Bachelor's Thesis

Grass Snake and Gray Line, Buried Thread Thousands of Miles
: Research on Narrative Techniques Based on Text Mining and Social Network Analysis

Xue Xiumei

School of Statistics

2015201589

Supervisor:

Huang Danyang   19.04.2019

Abstract

The method of *grass snake and gray line* is a narrative technique originating from Chinese classical novels, which connects events, appearing successive characters, forming readers'expectation horizons, and suggesting the trend of plots. It has been used by a lot of literary creators to write unexpected but reasonable articles. *Journey under the midnight sun*, considered as the peak work of the famous Japanese detective novelist Keigo Higashino, is a modern variant of the the method of *grass snake and gray line.* Its enduring artistic charm shows that this technique is not out of date. Through text mining and social network analysis, this paper restores the novel's time axis, explores the characters' appearance frequency and sequence, and draws social network maps, which makes the role factions clear at a glance and verifies the strong connection between the two main characters. The study shows that text mining and social network analysis still have more values to be exploited in exploring complex literary works. This paper fills in the current research gap in this field, predicting that after introducing more modern statistical techniques, literary research will break through the restrictions of qualitative analysis and open up a new paradigm for studying literary works.

Key words：The method of *grass snake and gray line*　　*Journey under the midnight-sun*　text mining　social network analysis　Two-mode network analysis

Contents

1 Introduction

When commenting on *All Men Are Brothers*, Jin Shengtan summarized the *Grass Snake and Gray Line* method. Later, critics of Ming and Qing's novels often used this method to comment on works, which is also a prominent feature of Chinese classical novels to attract readers. Its name refers to the arrangement of **clues** in the text, just like snakes slithering in the grass, faintly visible, while the traces left on the ground are intermittent, like loose ashes. In writing, this technique is often represented by the author's conscious repetition of characters, objects, time and other details, as well as the organic connection between one narrative segment and another in the structure, giving readers an unexpected but reasonable feeling. This technique has the functions of connecting events, following characters, forming readers' horizon of expectation, and suggesting the trend of the plot. Until now, people have consciously used this technique.

With the rapid development of information technology, we are in the era of information explosion, and text information is no exception showing explosive growth. How to control these huge and rich text resources has become a new research topic. At this time, text-mining technology came into being and received more and more attention. Text mining mainly includes text acquisition, preprocess, in-depth analysis, data visualization and other steps, covering various technologies such as data mining, information extraction, natural language processing and statistical data analysis, and is committed to obtaining valuable information and knowledge from text data.

In the field of literary research, in the face of the increasingly electronic publications, scholars also began to try to use quantitative research methods like word statistics and information extraction, which initially showed great power in the research on author's attribution, literary style and literary genre analysis, and greatly broke through the traditional paradigm of qualitative analysis of literary works. In particular, due to the particularity of the novel as a literary genre, it is the core of novel text research to explore the relationship between characters by means of social network analysis.

Previously, scholars at home and abroad paid less attention to the application of text mining technology in the study of novel techniques, and the **dual-mode social network** is also a new perspective in social network analysis. Based on this, this paper chooses to focus on the *Grass Snake and Gray Line* method, a common literary technique in Chinese classical novels, and combines the current specific novel text, through the methods of text mining and social network analysis, to make the original **obscure** clues become **clear**. This research fills in the previous research gap in this field, which is of great significance for further expanding the scope of application of text mining and social network analysis, and also provides more possibilities for future literature research methodology, which reflects the strong interdisciplinary integration ability and broad application prospects of statistical data analysis.

2 Literature Review and introduction of methods

2.1 Grass Snake and Gray Line Method

It was in the Qing Dynasty that we first discovered this artistic technique and gave it a theoretical summary, with the name of "Grass Snake and Gray Line method" by Literary critic Jin Shengtan. In reading the Fifth Talent Book, his comment on All Men Are Brothers said: *There is Grass Snake and Gray Line method. Such as Jingyang Gang frequently describes many "whistle stick"word, Purple Stone Street even writes a number of "curtain" word and so on. At a glance, it looks like nothing; when you look out, there is a clue.*

Obviously, because this concept is descriptive, figurative, and there is no strict definition, different people have different understandings of it. In critics, knowledge of this technique varies. Liang Guizhi (2001), a professor of the School of Literature of Liaoning Normal University, holds that the Grass Snake and Gray Line method is the implication and foreshadowing for the development of the later plot left everywhere in the novel writing, so he says "buried thread thousands of miles". In addition to the mainstream understanding of identifying it as the long-term thread, in the Chinese classical novels, such as *The Plum in the Golden Vase*, *Romance of The Three Kingdoms* and *A Dream of Red Mansions*, the method even reflects the complex and rich connotations such as structural clues, foreshadowing care, metaphors and symbols.(Yang Zhiping, 2008)

In terms of its function and significance, the Grass Snake and Gray Line method mainly connects up and down, echo each other to realize the complete unity of artistic structure, enhances the organic of the plot, and suggests the fate of characters or the direction of events (Ruan Fang, 2008). However, because modern novels are no longer confined to the mode of telling only a story line at the same time, the role of this technique in continuing characters or plots in classical novels has been weakened. Though, in modern novels, the Grass Snake and Gray Line method can still be used repeatedly to make the readers have a specific psychological expectation for it, so as to lead to the climax of the plot and produce a fascinating artistic effect. For example, Qian Zhongshu wrote in *Fortress Besieged* that Fang Hung-chien found the personnel of Sanlu University dirty, which involved the secret contact between the president Gao Songnian and the wife of Wang Chuhou, the director of the Chinese language department. Careless readers may not have noticed it, but the author has already used the Grass Snake and Gray Line method to imply it many times (Zhou Yunlong, 2006).

Although the Grass Snake and Gray Line method can be traced back to hundreds of years ago, but its successful experience in classic works, is undoubtedly a good way to layout the writing. In modern novels, the Grass Snake and Gray Line method can still play its role, helping people to write the unexpected but reasonable excellent literary works.

2.2 Social Network Analysis

It is generally believed that the study of social networks originated in British anthropology. The British anthropologist A.R.Racliffe-Brown used the concept of "Social Networks" for the first time. At this time, it mainly describes the social structure with the network, which is just a metaphor here. By the 1850s, some anthropologists such as S.F.Nadel and J.A.Barnes began to systematically develop the concept of network, and Elizabeth Bott developed the first explicit measuring tool for network structure, knit. With the series of achievements, network analysis has attracted the attention of many international scholars. By the mid-1970s, network analysis had become a more influential field in Sociology.

Nowadays, the research field of Network Analysis is not limited to Sociology, and much of the data in real life is expressed as the network form of interconnected objects. Social networks are one of the main categories of data networks, and the common forms are shown in Table 1.

Table 1: Common Examples of Social Networks

Examples	Appliactions
Friendship networks	Student, organization, Facebook
Follower networks	Twitter, LinkedIn, etc
Interaction networks	Phone calls, SMS, email, etc
Spread networks	Infectious diseases, information, gossip, etc

The research object of Social Network Analysis(SNA) is the underlying structure of this network, which aims to explore the interaction mode between the participants in the social network and analyze the relationship structure and attributes of the social network. Common tasks in SNA include the use of statistical indicators to identify the most influential, prestigious or core participants, and the use of community detection technology to find community(Shazia, 2018). Due to the huge potential of this approach, SNA has become a popular method in many fields ranging from biology to business. For example, some companies target through SNA for customers with higher network value, namely those with higher influence and support (Domingos & Richardson, 2006).

Social networks can be constructed with relational data, consisting mainly of a set of social entities and the relationships between them. These networks are usually modeled graphically, and the graph consists of two basic units: "vertices" and "edges", where the vertices represent the social entities and the edges represent the connections between them. The edges can be directed or undirected, depending on whether the nature of the relationship is symmetric. As shown in Figure 1, between B and D is the undirected line segment, namely the symmetry relationship.
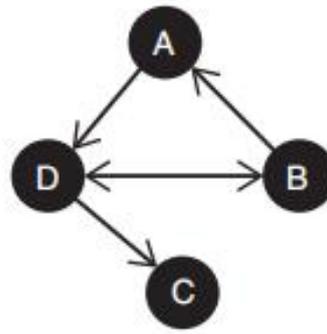
Figure 1: Example of the Network Diagram.

In order to compare different networks, the normalization of some indicators of networks is also a necessary prerequisite for social network analysis. For example, a node with higher out-in degrees has a higher degree centrality:

$$C_{ADi} = d(i) = \sum_j x_{ij} \text{ (absolute)}$$

$$C_{RDi} = d(i)/(n-1) \text{ (relative)}$$

A network with a higher overall centrality has a higher node-degree centrality:

$$C_{AD} = \frac{\sum_i (C_{AD\max} - C_{ADi})}{(n-1)(n-2)} \text{ (absolute)}$$

$$C_{RD} = \frac{\sum_i (C_{RD\max} - C_{RDi})}{n-2} \text{ (relative)}$$

There are also node's degree, network density and other indicators to be further elaborated later. Through these indicators, social network analysis can discover more valuable information from the network graph, and give it a powerful application value.

In recent years, social network analysis has also been widely used in the study of literary works. Elson et al.(2010) have extracted social networks from the 19th-century English literature to provide an intuitive connection between these literary works. Agarwal et al.(2013) discuss how to accurately extract social events and social networks by using *Alice in Wonderland* from the perspective of technology. Thtough analyzing Shakespeare's plays, Nalisnick and Baird (2013) discovered small-world effects in the dramatic role network.

2.3 Dual-mode Network Analysis

Usually the network consists of a group of participants or nodes, and the relationship between them. This relationship can be any meaningful connection between a pair of actors, which can be described by directed or undirected line segment: such as "friendship" positive contact, or "communication" or "know" neutral contact, even can be "hate" or "dislike" and other negative links. A network composed of a pair of mode data is called a one-mode network (single-mode network). Accordingly, when we have data with two pairs of mode, a slightly more complex situation occurs, namely the so-called dual-mode network (Everett, 2016).

Dual-mode network has been studied for a long time. Davis et al. (1941) first collected double mode data in their classic study on women participating in social activities in the south, then Stokman et al. (1985) studied the interlocking board since the 1970s, Breiger (1974) systematically introduced how to analyze the network, and introduced the concept of duality between people and groups. The Breiger method is now usually called projection or conversion, which is the more popular dual-mode network analysis method. Further, Borgatti and Everett(1997) proposed a direct analysis of the dual-mode data. In recent years, there has been increasing interest in dual-mode networks due to increasing amounts of data having this form.

Dual-mode social network (Two-mode social network) data consists of two different sets of nodes that can be called "actors" and "events", as well as the "relation" between actors and events. Such as the "attendance" between students and the course, the "published" between the author and the journal, etc. These data are usually expressed in the form of an affiliation matrix, where rows are actors and columns are events.

Table 2: The Affiliation Matrix A.

	Red Sorghum	Farewell My Concubine	Crouching Tiger, Hidden Dragon	Let the Bullets Fly	The Grandmaster
Zhang Ziyi	0	0	1	0	1
Leslie Cheung	0	1	0	0	0
Gong Li	1	1	0	0	0
Chang Zhen	0	0	1	0	1
Jiang Wen	1	0	0	1	0
Ge You	0	1	0	1	0

The general form of the affiliation matrix is a 0-1 adjacency matrix, just like the matrix A shown in Table 2. Matrix A is a matrix of 6 rows x 5 columns: the performance actor group, including Zhang Ziyi, Leslie Cheung, Gong Li, Chang Zhen, Jiang Wen, and Ge You; and the list is the film group, including five films from *"Red Sorghum"* and "*Farewell My Concubine*" to "*Crouching Tiger, Hidden Dragon*", "*Let the Bullets Fly*" and "*The Grandmaster*". Matrix $A = (a_{ij})_{6x5}$, where

$$a_{ij} \in \{0,1\}$$

$$1 \le i \le 6, \quad 1 \le j \le 5$$

Defining 0 for "not acting" and 1 for "acting", matrix A represents a group of dual-mode network relationships of "acting" between "actor-film".

For this form of data, there are two methods of analysis. The first is the projection method, which transforms the double mode network data into single mode data and then analyzes it according to the general social network method. For actors, we can construct a matrix to represent the relationship that "attended the same event"; and for events, we can similarly construct a data matrix where the relationship is "at least one person participated in two events simultaneously". Further, we can even construct a matrix to represent the number of times a pair of participants participate in an event together, and how many co-actors participate simultaneously between the two events. These are given by the matrix product sum AA^T and $A^T A$, where A is the affiliation matrix and A^T is the transpose of A. A classic article by Breiger (1974) explores the relationship between these matrices. As shown in the matrix A of Table 2, the matrix indicating "whether to participate" between "movies and actors" can be obtained after transposition, and the two multiplication can obtain the matrix product sum AA^T and $A^T A$, that is, to complete the process of converting the dual-mode network data into single-mode, which AA^T is the actor matrix and $A^T A$ is the movie matrix.

The second method of dual-mode network analysis is to analyze the dual-mode data directly as a special network. Borgatti and Everett (1997) advocated this approach, and it will not be described here because few relevant research results are available.

In this paper, we mainly use the projection method to transform the dual-mode network data into a single-mode network for analysis.

3 A modern variant of the Grass Snake and Gray Line method: Take *Journey Under The Midnight Sun* as an example

3.1 Background Introduction

Keigo Higashino, a famous Japanese mystery writer, is good at depicting human nature, whose works are famous for their plot twists and exquisite story structure. *Journey Under The Midnight Sun*, as a novel written by Keigo Higashino in 1999, it vividly embodies his flexible narrative techniques. It is regarded as the peak of Keigo Higashino and the uncrowned king in Keigo's works in the eyes of many readers. With its excellent artistic charm, the novel has become popular among thousands of readers around the world, and has the most adaptations in Keigo's novels. In 2006, the novel was adapted into a TV series of the same name, winning four awards at the 48th Japanese Drama Academy. Recently, the musical of the same name opened in many places in China, attendance of which is much high.

Journey Under The Midnight Sun is set in the period of the Japanese bubble economy. The bursting of the bubble economy caused Japan's domestic economy to fall from the peak to the bottom, with a large number of enterprises closed down and the unemployment rate increased sharply. According to statistics, nearly a third of Japan's population during this period was unemployed, their families were burdened with huge debts, and the overspending Japanese began to realize the importance of money. In order to survive and have a sense of security, money has become the thing above family, friendship, love, etc., when human nature is lost in the pursuit of money, personal standard and social innocence are in their age. Japanese sociologists generally believe that this has fundamentally changed the values and world view of the Japanese people, especially the young generation, and the spiritual and moral crisis hidden under the social order began to appear.

The novel is set in such a social background, narrating around a pair of primary school students with unusual feelings. In 1973, a man's body was found in an abandoned building in Osaka, and during 19 years later, the suspect's daughter Yukiho, and the victim's son Ryoji, were secretly connected. They seem to have taken a completely different path: one in the upper, one at the bottom, but the people around them have suffered one after another strange death. After 19 years of hard tracking of the police Sasagaki, the truth finally came to light. The author perfectly combines the hopeless but persistent sad love and the persistent and careful calm reasoning, which makes the readers lament about the love tragedy caused by the indifference of the family and the distortion of human nature.

3.2 Modern Variant of the Grass Snake and Gray Line Method

Why is *Journey Under The Midnight Sun* so popular? In addition to the deep human nature that is worth recalling, the exquisite narrative structure of the novel is also an important reason for its extremely high artistic charm.

Journey Under The Midnight Sun consists of 13 chapters, among which more than **60** characters appeared and lasted nearly 20 years, and 14 major cases occurred. The text capacity and story density are not small. What is even more amazing is that in the process of writing, the author transforms narrative perspectives for several times and buries several foreshadowing hints. It lets the readers feel seem to catch something but in fact no clue until the end, bringing a kind of unexpected but reasonable feeling.

Unexpected but reasonable, this is the most prominent artistic effect of the Grass Snake and Gray Line method. In *Journey Under The Midnight Sun*, Yukiho and Ryoji never talk directly from the beginning to the end, but the people around them suffer from misfortune one after another. Readers may doubt this rare coincidence, but without the words of the characters and the author, it is not so clear until the end that their childhood misfortunes have kept their lives close in secret. Looking back, this dark secret in *Journey Under The Midnight Sun* is hidden in many hidden lines, just like many previous literary works using the Grass Snake and Gray Line method.

For example, the novel does not specify that Yukiho has always been a deep one, but through the "bell", this hidden line, it can be seen that the death of Yukiho's birth mother can be largely attributed to her deliberate failure to timely rescue.

In Chapter 1, when Yukiho came home from school, she found herself without the key to find the landlord Tagawa to open the door:

"*Each movement, a tinkling bell came about her. Tagawa was curious about what bell it was and looked at it with his heart, but he could not see it from the appearance. ...She nodded gently and heard another jingle bell.*" When she finally got home, she found that her mother, Fumiyo nishimoto, was on the verge of death and finally died on the way to the hospital, "*...life may can be saved half an hour early*".

At the end of Chapter 4, Yukiho talks about this story with the tutor Masaharu:

"*'If I had brought the key, I could have found her earlier without looking for the landlord... so I would definitely carry the key with myself now. Look, just like this.'*
'What an old key ring', Masaharu said.
'Yes, this, and then there was the key to the house. But on that same day, I left it at home and forgot to bring it.'And so she put the key in her pocket.
The little bells on the key ring dingled."

It can be seen that the Yukiho was with the door key at that time, but in order to become Karasawa's adopted daughter and live a better life, she gave up the opportunity to rescue her mother. It can be seen that Yukiho's heart is very hard.

Although she was still a primary school student at that time, she quite had the shadow of the adult who would do anything to achieve her goal. If the reader could have realized this line, the rest plot might not be so surprising. Grass Snake and Gray Line method, which is derived from the classical Chinese novels with outstanding artistic effect, has been consciously applied by the ancient and modern literary creators. *Journey Under The Midnight Sun* is a typical example of its modern variant.

Next, this paper will conduct a text mining analysis of *Journey Under The Midnight Sun*, reorganize the events according to the occurrence time, restore the important characters, and clarify the important clues in the novel. Further, I will construct the character network, and draw the social network map. Due to the rare joint appearance of Yukiho and Ryoji, a two-mode network analysis will also be introduced to verify the strong connection between the two.

4 Text Mining Analysis : Take *Journey Under The Midnight Sun* as An Example

4.1 Text Preprocess

The text was first pre-processed. Import the novel document in R, to find out the important characters and important terms in each chapter through the word segmentation, such as the company name, location name, etc.

There are 65 characters in *Journey Under The Midnight Sun*, and one character may have multiple different titles, having to extract all their names and establish a character list. Take "Ryoji kirihara" as an example, sometimes called "kirihara", sometimes called "Ryoji" and sometimes "Ryo". After chapter 10, Ryoji even disguised as his middle school classmate who appeared in the previous chapter, named "Yuichi akiyoshi". Therefore, the list of characters actually needs to be phased into, namely, chapters 1,2 to 9, and 10 to 13. A partial example of the character list is shown in Figure 2, where a row represents a character, and the different names of the same character are separated by comma.

Table 3: Example of the List of Characters.

Order	Character
1	Yaeko,Yaeko kirihara,Fujin kirihara
2	Ryoji kirihara,Ryoji,Ryo,kirihara
3	Fumiyo nishimoto,Fumiyo,Fujin nishimoto
4	Yukiho nimoto,Yukiho,Yukiho karasawa,karasawa,Karasawasan,Yukiho shinoduka

In addition, there are 8 murder cases, 3 assault cases, and 3 leak cases in the novel. The characters and important nouns in each case are different, which can be used as a **clue** to restore the timeline. For example, by locating the important term Submarine(name of a game software) related to the case, it is obvious from Table 4 that the Submarine leak case occurred at the time node of chapters 3 to 4.

Table 4: Example of Important Terms.

Terms	Appearance	Scene
Submarine	Chapter 3	the main product of Ryoji's company *Unlimited Planning*
	Chapter 4	The game studied by Yukiho's tutor Masaharui

At the same node of the murder incident, it is easy to get a conclusion from the statistics of the character appearance frequency in chapter 3.
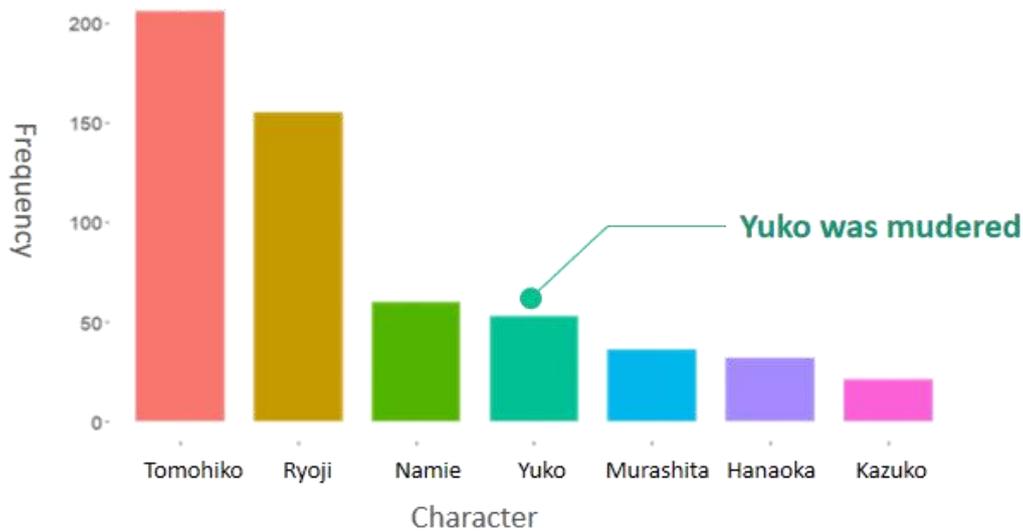

Figure 2: Statistics of character appearance frequency(Take Chapter 3 as an example)

Therefore, the timeline of the past 20 years can be restored, and the time and main contents of the 14 cases can be indicated, so as to provide conditions for us to further understand the experiences of these characters and clarify the important clues.
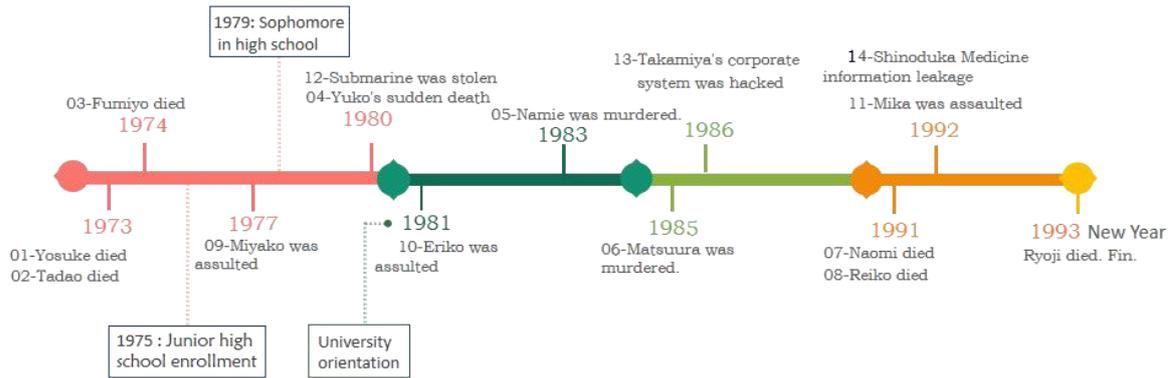
Figure 3: The Timeline of *Journey Under The Midnight Sun.*

4.2 Character Debut

As the key element of the novel, character plays an important role in connecting the events and promoting the development of the plot. Through the statistics of the appearance frequency of each character (Figure 4), it can be found that Ryoji and Yukiho are the two protagonists in *Journey Under The Midnight Sun*. They appear about 800 times in the novel, while Police Sasagaki, as the finder of the truth, appears about 600 times, ranking the third.
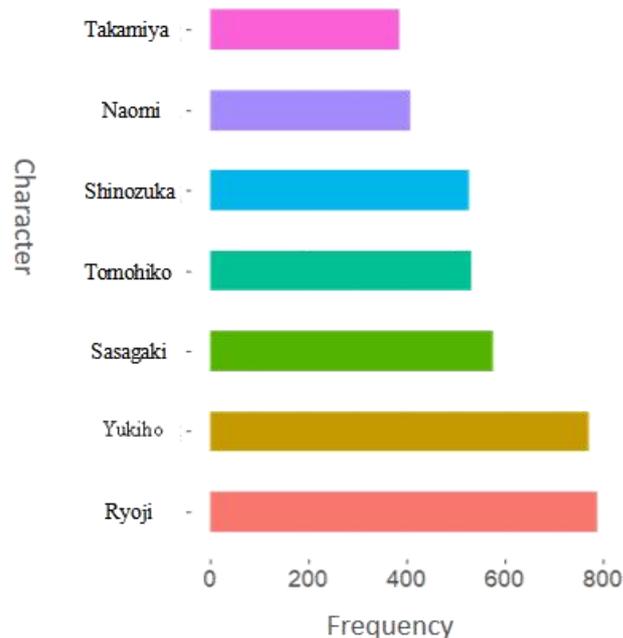

Figure 4: Chart of the Appearance Frequency of Each Character.

Secondly, through the analysis of the appearance order of each character, we can get the appearance relationship of the main characters such as Ryoji, Yukiho and Sasagaki as shown in Figure 5. It can be seen that the appearance of Ryoji and Yukiho has been in a state of "fluctuation" for a long time, and when the two appear in the same chapter, it is always that Yukiho in the light and Ryoji in the dark (the appearance

frequency of Yukiho is always higher than that of Ryoji). For the Police Sasagaki, his appearance presents the characteristics of the previous introduction and later focus, it is not difficult to conclude that in the later stage of the novel, it is necessary to focus on Police Sasagaki and his efforts to exploring the truth finally.
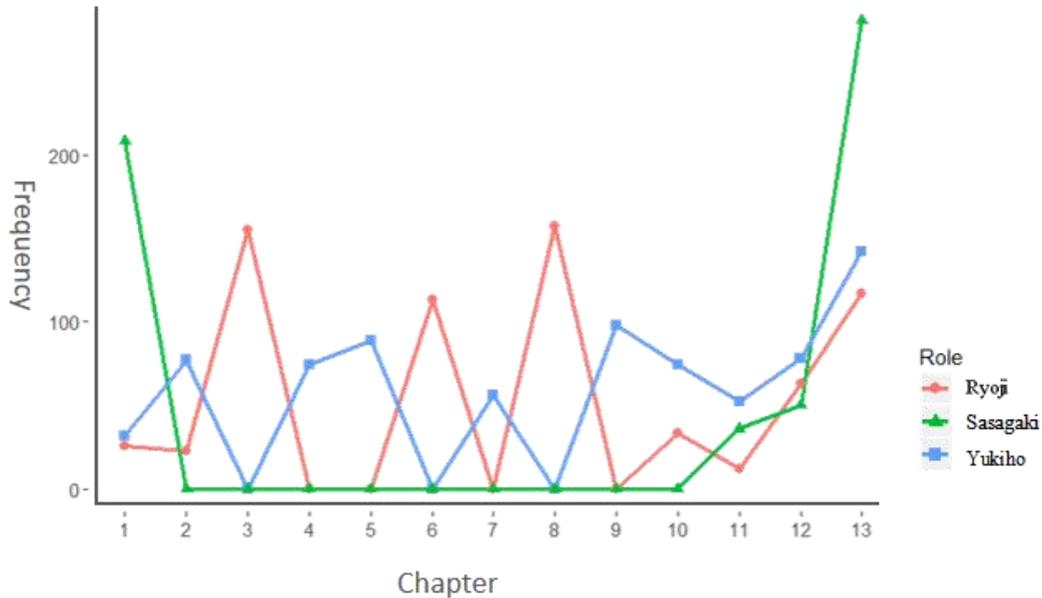

Figure 5: Statistics of the Appearance Sequence of the Main Characters.

The above analysis shows the author's exquisite setting of clues and narrative perspectives, which make the direct relationship between Yukiho and Ryoji seem to be untraceable. To uncover the implicit strong connection, more clues need to be identified.

The intimacy value matrix (Figure 6) also shows that the intimacy value between Ryoji and Yukiho are low. To place the seven most frequently appeared characters in the novel on the rows and columns and the cell number corresponding to each pair of characters represents the number of paragraphs where the characters appear together, thus the intimacy value matrix of the main roles is structured. The darker the cell color, the larger the number in the cell, and the closer role relationship. Obviously, Ryoji has the highest intimate value with his friend Tomohiko, and her first husband Takamiya is Yukiho's most close contact. Police Sasagaki not only pays close attention to the two people, but also gets close to Shinozuka kazunari, private investigator Naomi imada. It can be said that "When a mantis is trying to catch a cicada, a canary is behind it."

Figure 6: The Character Intimacy Value Matrix.

How to find more clues? We decided to combine the appearance order of important nouns with the context, to get the hidden clues between Ryoji and Yukiho. In addition, through other hidden lines such as potassium cyanide(KCN), we can get such association like "Ryoji(Pretended Yuichi)- Noriko(Pretended Yuichi's girlfriend)- **potassium cyanide(KCN)**- Detective Naomi- Takamiya- Yukiho" , namely through the important nouns and other roles, we also can get the connection between the two protagonists. Sort all these clues out.

4.3 Social Network Analysis

The main part of the text analysis of *Journey Under The Midnight Sun* is to excavate the relationship between the characters in the novel with the help of social network analysis, especially to verify the strong correlation between Ryoji and Yukiho.

4.3.1 Single-mode Network Analysis

In ordinary single-mode social network, the relationship between characters are determined by the joint appearance of characters, from which the social network map can be drawn.(Figure 7) In Figure 7, the nodes representing the main characters are marked, and the different colors represent the nodes belong to different camps.
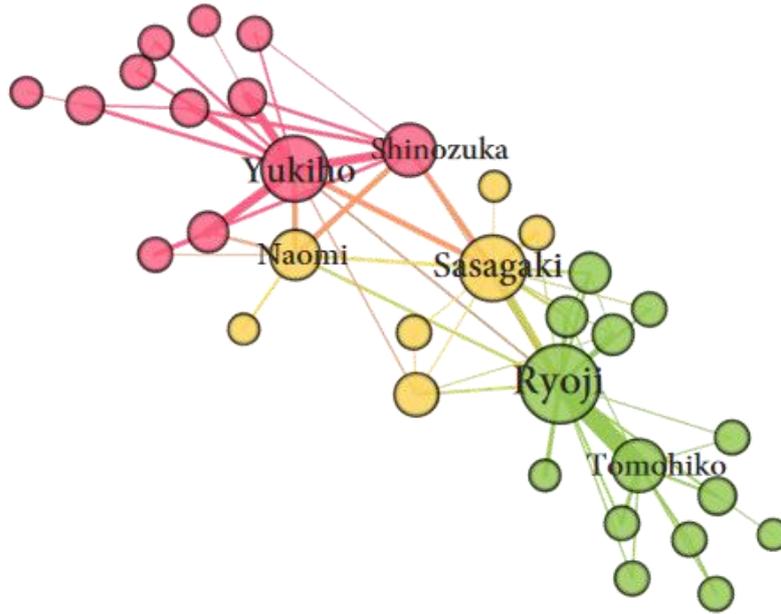
Figure 7: The Initial Social Network Map.   *Drawn by Gephi*

The social network diagram is drawn by marking two people who appear in a paragraph> 10 times as "related", and the number of times determines the thickness of the edges while the number of edges determines the size of the nodes. There are 33 characters and 63 edges, and the network density is 0.119.

In abstract, the density of a network graph is a measure of the completeness of the graph. The so-called completeness refers to the degree of adjacency between the points in the graph. For example, a complete graph means that the nodes in the graph are adjacent each other. The density of a network graph partly symbolizes the number and complexity of relationships in the network: In a network of density 1, each individual is associated with all other individuals, instead, there is no relationship between individuals in a network of density 0.

The network density is generally defined as the ratio of the actual number of edges in the network and the maximum possible number of edges in theory. The number of recorded edges is n, and the number of vertices is d. For the undirected graph, the maximum theoretical possible value of included relationships is n (n-1) / 2, so the network density

$$ND = \frac{d}{n(n-1)/2}$$

Considering that the novel includes all of the characters, the social network diagram is relatively sparse. At the same time, it can be seen that the characters in *Journey Under The Midnight Sun* are mainly divided into three camps: Yukiho, Ryoji and Sasagaki are the core figures of each camp. Police Sasagaki maintains a certain correlation with both Ryoji and Yukiho, but the relationship between Yukiho and Ryoji is not obvious.
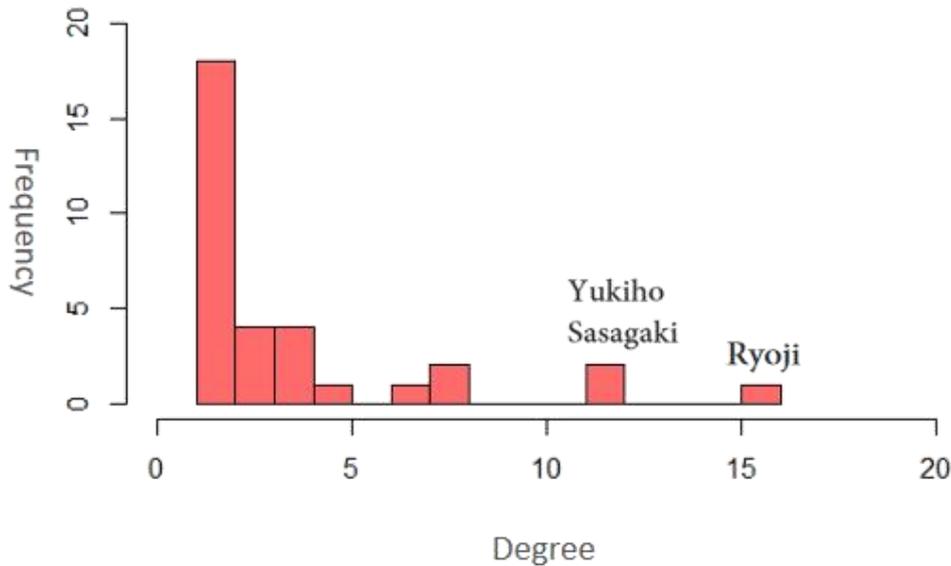
Figure 8: Degree Distribution Diagram.

Another indicator in social network analysis is "degree", which is the number of edges elicited by a node. Here the degree of a character is the sum of the number of other characters that appear in any paragraph>10 times at the same time. From the degree distribution diagram(Figure 8), it can be seen that the average correlation degree of each character is about 4, among which Ryoji relates to 16 people and both Yukiho and Sasagaki have 12 people. More than a half of the characters' degree is 2. The social network approximately fits the power-law characteristics.

4.3.2 Dual-mode Network Analysis

Because ordinary social network analysis cannot dig out the strong connection between Ryoji and Yukiho, we try to conduct dual-mode network analysis. As we know from the previous introduction, dual-mode social network data consists of two different sets of nodes, called "actors" and "events", as well as the "relation" between actors and events. And these data are usually expressed in the form of an affiliation matrix, where rows are actors and columns are events.

In this analysis, the key material evidence, company name and other clues of the previous cases are taken as events, and the roles related with them are taken as actors, so as to construct the "correlation" relationship between the case clues and different roles. By marking "related" between the clues and the roles who appear simultaneously in a paragraph more than or equal to one time, the affiliation matrix is obtained, which is partially shown in Table 4.

Table 4: A Partial Schematic Diagram of Matrix A.

	Library	Cloth	Submarine	Credit card	20 million
Yukiho	1	1	1	1	1
Tagawa	0	0	0	0	0
Ryoji	1	1	0	1	0
Namie	0	0	0	0	1
Tomohiko	0	0	1	1	0

In the role-clue network, there are 17 main clues and 12 implicit clues. The matrix A_{17x12} represents the network, the row represents the role, the column represents the cue, a_{ij} indicates whether the clue is related to the role, 0 represents "irrelevant" and 1 represents "related". That is to say

$$A_{17x12} = (a_{ij})_{17x12}$$

$$a_{ij} \in \{0,1\}$$

$$1 \le i \le 17, \ 1 \le j \le 12$$

Since in this study we mainly focus on the relationship between roles and roles, we construct $B = AA^T$ and obtained the role-role matrix B, where

$$B_{12x12} = (b_{ij})_{12x12}$$

$$b_{ij} = \{ \text{The number of clues associated both with character i and j} \}$$

$$1 \le i \le 12, \ 1 \le j \le 12$$

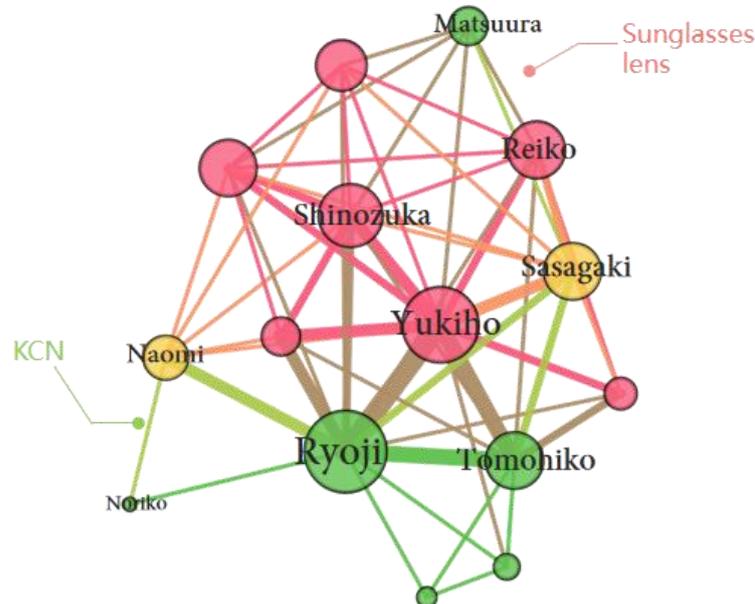Thus, the two-mode network data can be transformed into common single-mode networks for analysis by projection method.

Figure 9 shows a new social network after the projection. As can be seen from the figure, the new network has 15 characters and 55 edges, and the network density is 0.533, which is significantly improved compared with the original one. At the same time, the roles are no longer quite distinct from each other. Compared with the original network, people in different camps establish a connection through the form of "role a- clue 1- role b", among which Yukiho and Ryoji are the most closely connected. Thus, the symbiotic relationship between the two people is finally verified, just like the relationship between shrimp and goby in Sasagaki's words.

It can be seen that dual-mode network analysis can utilize the relationship between several node groups to dig the correlation which can be neglected easily in single node group, which is an important method in social network analysis. When the relationship between objects is vague, we can pay attention to the clues connecting objects and explore deeply through the dual-mode social network method.

5 Summary

Although the Grass Snake and Gray Line method originated from Chinese classical novels, it is not out of date as a narrative technique. As a modern variant of the Grass Snake and Gray Line method, the enduring artistic charm of *Journey Under The Midnight Sun* shows that today's literary creators can still use this narrative technique.

Through this analysis of *Journey Under The Midnight Sun*, it can be seen that text mining and social network analysis have quite strong application value for exploring complex literary works. This study fills in the previous research gap in this field, which is of great significance for further expanding the application scope of text mining and social network analysis. At the same time, it can be predicted that after introducing more modern statistical techniques, literary research will break through the restrictions of qualitative analysis and open up a new paradigm for studying literary works.

Bibliography

[1]Liang Guizhi. The interpretation of Grass Snake and Gray Line: By two comments of the Qing Dynasty people on a Dream of Red Mansions[J]. Studies On a Dream of Red Mansions,2001(02):262-276.

[2]Yang Zhiping.Comments on "Grass Snake and Gray Line" and ancient Chinese novels[J]. Seeking Truth,2008(01):108-113.

[3]Ruan Fang. Grass Snake and Gray Line, Buried Thread Thousands of Miles: A unique structural technique of Chinese classical novels[J].

Journal of Hubei Radio & Television University,2007(03):80-82.

[4]Zhou Yunlong. On the "Grass Snake and Gray Line" in Chinese classical novels[J]. Journal of University of International Relations,2006(01):67-72.

[5]Zhou Jilei, Zhou Jing, Li Ji. Social Media Marketing of "IP Drama": a Research based on text mining and Social Network Analysis[J]. Modern Management Science,2018(07):99-101.

[6]Li Hui, Xiang Huating, Tang Qiang.The Wikipedia author trust model based on the dual-mode social network[J]. Library And Information Service,2015,59(12):78-84.

[7]Adamic, L. A., & Adar, E.. Friends and neighbors on the web. Social Networks, 2003,25(211–230).

[8]Abraham, A., Hassanien, A.-E., & Sná, V.. Computational social network analysis: Trends, tools and research advances. London: Springer Science & Business Media,2009.

[9]Filip Agneessens,Martin G. Everett. Introduction to the special issue on advances in two-mode social networks[J]. Social Networks,2013,35(2).

[10]M.G. Everett,S.P. Borgatti. The dual-projection approach for two-mode networks[J]. Social Networks,2013,35(2).

[11]Shazia Tabassum,Fabiola S. F. Pereira,Sofia Fernandes,João Gama. Social network analysis: An overview[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,2018,8(5).