Analysis of Social Network Constructed from Literary Fictions

Abstract—Social Network is a useful model to reveal hidden structure and meaningful knowledge among social atoms interacting each other in a complicated way. Till now lots of work have focused the real data of social space such as email, twitter-like instant messages, and human communities. In this paper, we try to construct another social network from a fictional space using well-known fiction literature including Harry Potter, War and Peace and Three Kingdoms (三國志演義). In every fiction, a set of virtual characters work together to maintain the main plot of a novel, so it is quite natural to reconstruct the social network by analyzing the textual structure of subject fictions. The main process of constructing the social network from novels consists of three steps (1) locating characters in textual space of fiction (2) computing the closeness(interaction degree) of all character pairs based on textual distance (3) constructing the Social Network from Fiction (SNF) by mapping characters to vertices and their strength of relations to edges between characters. For experiment, We selected more than 20 novels in plain text including Tolstoy’s “War and Peace”, J.K. Rowling’s “Harry Potter”, a well-known Chinese novel “Three Kingdoms (三國志演義)”. It is very interesting that the topological structure of SNF is quite similar to the common social network based on real persons in terms of power-law distribution of node degree and path distance. Also we showed that it is possible to classify all characters according to the importance level in fictions by applying the minimum spanning tree from SNF. Main contribution of this paper is that our SNF model would be exploited to reveal the deep structure of fictions by graph topology rather than the traditional category such as a short or long novel.

Keywords—Social Network, Virtual characters, Graph theory, Minimum spanning tree, Literature analysis

I. MOTIVATION

In generally, linguistic analysis such as studying contextual meaning of words are used for identifying features of literatures. However, analysis of literature using semantically-oriented techniques have the difficulty in automatically determining meaningful interpretations. In order to analyze the subtle meaning of words, it is necessary to read literature from beginning to the end by human. Therefore, literature analyzing with the linguistic approach is complicated and time consuming. In our previous paper [8], we numerically calculated the relationship between agents that appeared in literature, and construct the Social Network from Fiction (SNF). The agent can be any words that were used in literature. For example, we can construct SNF from relations of “Me” and “House”, “Movie” and “coke”, “eat” and “diner”, etc. This method does not require any complicate linguistic analysis, so it is possible to construct in a short time by computers.

In this paper, we only considered characters as the agent. A network that made up of the characters can be said the social networks of the virtual world, so many existing social network analysis methods can be applied. For example, we can extract major relations from networks by Minimum distance spanning tree of social network. Also, we can measure the importance level of characters by analyzing depth of the spanning tree.

Our approach has 3 processes, (1) locating characters in textual space of fiction (2) computing the closeness of all character pairs based on textual distance (3) constructing the Social Network from Fiction (SNF) by mapping characters to vertices and their strength of relations to edges between characters, with edge weights corresponding to the frequency and position of their appearance.

We carried out this work on a corpus of more than 20 novels, including Tolstoy’s “War and Peace”, J.K. Rowling’s “Harry Potter”, a well-known Chinese novel “Three Kingdoms (三國志演義)”. Our results provide evidence that the feature of literatures can be obtained from SNF graph through a graph theoretical approach.

In the following sections, we survey related work on studies of literature and social networks. We then describe our methods we use to extract characters and construct SNF graph, along with graph theoretical approach to analyzing their characteristics. After then, we present experimental results and analyze them.

II. RELATED WORK

It is general that the computer-based literary textual analysis has typically performed at the word level. This work has focused the discover authorial style and the lexical patterns of word use [3]. Recently, researchers started to study properties of co-occurrence words in natural language space [7], [10], [11]. Since all words in a text are closely related, so the co-occurrence pattern of some pair of words may reveal the important features hidden in the text. One interesting application of this co-occurrence analysis gave