*Review*

Social Science Computer Review
2022, Vol. 40(4) 1014–1034

DOI: 10.1177/08944393211028182
journals.sagepub.com/home/ssc
SAGE

# Analysis of the Social Network and the Evolution of the Influence of Ancient Chinese Poets

**Jiaqi Zhao[1], Yuting Wei[1], and Bin Wu[1]**

## Abstract

Chinese classical poetry occupies an important position in ancient Chinese literature. However, the existing research on Chinese classical poetry is usually limited to a certain poet or dynasty to analyze its historical and cultural influence and lacks comprehensive research on the process of ancient poetry from the perspective of time and space. We integrate multisource data and information, including poets' biographies and Chinese classical poetry, to build a relatively complete social network of 41,310 poets. Based on this network, we use natural language processing and social network analysis techniques to research the relationships between poets. For example, how poets of different dynasties and different schools relate to each other. In order to quantitatively analyze the changing process of poets' influence over a period of time, we propose a new method—time-series entropy weight method—to calculate the dynamic changing process of poets' influence over time. Besides, we evaluate and discuss the method of calculating the influence of poets by means of propagation dynamics model and verify the effectiveness of the proposed method. Through the study of the complex social relations of poets and the quantification of their influence, we can assist in the study of the development of different schools and styles of ancient poetry. Our work offers a new, data-driven, long-run perspective on the evolution of Chinese poetry for historical researchers and enthusiasts to understand the complex relationships among historical figures.

## Keywords

Chinese classical poetry, complex network, social network analysis, node importance, entropy weighting method

The historical civilization of the Chinese nation for thousands of years is the precious cultural heritage of the world. Chinese classical poetry, as an important part of Chinese civilization, not only is a carrier for poets to express their feelings but also reflects the social atmosphere and cultural

[1] Beijing University of Posts and Telecommunications, China

**Corresponding Author:**
Bin Wu, Beijing University of Posts and Telecommunications, Beijing 100088, China.
Email: wubin@bupt.edu.cn

outlook of ancient China. In ancient China, poetry was an indispensable part of life. The literati often held various forms of poetry competitions and composed poems impromptu as a kind of entertainment at family or social gatherings. In the heyday of Chinese literature, there were poets from all walks of life. From emperors to peasants could recite poems. In addition, poetry was also used as a social tool in ancient China. Poets liked to give each other poems to show their affection for their friends and families. For example, Li Bai (李白) wrote "Peach Blossard Pond water is a thousand feet deep, not as deep as Wang Lun to send me affection" (`桃花潭水深千尺, 不及汪伦送我情') in *To Wang Lun* (赠汪伦), describing the friendship between Wang Lun (汪伦) and himself. Liu Yuxi (刘禹锡) once sent to Bai Juyi (白居易) a poem *Rewarding LeTian at the First Meeting in Yangzhou* (酬乐天扬州初逢席上见赠), and Bai Juyi also returned Liu Yuxi a poem *Drunken Gives to envoy Liu ErShiBa* (醉赠刘二十八使君). Liu Yuxi and Bai Juyi wrote more than 100 poems to each other all their lives. Poetry had a great influence on classical literature and the daily life of ancient society. We hope to explore the historical changes and cultural development from these surviving poems and the complex social relationships among poets. However, the relationships between poets are often overlapping and complex, so it is necessary to use computer technology for analysis.

As an important literary form of expressing emotions, poetry has developed into different styles and schools. If researchers want to study how the genre of poetry is developing over a long period of time, it is necessary to study not only the social relationships of poets but also the influences they exerted on their descendants. Historical biographies and poetry works often record this information. For example, Tao Yuanming (陶渊明, WeiJin Dynasty), known as the originator of pastoral poetry, had a great influence on later generations. Meng Haoran (孟浩然, Tang Dynasty), a pastoral poet, had a strong admiration for Tao Yuanming. He said in his poem, "Reading *the Biography of Gentlemen*, Tao Zhengjun (another name of Tao Yuanming) is the best. Although the stars are shining, you are the only one for me" ("赏读《高士传》，最佳陶征君，目耽田园趣，自谓羲皇人。"). In terms of poetry creation, Meng Haoran inherited Tao Yuanming's writing style in terms of subject matter, content, and expression. Xin Qiji (辛弃疾), a patriotic poet in the Southern Song Dynasty, also took Tao Yuanming as his confidant. He left more than 600 poems, of which 60 were intoned, mentioned, or quoted Tao Yuanming.

In recent years, the research on ancient Chinese literature has made great achievements in platform construction, text analysis, and visualization by using digital humanities technology to explore new ways and expand new space Jian (2016). *Garden of Tang poetry* (http://tsby.e.bnu.edu.cn/), developed by National Engineering Laboratory for Intelligent Technology and Application of Internet Education, provides semantic retrieval function. It also adopts visualization technology to display the social network of poets, migration map of poets, and regional hotspot map of poetry in the Tang Dynasty. Xinhuanet and Zhejiang University built a platform—*Using Song poetry to draw the world* (http://www.xinhuanet.com/video/sjxw/2018-09/07/c_129948936.htm), which analyzed the travel routes and lives of poets of the Song dynasty and the common imageries of *All Song poetry*. Knowledge graph of Academic Inheritance in the Song Dynasty (H. Yang & Wang, 2019), based on Chinese Biographical Database (CBDB), visualizes the academic teacher–student relationships and some kinship relationships of scholars in the Song Dynasty. *Chronological map of Tang and Song literature* (https://sou-yun.cn/PoetLifeMap.aspx), sponsored by Professor Zhaopeng Wang, constructed a dating map platform for the Tang Dynasty and the Song Dynasty literature to visually display the space–time track distribution information of poets. Dong and Liu (2020) built a social network of the Tang Dynasty poets based on *All Tang poetry* and inferred the social relationship of the Tang poets by Poisson graphical lasso. These studies have made some achievements, but they are still limited to the analysis of one aspect in the Tang or Song Dynasties, failing to present a comprehensive picture of ancient Chinese literature in a larger time and space span.

With the development of computer technology, social network analysis has gradually become a mature discipline. Social network analysis is a theory integrating methods originating from informatics, mathematics, sociology, management, psychology, and related disciplines, which provides a computable analysis method for understanding the formation of various social relationships. Schich et al. (2014) constructed a figure of people's migration according to the information of the places of birth and death of European historical figures in the 18th century, taking cities as nodes and constructing an edge between the city of birth and the city of death of the figures. Based on PageRank algorithm, the nodes in the network were analyzed, and the most influential cultural center in European cities was calculated. Shen and Barab ási (2014) studied the scientific research cooperation network and proposed a method to quantify the scientific impact of coauthors. Fraiberger et al. (2018) constructed a collaborative exhibition network between artists and museums, taking museums as nodes. If the same artist had exhibited works in both museums, an edge will be constructed between the two nodes. According to the network, the influence of artists was calculated and analyzed, and whether artists would exhibit their works in famous museums was predicted. Wu et al. (2019) analyzed more than 65 million papers, patents, and software products from 1954 to 2014; built a paper citation network; and analyzed the relationship between the importance of papers and the size of the author teams. Chowdhury et al. (2019) constructed the social network of characters in Harry Potter books and analyzed its network characteristics. *Graphing the history of philosophy* (http://coppelia.io/2012/06/graphing-the-history-of-philosophy/) analyzes the formation and development of different philosophical schools by building social networks among western philosophers.

With the development of digital humanities, complex network analysis technology has been applied to social humanities research more frequently. In order to study the interaction between poets and quantify the influence of poets, we built a complete social network of poets from the Pre-Qin to the Qing Dynasty (from 770 B.C. to 1800). The evolution of poets' influence is of great significance in assisting the study of the changes of ancient philosophy, the development of different poetry schools, and the analysis of the appearance of ancient society. For example, how did the style of the Tang poetry develop and change and how did the different schools of the Song Poetry come into being and developing. According to the poets' social network diagram, we can study the continuous mature creation process of one poet, which predecessors influenced him and which younger generations he influenced.

The main contributions of this article are as follows:

1. We collate multisource heterogeneous data information and construct a poets' social network with 41,310 nodes and 114,697 edges, which spans more than 2,000 years, from 770 B.C. to 1800. This network includes poets of all dynasties in Chinese history, which is the most complete network of poets with the largest span of time. Based on this network, we can study how poets of different dynasties and different schools relate to each other.
2. In order to deeply study this social network and explore the dynamic evolution process of poets' influence, we put forward the time-series entropy weight method (TSEWM) to quantify the influence of poets in a period of time. We evaluate the ranking results of node importance by using the propagation dynamics model and verify the accuracy of TSEWM in measuring node importance from two aspects of propagation range and propagation rate.

## Method

### *Poets' Social Network*

There are many challenges in constructing the poets' social network: (1) Poets in ancient times often had many aliases. Ancient poets seldom used other poet's name when writing poems but more used

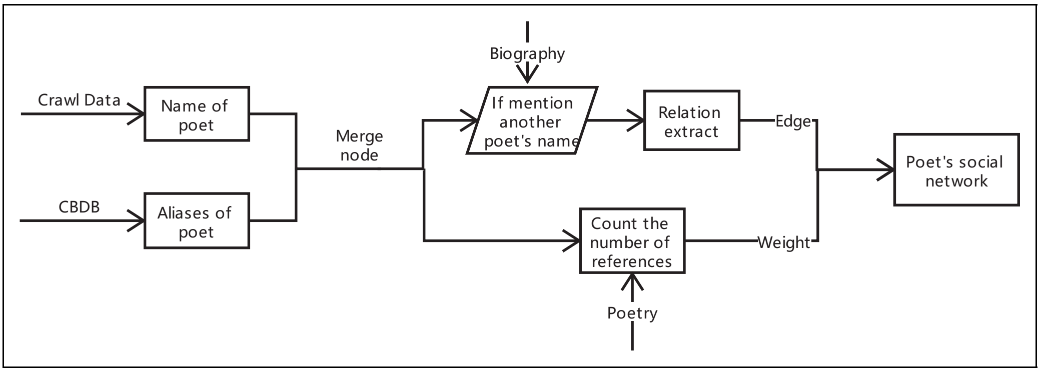


**Figure 1.** A flowchart of the poets' social network.

their alias. For example, Le Tian (乐天) means Bai Juyi in *Rewarding Le Tian at the First Meeting in Yangzhou* (酬乐天扬州初逢席上见赠), Li Shi Er Bai (李十二白) is another name for Li Bai in *Send Li Shi Er Bai Twenty Rhymes* (寄李十二白二十韵), and Du Erfu (杜二甫) is another name for Du Fu in *Farewell Du Erfu at East Shimen of Lu County* (鲁郡东石门送杜二甫). Therefore, in order to extract the relationship between poets from a large number of text data, the first thing to do is to match the names of poets with their aliases; (2) Multisource heterogeneous data. There are a lot of data about poets and poetry on the web, but most of them are fragmented and in the form of unstructured texts. So, we need to collect and process a lot of data on poetry and biographical information about poets; (3) Poetry and ancient texts are difficult to process with conventional natural language processing techniques and therefore require a lot of human involvement. The construction process of the poets' social network is shown in Figure 1. We will describe the construction process of the poets' social network from three aspects: data processing, nodes building, and edges building.

*Data processing.* There are three main sources of our data set:

- Poetry data: We crawl the poetry data from several ancient poetry platforms (*Souyun* [https://sou-yun.cn/PoetLifeMap.aspx], *ancient poetry and prose website* [https://www.gushiwen.org/], etc.), and sort out 905,675 works of 41,310 poets from the Pre-Qin Dynasty to the Qing Dynasty.
- Poet biographical data: We have collected biographical information of people from many ancient poetry platforms and historical books such as *Souyun*, *ancient poetry and prose website*, *Wikipedia*, *Dictionary of Chinese People's Names in Past Dynasties*, *Comments on Tang Poems* and *Poems of the Whole Song Dynasty*. We use named entity recognition and relationship extraction techniques to extract the social relationships of poets from their biographies. The relationship includes kinship, friendship, love, teacher–student, and admiration.
- Poet aliases data: CBDB (https://projects.iq.harvard.edu/cbdb) established by Harvard University is a relational database, which contains biographies of about 471,000 people, mainly from the seventh century to the 19th century, including biographies of the Tang Dynasty, the Song Dynasty, the Ming Dynasty, and the Qing Dynasty. We match the aliases of the poets from the CBDB. Since there is only information about historical figures after the seventh century in CBDB, we also use named entity recognition technology to identify the poet aliases from Wikipedia.

We use the open-source toolkit—OpenNRE (https://github.com/thunlp/OpenNRE)—in the named entity recognition and relation extraction in the experiment. In order to ensure the accuracy of relational extraction results, we have carried out partial manual correction on the extraction results. Table 1 shows examples of data sets and relationship extraction.

**Table 1.** Data Set and Relational Extraction Examples.

| Data Set | Relationship | Define | Content | Relation Extraction |
|---|---|---|---|---|
| Introduction of the poet | Kinship | It means a relative, especially a person who is related by blood, immediate or collateral relatives. | *Su Zhe* was born in Meishan, Meizhou, the Song Dynasty. His courtesy name was Ziyou, Tongshu, and his another name was YingbinYilao. He *is the son of Su Xun* and *the brother of Su Shi*. | Su Zhe, Su Xun, kinship; Su Zhe, Su Shi, kinship |
| | Love | The person you love, the object of love, or the object of marriage. | *Li Qingzhao*, a native of Zhangqiu in Qizhou of the Song Dynasty, was named Yi'an Ju Shi. She was *the daughter of Li Gefei* and *the wife of Zhao Mingcheng*. | Li Gefei, Li Qingzhao, love; Li Qingzhao, Zhao Mingcheng, love |
| | Friendship | A friend is a person who has a relationship with each other, people who want to be good to each other. | *Bai Juyi* is as famous as *Yuan Zhen* in poetry and is known as *Yuan and Bai* in the world. In his later years, he wrote poems with *Liu Yuxi*, also known as *Liu and Bai*. | Bai Juyi, Yuan Zhen, friends; Bai Juyi, Liu Yuxi, friendship |
| | Teacher–student | Teacher and student. | *Zhou Dunyi* was the founder of Daoism, and *Cheng Hao and Cheng Yi* all *learned from him*. | Zhou Dunyi, Cheng Hao, teacher–student; Zhou Dunyi, Cheng Yi, teacher–student |
| | Admiration | Adore and show appreciation for each other. | *Luo Dunyan*, a native of Shunde, Guangdong Province in the Qing Dynasty, was a scholar of Confucianism. He *wrote some Confucianism books* such as "Ji Yi Bian" *The Collection of Confucius*. | Luo Dunyan, Confucius, admiration |
| | Admiration | Adore and show appreciation for each other. | According to Feng Zhi's *Yun Xian San Lu*, the late Tang poet *Zhang Ji* was *infatuated* with *Du Fu's poems*. | Zhang Ji, Du Fu, admiration |
| | Partner | People who are in charge of the same thing or work with each other. | *Cui Xiyi* once served as the Hexi Frontier Envoy, and *Wang Wei assisted him*. | Cui Xiyi, Wang Wei, partner |
| Poetry | Friendship | A friend is a person who has a relationship with each other, people who want to be good to each other. | Peach Blossard Pond water is 1,000 ft deep, not as deep as *Wang Lun* to send me affection.—*Li Bai*, *To Wang Lun*. | Li Bai, Wang Lun, friendship |
| | Admiration | Adore and show appreciation for each other. | Only by drinking can we be relieved, only by reciting poetry can we contemplate. Only *Tao Qian* can understand me. It's a pity that I was born later than you.—*Du Fu*, *Pity*. | Du Fu, Tao Qian, admiration |

**Algorithm 1.** Poet Name Disambiguation Strategy.

**Input:** The name of poet, $name$; The beginning of the poet's dynasty, $T_{ds}$; The end of the poet's dynasty, $T_{de}$;
**Output:** Aliases of the poet, $result$;

List of Candidates, $S$
Search $name$ in CBDB database to get a list of people with $name$ $\{x_1, x_2 \ldots, x_n\}$, and the time of each person's birth $T_{is}$ and death $T_{ie}$;
**for** $x_i$ **do**
  **if** $T_{ds}$<$T_{is}$ or $T_{ie}$<$T_{de}$ **then**
    Add $x_i$ to $S$;
  **else**
    continue;
**if** the length of $S$ is 1 **then**
  Search aliases of the $x_i$ in CBDB, and we assign it to $result$
**else**
  Find fail, the poet is thought to have no alias
**return** $result$

*Note*. CBDB = Chinese Biographical Database.

*Nodes building.* A node in poets' social network represents a poet. In order to extract the relationship between poets completely and accurately, the first step is to match the poet's name with his alias. We use the CBDB to eliminate the ambiguity of poets' names. CBDB consists of many tables, each of which records different information of characters. We only use two tables, the main information table of characters—BIOG_MAIN—and the alias table of characters—ALT *name* _DATA. At first, the person number c_personid is found out from BIOG_MAIN, and then, the alias is found out from ALT *name* _DATA by using c_personid. Because there are too many historical figures collected by CBDB and the phenomenon of duplicate names is very serious, the following strategies are used to eliminate the ambiguity of poets' names for every poet. According to Algorithm 1, the ambiguity of the poet's name is eliminated and the aliases of the poet are matched:

Since there is only information about historical figures after the seventh century in CBDB, we also use named entity recognition technology to identify the poets' alias and profiles in Wikipedia. Taking poet Bai Juyi as an example, he has seven aliases—Wen ("文"), Bai Fu ("白傅"), Bai Wengong ("白文公"), Zui Yin Xian Sheng ("醉吟先生"), Xiang Shan Ju Shi ("香山居士"), Le Tian ("乐天"), and Bai Er Shi Er ("白二十二").

Further data cleaning is needed for the extracted poet. First of all, different poets may have the same alias. For example, the poet Bai Juyi and the poet Pi Rixiu (皮日休) both have the alias "Zui Yin Xian Sheng," which is removed for accuracy. Second, some poets' aliases belong to common words in poetry. For example, "Haoran" ("浩然") and "Qianli" ("千里"), which are very common in ancient poetry and mostly related to poets, so such aliases will be removed. In addition, we also removed the aliases with only one Chinese character because it is difficult to judge whether a character is a poet's alias or an ordinary character in poetry. After the above data cleaning process, poets and their aliases are saved for the next step to extract the reference relationship between poets.

*Edges building.* The social relations among poets are extracted from two kinds of data: poetry and biographies of poets. Biographies of poets record some of the poets' life experiences and often record their social relations. For example, according to Feng Zhi's *Yun Xian San Lu*, Zhang Ji, a poet in the late Tang Dynasty, once burned Du Fu's poems because he was infatuated with them. He mixed the ashes with honey and ate three spoonfuls of them every morning. From such data, we can extract the relationship of admiration between Zhang Ji and Du Fu. Because biographies of poets are often long paragraphs, it is difficult to extract the relationship directly. The processing process of biographical information of characters is as follows. First, we split the paragraphs into sentences and clean the data. If the name of Poet B appears in a sentence of the profile of Poet A, this sentence is entered into the relational extraction model and an undirected edge is constructed between A and B. Such social networks can reflect certain poets' social relationships.

However, edges that are built by biographies without weight cannot reflect the strength of different social relationships. For example, Lu Guimeng (陆龟蒙) and Pi Rixiu mentioned each other more than 100 times in their poems, so they are much closer than other people. Therefore, we use the above list of poets and their alias to search for the citation relationship between poets in all collected poems. The rule is if the poem mentions the other party in the title and body of the poem, the citation relationship between the two parties is added 1. If the poem mentions the other party more than once, only one citation is counted. Finally, we take the number of reference relations as the weight of the edges in the social network to get a social network.

## Node Influence Analysis

*Node importance.* The relevant people information of each poet comes from *Great Dictionary of Chinese Names through the Ages*, *The Collection and Evaluation of Tang Poetry*, *All Song Poetry*, and other books. If the name of Character B appears in the profile of Character A, then Character B becomes the relevant people of A, and A is also the relevant people of B. Literary activity is not the behavior of isolated individuals but the interaction of groups (Wang & Shao, 2020). The purpose of constructing poets' social network is to observe the development of literature more carefully. Modern people generally think that Li Bai and Du Fu are poets with the highest literary achievements in the Tang Dynasty. Who were the most influential poets in the eyes of the Tang Dynasty? We need to return to the historical site for dynamic investigation and restoration. Sociologist Rashotte (2007) defined influence as the phenomenon that an individual changes his own thoughts, feelings, attitudes, or behaviors when interacting with others or groups. Node influence is what causes a person's behavior to change when he communicates with people who are better than him or who share similar interests. With the help of the node influence analysis of poets' social networks, literary lovers can quickly understand the literary style and social outlook of a certain period.

The emergence of social networks provides a quantitative basis for defining and studying node influence, and the ranking index of node importance in social networks can be used to measure node influence. Degree centrality (Bonacich, 1972), betweenness centrality (Freeman, 1977; Newman, 2005), close centrality (Sabidussi, 1966), and clustering coefficient (Ren et al., 2013) of nodes all can measure node influence to a certain extent. K-shell decomposition (Chen et al., 2013) is also an index to measure the importance of nodes, where the main idea is to decompose nodes in topological structure into different levels from edge to core. In addition, PageRank (Berkhin, 2005; Bryan & Leise, 2006), Hyperlink-Induced Topic Search (Kleinberg, 1999), Leader Rank (Lü et al., 2011), and other random walk algorithms distinguish the influence of nodes by the sorting result of the node scores.

Topological structure in complex network is a recognized index to measure network structure. Topological structure can describe the influence of nodes from a macro level, which is easy to obtain. Therefore, it is a common practice to measure the influence of nodes by topological structure. However, there are some limitations in measuring the importance of nodes with a single index, so we use the entropy weight method (EWM; Y. Yang et al., 2019; Zou et al., 2006) to comprehensively consider a variety of node importance ranking indexes to measure the influence of poets.

*TSEWM.* According to the definition of information entropy, entropy value can be used to judge the dispersion degree of a certain index. The smaller the information entropy value is, the greater the

**Table 2.** The Evaluation Index of Entropy Weight Method.

| Factors of Influence Evaluation | Types of Influence | Evaluation Index | Equation | Definition |
|---|---|---|---|---|
| Network structure information | Directness | Degree centrality | $DC(i) = \frac{k_i}{n-1}$ | $k_i = \sum_i a_{ij}$, $a_{ij}$ is the element in row $i$ and column $j$ of the adjacency matrix, $n$ is the number of network nodes, and the denominator $n - 1$ is the maximum possible value of nodes. |
| | Indirectness | Closeness centrality | $CC(i) = \frac{1}{d_i} = \frac{n-1}{\sum_{j \neq i} d_{ij}}$ | $d_i = \frac{1}{n-1} \sum_{j \neq i} d_{ij}$, $d_i$ represents the shortest average distance between node $v_i$ and other nodes in the network. The smaller $d_i$ is, the closer node $v_i$ is to other nodes in the network. Therefore, the reciprocal of $d_i$ is defined as the proximity centrality of node $v_i$ |
| | | Betweenness centrality | $BC(i) = \sum_{i \neq s, i \neq t, S \neq t} \frac{g_{st}^i}{g_{st}}$ | $g_{st}$ is the number of shortest paths from node $v_s$ to $v_t$, and $g_{st}^i$ is the number of shortest paths from node $v_s$ to $v_t$ through $v_i$. |
| | | Eigenvector centrality | $EC(i) = x_i = c \sum_{j=1}^{N} a_{ij} x_j$ | $c$ is a proportionality constant, $x = [x_1, x_2, \ldots, x_N]^T$, and it can be written in matrix form when we reach steady state after many iterations: $x = cAx$, where $A = (a_{ij})$ is the adjacency matrix of the network. $X$ is the eigenvector for the eigenvalue $c^{-1}$ of matrix $A$. |
| Node information | Directness | Number of poets' poetry | $P(i)$ | Number of poet's work |

dispersion degree of the index is, and the greater the influence (i.e., weight) of the index on the comprehensive evaluation will be. If all the values of an index are equal, the index will not play a role in comprehensive evaluation. Therefore, information entropy can be used to calculate the weight of each index, which provides a basis for multi-index comprehensive evaluation.

As for network structure information, we measure the poets' influence network from two perspectives of directness and indirectness. In terms of directness, we use degree centrality to measure the poet's direct influence. The higher the degree centrality of nodes, the greater the number of nodes directly affected by the poet and the higher the direct influence of the poet. In terms of indirectness, we use closeness centrality, betweenness centrality, and eigenvector centrality to measure the indirect influence of poets. Closeness centrality refers to the sum of all the shortest distances that can establish an indirect influence relationship with a poet. The higher the value is, the more significant the indirect influence is. Betweenness centrality refers to the sum of times that a poet acts as a bridge in the indirect influence between other nodes, and the higher the value, the more significant the influence of the bridge. Eigenvector centrality means that the importance of a poet depends not only on the number of its neighbors but also on the importance of its neighbors. The higher the value is, the more significant the indirect influence is. Besides, for the node information, we mainly use the quantity information of the poet's poetry we collected. Generally speaking, the greater the number of a poet's poetry, the higher his direct influence. The specific calculation method of the evaluation index is shown in Table 2.

The influence of social network nodes is dynamic, but most of the existing studies examine the influence of users on the static network topology. However, as the social network topology of users changes, so does their influence. The formation of poets' influence is a long process, in other words, their influence is the result of continuous accumulation and continuation. With the change of time, the poets' social network is changing, and the poets' influence is also changing. The method of calculating influence based on network snapshot is to calculate node influence according to static network topology in principle, which cannot save the network structure information of the previous moment, is sensitive to data changes. Hence, based on the constructed poets' social and influential network above, we propose the TSEWM to analyze the evolution of poets' influence. TSEWM can reduce the sensitivity of entropy weight and poets influence change, which enables poets to inherit their previous influence and enhance the robustness of the model. This method comprehensively evaluates the influence of the poet from the two dimensions of the network structure and the poet's own information. The main evaluation indicators involved are degree centrality, proximity centrality, betweenness centrality, eigenvector centrality, and the number of poets' poems. The specific algorithm process is as follows:

1. Data extreme standardization:

$$X_{ij}(t) = \frac{x_{ij}(t) - \min\{x_{1j}(t), \ldots, x_{nj}(t)\}}{\max\{x_{1j}(t), \ldots, x_{nj}(t)\} - \min\{x_{1j}(t), \ldots, x_{nj}(t)\}}. \tag{1}$$

The $x_{ij}(t)$ denotes the original value of index $j$ of poet $i$ in different dynasty $t$. The $X_{ij}(t)$ denotes the value of index $j$ of poet $i$ in different dynasty $t$ after data extreme standardization. The total number of poets in different dynasty $t$ is $n(t)$, $(i = 1, 2, \ldots, n(t); j = 1, 2, 3, 4, 5)$.

2. Calculate the evolutionary weight $W_j(t)$ of index $j$ in different dynasty $t$ by information entropy:

$$\begin{cases} e_j(t) = -\dfrac{1}{\ln(n(t))} \sum_{i=1}^{n(t)} p_{ij}(t) \ln(p_{ij}) 1.5ex, \\ p_{ij}(t) = x_{ij}(t) / \sum_{i=1}^{n(t)} x_{ij}(t) 1.5ex, \\ w_j(t) = \dfrac{1 - e_j(t)}{\sum(1 - e_j(t))} 1.5ex, \\ \Delta w_j = w_j(t) - W_j(t-1) 1.5ex, \\ \mathrm{u}_j = \dfrac{\ln \Delta \mathrm{w}_j}{\sum \ln \Delta \mathrm{w}_j} 1.5ex, \\ W_j(t) = w_j(t) - u_j \cdot \Delta w_j 1.5ex, \\ W_j(0) = w_j(0) = 0. \end{cases} \tag{2}$$

The $e_j(t)$ denotes the information entropy value of index $j$ in different dynasty $t$. $w_j(t)$ denotes the original weight of index $j$ in different dynasty $t$ without evolution. $W_j(t)$ denotes the weight of index $j$ in different dynasty $t$ after considering the time factor.

**Table 3.** Nodes and Edges Information of Poet Social Network.

| Dynasty | Nodes | Edges (Same Dynasty) | Edges (Different Dynasty) | Edges |
|---|---|---|---|---|
| XianQin | 210 | 208 | 81 | 290 |
| Qin | 43 | 42 | 101 | 143 |
| Han | 615 | 690 | 352 | 1,042 |
| WeiJin | 505 | 594 | 743 | 1,337 |
| Sui | 240 | 231 | 451 | 682 |
| Tang | 4,952 | 11,629 | 2,796 | 14,425 |
| NanBei | 961 | 1,351 | 470 | 1,821 |
| Song | 8,958 | 25,004 | 4,499 | 29,503 |
| Liao | 26 | 25 | 8 | 33 |
| Jin | 615 | 262 | 353 | 705 |
| Yuan | 1,707 | 2,855 | 2,208 | 5,063 |
| Ming | 7,446 | 27,060 | 7,025 | 34,085 |
| Qing | 11,253 | 23,508 | 8,135 | 31,643 |
| Sum | 37,531 | 93,459 | 27,222 | 120,772 |

3. Calculate the influence value $influence_i(t)$ of poet $i$ in different dynasty $t$.

$$Influence_i(t) = \sum_{j=1}^{5} W_j(t)œXij(t). \tag{3}$$

# Results

## Visualization of Social Networks

Table 3 shows the data information of the poets' social network after the data processing in the above steps. Apart from having social connections with people in their own dynasties, poets also have a long-term impact on later generations. Our poets' social network data include some of the poets' long-term influence on later generations. The influence of poets over a long period of time can be measured quantitatively according to these data. Figure 2 uses Confucius as an example to show how he connected with people in other dynasties.

Although the social network constructed is not fully connected, the maximum connected subgraph covers 99.7% of nodes, so it is possible to use the maximum connected subgraph to measure the influence of poets. The following data results are calculated using the maximum connected subgraph. Table 4 shows some characteristic parameters of the network. The average shortest path length of the network is close to 6, indicating that the six degrees theory is also applicable to the social network of ancient poets. It shows that the social network of ancient poets is also a small-world network.

*Gephi* (https://gephi.org/) is an open-source network analysis and visualization package, which is often used in complex network research projects. The social network can be visually analyzed by importing the edge and node information of the constructed network into Gephi. Figure 3 shows the social networks of poets from the Pre-Qin Dynasty to the Qing Dynasty. For clear demonstration, only nodes with a degree of 10 and above are shown in the figure. We add a birth time attribute for each poet, resulting in a social network of poets that grows over time. It can be seen that the four dynasties of Tang, Song, Ming, and Qing accounted for more than 80% of the network, and the poets of the Tang Dynasty (indicated by orange) occupied a large part in the picture, with the most

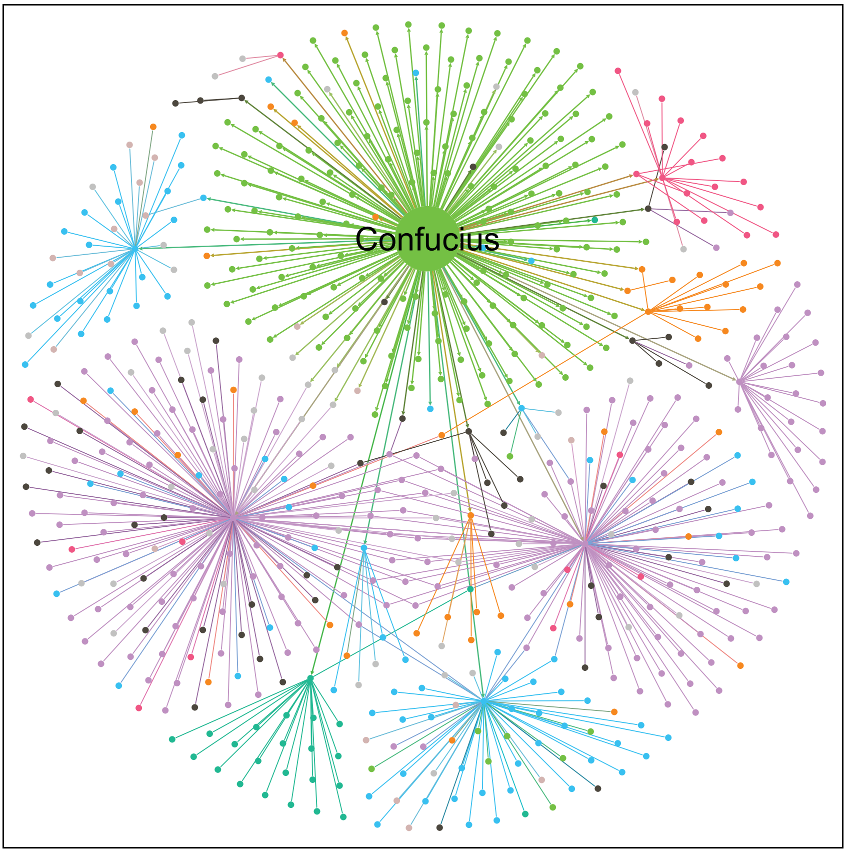


**Figure 2.** Confucius' social graph. *Note.* The green nodes in the picture represent the people who have direct social contact with Confucius in the same Dynasty, and the other colored nodes represent the people who have contact with Confucius in other dynasties.

**Table 4.** Topological Parameter of the Poets' Social Network.

| Type of Network | Number of Nodes | Number of Edges | Sparsity | Average Degree | Average Clustering Coefficient | Average Shortest Path Length |
|---|---|---|---|---|---|---|
| Network | 41,310 | 114,697 | .000134 | 5.553 | .057 | — |
| Maximum connected subgraph | 40,359 | 114,156 | .00014 | 5.657 | .058 | 6.454 |

complicated social relations. In addition, poets of the Tang Dynasty were most fond of using poetry as a social tool, and the two poets with the closest contacts are Lu Guimeng and Pi Rixiu.

## *The Ranking of Poets' Influence*

After calculating the influence of each dynasty's poets by the EWM, we finally got a ranking result as shown in Figure 4. From the ranking results, the overall influence of poets in the Tang and Song

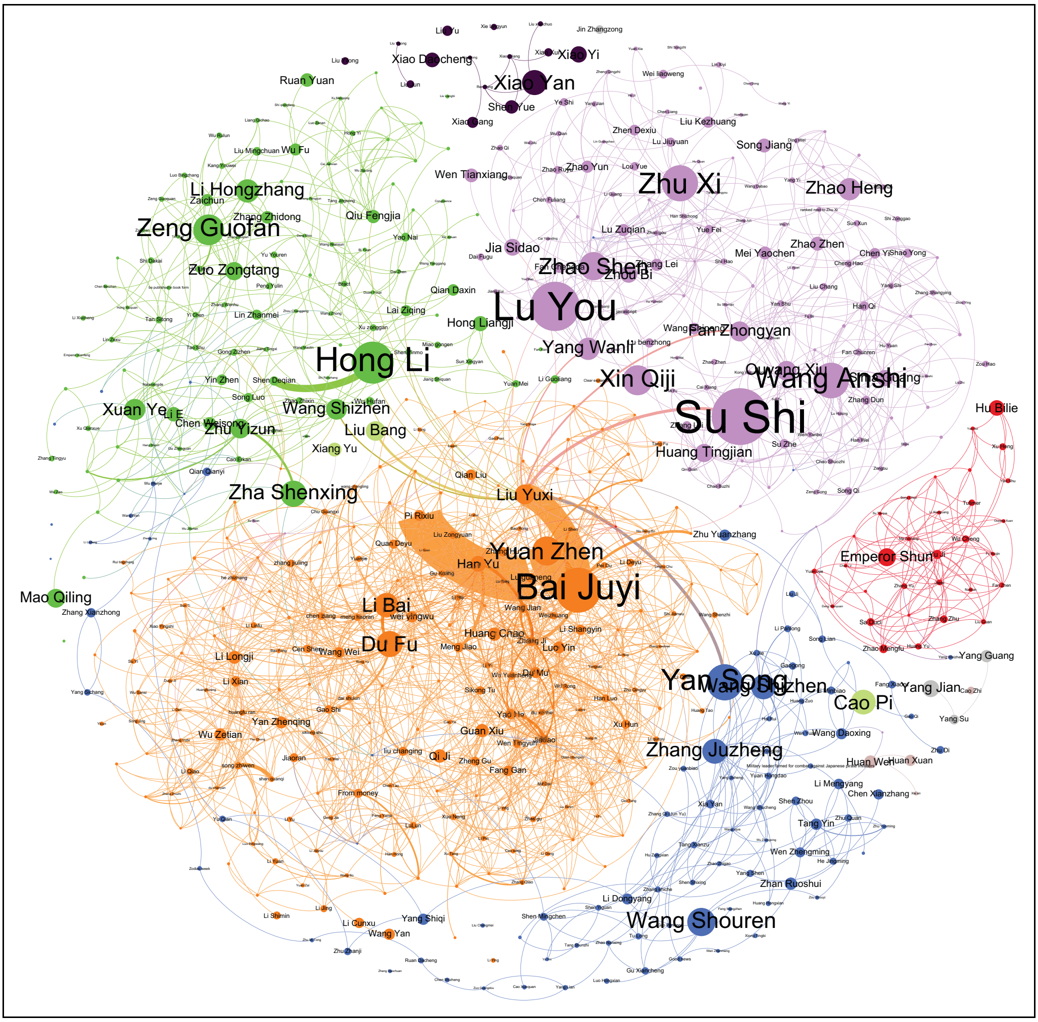


**Figure 3.** The social network of poets. *Note*. The different colors of the nodes in the picture represent the different dynasties of the poets (orange: Tang Dynasty; purple: Song Dynasty; blue: Ming Dynasty; green: Qing Dynasty), the size of the node indicates the poets' influence, and the weight of the edges indicates how closely the two poets communicated with each other.

Dynasties is at a high level. Generally speaking, the larger nodes are the more famous poets in history. But the figure also shows some poets whose fame was not so great in later generations. They made certain achievements in their own time. But later generations often ignore them, which attracts us to get to know these poets in history. We tend to overlook some poets' achievements because of their special status. For example, Yan Song (严嵩) was regarded as a treacherous court official in history, but after our analysis, we find that he also has a certain status in the poetry circle; Hong Li (弘历) was a famous emperor in history, but we often ignore that he also created a large number of poems. For example, Meng Haoran (孟浩然) did not like to be an official. Therefore, although he was regarded as a famous poet of pastoral poetry by later generations, he was not a popular poet at that time.

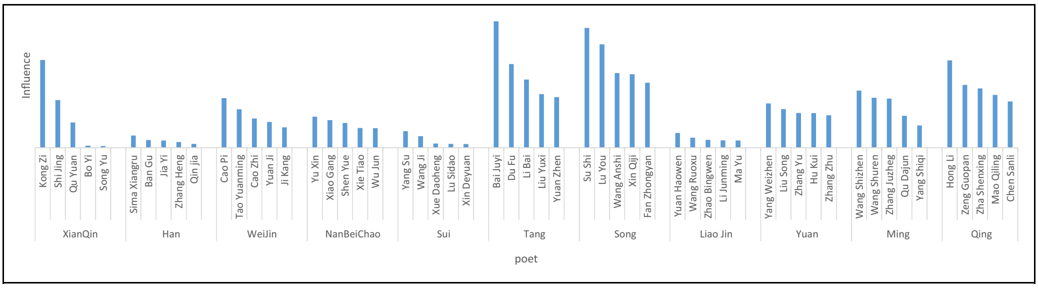


**Figure 4.** Ranking of poets' influence in different dynasties. *Note*. The top five influential poets of different dynasties are selected. The abscissa is the poet's name, which is sorted by dynasty time from left to right, and the ordinate represents the poets' influence. Because of the short period of Qin Dynasty and the backward development of literature, it is not shown here.

## *Evolution of Poets' Influence*

The evolutionary process of a poet's influence over time can be analyzed by TSEWM. Figure 5 shows the evolution of the influence of the Confucian master Confucius (□□) over time. In the calculation process, the time is calculated in the unit of dynasty. In order to make the result more accurate, the time unit can be smaller. The influence of Confucius reached its lowest point in the Qin dynasty, which was related to *the burning books and burying Confucianism* (The rulers of the Qin Dynasty suppressed Confucianism represented by Confucius; Jiang, 2007) in history. In the Han Dynasty, Emperor Wudi carried out the policy of "banishing all schools of thought and respecting Confucianism only" (Jiang, 2007; Emperor Wudi of the Han Dynasty advocated Confucianism represented by Confucius), which made Confucianism represented by Confucius become popular again. Then, the influence of Confucius gradually rose to the highest point in the Song dynasty, indicating the continuous development of Confucianism reached the peak in the Song dynasty. Literature flourished during the Song Dynasty, with many thinkers like Zhu Xi (朱熹), pushing Confucianism to new heights.

According to the time points in history, we divided the Tang Dynasty into four periods: the early Tang Dynasty, the prosperous Tang Dynasty, the middle Tang Dynasty, and the late Tang Dynasty, and analyzed the changing trend of the influence of crucial poets in different times. The development of the Tang poetry was closely related to the prosperity of the dynasty. It can be seen that the Tang poetry was formed in the early Tang Dynasty and reached its peak in the prosperous and middle Tang Dynasties as shown in Figure 6. With the decline of the Tang Dynasty, the influence of poets in the late Tang Dynasty was generally low. Generally speaking, most poets were most influential in their time. New poets were constantly emerging, and only those poems that truly reflected the social and historical faces of the time could have a high influence, such as Bai Juyi, whose poems reflected the human suffering in the late Tang Dynasty, and thus could maintain a high influence.

Generally speaking, the Song poetry can be divided into two schools: *the graceful and restrained school* and *the bold and unconstrained school*. We analyzed the influence of the representative poets of each school. Figure 7 sorts the poets' influences of the two schools in the Song Dynasty, among which the top three are poets of the bold and unconstrained school, and their influences far exceed those of the following poets of the graceful and restrained school. But poets of the graceful and restrained school are more numerous and more popular with the public. Figure 8 analyzes the influence of poets of the graceful and restrained school and the Bold and unconstrained school in different periods of the Song Dynasty. The influence of each school in each period is calculated by the sum of the influences of the representative poets of each school. From the results, the peak value of the bold and unconstrained school is higher than that of the graceful and restrained school, but the

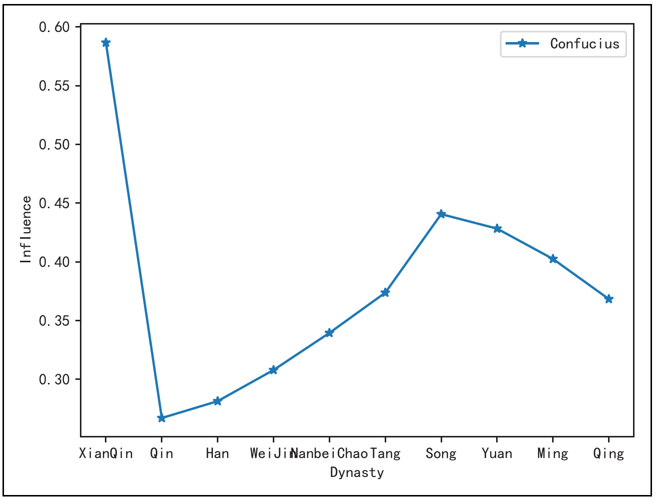


**Figure 5.** Evolution of Confucius' influence over time.

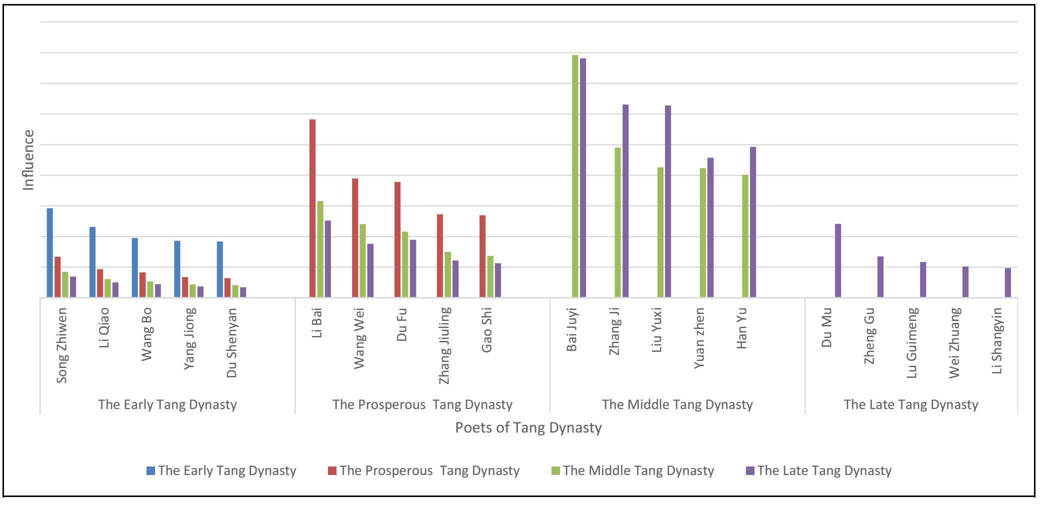


**Figure 6.** The evolution of poets' influence in different periods of the Tang dynasty. *Note*. The abscissa divides the Tang Dynasty into four periods, and the ordinate represents the poets' influence. The figure shows how the influence of the poets in each period has changed over time.

graceful and restrained school started earlier. The style of the Song poetry was originally graceful, to a certain extent, and the bold style was developed on the basis of the graceful style.

Although the poets of the Song Dynasty can be mainly divided into the graceful and restrained school and the bold and unconstrained school, the poets of these two schools are not completely separated from each other, and there are complex social connections between them. Figure 9 shows the social connection between them and also shows that the poetry of these two schools was constantly developing in the process of mutual communication.

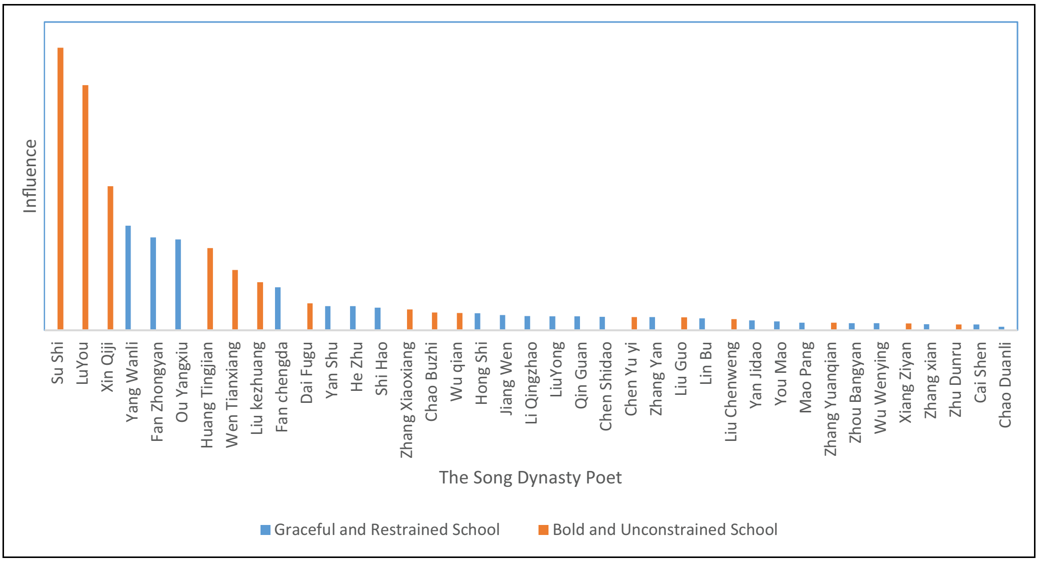


**Figure 7.** The influence of the graceful and restrained poets and the bold and unconstrained poets in the Song dynasty. *Note*. The abscissa is the poet's name, and the ordinate is the poet's influence. Blue stands for *the graceful and restrained school*, and orange stands for the *bold and unrestrained school*.

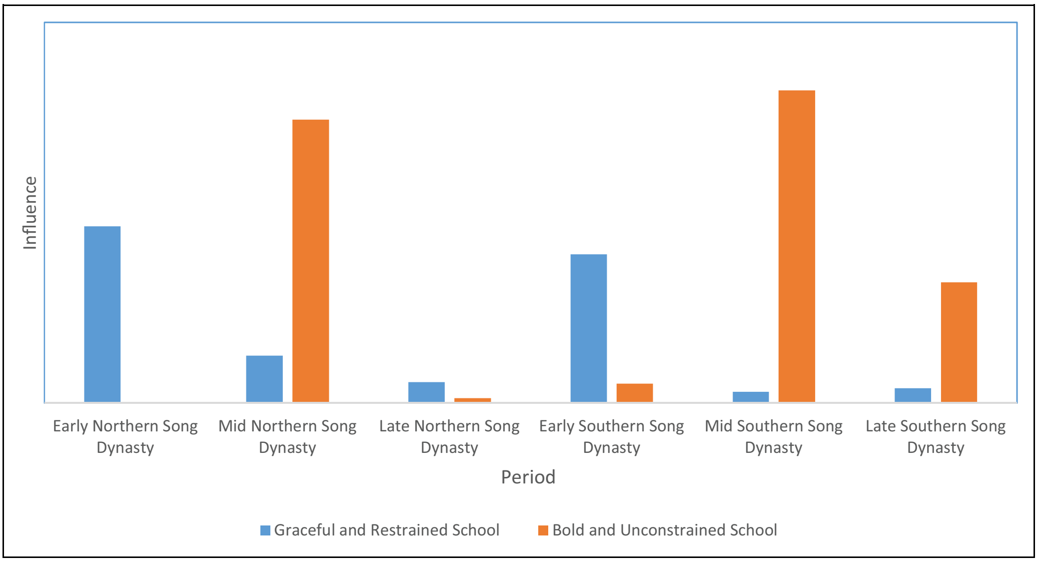


**Figure 8.** The evolution of the influence of *the graceful and restrained school* and the *bold and unrestrained school* in different periods of the Song Dynasty. *Note*. Blue stands for the graceful and restrained school, and orange stands for the bold and unrestrained school.

## *Verification Experiment*

In the research of node importance, the ranking results of node importance are usually evaluated by using the propagation dynamics model (Han et al., 2017), and the susceptible–infected (SI) model

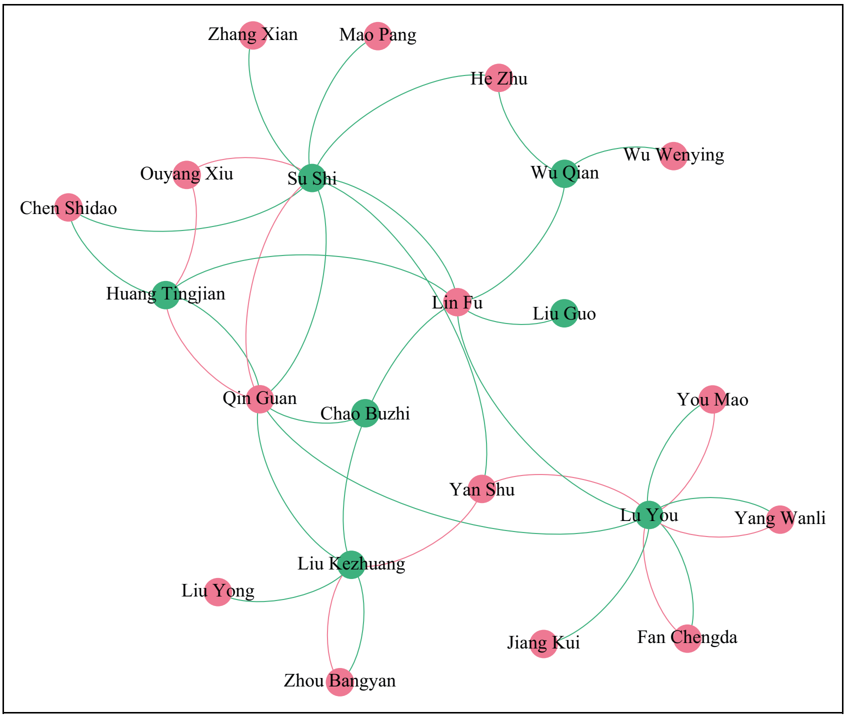


**Figure 9.** The social connection between poets of the graceful and restrained school and poets of the bold and unconstrained school. *Note.* The green node represents the bold and unconstrained school, and the pink node represents the graceful and restrained school.

(Shi et al., 2009; Zhou et al., 2006) is one of them. In the SI model, a node has two states: S (susceptible) indicates that the node may be infected by an infected node among its neighbors, and I (infected) indicates that the node is infected. Once a node is infected, it will always be infected. In the SI model, there are generally two criteria for evaluating the importance of nodes: one is the average propagation range of nodes, that is, the number of infected nodes when reaching steady state. The larger the number is, the more important the nodes are. The other is propagation rate, which can be judged from the time taken to reach the steady state of infection. The less time it takes, the more important the node is.

We have conducted three different experiments on social network data sets of four different dynasties. The first two groups of experiments evaluated each measurement method from two aspects of propagation range and propagation rate to verify the accuracy of the proposed algorithm in measuring the importance of nodes. The third group of experiments is the ablation experiment. The experiment is carried out with the EWM without time series as the control, which verifies the scientific of TSEWM. Table 5 shows the data set information. The comparison methods in the experiment are degree centrality, PageRank, and k-shell.

*Experiment 1*. By observing the propagation range and propagation rate of the SI propagation curve, the accuracy of measuring the node importance of each method is evaluated. In the four networks, the measures of node importance are sorted in descending order according to each method, and the first 1% nodes are selected as the initial infection source nodes. The experimental results are shown in Figure 10, in which the abscissa represents the number of time iteration steps, and the ordinate

**Table 5.** Data Set Information.

| Data Set | Node Number | Edge Number |
|---|---|---|
| Han Dynasty | 601 | 689 |
| Tang Dynasty | 4,830 | 8,713 |
| Song Dynasty | 6,763 | 10,293 |
| Qing Dynasty | 6,231 | 7,749 |

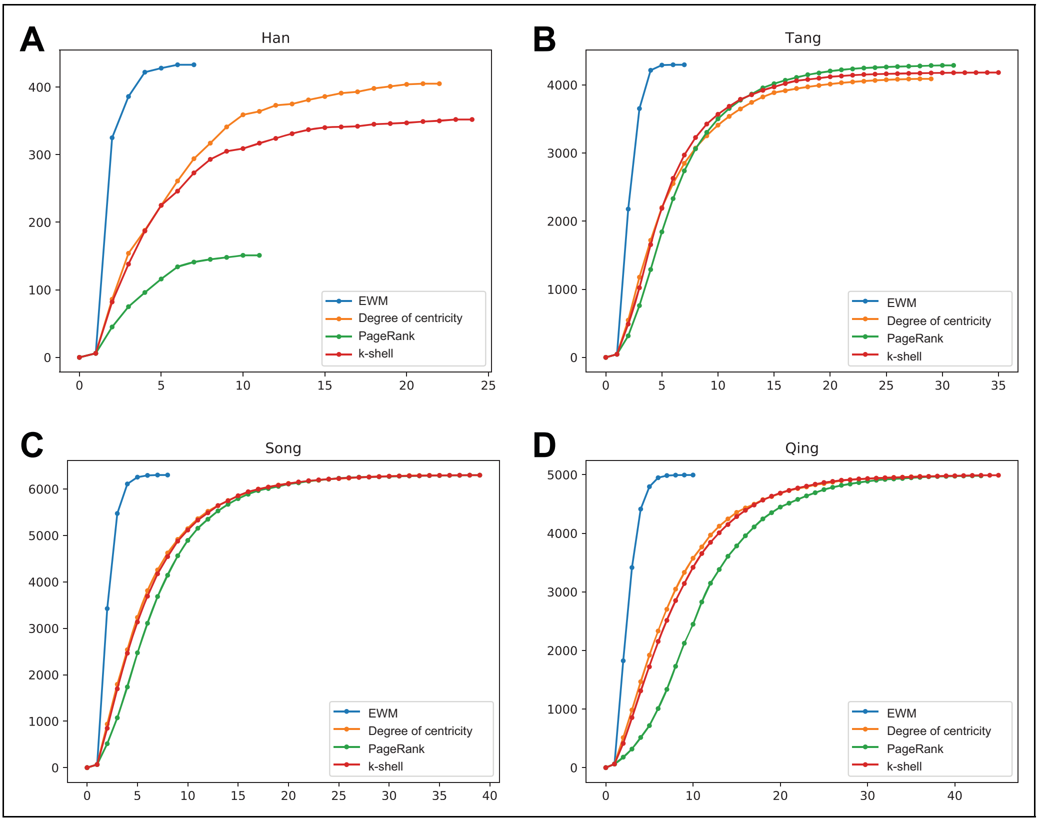


**Figure 10.** Susceptible–infected transmission curve when the top 1% nodes obtained by each measurement method are the nodes of the initial source of infection. *Note*. The abscissa represents the number of time iteration steps, and the ordinate represents the number of infected nodes when reaching steady state, which illustrates the propagation range of infected nodes. Panel A: Han Dynasty. Panel B: Tang Dynasty. Panel C: Song Dynasty. Panel D: Qing Dynasty.

represents the number of infected nodes when reaching steady state, which illustrates the propagation range of infected nodes. The slope change of curve illustrates the propagation rate of infected nodes. Compared with other methods, EWM can show better results on four networks in both propagation time and propagation range.

*Experiment 2*. In order to further discuss the correlation between the measured values of nodes of each method and their transmission ability, and evaluate the accuracy of each method in measuring the importance of nodes, Experiment 2 sorts the importance of nodes according to some

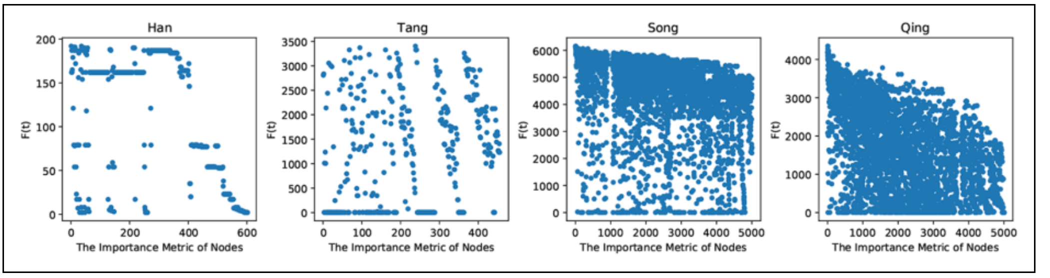


**Figure 11.** The correlation between degree centrality measure and number of infected nodes. *Note.* The abscissa represents the sorting of nodes, the ordinate represents the number of infected nodes when iteration steps is 5, and each data point represents a node in the network.

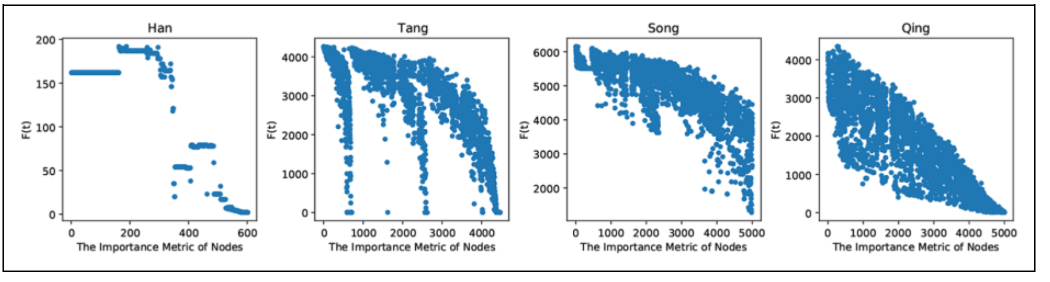


**Figure 12.** The correlation between PageRank measure and the number of infected nodes. *Note.* The abscissa represents the sorting of nodes, the ordinate represents the number of infected nodes when iteration steps are 5, and each data point represents a node in the network.

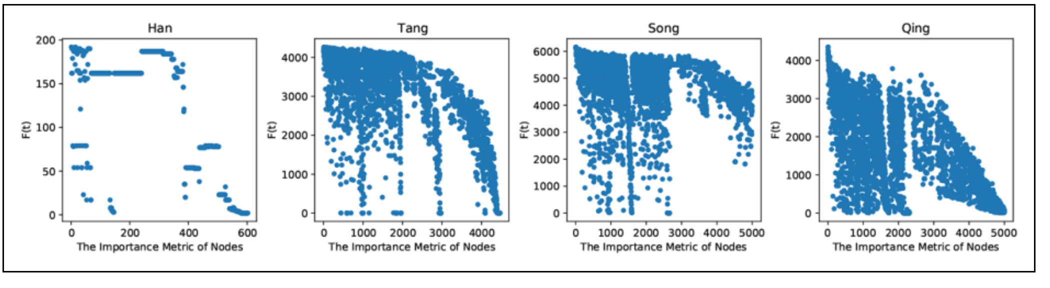


**Figure 13.** The correlation between k-shell measure and the number of infected nodes. *Note.* The abscissa represents the sorting of nodes, the ordinate represents the number of infected nodes when iteration steps are 5, and each data point represents a node in the network.

measurement method and then studies the correlation between node ranking and the number of node infections in the process of SI propagation. In this experiment, the importance of nodes is measured by calculating the propagation range of nodes after the same number of iterations. Theoretically, the images of better performance metrics should show a decreasing relationship. In other words, as the importance of nodes decreases, the number of corresponding infected nodes will also decrease. In this experiment, the number of iteration steps is set at $t = 5$. Figures 11–14 show experiment results of different methods on four data sets, in which the abscissa represents the sorting of nodes, the ordinate represents the number of infected nodes when $t = 5$, and each data point represents a node in the network. The experimental results show that the images of EWM show obvious negative

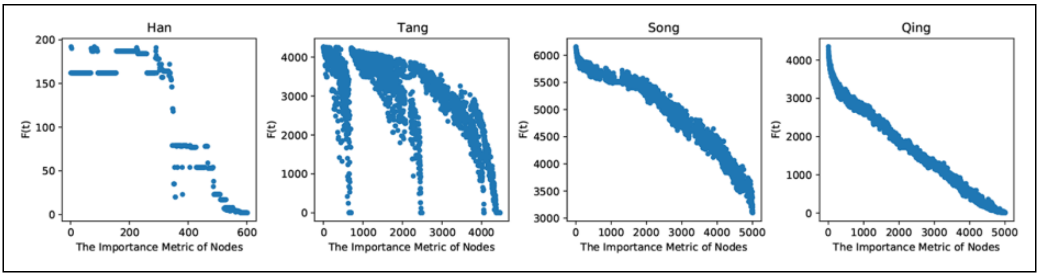


**Figure 14.** The correlation between entropy method measure and the number of infected nodes. *Note*. The abscissa represents the sorting of nodes, the ordinate represents the number of infected nodes when iteration steps are 5, and each data point represents a node in the network.

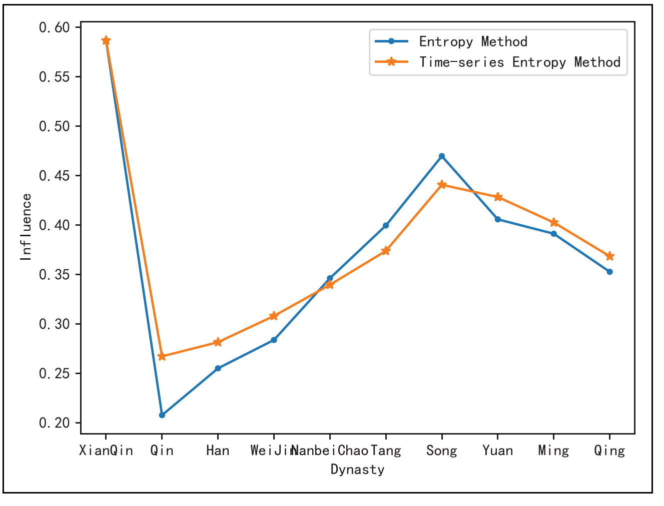


**Figure 15.** Comparison chart of influence evolution calculation based on entropy weight method (EWM) and time-series entropy weight method. *Note*. The blue line in the figure represents the evolution process of Confucius' influence calculated by the unmodified EWM, and the orange line represents the evolution process of Confucius' influence calculated by the EWM with time series.

correlation trend on four data sets, and the larger the data set range is, the more obvious the negative correlation trend is. PageRank and k-shell methods perform slightly worse.

*Experiment 3*. In order to verify the scientific of TSEWM, we take the EWM without time series as the control. Still taking the evolution of Confucius' influence as an example, the experimental results are shown in Figure 15. From the experimental results, the TSEWM can make the poets' influence curve more stable, keep the continuity of the influence of poets, and make the change of influence cumulative.

## Conclusions

In this article, a social network of ancient Chinese poets spanning more than 2,000 years is constructed by using the complex network and natural language processing technology. Based on this network, we can study how poets of different dynasties and different schools relate to each other. Besides, the evolution process of the poet's influence is analyzed.

First, we integrate multisource data and information about poets and poetry and collect them into a structured database. Then, we eliminate ambiguity of poets' names through poet alias mapping. Finally, we build a complete poet social network by querying the citation relationship among poets and the results of relation extraction. This network includes poets of all dynasties in Chinese history, which is the most complete network of poets with the largest span of time. Based on the network topology and node information, we quantitatively calculate the influence of poets. The TSEWM used in this article can take more factors into account in the measurement model of node influence, making the model more close to the influence in the real society and improving the effect of the model on the premise of low time complexity.

In addition, we propose a new model to calculate the evolution of poets' influence, which can also be applied to society, culture, and other aspects and has very important theoretical and practical values. We propose TSEWM to comprehensively evaluate the evolution of poets' influence and use SI propagation model to evaluate the algorithm model and verify the accuracy of the ranking results of influence. The research and modeling of the evolution of poet influence can help people to further understand the evolution of poets' individual and group behaviors in social networks, the changes of ancient philosophy, the development of different poetry schools, and even the analysis of the appearance of ancient society.

Our future work will focus on two aspects. On the one hand, expand the scope of the data set and integrate other data sources to analyze the location information of the works such as poet biography and local chronicles. On the other hand, the research of this article mainly focuses on the influence of individuals on the whole, and measuring the influence and influence mechanism of poets among groups will be our next research goal. Our ultimate goal is to build a complex network of ancient literature based on time and space, which is convenient for people to use it to study the development of ancient literature in depth.

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by the National Natural Science Foundation of China (No. 61972047), National Key Research and Development Program of China (2018YFC0831500), and NSFC-General Technology Basic Research Joint Funds (U1936220).

## References

Berkhin, P. (2005). A survey on PageRank computing. *Internet Mathematics*, *2*(1), 73–120.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*(1), 113–120.

Bryan, K., & Leise, T. (2006). The $25,000,000,000 eigenvector: The linear algebra behind Google. *SIAM Review*, *48*(3), 569–581.

Chen, D. B., Gao, H., Lü, L., & Zhou, T. (2013). Identifying influential nodes in large-scale directed networks: The role of clustering. *PLoS ONE*, *8*(10), e77455.

Chowdhury, T., Muhuri, S., Chakraborty, S., & Chakraborty, S. N. (2019). Analysis of adapted films and stories based on social network. *IEEE Transactions on Computational Social Systems*, *6*(5), 858–869.

Dong, A., & Liu, X. (2020, February). A statistical method for constructing tang poet social networks. *2020 IEEE 14th international conference on semantic computing (ICSC)* (pp. 101–107). https://ieeexplore.ieee.org/document/9031489

Fraiberger, S. P., Sinatra, R., Resch, M., Riedl, C., & Barabási, A. L. (2018). Quantifying reputation and success in art. *Science*, *362*(6416), 825–829.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*(1), 35–41.

Han, Z., Chen, Y., Liu, W., Yuan, B., Li, M., & Duan, D. (2017). Research on influence analysis of social network nodes. *Journal of Software*, *28*(1), 84–104.

Jian, O. (2016). Large-scale visual analysis and mining of ancient texts for digital humanities research. *Journal of Library Science in China*, *42*(02), 66–80.

Jiang, L. (2007). From "burning books and burning Confucianism" to "adoring huang-lao" and then to "respecting Confucianism alone"—academic thought and imperial civilization in Qin and Han Dynasties (in Chinese). *Zhejiang Social Sciences*, *2007*(1), 152–158.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, *46* (5), 604–632.

Lü, L., Zhang, Y. C., Yeung, C. H., & Zhou, T. (2011). Leaders in social networks, the delicious case. *PLoS ONE*, *6*(6), e21202.

Newman, M. E. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, *27*(1), 39–54.

Rashotte, L. (2007). Social influence. In G. Ritzer (Ed.), *The Blackwell encyclopedia of sociology.* Wiley.

Ren, Z. M., Shao, F., Liu, J. G., Guo, Q., & Wang, B. H. (2013). Node importance measurement based on the degree and clustering coefficient information. *Acta Physica Sinica*, *62*(12), 128901.

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, *31*(4), 581–603.

Schich, M., Song, C., Ahn, Y. Y., Mirsky, A., Martino, M., Barab ási, A. L., & Helbing, D. (2014). A network framework of cultural history. *Science*, *345*(6196), 558–562.

Shen, H. W., & Barab ási, A. L. (2014). Collective credit allocation in science. Proceedings of the National Academy of Sciences, *111*(34), 12325–12330.

Shi, X., Zhu, J., Cai, R., & Zhang, L. (2009). User grouping behavior in online forums. *In* Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 777–786). https://doi.org/10.1145/1557019.1557105

Wang, Z., & Shao, D. (2020). The preliminary practice and academic significance of digital humanism in the study of ancient literature (in Chinese). *Social Sciences China*, *8*(08), 108–129.

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, *566*(7744), 378–382.

Yang, H., & Wang, J. (2019). Construction and visualization of knowledge graph of academic teachers in Song Dynasty (in Chinese). *Data Analysis and Knowledge Discovery*, *3*(6), 109–116.

Yang, Y., Yu, L., Wang, X., Zhou, Z., Chen, Y., & Kou, T. (2019). A novel method to evaluate node importance in complex networks. *Physica A: Statistical Mechanics and its Applications*, *526*, 121118.

Zhou, T., Liu, J. G., Bai, W. J., Chen, G., & Wang, B. H. (2006). Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity. *Physical Review E*, *74*(5), 056109.

Zou, Z. H., Yi, Y., & Sun, J. N. (2006). Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment. *Journal of Environmental Sciences*, *18*(5), 1020–1023.

## Author Biographies

**Jiaqi Zhao** is a master candidate in Beijing University of Posts and Telecommunications. Her main research interests include knowledge graph, natural language processing, computational social science and complex network.

**Yuting Wei** is a doctoral candidate in Beijing University of Posts and Telecommunications. Her main research interests include knowledge graph, natural language processing, data mining and computational social science.