

Character Interaction Network Analysis of Chinese Literary Work- A Preliminary Study

Bikun CHEN

Department of Information Management, Nanjing
University of Science and Technology
No. 200, Xiaolingwei, Nanjing, 210094, China
chenbikun@njjust.edu.cn

Yuefen WANG

Department of Information Management, Nanjing
University of Science and Technology
No. 200, Xiaolingwei, Nanjing, 210094, China
yuefen163@vip.163.com

ABSTRACT

Considering the shortages of character interaction network analysis from literary work by literary theorists and computer scientists, this paper combined their methods and conducted qualitative and quantitative approach to extract and analyze character interaction network from literary text. Firstly, it proposed the analysis framework of character interaction network from literary work: one-mode and two-mode network analysis. One-mode network analysis was applied in macro-level (chapters), meso-level (social events) and micro-level (character dialogues). Two-mode network analysis was applied in meso-level (social events). Then two literary scholars manually extracted social events and character dialogues from the fiction “Journey to the West Prequel” and manually recognized characters in different levels. Then it constructed the one-mode and two-mode matrices based on the proposed standard formula which was based on the idea of co-occurrence. Different matrices were computed and visualized by SNA software. The results validated by literary scholars showed that the proposed analysis framework was applicable in interpreting and revealing the contents of the fiction. Finally, the discussions and further researches were summarized.

Keywords

Literary work, social network analysis, sna, in-depth reading, gephi.

INTRODUCTION

As a classical method, SNA has been affecting a wide range of academic disciplines and practical applications: physics, management, scientometrics, literary theory, social media, entertainments and so on. In the past, social networks were

constructed through interviews, surveys and experiments. With the advent of the internet and online social networks, researchers have started constructing networks using metadata that reflects interactions, such as self-declared friendship linkages, sender-receiver email linkages, comments on a common blog-post, etc. (Agarwal, Kotalwar & Rambow, 2013).

In recent years, more and more researchers begin to extract social networks from literary texts, news stories and historical texts. Their methods can be summarized as follows. In the first one, literary theorists tend to perform in-depth qualitative studies examining the intricacies of plot structure and character interactions in the literary works. This kind of detailed analysis allows for deep conclusions to be drawn about literary works, but it does not scale well to large numbers of texts due to the significant amount of time required for a human to read, understand, and thoroughly analyze a piece of literature (Gil, Kuenzel & Suen, 2011). For example, Moretti plotted the character interaction networks of “Hamlet”, “Our Mutual Friend” and “The Story of the Stone” and performed a deep analysis of the plot structures (2005). Sparavigna applied the method of Moretti’s to plot the character interaction networks of “Harry Potter” (2013).

Compared with literary theorists, computer scientists typically take a very different approach to the task of analyzing literature by focusing on frequencies, statistics and algorithms. For instance, Stiller and Hudson drew the character networks of ten Shakespeare plays and found that the network exhibited small-world properties (2005). Rydberg-Cox created an application to visualize and explore social networks of Greek tragedies (2011). Sparavigna and Marazzato apply the Graph Visualization Software to visual the character networks of two Shakespeare’s play (2014). Marazzato and Sparavigna proposed a tool CHAPLIN (CHARacters and PLaces Interaction Network) to gather information on social networks from narrative texts (2014).

Basically, they extract social networks from text by mining interactions between people expressed linguistically in unstructured text signaled by quoted speech (e.g. Elson and

[This is the space reserved for copyright notices.]

ASIST 2016, October 14-18, 2016, Copenhagen, Denmark.

[Author Retains Copyright. Insert personal or institutional copyright notice here.]

Mckeown, 2010(a), 2010(b); He, Barbosa & Kondrak, 2013), lexical structure (Park, Kim & Cho, 2013) and social events (e.g. Agarwal & Rambow, 2010; Agarwal, Kotalwar & Rambow, 2013). Also, their studies mainly focus on micro-level (quoted speech or lexical structure) and meso-level (social event) and most networks are only one-mode, neglecting the relations between characters and social events. Considering the shortages of literary theorists and computer scientists, this paper combined their methods and conducted one-mode and two-mode character interaction network analysis together. Specifically, one-mode network analysis was applied to reveal the “character-character” relations in macro-level (chapters), meso-level (social events) and micro-level (character dialogues) and two-mode network analysis was applied to show the “character-event” relations in meso-level (social events).

DATA AND METHODS

Data

This paper chose the literary fiction “Journey to the West Prequel” as the raw data. As we know, “Journey to the West” is one of the Four Great Classical Novels of Chinese literature. It is popular in the world and the novel was made into different TV series, movies and comedy books, but readers or the audience were puzzled by lots of plots in the novel, TV serials or movies. In order to solve the puzzles, Zhi Ning, another Chinese scholar, writes the literary fiction “Journey to the West Prequel” which is the prequel to “Journey to the West”. It systematically reveals the earthshaking, complicated and mysterious backgrounds before the journey to the West, including the main characters, origins of different stories, causes and effects of important events and so on (Zhi, 2015).

The main reason of selecting “Journey to the West Prequel” was that we were familiar with the main characters and important events in “Journey to the West”, which made it easier to test the applicability of our methods.

Methods

One-mode network extraction

• Macro-level

The literary works usually consist of chapters which are carefully set by authors. Each chapter is an independent part, especially in literary fiction. So, the chapter is set as the coarsest unit to extract the relations of characters. In order to ensure the accuracy of network extraction, the characters in each chapter are manually recognized by two literary scholars with several rounds of in-depth readings. And the standard of measuring the relations are as follows: when any character i and j co-occur in any chapter C , their co-occurring value is number one. The total interaction intensity ϕ between character i and j is the sum of character i and j co-occurring in any chapter C .

$$\phi_{ij} = \sum_C One(C_i, C_j) \quad (1)$$

• Meso-level

Compared with the macro-level, the meso-level has a finer granularity. Because each chapter usually contains several social events. So, the social event is set as the unit to extract the relations of characters. Also, social events and the characters in each social event are manually recognized by literary scholars. And the standard of measuring the relations are as follows: when any character i and j co-occur in any social event E , their co-occurring value is number one. The total interaction intensity ϕ between character i and j is the sum of character i and j co-occurring in any chapter E .

$$\phi_{ij} = \sum_E One(E_i, E_j) \quad (2)$$

• Micro-level

Compared with the meso-level, the micro-level has the finest granularity. Also, the characters in each dialogue scene are manually recognized by literary scholars. And the standard of measuring the relations are as follows: when any character i and j co-occur in any dialogue scene Ψ , their co-occurring value is the minimum number of dialogues. The total interaction intensity ϕ between character i and j is the sum of character i and j co-occurring in any scene Ψ .

$$\phi_{ij} = \sum_{\Psi} \text{Min}(\Psi_i, \Psi_j) \quad (3)$$

Two-mode network extraction

In order to identify the contents of every social event, this paper also manually analyzes the events one by one and summarizes them into six categories: The Creation, Buddhist Rise, Heavenly Affairs, Buddhist VS. Taoist, Buddhist Affairs, Fight of Sutra-seeking. The standard of measuring the relations of characters and social events are as follows: any character that appeared in one social event can be seen as engaging in the event once, then sum up the number any character appears in any social event.

Matrices generation and network analysis

In terms of the formulas above, four different matrices were generated by VBA programming in Excel software. Then the matrices were imported to the SNA software to plot the network and calculate the network indicators. Finally, two literary scholars were invited to interpret the plots.

RESULTS

In Gephi, character interaction networks are visualized with ForceAtlas2 layout algorithm (Jacomy Venturini & Heymann, 2014). In the networks, every node signifies a character or social event, the size of every node means its degree, and the thickness of the line between two nodes signifies its interaction intensity. In addition, different color signifies different clusters (obtained by Modularity Class) (Dblondel, Guillaume & Lambiotte, 2008).

One-mode network analysis

Macro-level

From Figure 1, it shows the relations among different fractions in the fiction: the Taoist, the Heaven, the Buddhist

and the Devil. “Daode Tianzun”, “Jade Emperor”, “Buddha”, “White Elephant” and “Azure Lion” are the leaders of different fractions.

Meso-level

From Figure 2, it also shows the relations among different fractions in the fiction: the Taoist, the Heaven, the Buddhist and the Devil. “Daode Tianzun”, “Jade Emperor”, “Buddha”, “Great Peng” are the leaders of different fractions. Compared with Figure 1, the leader of the Devil in Figure 2 is “Great Peng”. In the novel, “Great Peng” is the true leader in the Devil fraction before the event Sutra-seeking.

Micro-level

From Figure 3, it shows more clusters than Figure 1 and 2. Also, there is an interesting result in the Figure 1 and Figure 2 that “Puti” is in the Taoist fraction and “Golden Headed Immortal” is in the Buddhist fraction. But in Figure 3, “Puti” is in the Buddhist fraction and “Golden Headed Immortal” is in the Heaven fraction. In fact, “Puti” is the “boss” behind the Buddhist and “Golden Headed Immortal” is the undercover of the Taoist and Heaven in the fiction. The macro-level and the meso-level cannot detect their true identities, but the micro-level reveal them.

Two-mode network analysis

In Figure 4, “Fight of Sutra-seeking” and “Buddhist VS. Taoist” have the highest degree centrality, then “Heavenly Affairs” and “The Creation”, “Buddhist Affairs” and “Buddhist Rise” are the least, which indicate that “Fight of Sutra-seeking” and “Buddhist VS. Taoist” are the core contents in the fiction, especially for the “Fight of Sutra-seeking”. In order to spread Buddhism and then control the world, the Buddhist proposed the plan “Sutra-seeking”. As the founder of the world, the Taoist can’t accept the plan of the Buddhist and spare no efforts to ruin it. The Heaven, founded by the Taoist, want to be independent from the control of the Taoist and support the Buddhist secretly. So, the core contents of the fiction are “Buddhist VS. Taoist” in essence and “Fight of Sutra-seeking” is the most important event between them.

CONCLUSION

In this study, the idea and method of SNA is applied to character network extraction and analysis and the results are consistent to the fiction contents, which proves that the SNA is applicable to the literary research (the dialogue-based and event-based approaches are better than the chapter-based approach). Also, SNA, combined with additional in-depth reading by literary scholars, can quantitatively, visually and comprehensively reveal the themes and contents of literary fiction. Besides, the readers and researchers can refer the results in this paper to interpret the puzzles of the classical fiction “Journey to the West”. In the further research, the automation method will be applied to extract character interaction network in

macro-level (chapters), meso-level (social events) and micro-level (dialogues).

ACKNOWLEDGMENTS

The research was supported by National Social Science Foundation of China (No. 15CTQ035) and National Natural Science Foundation of China (No. 71373124).

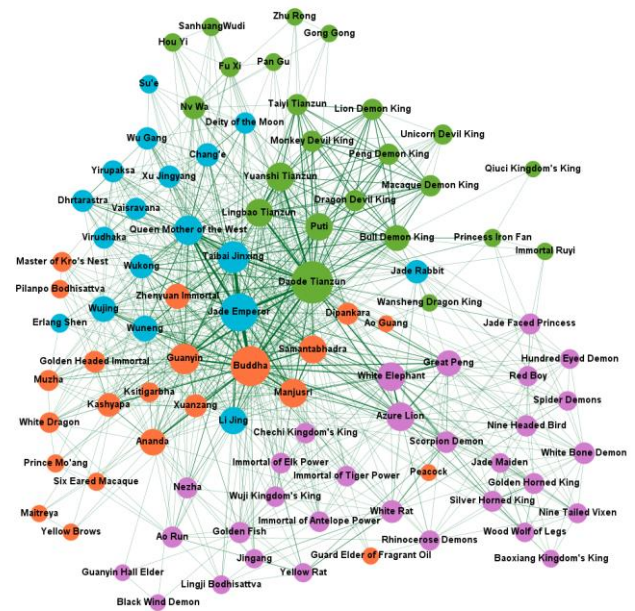


Figure 1. Chapter-based Network

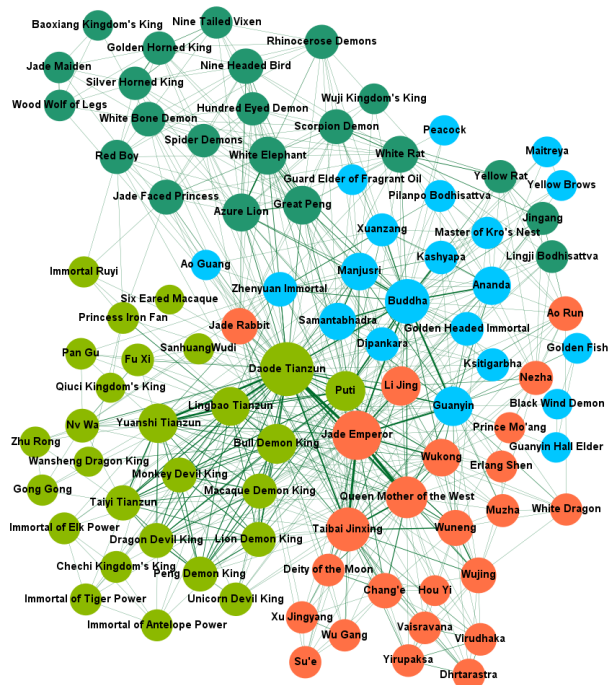


Figure 2. Event-based Network

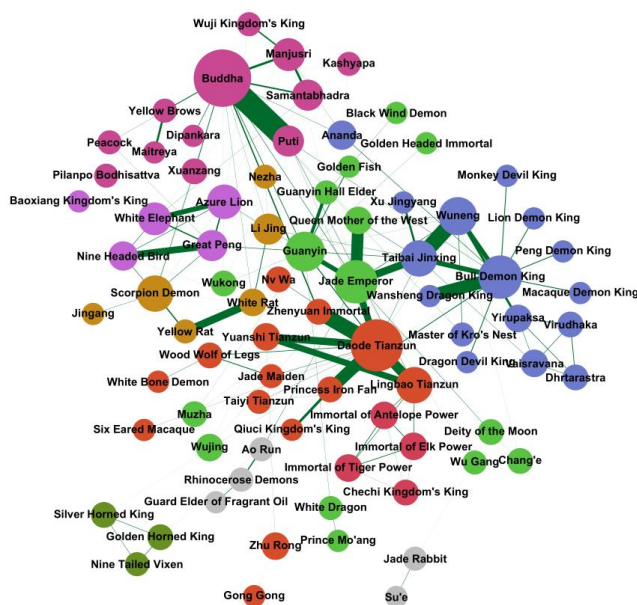


Figure 3. Dialogue-based Network

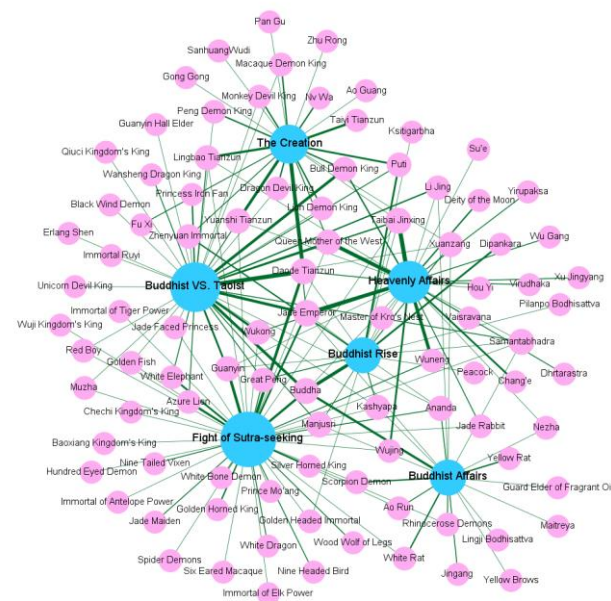


Figure 4. Character-Event Network

REFERENCES

Agarwal A., Kotalwar A. & Rambow O. (2013). Automatic extraction of social networks from literary text: a case study on Alice in Wonderland. In *Proceedings of the 6th International Joint Conference on Natural Language Processing* (pp. 1202-1208), Nagoya.

Agarwal, A., & Rambow, O. (2010). Automatic Detection and Classification of Social Events. *Empirical Methods in Natural Language Processing*.

Dblondel V., Guillaume J. & Lambiotte R. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, (10).

Elson, D. K., Dames, N., & McKeown, K. R. (2010). Extracting social networks from literary fiction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp.138-147). Uppsala: ACM.

Elson, D. K., & McKeown, K. R. (2010). Automatic Attribution of Quoted Speech in Literary Narrative. *The 24th AAAI Conference on Artificial Intelligence* (pp. 1013-1019). Atlanta: AAAI.

Gil S., Kuenzel L., Suen C. (2011). Extraction and analysis of character interaction networks from plays and movies. Retrieved June 15, 2016 from <http://stanford.edu/~cysuen/projects/GilKuenzelSuen-CharacterInteractionNetworks.pdf>.

He, H., Barbosa, D., & Kondrak, G. (2013). Identification of Speakers in Novels. *Meeting of the Association for Computational Linguistics* (pp. 1312-1320). Sofia.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLOS ONE*, 9(6).

Marazzato, R., & Sparavigna, A. C. (2014). Extracting networks of characters and places from written works with CHAPLIN. *Computer Science*.

Moretti, F. (2005), *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso Press.

Park, G., Kim, S., & Cho, H. (2013). Structural analysis on social network constructed from characters in literature texts. *Journal of Computers*, 8(9): 2442-2447.

Rydberg-Cox J. (2011). Social networks and the language of greek tragedy. *Journal of the Chicago Colloquium on Digital Humanities and Computer Science*, 1(3).

Sparavigna A. C. (2013). On social networks in plays and novels. *International Journal of Sciences*, 2(10):20-25.

Sparavigna A. C., Marazzato R. (2014). Graph visualization software for networks of characters in plays. *International Journal of Sciences*, 3(2):69-79.

Stiller J., Hudson M. (2005). Weak links and scene cliques within the small world of Shakespeare. *Journal of Evolutionary Psychology*, 3(1):57-73.

Zhi N. (2015). *Journey to the West prequel*. Shanghai: Shanghai Joint Publishing Company.