An Improved Method of Character Network Analysis for Literary Criticism: A Case Study of <Hamlet>

Ho-Chang Kwon
Department of Cinema Studies
Korea National University of Arts, Seoul, 02789, Korea

Kwang-Hyun Shim
Department of Cinema Studies
Korea National University of Arts, Seoul, 02789, Korea

ABSTRACT

As a computational approach to literary criticism, the method of character network analysis has attracted attention. The character network is composed of nodes as characters and links as relationship between characters, and has been used to analyze literary works systematically. However, there were limitations in that relationships between characters were so superficial that they could not reflect intimate relationships and quantitative data from the network were not interpreted in depth regarding meaning of literary works. In this study, we propose an improved method of character network analysis through a case study on the play <Hamlet>. First, we segmented the character network into a dialogue network focused on speaker-to-listener relationship and an opinion network focused on subject-to-object relationship. We analyzed these networks in various ways and discussed how analysis results could reflect structure and meaning of the work. Through these studies, we strived to find a way of organic and meaningful connection between literary criticism in humanities and network analysis in computer science.

Key words*: Social Network Analysis, Dialogue Network, Opinion Network, Sentiment Analysis, Quantitative Narrative Analysis.*

1. INTRODUCTION

Literatures reflect human and society through a unique story-world with characters, events, and temporal-spatial backgrounds. Analysis of the structure, mechanism, and meaning of these literatures has been mainly handled by humanities such as narratology and literary theory. Although the humanistic approach enabled deep insights based on philosophy, inaccuracies and ambiguities in terminology and methodology made it difficult to expand and develop literary studies [1]. To overcome these limitations, attempts have been made to combine developed digital technology with literary studies [2].

Among these computational approaches, character network has been widely used for analysis of literary works. The character network is made under the assumption that the story-world of a literary work is a society, and composed of nodes as the living characters and links as the relationship between characters. The method of character network analysis is to quantitatively analyze the character network using the graph theory, and to reveal characteristics of the literary work. This method seems to be a general approach for systematic analysis of literary works, however, there are two important limitations. First, links in a character network are generally made on the basis of explicit relationships (e.g. co-appearance), which do not reflect intimate relationships including judgements or emotions. Second, quantitative data extracted from the character network were not be interpreted sufficiently in relation to the meaning of literary works.

In this paper, we propose an improved method of character network analysis through a case study on the play <Hamlet>. We first segmented the character network into a dialogue network focused on speaker-to-listener relationship and an opinion network focused on subject-to-object relationship. After that, we analyzed these networks in various ways and discussed how the analysis results could reflect the structure and meaning of <Hamlet>. Franco Morretti [3] argued that network analysis should go beyond 'just showing' the structural characteristics of the network in a literary work and emphasized the need to understand and actively use existing literary theory and criticism for a deeper analysis. In this paper, we tried to find a way of organic and meaningful connection between literary criticism in humanities and network analysis in computer science.

2. RELATED WORKS

The study of Stiller et al. [4] is considered a pioneering work in character network analysis. They selected 10 Shakespeare plays and built each character network, and compared it with the real world social network. They analyzed the size, distribution and density of character network, and found that these are similar to the characteristics of network of 'hunting-gathering society'. Newman et al. [5] analyzed the character network of the novel <Les Miserables> as a case study, suggesting a new algorithm to find community structures on a network. They showed that <Les Miserables> is composed of eleven communities and that this structure clearly reflects the plot structure of the work; the protagonist Jean Valjean and his nemesis, the police officer Javert, are central to the network and form the hubs of communities composed of their respective adherents. Choi [6] and Kim [7] built the character networks of Greek mythology and the novel <Togi>, respectively. They used the existing biographical dictionary of each work, analyzed the structural properties of the character networks. Lee et al. [8] pointed out that the existing studies analyzed the character network in static manner, and analyzed the dynamic changes of the character network of the novel <Three Kingdoms>. They divided the period covered by the novel into four periods, analyzing the division and unification of the nation, the meeting and parting of characters, and the changes in the relationship of major characters. Kim [9] analyzed the dynamic changes of character network of the novel <Togi>. As gradually extending one section to another one, he constructed 4 different extended networks and compared the changes.

The studies of character network generally assume two; the story-world of a literary work is a society and the relationship between characters is key to the story. The character network is built in a similar way to describe characters as nodes and their co-appearances in the same scene as links. The character network is quantitatively analyzed for indices (i.e., size, density, centrality, diameter, clustering coefficient, etc.) using graph theory [10]. The character network has been used to analyze the literary works systematically, however, there were limitations in that the relationships between characters were so superficial that they could not reflect intimate relationships and the quantitative data from the network were not be interpreted in depth with regard to the meaning of literary works. In this paper, we propose an improved method focusing on the semantic analysis of literary works while accepting achievements of the existing studies.

3. DIALOGUE NETWORK AND OPINION NETWORK

We selected one literary work and conducted a detailed case study on this work. The selected work is Shakespeare's play <Hamlet> for the following reasons. First, as a representative work of English literature, there have been various studies on this work, which can be compared with the founding in this paper. Second, drama is a semi-structured text unlike novel, which facilitates the computational processing.

We first downloaded the works in HTML format from the 'Gutenberg Project' [11] and then preprocessed the text to build the character network. In drama, most of information transmission and emotional expression are done through the character's lines. There are descriptions and explanations, but these are generally related to the stage setting or the behavior of actors, then not directly related to the contents of the work. Therefore, we have built two different networks focusing on character's lines.

The first is a dialogue network. In the exciting studies, the links were generally made on the basis of co-appearance, however, this method can degrade the quality of analysis since the characters may appear together in the same scene even if there is no significant relationship between them [12]. In the dialogue network, the node is a speaker and a listener and the link is formed when a dialogue is made between the two. The node and link are weighted according to frequency, and the link is directional. When building a dialogue network, the key is to detect the listener. In the case of each speaker, because of the nature of play, it is tagged explicitly, but a separate task is required to detect each listener. In this paper, we set the following three rules to detect the listener. 1) The listener of each dialogue is the next speaker to appear immediately. 2) The dialogue is made on a scene-by-scene basis. In each scene, the listener of the last dialogue is the speaker just before that. 3) If each dialogue includes 'aside' tag, the listener and speaker are the same.

We detected the listener based on these rules and evaluated the accuracy based on the expert's tag. In <Hamlet>, the whole dialogue is 1260 lines, and three English majors tagged the listener for each line. The accuracy was 87.2% based on binary classification; true if the expert's tag matches the answer found by the model and false if the answer is not matched. We examined the false part and checked two characteristics; there were lots of errors when three or more people were talking each other, and it is difficult to specify the monologue (that is, the speaker and listener are the same) because there is no separate tag like 'aside'.

The second is an opinion network that represents the opinions expressed by one character on another. The opinion is important for the semantic analysis of literary works because it reflects intimate relationships between characters such as judgment and emotion. It is assumed that the expression of opinion is done through lines, so the subject of opinion is the speaker of lines. The object of opinion expressed in the line becomes another node. The node and link are weighted according to frequency, and the link is directional. When building an opinion network, the key is to find the character (i.e., the object of opinion) who is mentioned in the line. This is the same issue in the field of opinion mining, detecting an opinion target. The characters mentioned in the line may be explicit (e.g. in proper title or name), however, in most cases appear only contextually. In order to find the characters mentioned in the line, we have used several formal methods such as POS tagging and character relation matrix. And we compared the result with the expert's tag, but the accuracy was too low (i.e., 32.4%). Therefore, we built the opinion network based on the expert's tag.

4. ANALYSIS AND DISCUSSION

4.1 Overall structural characteristics

We first analyzed the structural characteristics of the dialogue network. As described in the previous study [13], the characteristics of the power-law and the small world distribution (i.e., diameter: 4, average geodesic distance: 2.24) were confirmed. Fig. 1 shows the dialogue network of <Hamlet>.
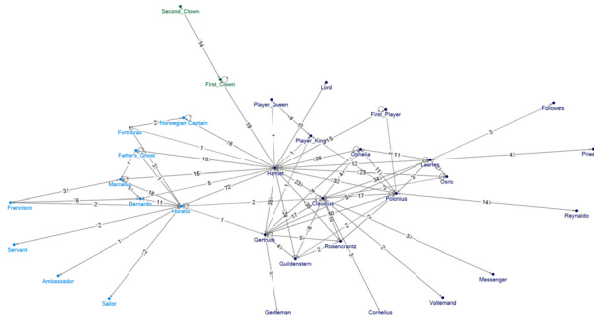

Fig. 1. Dialogue network of <Hamlet>

As clustering with the Girvan-Newman algorithm [14], the entire network was divided into three groups, where we could find an interesting point of <Hamlet>. The two central nodes of the largest group were Hamlet and Claudius, which shows that the main plot of the work is a confrontation between these two characters. The second group was centered on Laertes, Polonius and Ophelia, who are members of a family and the characters of the subplot. The third group structurally showed one of the important features of <Hamlet>. In this work, there is a play within a play called 'the murder of Gonzago', in which the characters (i.e., clowns, theater actors) were the central nodes of the third group. These results show that the character network can be used effectively in analyzing the structure and plot device of literary works.

Another interesting point is the role of Horatio as a notable node. Horatio is Hamlet's friend, an auxiliary character who does not play a major role in main conflicts or events throughout the work. However, in the dialogue network, Horatio has a higher centrality than other major characters. This indicates that Horatio is a functional character as a connector that connects different characters and smoothens the story development [15].

The overall structure of opinion network is similar to the dialogue network, however, its meaning is quite different. In the opinion network, a node with out-degrees is the character who expresses opinions and is the same as a speaker in the dialogue network. However, a node with in-degrees is the character mentioned by other character, regardless of appearance. Because of this difference, the result of clustering is different. Although clustered into three groups like the dialogue network, the central nodes of each group are different. The central nodes of the third smallest group were Laertes, Polonius and Ophelia, which were the same as the dialogue network. However, the central nodes of the first and second groups differed from the dialogue network. The central nodes of the first group were Hamlet and King's ghost, and the

central nodes of the second group were Claudius and Gertrude. This result shows that the opinion network reflects the more intimate relationship between characters including emotions, while the dialog network reflects the narrative structure focused on the action.

In this sense, the interesting node in the opinion network is King's ghost. The ghost does not appear much as a speaker or listener in the whole of the work, and only talks with Hamlet. However, the in-degree of the ghost in the opinion network is relatively high, which means that the ghost is mentioned a lot by other characters. This is related to the atmosphere of death, which is one of the characteristics of <Hamlet> [16].

4.2 Sentiment analysis

For more in-depth semantic analysis, sentiment analysis was executed on the opinion network. There are various techniques in sentiment analysis, and we adopt SentiWordNet method based on WordNet [17]. We used Python's NLTK (Natural Language Toolkit) package in preprocessing all of the lines and scored each word base on the lexical resource of SentiWordNet. And we combined the scores and assigned polarity (i.e., positive and negative) and sentiment value for each line [18].

We have divided the whole opinion network into the positive network (shown in Fig. 2) and the negative network (shown in Fig. 3).
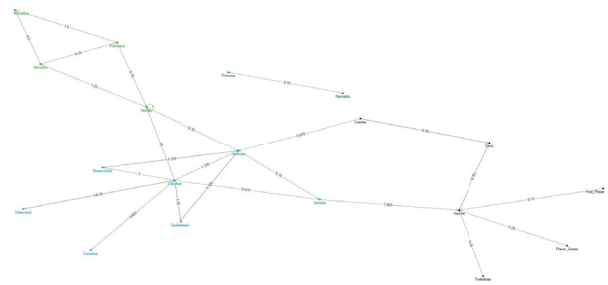

Fig. 2. Positive opinion network of <Hamlet>

The negative network is more dominant in the opinion network, which seems to be related to the characteristics of <Hamlet> as a tragedy. To make this clearer, more works divided into comedy and tragedy should be analyzed.

Through the sentiment analysis, we have found another interesting point based on the character model proposed in narratology. Greimas proposed the actantial model that considers a character as divided into six facets, called actants; subject, object, helper, opponent, sender, and receiver [19].
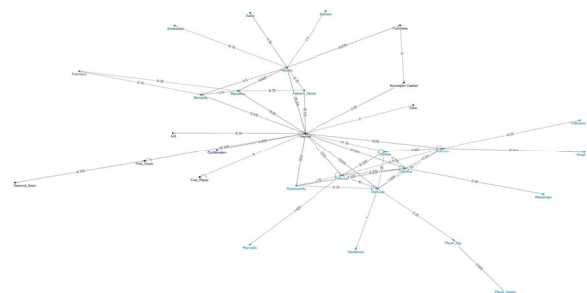

Fig. 3. Negative opinion network of <Hamlet>

In this model, the subject pursues the object, and the helper and opponent have a positive or negative impact, creating a fluctuation in the process (i.e., a journey of the protagonist). If Hamlet is the subject in the work, it can be assumed that a character who mentions Hamlet positively is the helper, and a character who mentions Hamlet negatively is the opponent. Under this assumption, we have identified the helper and opponent based on the accumulated sentiment value, and arranged the results in accordance with the actantial model as shown in Fig. 4.
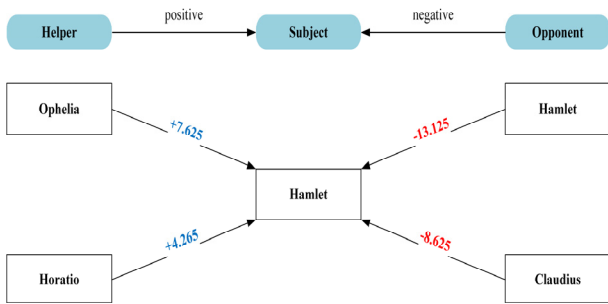

Fig. 3. Classification of sentiment value for Hamlet

The most interesting point here is that Hamlet himself is the most antagonistic character. This reflects the character of Hamlet, skeptical self-contradictory character, which is consistent with the analysis of many literary critics and psychoanalysts of <Hamlet> [20].

4.3 Dynamic analysis

In order to capture the changes according to the story development, we have built sub-networks by dividing the whole story into 5 Acts. Table 1 shows various property values analyzed in the sub-graphs of the whole dialogue network.

Table 1. Property values of each sub-graph

Act	Line	Vertex	Edge	Average geodesic distance	Graph density	Average clustering coefficient
I	945	12	42	2.01	0.28	0.649
II	735	11	35	1.63	0.29	0.464
III	944	6	7	1.53	0.2	0
IV	721	13	23	2.03	0.12	0
V	729	11	22	2.34	0.18	0.07

The plot structure model formalizes the change of the narrative contents or the recipient's emotion according to the event development. It derives from the three acts structure introduced in Aristotle's Poetics, and through the modernization by Freytag, today there are various models. In general, it divides narrative discourse into 5 or 7 stages, and specifies the characteristics of each stage or turning point. The plot structure model has become one of the key guidelines for narrative creation or analysis. Fig. 5 shows the line graph after normalizing the values presented in Table 1 as a percentage.
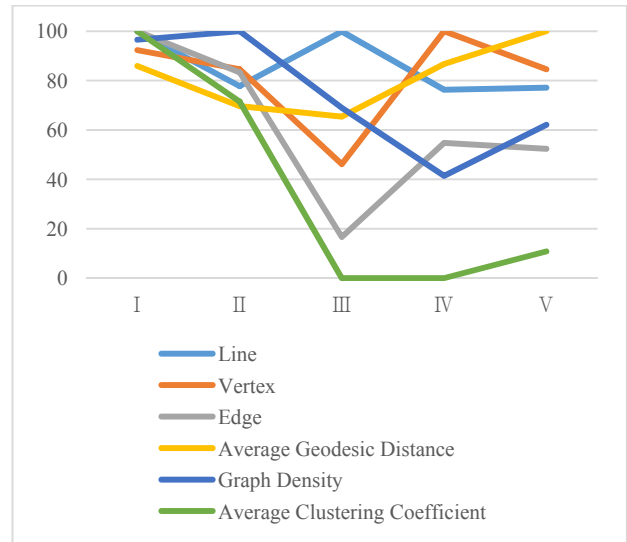

Fig. 5. Changes of the dialogue network

Further research is needed on how the patterns in this graph correspond with the plot structure model. But this method can be used to evaluate quantitatively narrative complexity or tension, which are the core indices of the plot structure model.

Table 2 shows the node with the highest frequency in the opinion network. The node with highest out-degree is the most appearing character and the node with the highest in-degree is the most mentioned character.

Table 2. Main node in the opinion network

Act	The highest out-degree	The highest in-degree
I	Horatio	King's Ghost
II	Claudius	Hamlet
III	Hamlet	Hamlet
IV	Laertes	King's Ghost
V	Hamlet	Claudius

Hamlet is the most appearing and most mentioned character in the whole work. However, dividing the work by Acts, it is confirmed that the central characters in each Act are continuously changing. For example, in Act 4, the most appearing character is Laertes and the most mentioned character is King's Ghost. This analysis allows us to understand the structure and development of the subplots, which can be used to summarize the story.

We analyzed the change of emotion between characters based on the sentiment analysis. Table 3 shows the accumulated sentiment value of Hamlet to other characters in each Act. In the case of blank, the sentiment value was not analyzed because there was no mention on the character.

Table 3. Hamlet's sentiment value to other characters

Act	to Claudius	to Ophelia	to Hamlet
I	-1.375	2.75	-0.875
II	n/a	-1	-7.375
III	-5.25	-4.25	-8.875
IV	-2.625	3.625	n/a
V	-5.375	n/a	0.125

In this table, the change of each value has a different meaning depending on the target character. The change in Hamlet's sentiment value to Claudius is related to his motivation of revenge. In general, antagonism has been strengthened, and the sentiment value increases in Act 3, which Hamlet asserts Claudius' crime, and Act 5, which Hamlet actually takes revenge. And with regard to Ophelia, the change of the polarity of sentiment value shows Hamlet's emotions of love and hatred toward her. In addition, the change of sentiment value to Hamlet himself reflects that the contempt for his indecisiveness is strengthened, and then resolved after revenge in Act 5 [21]. The change of sentiment value in the opinion network corresponds to the character arc which means the transformation of a character's motivation or emotion.

5. CONCLUSION

In this paper, we proposed an improved method of character network analysis through a case study on the play <Hamlet>. We segmented the character network into the dialogue network and opinion network, and analyzed these networks in various ways with regard to the structure and meaning of the work. The first consideration in this paper was the organic and meaningful connection between literary criticism and network analysis. We proposed the concept of dialogue network and opinion network to reflect intimate relationship between characters. And we demonstrated how to analyze literary works more sophisticatedly and systematically using these networks.

There are several limitations and these should be complemented by subsequent studies. First, the two networks proposed in this paper should be able to be built automatically and accurately. In particular, in the opinion network, it is necessary to employ the method of opinion target detection in the field of opinion mining. Second, the network analysis method should be complemented in two directions; humanities and computer science. In other words, the method should be connected to the main concepts of literary criticism; story and discourse, agent's viewpoint and knowledge, focalization, etc. And the method should be complemented by the techniques in graph theory and sentiment analysis, especially NLP for semantic analysis. Third, extensive case studies should be conducted to analyze various literary works. The works will be selected considering both the media type (e.g. play, novel, or screenplay) and the genre which is a strong classification system. Analyzing these works using the proposed method, we will compare the analysis result with the characteristics of media or genre. These follow-up studies may lead to a more complete and sophisticated method of character network analysis. We expect that this method will effectively help writers, readers, and critics to create and understand the literary works.

REFERENCES

[1] David M. Berry, "Introduction: Understanding the digital humanities," Understanding digital humanities, Palgrave Macmillan UK, 2012. pp. 1-20.

[2] Sarah Allison, Ryan Heuser, Matthew Jockers, Franco Moretti, and Michael Witmore, "Quantitative Formalism: an Experiment," Stanford Literary Lab Pamphlet 1, 2011.

[3] Franco Moretti, "Quantitative Formalism: an Experiment Network Theory: Plot Analysis," Stanford Literary Lab Pamphlet 2, 2011.

[4] J. Stiller, D. Nettle, and R. Dunbar, "The small world of Shakespeare's plays," Human Nature, vol. 14, no. 4, 2003, pp. 397-408.

[5] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," Physical Review E, vol. 69, no. 2, 2004, p. 026113.

[6] Y. M. Choi, "Greek Myth as a Complex Network," New Physics: Sae Mulli, vol. 49, no. 3, 2004, pp. 298-302.

[7] S. R. Kim, "Complex network analysis in literature: togi," New Physics: Sae Mulli, vol. 50, no. 4, 2005, pp. 267-271.

[8] Y. K. Lee, H. I. Shin, J. E. Ku, and H. Y. Kim, "Analysis of Network Dynamics from the Romance of the Three Kingdoms," The Journal of the Korea Contents Association, vol. 9, no. 4, 2009, pp. 364-371.

[9] H. Y. Kim, "Analysis of Network Dynamics from the Roman-Fleuve, Togi," The Journal of the Korea Contents Association, vol. 12, no. 11, 2012, pp. 519-526.

[10] H. C. Kwon, H. T. Kwon, and W. C. Yoon. "An information-theoretic evaluation of narrative complexity for interactive writing support," Expert Systems with Applications, vol. 53, 2016, pp. 219-230.

[11] http://www.gutenberg.org/ebooks/1524

[12] S. B. Park, K. J. Oh, and G. S. Jo. "Social network analysis in a movie using character-net," Multimedia Tools and Applications, vol. 59, no. 2, 2012, pp. 601-627.

[13] J. Stiller, D. Nettle, and R. Dunbar, "The small world of Shakespeare's plays," Human Nature, vol. 14, no. 4, 2003, pp. 397-408.

[14] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," Proceedings of the National Academy of Sciences of the United States of America, vol. 99, no. 12, 2002, pp. 7821-7826.

[15] G. Egan and M. de Grazia, "Hamlet without Hamlet," New Theatre Quarterly, vol. 23, no. 4, 2007, pp. 429-446.

[16] B. L. Rhee, "The Ghost in Shakespeare's Hamlet and Yeats's The Only Jealousy of Emer," The Yeats Journal of Korea, vol. 43, 2014, pp. 121-131.

[17] http://sentiwordnet.isti.cnr.it/

[18] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," LREC, vol. 10. 2010.

[19] G. Prince, *A dictionary of narratology*, University of Nebraska Press, 2003.

[20] J. Lacan, J. A. Miller, and J. Hulbert, "Desire and the Interpretation of Desire in Hamlet," Yale French Studies, vol. 55, 1977, pp. 11-52.

[21] S. Greenblatt, *Hamlet in purgatory*, Princeton University Press, 2013.

Ho-Chang Kwon
He received the M.S. in cinema studies from KARTS in 2010, and also Ph.D. in knowledge service engineering from KAIST in 2017. He is teaching narratology and new media at KARTS. His main research interests include digital storytelling and HCI.

Kwang-Hyun Shim
He received the B.S. in German language, and completed his doctorate course in aesthetics in Seoul National University. He is a professor working for Department of cinema studies at KARTS. His main research interests include media theory, cognitive science, and cultural studies.