} n \\ k \end{bmatrix} P^k(1-P)^{n-k}, \left(k=0,1,2,3\cdots n\right) \tag{7}$$

(3) multinomial distribution: just as the probability of shaking the six points of the lottery, multinomial distribution will also appear a variety of results. Suppose an experiment has K possible scenarios as $A_1, A_2 ... A_k$, then it happens n times when A_1 occurs n_1 times... A_k occurs n_k times, then $P\left(A_1, A_2, A_3 ... A_k\right)$ has the probability shown in Equation (8):

$$P\left(A_1, A_2, A_3 \cdots A_k\right)=\begin{cases} P^{n_1}P^{n_2}\cdots P^{n_k}\dfrac{n!}{n_1!n_2!\cdots n_k!}, \sum_{i=1}^{k} n_i = n \\ 0, \text{otherwise} \end{cases} \tag{8}$$

(4) Conjugate Prior: Understanding conjugate prior requires that we first know the concepts of the words prior distribution, posterior distribution, likelihood estimation, and so on. Which the a priori distribution is used to describe the degree of uncertainty of an uncertain quantity. The posterior distribution is the result of the product of the prior probability and the likelihood function. It refers to the conditional probability of an uncertainty after the data and evidence have been confirmed, and the prior and posterior distributions that have the same functional form also become the conjugate prior. Where the prior distribution probability is $P(\theta)$, the posterior distribution is $P(\theta \mid X)$, and the likelihood function probability is $P(X \mid \theta)$.

(5) Beta distribution: Beta distribution takes the value interval of $\left[0,1\right]$ continuous probability distribution. beta distribution is also the binomial distribution of the conjugate distribution. Its parameters α, β have values greater than 0, and its probability density formula is expressed as equation (9):

$$Beta(P \mid \alpha, \beta)=P^{\alpha-1}(1-P)^{\beta-1}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \tag{9}$$

where τ is the Gamma function satisfying $\Gamma(x)=(x-1)!$.

(6) Dirichlet distribution: also known as the multivariate Beta distribution, is its distribution in higher dimensions and is a k -dimensional conceptual function expressed as equation (10):

$$Dirichlet(P \mid \alpha)=\prod_{i=1}^{k} P_k^{\alpha_k-1}\frac{\Gamma\left(\sum_{i=1}^{k}\alpha_k\right)}{\prod_{i=1}^{k}\Gamma(\alpha_k)} \tag{10}$$

Its covariance with the multinomial distribution also exists and can be expressed as equation (11).

$$Dirichlet\left(P \mid \alpha\right) + MultiCount\left(d\right) = Dirichlet\left(p \mid \alpha + d\right) \tag{11}$$

(7) Gibbs sampling: For the estimation of document-topic parameter θ and topic-word parameter φ, we use Gibbs sampling method to solve the problem, we need to solve the conditional probability distribution of each feature dimension, then we need to find out the joint distribution of topic $Z_{m,n}$ and word $W_{m,n}$ first, then we need to The probability distribution $P(W)$ of a word under the corresponding topic is obtained until Gibbs sampling converges, at which time we will find the topics of all words in turn and obtain the distribution of each topic word, and then calculate the number of topics to obtain the distribution of topics for each document.

LDA in the PLSA document - word, word - topic of the two generation process set of layers of Bayesian framework, but also in the topic distribution and word distribution on the addition of two Diracke Ray prior distribution, unlike PLSA, it is considered that the topic and word distribution is uncertain, and randomly extracted from its model. lda model is shown in Fig. 1:

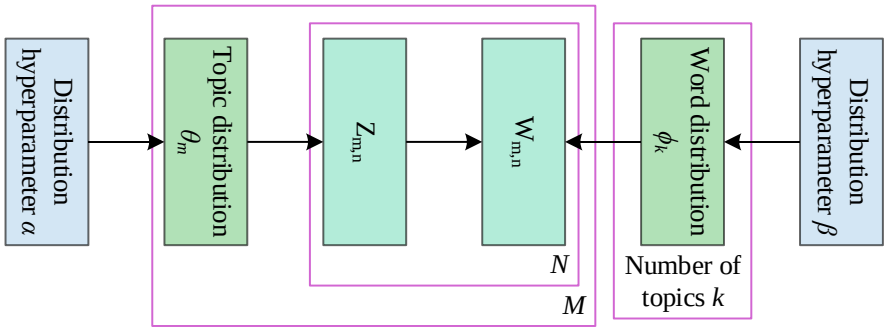


Figure 1: LDA illustraton of model

The overall process of LDA document generation is as follows:

(1) Generate the topic distribution θ_m of document M by sampling from the Dirichlet distribution α, where $\theta_m = Dirichlet(\alpha)$ and α is the distribution hyperparameter.

(2) Generate the n th word topic of document m by random sampling from the polynomial distribution of topics θ_m: $Z_{m,n}$, where $Z_{m,n} = Multi(\theta_m)$.

(3) Generate a random sample of topics $Z_{m,n}$ from the Dirichlet distribution β and determine the corresponding word distribution.

(4) The $\varphi_{m,n}$ sampling process ultimately generates the word $W_{m,n}$. The word $W_{m,n}$ is generated based on the parameter $Z_{m,n}$ and η, where $W_{m,n} = Multi(\varphi_{m,n})$. Here, $\varphi_{m,n} = Dirichlet(\eta)$, and η is a V -dimensional vector (representing the total number of words in the vocabulary) and also a distribution hyperparameter. This process repeats in the order of document - topic - word until all the words in the document are included and the document is finally generated.

In summary, LDA as the latest proposed theme model, compared to other theme models also has its unique advantages, compared to other models, the computational complexity is lower computational efficiency, can reduce the dimensionality and refinement, the full use of a priori knowledge, there is a more powerful mathematical and statistical knowledge as a support, and the scope of the application and application of the most widely used, so this paper selects the LDA theme model to achieve a better classification of the subject matter. Therefore, the LDA topic model in this paper can achieve a better effect of topic classification.

4 Empirical analysis

4.1 Word frequency statistics and feature word extraction

4.1.1 Word Frequency Statistics for Character Clothing

The text segmentation method used in this paper is based on the R software, which is realized by calling the jiebaR installation package. jiebaR package has its own systematic lexicon and is quite rich in vocabulary, which is enough to be used in this text analysis. After the text is divided into words, the whole text is divided into 2975 words, but in fact, these words include intonation auxiliaries, adverbs, prepositions, conjunctions and so on, which do not have much significance for analysis, but appear with high frequency, such as “get, what about, got, also, so, so” and so on. In order to avoid adding a lot of noise when counting the word frequency at a later stage, so these words are usually filtered. In this paper, we use the deactivated words of HIT, and the remaining 1800 words after sifting out the deactivated words. The extracted word frequencies are shown in Table 1 after the process of word segmentation and de-duplication. Among them, “jacket” 371 times is the daily dress of Ximen Qing's female family members in the novel “Jin Ping Mei”, and the word frequency of “hat” 174 times reflects the social roles, such as the official with a “sarong”.

Table 1: Top 24 words frequency statistics

Keywords	Word frequency	Keywords	Word frequency
A short Chinese-style coat or jacket	371	Shirt	22
Skirt	195	Jin jacket	33
Robe	183	Fur-lined jacket	17
Cap	174	Priest frock	29
Hairpin	93	Undershirt	32
Loop	85	Luo Shan	38
Belt	77	Sleeveless coat	15
Bijia	56	Battle skirt	9
Skirt dress	43	Cloak	17
Skirt and jacket	47	Lined short gown	21
throw on a gown	54	Overall	16
Ling Ao	41	A pleated skirt	34

4.1.2 Character dress feature word extraction

A very important part of this process is the extraction of feature words, and then by the IDF to calculate the weight of each word, the higher the frequency of the word is the greater the IDF value. After obtaining the “Term Frequency” (TF) and “Inverse Document Frequency” (IDF), the values are multiplied to obtain the TF-IDF value of the word. The importance of a word to the article is directly proportional to the size of the word's TF-IDF value. Finally, we only need to select the words with the top TF-IDF values, which are the feature words of the article. According to the algorithm, the text feature words are extracted and roughly divided into 4 categories as shown in Table 2.

Table 2: Attribute feature word

Attribute	Feature word	Meaning
Material grade	Silk, satin, brocade, gauze, and silk fabric	Reflecting social class and consumption power
Identity politics	Shoes and socks, bra, sweat towel, nail clipper	Closely related to narrative
Identity	Sweat cap, Taoist robe, and crown	Reflect social role
Material details	Gold, embroidery, sable fur, hairpins...	Echoing the theme of the novel

4.2 Ming Dynasty Novel Dress Network Construction and Analysis

4.2.1 Ming Dynasty Novel Costume Network Construction

Using the same chapter principle and a time interval of every 20 chapters as an example, the character relationship networks of chapters 1-20, 21-40, 41-60, 61-80, and 81-100 are constructed:

(1) Chapters 1-20. Core main line: Simon Qing-Pan Jinlian; ambiguous line: Simon Qing-Li Bier; branch lines: Simon Qing-Wu Yue Niang, Pan Jinlian-Wusong.

(2) Chapters 21-40. Core advancement line: Simon Qing-Li Bier; main line of internal struggle: Pan Jinlian-Li Bier; branch lines: Simon Qing-Chen Economy, Simon Qing-Officials, Wu Yueliang-Concubines.

(3) Chapters 41-60. Tragedy Main Line: Pan Jinlian-Li Bier-Officials; Sub-Lines: Simon Qing-Pang Chunmei, Simon Qing-Economy/Officials, Wu Yue Niang-Simen Qing.

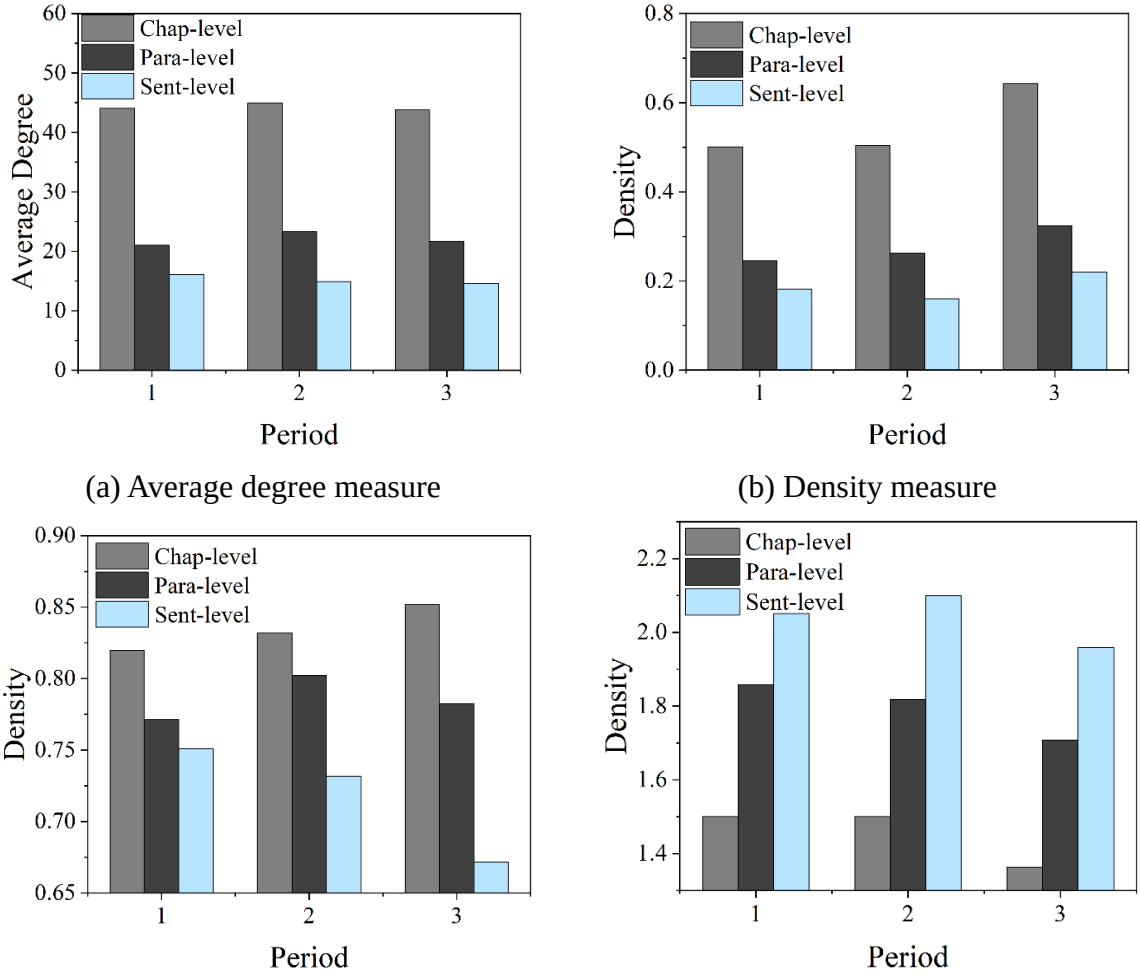
(4) Chapters 61-80. Core turning point: death of Simon Qing; incest main thread; Pan Jinlian - Pang Chunmei - Chen Economy; sub threads: Wu Yueliang - concubines/Chen Economy, Meng Yulou - Sun Xue'e

(5) Chapters 81-100. Revenge Main Line: Pan Jinlian-Wusong; Tragedy Spur: Pang Chunmei-Chen Economy; Spur: Wu Yueliang-Hsiao, Other Characters.

4.2.2 Analysis of network topology properties

Based on the time interval of every 20 chapters of the network for network basic statistical analysis, including average degree statistics, network density statistics, average clustering coefficient statistics and average path length statistics. Figure 2 shows the results of the four kinds of network statistical information under different connecting edge principles, with the horizontal coordinate representing time and the vertical coordinate representing the corresponding statistical index.

It is observed that in the latter 20 times, the network density increases, the number of nodes and edges decreases, and the average path decreases while the average degree remains unchanged, which indicates that the characters' costumes are clustered and concentrated in the portion of the author's description, the scope of the description is narrowed down, and the plots are more closely connected. At the same time, it can be found that the proportion of edge nodes in the network is also gradually reduced, filtering the network, the proportion of nodes with statistical degree 15 or less evolves as follows: 10.14% - 7.73% - 4.39%, indicating that the author's ink for the edge of the characters less and put more focus on main characters writing.



(a) Average degree measure

(b) Density measure

(c) Average clustering coefficient measure

(d) Average path length measure

Figure 2: Statistics of basic information from networks

Based on the time interval for every 10 chapters with the same sentence principle and the time interval for every 1 chapter with the same sentence as well as the same paragraph principle of the network we performed the node and the number of connecting edges statistics as shown in Fig. 3 and Fig. 4.

Observing Figure 3, we find that the peaks of the connecting edges appear in the first 60 chapters written by the author, which are the 15th and 55th chapters, respectively, and the peaks imply that the depictions of the characters' costumes are more complicated than the other chapters in the specific chapters. On the contrary, the last 40 chapters are smooth, indicating that the author is weak in depicting complex characters' costumes. Figure 3 shows similar characteristics as Figure 4.

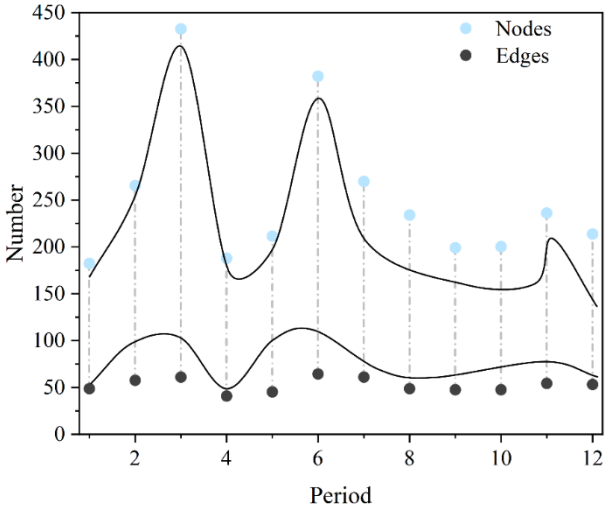
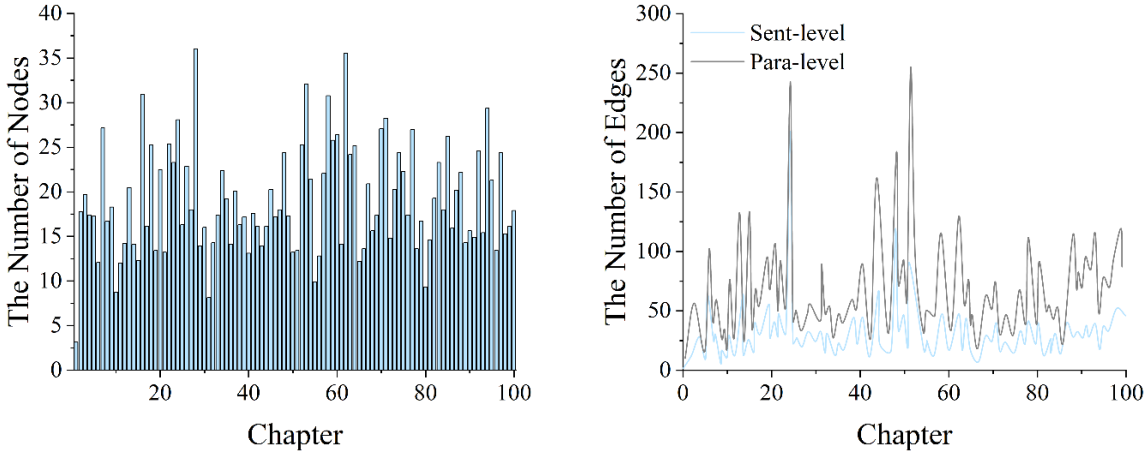


Figure 3: Change of the numbers of nodes and edges on sent-level every 10 chapters



(a) Change of the number of nodes (b) Change of the number of edges

Figure 4: Change of the numbers of nodes and edges on sent-level and para-level every chapter

4.2.3 Evolution of clothing color importance ranking

The main characters of “Jin Ping Mei” novel are Ximen Qing, Pan Jinlian, and Wu Yue Niang, and in this section we will focus on analyzing the costume metaphors of the main characters. Based on the node degree centrality, the changes in the importance ranking of the characters' costumes in 100 chapters are studied, and the results are shown in Figure 5. In the picture, in the 40th chapter of "Jin Ping Mei", Ximen Qing's original text reads: "He made a piece of animal-style dragon-pattern brocade robe." He borrowed the official's dragon pattern brocade collar as a reference, suggesting that after donating his position, his identity as a "merchant who has become a bureaucrat" was clinging to the loosening of social ethics. In the 24th chapter of "Jin Ping Mei", regarding Wu Yuming's clothing description: "Only Wu Yuming wore a large red and golden-striped robe, a fox-skin coat, and a flower-patterned skirt. On his head, there were piles of jewels and pearls, and the phoenix hairpin was half-unfastened." This highlights her family status and the social norms of "clothing hierarchy" in the Ming Dynasty. In the 1st chapter of "Jin Ping Mei", regarding Pan Jinlian's clothing description: "She wore a piece of a buttoned shirt, a white shirt with a fringe border, and a red and purple silk skirt below." This reflects the character of the person.

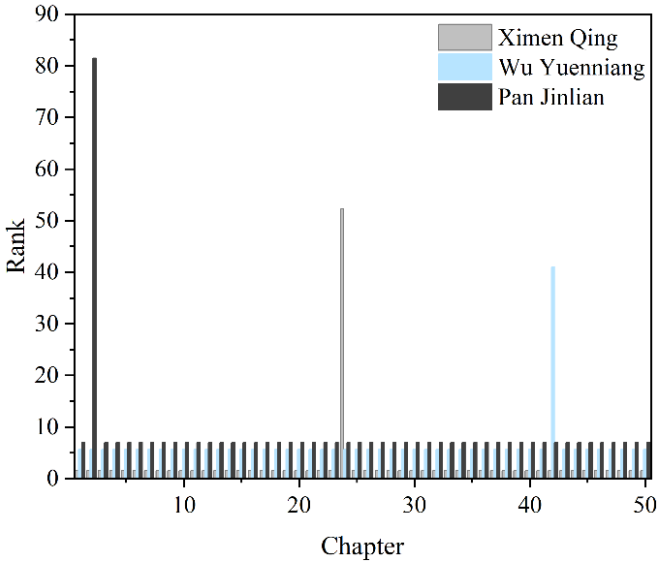


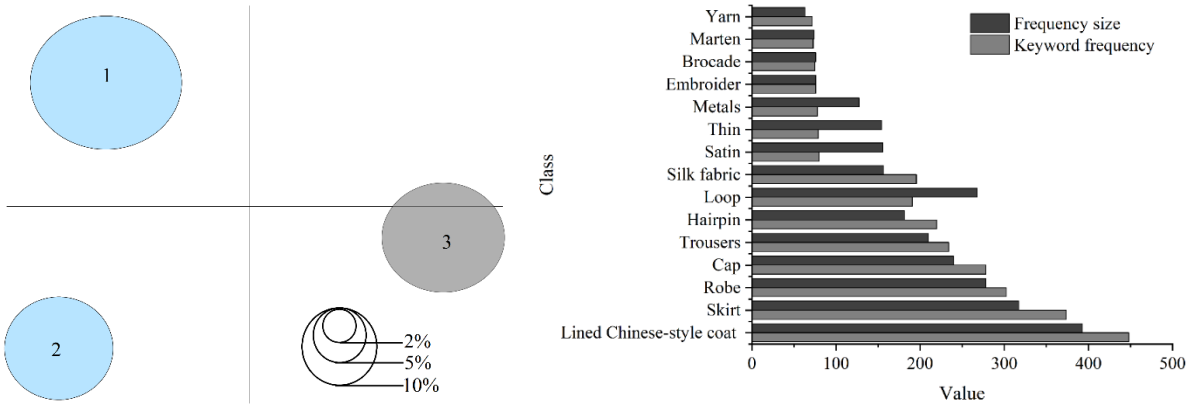
Figure 5: Importance Rank of in Each Chapter

4.3 Thematic Visualization of Dress Metaphors in Ming Dynasty Novels

4.3.1 Mining the Metaphorical Theme of Clothing in Ming Dynasty Novels

Based on the LDA model for positive and negative theme mining of the novel text, the number of iterations is set to 500, and the optimal number of themes is determined by multiple tuning comparisons and observing the pyLDAvis visualization effect, and the experimental results are shown in Fig. 6, taking the positive comments as an example.

Figure 6 shows that the circles indicating the 3 themes are clearly different and do not overlap, indicating a good cohesive effect of the themes. The column list will show the keywords in the topic, where the light-colored bar indicates the keyword word frequency, and the dark-colored bar represents the word frequency size of the keyword within the topic 3. With the help of visualization, it is possible to clearly and intuitively view the arrangement of keywords and the difference in distribution within each chapter.



(a) Intertopic distance map (b) Top-15 most relevant terms for topic3

Figure 6: Topic mining visualization

4.3.2 Thematic analysis

In order to further explore the themes of the characters' costumes in “Jin Ping Mei”, this study

uses R language to construct LDA theme models for 100 chapters. Gibbs sampling technique was used to determine the optimal number of theme categories as 4. The model results obtained 4 themes and the keywords corresponding to each theme, and the top 6 keywords and their probability distributions in each category of themes are shown in Table 3. The larger the probability value of a keyword, the stronger the relationship between the word and the topic. Based on the keywords with a high degree of relevance in each theme, the meaning of the words is analyzed and understood and summarized.

Table 3: Theme and corresponding core keywords

Theme1:Luxury women's clothing		Theme2:Daily clothing for women at the bottom of the social hierarchy		Theme3:Men's official and merchant attire		Theme3:Festive and special occasion attire	
Brocade	0.086	Plain cloth	0.051	Round collar	0.041	Bright red	0.033
Long-sleeved jacket	0.019	Shirt	0.026	Suede boots	0.035	Maze flowers	0.029
Feng Chai	0.011	Sarong	0.021	Waiyuanjin	0.027	Embroid	0.022
Jade pendant	0.010	Little jacket	0.019	Jade belt	0.022	Gold selling	0.017
Bijia	0.091	Strap	0.015	Luo Shan	0.018	Cloak	0.009
pearls and jade	0.075	Highlight	0.011	Pleat	0.015	Cloud shoulder	0.002

Theme 1: The keywords brocade, phoenix hairpin, and jewels and emeralds are direct symbols of the status of wealthy merchants and official women in the late Ming Dynasty. It is used here to describe the costumes of upper-class women such as Wu Yueliang, the main wife of Ximen Qing in the novel, suggesting the power struggle of women in the inner house through dress comparison.

Theme 2: This theme mainly highlights the lower class female group, which is in obvious contrast to the costumes of the noble women of the mansion, reflecting the class hierarchy of women in the society.

Theme 3: The keyword is the iconic costumes of Ming officials and merchants, which here corresponds to the change in social identity of Ximen Qing from a wealthy merchant to an official body, implying the social culture of the Ming dynasty that utilizes money for official transactions.

Theme 4: Big red, makeup and embroidery are a kind of costumes used for special occasions and festivals, and these costumes often appear in key novel plots, such as Li Bail's marriage into the Ximen Mansion and the banquet for Ximen Qing's birthday, which emphasize the importance of the plot through the costumes, and reflect the importance of folk costumes and rituals in the Ming Dynasty.

5 Conclusion

Taking the famous novel “Jin Ping Mei” of the Ming Dynasty as the research object, the study combines natural language processing technology and social network analysis method to extract the high-frequency words and themes of characters' costumes in the novel by extracting the novel's text data and constructing a corpus. The potential association between the characters' costumes and metaphors in the novels of the Ming Dynasty is revealed, and the textual themes in the novels are comprehensively mined to obtain four themes, and the visualization of each theme is presented to clearly show the class differences and life scenes in the Ming Dynasty society.

About the Author

Menghe Tian was born in Hebei, China, in 1997. She obtained a bachelor's degree from Donghua University in China, and a Master degree in University of Leeds. She currently studying at Collage of Fashion and Design, Donghua University. Her main research direction is Ming Dynasty Fashion and Social psychology of clothing.

Xiangyang Bian, professor of art and doctoral supervisor. Long-term commitment to the history of clothing, fashion ecosystem education, scientific research, social services, cultural inheritance and other work. He has published 19 books and won the second prize of the 8th Scientific Research Achievement Award (Humanities and Social Sciences) of the Ministry of Education. He has published 170 papers and an album of personal fashion criticism. He has presided over more than 70 scientific research projects such as art key projects of the National Social Science Foundation. The results of teaching reform won the second prize of national teaching achievement Award.

References

[1] Burton, M. E. (2019). Clothing as a Metaphor for the Classical Hebrew Semantic Domain of כבוד. Clothing and Nudity in the Hebrew Bible, 289.

[2] Vauzelle, L. (2020). Clothes with Metaphorical Names and the Representation of Metaphors in the Costumes of the Aztec Gods. Indigenous Graphic Communication Systems: A Theoretical Approach, 154.

[3] Begley, K. (2022). Language Disguises Thought: Uncovering the Origins of the Clothing Metaphor in Tractatus 4.002. Disputatio, 11(23), 215-242.

[4] Estrada III, R. G. (2023). The racial significance of Paul's clothing metaphor (Romans 13: 14; Galatians 3: 27; Ephesians 4: 24; Colossians 3: 10). Religions, 14(6), 684.

[5] LIU, X. J., & MA, Y. L. (2025). Metaphors on "Sleeve": A Cognitive Perspective on Chinese Characters and Traditional Clothing. Journal of Literature and Art Studies, 15(1), 11-20.

[6] Wei, L. (2018). A preliminary study of Mongol costumes in the Ming dynasty. Social Sciences in China, 39(1), 165-185.

[7] Feng, J. (2024). Exploring Archaeological Discoveries Through Ming Dynasty's Clothing: Impact on Realistic Style, Colour Matching, and Space Construction In Late Renaissance Painting. Mediterranean Archaeology and Archaeometry, 24(3), 253-270.

[8] Zhang, B., & Yang, P. (2024). The ideological background of ancient Chinese clothing culture. Trans/Form/Ação, 47, e0240061.

[9] Zhu, Y. (2021). Rank Badges of Official Costumes of Ming and Qing Dynasties from the Perspective of Social Semiotics. Language & Semiotic Studies, 7(2).

[10] Lu, C., & Ismail, N. H. (2023). THE DESIGN CHARACTERISTIC OF LATE 15TH CENTURY MING DYNASTY COSTUMES IN THE JIANGNAN REGION. Jurnal Tuah, 4(1), 179-188.

[11] Bian, X., Tian, M., & Zhou, L. (2025). From Sacred to Secular: Daoist Robes as Instruments of Identity Negotiation in Ming Dynasty Literature. Religions, 16(7), 903.