sustainability

Article

Discovering Social Desires and Conflicts from Subculture Narrative Multimedia

O-Joun Lee [1], Heelim Hong [2], Eun-Soon You [3,*] and Jin-Taek Kim [1,*]

[1] Future IT Innovation Laboratory, Pohang University of Science and Technology, 77, Cheongam-ro, Nam-gu, Pohang-si, Gyeongsangbuk-do 37673, Korea; ojlee112358@postech.ac.kr
[2] School of Integrative Engineering, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Korea; heelim.korea@gmail.com
[3] Department of French Language and Culture, Inha University, 100, Inha-ro, Michuhol-gu, Incheon 22201, Korea
* Correspondence: jiwony71@gmail.com (E.-S.Y.); jintaek@postech.ac.kr (J.-T.K.); Tel.: +82-54-279-8853 (J.-T.K.)

Received: 21 September 2020; Accepted: 6 December 2020; Published: 8 December 2020

Abstract: This study aims at discovering social desires and conflicts from subculture narrative multimedia. Since one of the primary purposes in the subculture consumption is vicarious satisfaction, the subculture works straightforwardly describe what their readers want to achieve and break down. The latent desires and conflicts are useful for understanding our society and realizing smart governance. To discover the social issues, we concentrate on that each subculture genre has a unique imaginary world that consists of inventive subjects. We suppose that the subjects correspond to individual social issues. For example, game fiction, one of the popular genres, describes a world like video games. Under game systems, everyone gets the same results for the same efforts, and it can be interpreted as critics for the social inequality issue. Therefore, we first extract subjects of genres and measure the membership degrees of subculture works for each genre. Using the subjects and membership degrees, we build a genealogy tree of subculture genres by tracing their evolution and differentiation. Then, we extract social issues by searching for the subjects that come from the real world, not imaginary. If a subculture work criticizes authoritarianism, it might include subjects such as government officials and bureaucrats. A combination of the social issues and genre genealogy tree will show diachronic changes in our society. We have evaluated the proposed methods by extracting social issues reflected in Korean web novels.

Keywords: social issue mining; computational narrative analysis; subculture multimedia; genre classification; genre genealogy tree

1. Introduction

Recently, smart city is a prospective research field to solve various urban problems [1,2] including city sustainability issues [3,4]. Context-awareness for cities is a basis of smart city applications and intelligent urban services [1,5,6]. Most of the smart city applications have been focused on urban ecology issues [3,4] and infrastructure management [1,7]. However, regarding cities as living spaces of citizens, we should also consider economic, political, and cultural aspects of city sustainability [8]. To solve the human-centric urban problems, we need contextual information intimately connected to citizens' lives. To extract the information, various studies have been conducted on discovering social issues mainly from social media [1,5,6,9]. The social issues vary from citizens' mutual interest (e.g., iconic movie directors or big sports games) to serious social problems (e.g., housing shortage and real estate bubble). However, those data sources have limitations for reflecting inner minds and feelings of citizens. First, flooding rumors and fake news on social media cause uncertainty for input data [10]. Also, since social media are public places, users cannot ignore the public eyes during posting [11,12].

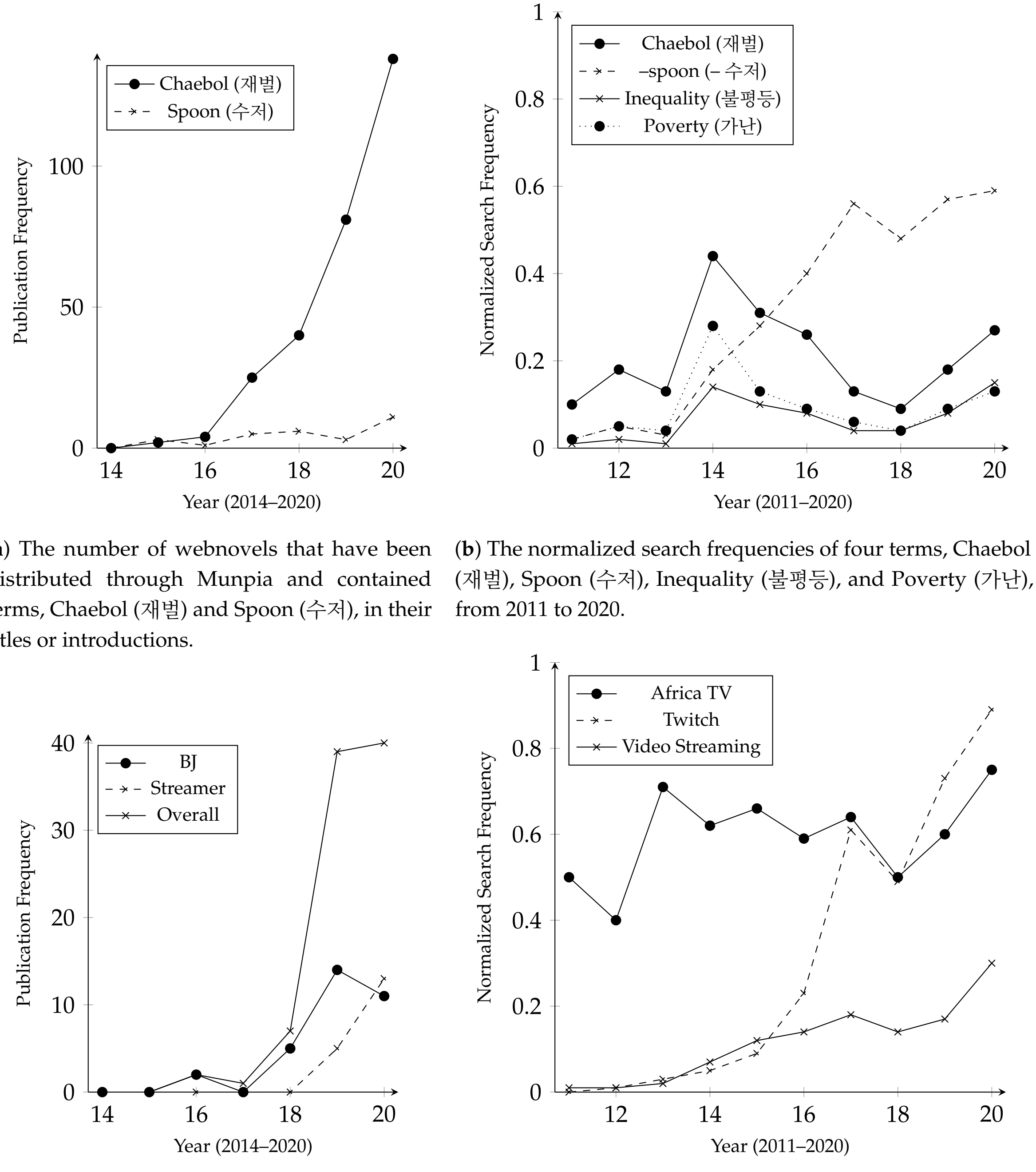
This study attempts to extract social issues from citizens' consumption of narrative multimedia (e.g., movies, TV series, and novels) instead of social media. Narrative multimedia reflects social phenomena, and various studies attempted to analyze the reflected social issues computationally [13–15]. Although we cannot analyze purchase histories or playlists of each citizen, context of the narrative multimedia industry (e.g., trends of genres or topics) will reveal subjects or issues that the citizens empathize with. Also, according to Jean Baudrillard [16], consumption is the way modern society speaks for itself. However, these studies focused on a few particular issues and did not present methods for extracting general social issues automatically from large corpora. Also, we cannot merely apply the existing social issue discovery methods [9]. These methods are mostly based on term frequencies and sentiment analysis. However, most of the proper nouns and named entities in narratives are imaginary things, and diverse literary, figurative, and rhetorical expressions in them hinder accuracy of the sentiment analysis.

We improve these problems by restricting our data sources into subculture narrative multimedia (e.g., web novels, webtoons, etc.). For social issue discovery, subculture narratives have two strong points, (i) frequent interactions between authors and customers and (ii) distinct genres, compared to the conventional ones. Most of the subculture works are published serially through web platforms (e.g., NAVER SERIES (https://series.naver.com/), Kakao Page (https://page.kakao.com/), and so on) that provide community spaces for their users (both authors and consumers). The users discuss each episode's content, and the discussions are a part of subculture authoring [17–19]. Some authors ask consumers' comments for stories, explain characters' behaviors, give hints for further stories, and open contests for titles of their works or character names. The authors are also enthusiastic consumers of other authors, and some readers try to author and post their drafts. These points make the subculture multimedia reflect social issues more agilely and vividly than the conventional narrative multimedia.

Second, genres of subculture works are more distinct than of the conventional multimedia [20,21]. Subculture works describe imaginary worlds, and each genre shares common imaginings (subjects). The authors create their works by adding a few inventive subjects to the shared ones. Sometimes, a good blending becomes a new genre. The authors state genre characteristics included in their works, and their readers also ask the authors to follow practices of the genres.

To attract readers' attention, authors attempt to put subjects related to popular issues in the real-world. The distribution environment makes this reflection of social issues close to real-time. For example, a Korean movie director, Bong Joon-ho, won Academy awards in January 2020, and his winnings were a massive issue in Korea. In Munpia (https://www.munpia.com/), one of the most popular web novel platforms in Korea, we found ten web novels that employ movie directors as protagonists. Seven of them started being published in 2020, and only three web novels started before the winnings. These novels are included in a genre, 연예계물 (Entertainment Fiction).

'재벌 (chaebol)' and '–수저 (– spoon)' are symbolic terms of inequality issues in Korea. Chaebol indicates affluent families that control large conglomerates. '– spoon' is originated from an English expression, silver spoon, that indicates inherited wealth. Korean citizens use this term to describe their inherited social classes like 금수저 (golden spoon) and 흙수저 (dirt spoon) (https://www.nytimes.com/2019/10/21/world/asia/south-korea-cho-kuk-gold-spoon-elite.html). Figure 1a presents the number of web novels that have been distributed through Munpia and contained the two terms in their titles or introductions each year (2014–2020). Figure 1b shows normalized search frequencies of 재벌 (chaebol), –수저 (– spoon), 불평등 (inequality), and 가난 (poverty) in Korea, which were collected from Google Trends (https://trends.google.com/trends). The search frequencies of 'chaebol,' inequality, and poverty showed peaks in 2014, and the frequency of '– spoon' was steadily increasing since 2013. Web novels containing the two terms also appeared around 2014 and have been consistently increasing. Moreover, the search frequency of '– spoon' and the number of web novels containing '– spoon' commonly showed temporary decrements around 2018. 'Chaebol' is related to a genre, 기업물 (Business Fiction).



(**a**) The number of webnovels that have been distributed through Munpia and contained terms, Chaebol (재벌) and Spoon (수저), in their titles or introductions.

(**b**) The normalized search frequencies of four terms, Chaebol (재벌), Spoon (수저), Inequality (불평등), and Poverty (가난), from 2011 to 2020.

(**c**) The number of webnovels that have been distributed through Munpia and described their protagonists' jobs as BJ, Streamer, and merely video streaming creators.

(**d**) The normalized search frequencies of three terms, Africa TV (아프리카TV), Twitch (트위치), and Video Streaming (인터넷 방송), from 2011 to 2020.

Figure 1. Examples of social issues reflected by web novels distributed through Munpia. (**a**,**c**) present the number of web novels, including the social issue-related terms, published at Munpia each year. (**b**,**d**) show normalized search frequencies of the terms in Google in Korea each year.

With the commercial success of video streaming platforms (e.g., YouTube, Twitch, Africa TV, etc.), a new genre, streamer fiction (스트리머물), has been popularized. This genre uses video streamers as protagonists. The streaming platforms have different names for calling video streamers. Africa TV uses BJ (Broadcasting Jockey), Twitch uses Streamer, and so on. Also, web novel writers use these various names for calling video streamers. We have examined whether streaming platforms' popularity has correlations with usage frequencies of those names in web novels. Africa TV and Twitch were the first and second most popular video streaming platforms in Korea, and Twitch has overtaken Africa TV in terms of their search frequencies in 2019, as shown in Figure 1d. When video streaming started becoming popular, most of the streamer fiction called their protagonists 'BJ.' However, after Africa TV and Twitch got similar search frequencies, the web novel authors started avoiding usages of terms related to particular video streaming platforms, as shown in Figure 1c. In the same period, the usage frequencies of 'Streamer' have been increased dramatically and have overtaken 'BJ' in 2020.

These examples underpin that subculture narratives reflect social issues, and we can discover the reflected issues from trends of genres. Although the subculture is resistance against hegemony, according to Dick Hebdige [22], we do not focus on only critical issues. Since vicarious satisfaction is one of the significant reasons for subculture consumption, subculture works describe stories that their protagonists achieved social desires (e.g., becoming Chaebol, famous video streamer, professional gamers) as much as they suffered by and broke down social irregularities (e.g., inequality, bureaucracy, aristocracy, and so on). Thus, this study aims at discovering social issues that are reflected due to (i) interactions between authors and readers, (ii) pursuits of popularity, and (iii) vicarious satisfaction, not interpreting artistic intentions of authors.

The conflicts and desires can be guidelines for decision makings in various smart city applications. However, social issues discovered from subculture narratives will be extremely varied, as shown in Section 4. Thus, the issues should be distilled to be used for a particular smart city application. For example, it is not easy to find connections between the popularity of streamer fiction and the smart grid. Although this study does not cover refining the issues for a specific application, we can discuss candidate applications that can employ the discovered social issues.

In data-driven policymaking, occurrence frequencies of discovered social issues and emotional words appearing with the issues can reveal policy demand and urgency for handling the issues. For example, laws for regulating content of video streaming have been absent in Korea, despite its massive influence. Korean national assembly has started discussing the regulation after a national scandal for video streamers' undisclosed ads (https://www.koreatimes.co.kr/www/tech/2020/08/133_294254.html). However, as shown in Figure 1c, genre trends of web novels have said exponentially growing influences of video streaming since 2016. This point is also closely connected with market trend analysis and business intelligence. Figure 1c presents the rapid chase of Twitch against Africa TV, which is reflected in the webnovel production. Even possession fiction, a recently popular genre, contains criticisms for the subculture industry trends. Furthermore, the proposed methods can be used in other social science studies. There have been studies for analyzing social issues reflected in narrative multimedia [14,15,23]. However, their methods were not designed for the social issue discovery; e.g., the conventional Word2Vec [24]. This study will provide more sophisticated tools for dealing with narrative multimedia and social issues.

Therefore, this study proposes methods for discovering social issues by analyzing trends of subculture genres automatically. First, we discover key subjects of genres by concentrating on that subculture works in each genre share a unique imaginary world. Since the main subjects are key components of the imaginary world, terms indicating the subjects frequently occur in the subculture works. Also, subculture communities make neologisms or change meanings of existing words to call the subjects that are parts of the imaginings [20]. Thus, the terms for key subjects (keywords) have different meanings between subculture works and general texts. We use the frequency and semantic difference of words to distinguish keywords from the other words.

The two assumptions for discovering subjects can also be used to measure membership degrees of subculture works for each genre. If a subculture work is in a genre, keywords of the genre will occur in the subculture work frequently and have the same meaning in both the work and the other works in the genre. Using the membership degrees, we can conduct multi-label genre classification for subculture works. Also, subculture genres are continuously evolved and differentiated. Novel genres are born by adding new and attractive subjects to existing genres. Thus, by tracing the inheritance of subjects, we can build a genealogy tree of subculture genres. Correlations between imaginary worlds might be correlations of reflected social phenomena, and the genre genealogy reveals both correlations.

Finally, subjects that correspond to social issues come from the real world rather than imaginings. Although artistic narrative works employ symbols, figurative expressions, and mise-en-scène to criticize the issues (e.g., vertical movements symbolize social classes in 'Parasite (2019)'), subculture works are more straightforward than the conventional narrative works. Therefore, different from searching imaginary subjects, we discover subjects correlated with the social issues from a genre by searching for words that have the same meanings in both the genre and general texts. Using the discovered issues and the genre genealogy, we can detect social phenomena and understand changes in our society.

To verify the proposed methods, we conducted experiments on Korean web novels. In Korea, various platforms are distributing subculture works (mainly webtoons and web novels) online. Recently, these platforms (e.g., NAVER SERIES and Kakao Page) are extending their customer range to general users, not only subculture mania. Thus, the subculture multimedia reflect Korean society enough to evaluate whether the proposed methods can extract social issues in Korean society. We chose web novels among the two popular subculture multimedia, since extracting terms from textual media is much easier than from graphical one. Based on the case study in Korean web novels, we validate the following research questions:

- RQ 1. Word semantic difference between subculture works and general texts is effective to detect imaginary subjects of subculture genres.
- RQ 2. Inheritance of the subjects is useful to trace the differentiation of subculture genres.
- RQ 3. Contrary to RQ 1, semantic consistency of words can distinguish subjects satirizing social issues from the others.

The remainder of this paper is organized as follows. Section 2 presents related studies mainly conducted in the computational narrative analysis and the digital humanities. Section 3 proposes methods for (i) detecting main subjects of subculture genres, (ii) classifying subculture works into the genres, (iii) building the genre genealogy tree, and (iv) discovering social issues criticized by the genres. We evaluate the proposed methods and validate our assumptions in Section 4. Finally, Section 5 describes limitations and future directions of this study.

2. Related Work

To the extent of our knowledge, there have not been studies proposing automated methods for discovering social phenomena reflected in narrative multimedia. Content analysis of subculture multimedia has also not been frequently studied. Braincolla (https://braincolla.com/) is a company that provides a recommendation engine for one of the Korean subculture platforms, Joara (http://www.joara.com/main.html). Similar to this study, the recommendation engine statistically analyzes occurrences of words to examine subjects and backgrounds of subculture works. Although this engine has been operated on the commercial platform, there has not been any academic publication covering detailed procedures and performance evaluation of the recommendation engine.

A few studies based on Chinese web novels presented features that can be applied to our research. Su [25] analyzed code-switching (i.e., a word in a language appears among words in other languages) in Chinese web novels. He conducted a statistical analysis for usages of the code-switching according to web novel genres. The analysis shows that the code-switching frequencies are correlated with genres and historical backgrounds of web novels. Although he also presented examples for code-switching purposes, the purposes were closely connected with characteristics of the Chinese language and writing system. However, we could find frequent code-switching in Korean web novels, and it was mainly used for clarifying meanings of imaginary subjects by showing two languages together (e.g., "파이어볼 (Fireball)"). Although this feature's meanings might depend on languages, it can help improve the accuracy of genre classification and keyword extraction. Lin and Hsieh [26] attempted to find which factors make a web novel popular. They used mainly three features, keywords extracted based on TF-IDF, usages of function words, and lexical diversity. Although they could not find correlations of these features with popularity features (e.g., the number of hits, favorites, and comments), the proposed features were genres and writing styles of web novels.

Additionally, a study had a similar purpose to this study. Jo and Oh [23] attempted to discover web novel readers' inner desires from hashtags annotated on web novels. However, their method also cannot support the automated social issue discovery. They applied TF-IDF (Term Frequency-Inverse Document Frequency) to hashtags annotated by an online book store (Ridibooks) and conducted qualitative discussions based on the TF-IDF values.

The existing studies on the computational narrative analysis mostly focused on conventional narrative multimedia, such as movies, novels, and TV series [27–29]. Also, these studies aimed at analyzing storytelling methods rather than subjects and backgrounds; e.g., plot structures [29–31], character roles [32,33], major events [34–36], and so on. Although a few studies [37,38] attempted to classify narrative works into genres, they were also based on interactions among characters, which are related to narrative development rather than subjects. Since these studies aimed at analyzing individual narrative works, their approaches are not appropriate to discover macroscopic events from narrative multimedia.

On the other hand, various digital humanities studies attempted to analyze social phenomena reflected in narrative multimedia, despite absence of systematic and automatic methods. Michel et al. [13] analyzed word usages in millions of books to discover cultural changes. They interpreted temporal distributions of correlated words by comparing the distributions with each other. Peaks of the distributions revealed social events, and their differences between languages (or nationalities) showed cultural differences or information propagation (e.g., Tiananmen in Chinese and English). Their study was meaningful in terms of demonstrating effectiveness of quantitative multimedia analysis as a tool for analyzing social phenomena.

A few studies concentrated on a more specific problem: social equality. Grayson et al. [14] analyzed gender roles reflected in the 19th-century novels by using word embedding techniques. They examined semantic differences of words in the novels according to genders of authors, and the differences were significant in gendered pronouns. Although their study did not accompany a statistical validation, we referred their approach to compare genres. Chen and Cui [15] dealt with gender bias in movies by applying social network analysis (SNA) techniques to the movies. They analyzed social relationships and dialogues between characters, considering the characters' genders. Using the analysis results, they predicted whether movies can pass the Bechdel test and observed changes in gender bias according to time. According to their results, gender bias issues have been improved during the last few decades. Caty Borum Chattoo [39] analyzed Academy Awards winners to show gender, racial, and ethnic equality issues. Although the analysis did not cover movies' content, the significant gaps between genders and between races indicated that the media industry is sensitive to social phenomena for better or for worse. As Chen and Cui [15] said, narrative worlds in artworks are a microcosm of the real world.

There have also been studies attempted to analyze large-scale artwork corpora but not aimed at social issues. Although these studies are not directly correlated with our study, their approaches can be applied to further research on subculture multimedia. Chae et al. [40] have proposed methods for clustering artworks according to their content by using social tags. They built networks of artworks based on shared social tags between the artworks and applied network analysis techniques. Since subculture genres are also a kind of social tag for subculture works' content, this approach and our study have a common point: understanding the content of artworks by analyzing a combination of social tags. However, this approach does not cover what kinds of features in artworks are correlated with each social tag, while it is the most primary concern of our study.

Park et al. [41] measured novelty of artworks by analyzing their citation networks. Their method exhibited effectiveness in detecting paradigm-shifting between genres of classical music. Although it is difficult to detect references between artworks in other kinds of media, we can apply this approach to exploit the propagation of imaginary subjects between subculture works. The propagation will enable us to analyze genre differentiation microscopically and discover landmark pieces that laid cornerstones of each genre.

Jung et al. [42] attempted to analyze relationships among narrative works that share a common narrative world. They classified connections between narrative works according to how the works extend the narrative world, based on characters and events that commonly appeared in the works. Since subculture genre differentiation is led by adding fresh subjects into shared imaginary worlds of each genre, this approach can be used to observe births of new genres, similar to the above study [41]. However, it is not easy to distinguish imaginary subjects from named entities. Although our proposed methods dilute the problem by examining whether the subjects appear all over the genre, this solution is too rough to be applied to a single artwork.

In terms of analysis methods, there have been mainly two approaches: (i) information retrieval (IR) and simple natural language processing (NLP) techniques and (ii) SNA techniques. Word embedding and topic modeling techniques were frequently used to analyze the semantics of words in narrative artworks. Grayson et al. [43] conducted Word2Vec [24] for all the words in the artworks, including character names. They assumed that characters with similar roles have similar locations in the embedding space. Peng and Jung [44] applied LDA (Latent Dirichlet Allocation) to discover metaphorical expressions from Chinese poems. Similar to our proposed methods, they assumed that metaphorical expressions have different meanings between the poems and general texts. Also, sentiment analysis was applied to interpret the metaphorical expressions by searching for words with similar sentiments. Reagan et al. [29] used fluctuations in frequencies of emotional words to compare storytelling methods of narrative works. However, their study did not validate whether the fluctuations have correlations to readers' perception of the narrative artworks' content. For the similar purpose, Micha Elsner [31,45] employed occurrence frequencies of characters together.

Various studies have been conducted for extracting social networks in narrative artworks and discovering narrative features from the social networks [46]. These studies extracted the social networks from texts [47,48], videos [33,49,50], or both of them [51,52]. A few studies applied sentiment analysis to enrich the social networks [34,53]. Node centrality on the networks reflected character roles [33], structures of the social networks were correlated to major events in stories [34–36], changes in the structures had relevancy with fluency of stories [54], and occurrence frequencies of characters showed plot structures [30,31]. These extracted narrative features were applied to mainly summarization [50,55,56].

Over the analysis within a single story, a few studies attempted to compare narrative works in terms of their stories. Early studies [57,58] used conventional features in SNA, such as cohesion of communities and node centrality. However, the conventional features are not enough to reflect various structural characteristics of the social networks. Also, it is difficult to conjecture what kinds of narrative characteristics are reflected by the conventional features. Therefore, a few recent studies [27,30,59,60] applied graph embedding techniques to compare stories using simple vector similarity metrics.

Since our study currently does not cover narrative features of subculture works, only subjects, the existing studies in the computational narrative analysis do not have many points that can be applied to this study. However, stories are one of the most significant components composing narrative multimedia as much as subjects and physical expressions. Thus, if subculture works in the same genre share imaginary worlds, which are formulaic and typical, there will also be distinct common points in their storytelling methods. In further studies, the narrative features will be combined with the proposed genre analysis methods.

3. Discovering Social Issues from Subculture Multimedia

Narrative artworks reflect our society. Narrative worlds come from the real world, and events in the imaginary worlds reflect our social issues. A few studies [14,15,39] attempted to analyze social equality issues reflected in the conventional narrative multimedia (e.g., movies). However, there have not been methods for extracting social issues form narrative works automatically. Most of the narrative works do not directly state what they criticize. For example, realism authors (e.g., Stendhal and Honoré de Balzac) stand back and describe a sequence of events realistically. Nevertheless, the state-of-the-art techniques in the computational narrative analysis do not even come close to interpreting the events' superficial meanings (needless to say comprehension of what the events satirize and criticize). On the other hand, subculture narrative works are far more straightforward than the conventional ones. Their subjects and backgrounds are tightly correlated to what they criticize or desire.

Therefore, we first classify subculture works into their genres, which are groups of works sharing key subjects. Since most subculture works deal with imaginary worlds, the subjects can be discovered by searching for words used in different meanings from general texts. The subculture genres have short life-cycles and are differentiated according to time. Thus, we represent the differentiation of genres as a genealogy tree. Finally, social issues that correspond to each genre can be detected by examining words used in similar meanings to the general texts, since targets of desires and criticism should co-exist in both real and imaginary worlds. Moreover, we can trace diachronic changes in social issues and analyze correlations between the issues by combining the social issues and the genealogy tree.

3.1. Genealogy Tree of Genres

Subculture genres mostly start from a monumental work, thrive by numerous imitators, and lose popularity by other genres' advents. The new genres are differentiated from existing ones, and boundaries between the genres are not always distinct. However, we still have a clue. The subculture communities give names to the genres, and the authors and readers tag those genre names on each work. In Korean subculture communities, the names are in a format of '–물 (a suffix used to indicate an object, an entity, or a material)' [20]. For example, in return fiction (회귀물), protagonists go back to the past and try to live a perfect life by avoiding all mistakes made at their previous trials. Game fiction (게임물) uses famous video games (e.g., League Of Legends) as its background. This genre connects successes in the games to success in real life. These genres are commonly based on dissatisfaction for real life.

Similar to the examples, subculture works in the same genre share key subjects and backgrounds. The game fiction and its branches include game-originated subjects, such as 'quests,' 'levels,' 'dungeons,' and so on. As key components of their imaginary worlds, these terms have similar meanings in the genres and different meanings between the genres and the others. Based on the semantic difference, we define the subculture genre as follows;

Definition 1 (Subculture Genre)**.** *The genre indicates a group of artworks that share subjects, ambiances, backgrounds, types of narrative development, physical expressions, and so on. Although the genre covers various narrative characteristics, the subculture communities use this concept to distinguish subculture works according to their key subjects and backgrounds. Thus, subculture genres have a group of words, which are correlated with their shared subjects and have unique meanings in the genres. The relationship between subculture works and genres can be formulated as:*

$$\mathcal{G}_i = \left\{ \mathcal{N}_a \middle| \mu_{\mathcal{G}_i}(\mathcal{N}_a) \geq \theta \right\}, \;\; \mu_{\mathcal{G}_i}(\mathcal{N}_a) \propto \sum_{w_n \in \mathcal{K}(\mathcal{G}_i)} \frac{1}{\mathcal{D}(w_n, \mathcal{G}_i, \mathcal{N}_a)} \times f(w_n, \mathcal{N}_a), \tag{1}$$

where $\mu_{\mathcal{G}_i}(\mathcal{N}_a)$ indicates a membership degree of the a-th subculture work $\mathcal{N}_a$ for the i-th genre $\mathcal{G}_i$, θ denotes a minimum threshold for the membership degree, $\mathcal{K}(\mathcal{G}_i)$ refers to a set of words that are related to essential subjects of $\mathcal{G}_i$, $\mathcal{D}(w_n, \mathcal{G}_i, \mathcal{N}_a)$ is a function for measuring semantic difference between the n-th word w_n in $\mathcal{G}_i$ and w_n in $\mathcal{N}_a$, and $f(w_n, \mathcal{N}_a)$ indicates an occurrence frequency of w_n in $\mathcal{N}_a$.

Vicarious satisfaction is one of the significant reasons for subculture consumption. Thus, subculture works reflect desires or conflicts in the real world, and their subjects will be what authors (and readers) want to achieve or break down. Thus, the rise and fall of genres can show diachronic changes in social issues. To analyze the changes, we build the genealogy tree of genres to trace roots and branches of the genres and related social issues.

When a genre is differentiated from existing ones, the novel one inherits subjects of the existing ones. The inheritance reveals whether two genres are relatives, and their publication frequencies show which genre is an ancestor of the other. For example, possession fiction (빙의물) describes protagonists that suddenly became a fictional character of video games or web novels. Both possession fiction and game fiction commonly include lots of game-originated subjects. Their protagonists commonly get 'achievements,' raise their 'levels,' and clear 'quests.' The shared subjects let us know the two genres are relatives. However, recently, possession fiction has a higher publication frequency than game fiction. Also, game fiction was popularized much sooner than possession fiction (the 2000s and 2010s, respectively). Thus, we can assume that game fiction is one of the ancestors of possession fiction. Using the inherited subjects and publication frequencies, we define the genre genealogy tree as follows;

Definition 2 (Genealogy tree)**.** *The genre genealogy tree indicates a hierarchical, directed, and acyclic graph for presenting differentiation of genres. A genre is established by mixing and extending its parent genres. Thus, we can understand new genres using an assumption that the children inherit characteristics (e.g., subjects and backgrounds) of their parents. If two genres have a parent-child relationship, we can detect it from shared subjects between the two genres. Then, we designate a popularized genre sooner than the other one as a parent and vice versa. When $\mathcal{T}_\mathcal{G}$ is a genealogy tree of a genre set $\mathcal{G}$, $\mathcal{T}_\mathcal{G}$ can be defined as a matrix representing relationships between genres $\mathcal{R} \in \mathbb{R}^{|\mathcal{G}| \times |\mathcal{G}|}$ and a vector describing layers of the genres $\mathcal{L} \in \mathbb{N}^{|\mathcal{G}|}$. This can be formulated as:*

$$\mathcal{T}_\mathcal{G} = \langle \mathcal{R}, \mathcal{L} \rangle = \left\langle \begin{bmatrix} r_{1,1} & \cdots & r_{1,|\mathcal{G}|} \\ \vdots & \ddots & \vdots \\ r_{|\mathcal{G}|,1} & \cdots & r_{|\mathcal{G}|,|\mathcal{G}|} \end{bmatrix}, \begin{bmatrix} l_1 \\ \vdots \\ l_{|\mathcal{G}|} \end{bmatrix} \right\rangle, \tag{2}$$

where $|\cdot|$ refers to the set size, $r_{i,j}$ indicates a relationship of $\mathcal{G}_i$ for $\mathcal{G}_j$, and l_i means that $\mathcal{G}_i$ is on the l_i-th layer. When $\mathcal{G}_i$ is a parent of $\mathcal{G}_j$, $r_{i,j} = 1$ and $r_{j,i} = -1$; otherwise $r_{i,j} = r_{j,i} = \emptyset$. Also, if $r_{i,j} = 1$, $l_j < l_i$, and l_i is a natural number $\leq |\mathcal{G}|$.

Based on these definitions, the following sections introduce methods for (i) discovering key subjects of genres, (ii) classifying subculture works into the genres, and (iii) revealing roots and branches of the genres.

3.1.1. Extracting Genre Keywords

A considerable number of subculture works have genre annotations labeled by their authors (or editors). The annotations are significant to readers' first impressions of the works. Although the readers can evaluate writing styles or narrative development after reading the works, the genres let the readers know whether the subculture works deal with their preferable subjects. Also, we can label subculture works by using readers' comments in subculture communities or platforms. In this study, we have collected 200 web novels, and 117 of the 200 novels have genre annotations.

From the annotated novels, we can discover each genre's keywords, which are related to common subjects, backgrounds, or character types of the genre. The keywords correspond to components of shared imaginary worlds in the genre. Therefore, most of the keywords (i) are neologisms or (ii) have semantic differences between usages in general texts and in subculture works. Since a few genres (e.g., the apocalypse fiction or picaresque fiction) have unique ambiances, (iii) emotional words can also be the keywords.

Peng and Jung [44] estimated the semantic difference of words between poems and general texts using LDA. If we compare narrative works with general texts, occurrence frequencies of words, which are the basis of LDA, will clearly show their difference. However, the subculture genres are differentiated and evolved according to time. Keywords of the genres will be inherited to their children with little changes in their meanings. Therefore, we should observe semantic relationships between words more finely than document-level co-occurrences.

Therefore, we use the representation learning instead of topic modeling. If a word w_n has different meanings between a general text corpus and a subculture multimedia corpus, relative distances of w_n to other words in the embedding space will differ between the two corpora. Also, if the word w_n has consistent meanings only among artworks in a genre $\mathcal{G}_i$, we can assume that w_n is a keyword of $\mathcal{G}_i$. Finally, keywords of $\mathcal{G}_i$ have high occurrence frequencies in a subculture work $\mathcal{N}_a$, when $\mathcal{N}_a$ is a member of $\mathcal{G}_i$.

The Skip-Gram method of Word2Vec [24] is used for embedding words in both general texts and subculture works. Since Skip-Gram is well-known and widely-used, we do not explain its procedures in detail. We first embed words in a general text corpus (e.g., Wikipedia dump (https://dumps.wikimedia.org/backup-index.html) or news articles) and use word vectors as pre-trained vectors for embedding the words in subculture works. Thus, a word w_n gets multiple vector representations from embedding for the general texts ($\overrightarrow{w_{n,G}}$) to embedding for a subculture work $\mathcal{N}_a$ ($\overrightarrow{w_{n,a}}$). Additionally, we use introductions and comments for the subculture works collected from subculture platforms and written by their authors and readers, together for embedding words in each subculture work. The interactions in subculture platforms could contain more concise explanations for the imaginary worlds than literary expressions in the subculture works.

Semantic similarity between words within a corpus is estimated by using the cosine similarity of word vectors. Then, a semantic difference for a word between two corpora is measured based on the similarity. When $\mathcal{W}$ indicates a universal set of words, the semantic similarity between words can be represented as:

$$\mathcal{S}(G) = \begin{bmatrix} \mathcal{S}(G)_{1,1} & \cdots & \mathcal{S}(G)_{1,|\mathcal{W}|} \\ \vdots & \ddots & \vdots \\ \mathcal{S}(G)_{|\mathcal{W}|,1} & \cdots & \mathcal{S}(G)_{|\mathcal{W}|,|\mathcal{W}|} \end{bmatrix}, \ \mathcal{S}(\mathcal{N}_a) = \begin{bmatrix} \mathcal{S}(\mathcal{N}_a)_{1,1} & \cdots & \mathcal{S}(\mathcal{N}_a)_{1,|\mathcal{W}|} \\ \vdots & \ddots & \vdots \\ \mathcal{S}(\mathcal{N}_a)_{|\mathcal{W}|,1} & \cdots & \mathcal{S}(\mathcal{N}_a)_{|\mathcal{W}|,|\mathcal{W}|} \end{bmatrix}, \tag{3}$$

where $\mathcal{S}(G)_{n,m}$ and $\mathcal{S}(\mathcal{N}_a)_{n,m}$ indicate semantic similarity between w_n and w_m in the general text corpus G and in $\mathcal{N}_a$, respectively. When $\mathcal{S}(G)_{n,*}$ refers to the n-th row of $\mathcal{S}(G)$, we can compare the meaning of

w_n in G with in $\mathcal{N}_a$ by comparing $\mathcal{S}(G)_{n,*}$ with $\mathcal{S}(\mathcal{N}_a)_{n,*}$. A semantic distance between w_n in G and w_n in $\mathcal{N}_a$ can be formulated as:

$$\mathcal{D}(w_n, G, \mathcal{N}_a) = \frac{1}{2} \times |\mathcal{S}(G)_{n,*} - \mathcal{S}(\mathcal{N}_a)_{n,*}|. \tag{4}$$

To discover keywords of $\mathcal{G}_i$, we search words that have low semantic distances among subculture works included in $\mathcal{G}_i$ and high semantic distances between members of $\mathcal{G}_i$ and the other subculture works. This approach is similar to the feature selection for clustering. When $\mathcal{W}$ is a feature set, we have to find words that maximize inner-compactness of $\mathcal{G}_i$ and minimize outer-adjacency of $\mathcal{G}_i$. To reduce search spaces, we focus on (i) words that have not appeared in G (neologisms), (ii) nouns/noun phrases excluding pronouns, and (iii) emotional words/phrases (discovered by using a dictionary (https://github.com/park1200656/KnuSentiLex)). Also, the keywords of $\mathcal{G}_i$ should frequently occur in most members of $\mathcal{G}_i$, not in the other genres. An objective function for searching keywords consists of two parts. First, outer-adjacency of $\mathcal{G}_i$ for w_n can be measured as:

$$\begin{aligned}\mathcal{L}_A(w_n, \mathcal{G}_i) = & - \frac{1}{|\mathcal{G}_i|} \sum_{\mathcal{N}_a \in \mathcal{G}_i} \log \mathcal{D}(w_n, G, \mathcal{N}_a) - \frac{1}{|\mathcal{G}_i| \times |\mathcal{G}_i^C|} \sum_{\substack{\mathcal{N}_a \in \mathcal{G}_i, \\ \mathcal{N}_b \notin \mathcal{G}_i}} \log \mathcal{D}(w_n, \mathcal{N}_b, \mathcal{N}_a) \\ & + \frac{1}{|\mathcal{G}_i^C|} \sum_{\mathcal{N}_a \notin \mathcal{G}_i} \log \frac{f(w_n, \mathcal{N}_a)}{\max_{\forall \mathcal{N}_b} f(w_n, \mathcal{N}_b)}.\end{aligned} \tag{5}$$

The first and second terms assess semantic differences of w_n between general texts and $\mathcal{G}_i$ and between the other genres and $\mathcal{G}_i$, respectively. The last term examines whether w_n is a usual term regardless of genres. Also, the frequency can filter insignificant named entities (e.g., names of characters and places). Then, inner-compactness of $\mathcal{G}_i$ for w_n can be measured as:

$$\mathcal{L}_C(w_n, \mathcal{G}_i) = \frac{2}{|\mathcal{G}_i| \times (|\mathcal{G}_i| - 1)} \sum_{\mathcal{N}_a, \mathcal{N}_b \in \mathcal{G}_i} \log \mathcal{D}(w_n, \mathcal{N}_b, \mathcal{N}_a) - \frac{1}{|\mathcal{G}_i|} \sum_{\mathcal{N}_a \in \mathcal{G}_i} \log \frac{f(w_n, \mathcal{N}_a)}{\max_{\forall \mathcal{N}_b} f(w_n, \mathcal{N}_b)}. \tag{6}$$

The first term has a low value when w_n indicates the same concept in all subculture works of $\mathcal{G}_i$. The other term examines whether w_n has high occurrence frequencies on allover $\mathcal{G}_i$. Among words appeared in $\mathcal{G}_i$, we find w_n that minimizes $\mathcal{L}(w_n, \mathcal{G}_i) = \mathcal{L}_A(w_n, \mathcal{G}_i) + \mathcal{L}_C(w_n, \mathcal{G}_i)$.

When $\mathcal{K}(\mathcal{G}_i)$ is a keyword set of $\mathcal{G}_i$, defining a threshold is the simplest way for composing $\mathcal{K}(\mathcal{G}_i)$, such as $\mathcal{K}(\mathcal{G}_i) = \{w_n | \mathcal{L}(w_n, \mathcal{G}_i) \geq \theta\}$. However, genres evolve and branch according to time, and keywords of old genres will not show enough cohesion because of being inherited to their descendants. Therefore, we set individual thresholds for each genre by using distributions of $\mathcal{L}(w_n, \mathcal{G}_i)$; $\theta_i = \mu_i - \alpha_g \times \sigma_i$, when θ_i, μ_i, and σ_i are a threshold, arithmetic mean, and standard deviation of $\mathcal{L}(w_n, \mathcal{G}_i)$ for all the words, and α_g is a weighting factor. We have empirically searched α_g according to the accuracy of the genre classification in a range $[0.5, 5]$ with a step size $+0.5$. And, α_g has been set as 2.0.

3.1.2. Classifying Subculture Works into Genres

Based on the keywords of genres, we can classify unlabeled subculture works. Even if a subculture work has genre characteristics not annotated by its author (or editor), we can detect the omitted annotations. The same approaches to the keyword search are applied to the classification. Equations (5) and (6) assess whether the keywords have the unique and shared meanings in a genre. In the classification, if keywords of a genre have the same meanings in a subculture work, the subculture work will be a member of the genre. Also, the keywords should have enough frequencies in the subculture work.

Subculture works contain characteristics of multiple genres, and their membership degrees are not uniform for the contained genres. Thus, Equation (6) is modified to measure the appropriateness of subculture works for each genre, not of keywords. We compare the target subculture work with other works, based on explicit genre annotations given by their authors (or editors). The membership function can be formulated as:

$$\mu_{\mathcal{G}_i}(\mathcal{N}_a) = \frac{1}{|\mathcal{K}(\mathcal{G}_i)| \times |\mathcal{G}_i|} \sum_{w_n \in \mathcal{K}(\mathcal{G}_i)} \left[\sum_{\mathcal{N}_b \in \mathcal{G}_i} [1 - \mathcal{D}(w_n, \mathcal{N}_b, \mathcal{N}_a)] \frac{f(w_n, \mathcal{N}_a)}{f(w_n, \mathcal{N}_b)} \right]. \tag{7}$$

If $\mu_{\mathcal{G}_i}(\mathcal{N}_a)$ is bigger than a threshold $\in [0,1]$, we determine that $\mathcal{N}_a$ has a characteristic of $\mathcal{G}_i$. We assign independent thresholds for each genre by searching cut points that maximize the classification accuracy with a step size $+0.05$. The accuracy is assessed by F_2 score for emphasizing recall than precision. Subculture works contain characteristics of various genres. Authors might intend most of the genre characteristics, but some could be flowed in unintentionally while imitating other popular pieces. Thus, we focus on what we should find rather than what we have already known.

3.1.3. Building Genealogy Trees of Genres

Shared keywords between genres indicate that the genres share subjects. These genres might be ancestors and descendants, but we do not know exact relationships between them. We find the relationships of genres using the following heuristics;

- Semantic distance: Let suppose that $\mathcal{G}_i$, $\mathcal{G}_j$, and $\mathcal{G}_k$ share keywords. If the keywords have more similar meanings in $\mathcal{G}_i$ and $\mathcal{G}_j$ than in $\mathcal{G}_j$ and $\mathcal{G}_k$, the shortest path between $\mathcal{G}_i$ and $\mathcal{G}_j$ should have fewer hops than the shortest path between $\mathcal{G}_j$ and $\mathcal{G}_k$.
- Differentiation of genres: If $\mathcal{G}_i$ is an ancestor of $\mathcal{G}_j$, $\mathcal{G}_j$ will have more other keywords excluding inherited ones from $\mathcal{G}_i$.
- Life cycles of genres: The creation of subculture works is focused on trendy and popular genres. The temporal distribution of published works in each genre reveals when the genres are spotlighted and faded away. If $\mathcal{G}_i$ was popular sooner than the other ones, $\mathcal{G}_i$ would be their ancestors.

To measure the semantic distance between genres, we recompose vector representations of words. In Section 3.1.1, the word embedding was conducted for each subculture work. We compose word vectors for each genre by considering membership degrees of subculture works for genres (Equation (7)), since subculture works are mixtures of multiple genres. We aggregate word vectors for subculture works using a weighted average, where weighting factors are the membership degrees. This can be formulated as:

$$\overrightarrow{w_{n,\mathcal{G}_i}} = \frac{1}{|\mathcal{G}_i|} \times \sum_{\mathcal{N}_a \in \mathcal{G}_i} \overrightarrow{w_{n,a}} \times \mu_{\mathcal{G}_i}(\mathcal{N}_a). \tag{8}$$

By comparing $\overrightarrow{w_{n,\mathcal{G}_i}}$ with $\overrightarrow{w_{n,\mathcal{G}_j}}$, we can examine semantic distance between w_n in $\mathcal{G}_i$ and w_n in $\mathcal{G}_j$. The semantic distance is measured by the same method with the keyword extraction (Equations (3) and (4)).

To compare $\mathcal{G}_i$ with $\mathcal{G}_j$, common keywords between them $(\mathcal{K}(\mathcal{G}_i) \cap \mathcal{K}(\mathcal{G}_j))$ will be the most significant, and the other keywords $((\mathcal{K}(\mathcal{G}_i) \cup \mathcal{K}(\mathcal{G}_j)) - (\mathcal{K}(\mathcal{G}_i) \cap \mathcal{K}(\mathcal{G}_j)))$ and the remaining words $(\mathcal{W} - (\mathcal{K}(\mathcal{G}_i) \cup$

$\mathcal{K}(\mathcal{G}_j)$)) might be the second and last, respectively. Therefore, we apply individual weighing factors to the three groups. Semantic distance between genres can be measured as:

$$\mathcal{D}(\mathcal{G}_j,\mathcal{G}_i)=\sum_{w_n\in\mathcal{W}}\mathcal{D}(w_n,\mathcal{G}_j,\mathcal{G}_i)\times W(w_n,\mathcal{G}_j,\mathcal{G}_i),\;\; W(w_n,\mathcal{G}_j,\mathcal{G}_i)=\begin{cases}\alpha_c & \text{if } w_n\in\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_j),\\ \alpha_k & \text{else if } w_n\in\mathcal{K}(\mathcal{G}_i)\cup\mathcal{K}(\mathcal{G}_j),\\ 1 & \text{otherwise.}\end{cases} \tag{9}$$

where α_c and α_k are the weighting factors for the shared and remaining keywords, respectively. We have empirically searched α_c and α_k according to the accuracy of building the genre genealogy tree in a range $1 \le \alpha_k \le \alpha_c \le 3$ with a step size +0.5. The weighting factors have been set as $\alpha_c = 2.0$ and $\alpha_k = 1.5$.

To draw the genealogy tree, we first compose a network, which has genres as nodes and their semantic distance as edges. The tree is revealed by deleting unnecessary edges and assigning levels on the nodes. The temporal distributions of genres determine directions of the edges. We compose the Gaussian distribution of monthly publication frequencies of each genre: $\mathcal{T}(\mathcal{G}_i) = \langle\mu(\mathcal{G}_i),\sigma(\mathcal{G}_i)\rangle$. Then, if $\mu(\mathcal{G}_j) \ge \mu(\mathcal{G}_i)$, head of an edge between $\mathcal{G}_i$ and $\mathcal{G}_j$ is $\mathcal{G}_i$. Also, when $\mu(\mathcal{G}_j)$ is in a range $\mu(\mathcal{G}_i) \pm \sigma(\mathcal{G}_i)$, and $\mathcal{G}_i$ satisfies the same condition for $\mathcal{G}_j$, we suppose that $\mathcal{G}_i$ and $\mathcal{G}_j$ are in the same generation ($l_i = l_j$ in Equation (2)). Genres in the oldest generation are assigned on level 1, and the others are labeled according to the descending order of their generation.

The genre genealogy tree allows multiple parents and children. Thus, merely deleting edges except the smallest semantic distance is not an adequate approach. We focus on that genealogy trees cannot have cyclic paths. We first reduce edge density in the network by deleting edges with bigger semantic distances than median distance ($\mathcal{D}(\mathcal{G}_j,\mathcal{G}_i) \ge \text{median}_{\forall\mathcal{G}_k}\,\mathcal{D}(\mathcal{G}_k,\mathcal{G}_i)$). Subsequently, we search cyclic paths and delete edges with the largest semantic distance in the paths.

The subculture genres come from a few origins (e.g., myths, J. R. R. Tolkien, Jules Gabriel Verne, Lewis Carroll, Jin Yong, and so on) and are born under influences of multiple fore-parents. Therefore, the genealogy tree should allow multiples paths between two nodes. However, at the same time, complicated connections between neighboring nodes are not desirable. We solve this problem by using intersections between keyword sets and reserving longer paths as possible. First, we delete one-hop connections among genres on the same level. If two similar genres were popular in the same period, they might be branches of the same parents, rather than one gives birth to another. Then, we examine parents of each genre and common ancestors of the parents. When $\mathcal{G}_i$ has two parents $\mathcal{G}_j$ and $\mathcal{G}_k$, procedures for deleting meaningless connections are as follows;

1. Check whether $\mathcal{G}_j$ and $\mathcal{G}_k$ have common ancestors.
2. If $\mathcal{G}_j$ and $\mathcal{G}_k$ share common ancestors, search the closest common ancestor.
3. When $\mathcal{G}_l$ is the closest common ancestor, examine inheritance of keywords by comparing $\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_j)$, $\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_k)$, and $\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_l)$ with each other.
   (a) If $|\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_j)-\mathcal{K}(\mathcal{G}_k)| \le |\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_j)\cap\mathcal{K}(\mathcal{G}_k)|$ and $|\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_k)-\mathcal{K}(\mathcal{G}_j)| \ge |\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_j)\cap\mathcal{K}(\mathcal{G}_k)|$, delete the connection between $\mathcal{G}_i$ and $\mathcal{G}_j$, and vice versa.
   (b) If $|\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_j)-\mathcal{K}(\mathcal{G}_k)| \ge |\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_j)\cap\mathcal{K}(\mathcal{G}_k)|$ and $|\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_k)-\mathcal{K}(\mathcal{G}_j)| \ge |\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_j)\cap\mathcal{K}(\mathcal{G}_k)|$, accept both of them as parents of $\mathcal{G}_i$.
   (c) Otherwise, if $|\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_j)-\mathcal{K}(\mathcal{G}_l)| \le |\mathcal{K}(\mathcal{G}_i)\cap\mathcal{K}(\mathcal{G}_j)\cap\mathcal{K}(\mathcal{G}_l)|$, delete $\mathcal{G}_j$ from the path between $\mathcal{G}_i$ and $\mathcal{G}_l$ (same for $\mathcal{G}_k$).

These procedures check whether parent genres have remained a distinctive legacy to their children. We iteratively conduct the procedures, until there are no more edges to delete.

3.2. Correlations of Genres with Social Issues

To discover subculture genres and their genealogy, we have concentrated on words used in different meanings from the general texts, since subjects and backgrounds of subculture works are imaginary. However, targets of social desires or grievances will be the ones that exist in the real world. Authors, readers, and characters in the subculture works will show intense emotions for those targets, whether the emotions are positive or negative. For example, the game and professional gamer also exist in the real world, and they mostly have similar meanings in both game fiction (게임물) and the general text corpus. The game fiction supposes an imaginary world that video games become significant parts of the real world. Thus, their readers and characters desire successes in the games and based on the games. We extract the keywords of social issues by using the following heuristics;

- Semantic distance: Reflection of the social issues is more straightforward in subculture multimedia than in conventional narrative multimedia. Most of the subculture works describe social issues without complicated symbols, figurative expressions, or mise-en-scène. Therefore, keywords related to social issues have similar meanings in both of the subculture works and general texts.
- Emotional intensity: Vicarious satisfaction for social desires and conflicts is one of the primary purposes of subculture multimedia consumption, and it will accompany intense emotions. Thus, we search words that frequently co-occurred with emotional words.

Word2Vec [24] is based on the assumption that frequently co-occurred words are semantically correlated. In other words, the co-occurred words have close locations in the embedding space. Thus, if readers and characters frequently show emotional reactions for a word w_n, the emotional words will be located around $\overrightarrow{w_{n,\mathcal{G}_i}}$. Also, words related to the same social issue will have similar vector representations.

To find words that satisfy the above requirements, we first cluster words according to their vector representations (Equation (8)). We do not know the meanings or significance of components in word vectors, and narrative works contain various types of terms (e.g., proper nouns, neologisms, emotional words, and so on). Therefore, this task requires a clustering algorithm that is robust to noises and can find arbitrarily shaped clusters. Thus, the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm is used. Before clustering the words, we add a component to the word vectors to consider the semantic distance between the words in the general texts and the words in the subculture genre. Addition of $\mathcal{D}(w_n, G, \mathcal{G}_i)$ as the new component gathers semantically consistent words into the same cluster.

The clusters correspond to key subjects of the genre. Most of the subjects are imaginary or figurative, but some of them straightforwardly imitate the real world. Also, if the imitation is correlated to social desires or conflicts, readers and authors will show longing or anger. Therefore, each cluster is assessed by the average semantic distance of words in the cluster. Also, we conduct the sentiment analysis for sentences that include the words. Keywords of social issues will occur frequently and accompany intense sentiment. An objective function for ranking clusters can be formulated as:

$$\mathcal{L}_K(w_n) = \left[\frac{1}{\mathcal{D}(w_n, G, \mathcal{G}_i) + 1} + \frac{1}{|S(w_n, \mathcal{G}_i)|} \sum_{s_l \in S(w_n, \mathcal{G}_i)} \mathbb{S}(s_l)\right] \times \frac{f(w_n, \mathcal{G}_i)}{|\mathcal{G}_i|}, \tag{10}$$

where $S(w_n, \mathcal{G}_i)$ is a set of sentences that are in $\mathcal{G}_i$ and include w_n, s_l indicates the l-th sentence, $\mathbb{S}(s_l)$ refers to the degree of sentiment in s_l, and $f(w_n, \mathcal{G}_i)$ denotes the frequency of w_n in $\mathcal{G}_i$. $\mathcal{L}_K(\cdot)$ assesses a word in a cluster, and the cluster is evaluated by the average of $\mathcal{L}_K(\cdot)$ for all the words in the cluster. Using the objective function, we first obtain clusters that correspond to social issues and choose words in the clusters that can be labels of the social issues. For the sentiment analysis, we use a dictionary of emotional words and phrases (also used in Section 3.1.1). In the dictionary, sentiment degrees were annotated with five steps: -2, -1, 0, 1, and 2. We normalized them into $[0, 1]$.

4. A Case Study on Korean Web Novels

To validate the effectiveness of the proposed methods, we have examined the performance of the proposed methods for analyzing Korean web novels. For the research questions on Section 1, we first evaluate whether words with the semantic difference correspond to subjects of subculture works (RQ 1) using the accuracy of the genre keyword extraction. Also, the accuracy of the genre classification shows whether the subjects are criteria of genre boundaries. Subsequently, the genre genealogy tree's accuracy supports that we can trace genre differentiation using subject inheritance (RQ 2). Finally, correlations of words without the semantic difference with social issues (RQ 3) are validated by the accuracy of the social issue detection.

4.1. Data Collection and Pre-Processing

We collected 200 Korean web novels published from 2016 to 2020 and distributed through a subculture platform, Munpia (https://www.munpia.com/). There are four major web novel platforms in Korea, NAVER SERIES, Kakao Page, Munpia, and Joara. However, we could not crawl web novels from NAVER SERIES and Kakao Page. Also, a significant number of web novels distributed through Joara were adult content. Therefore, we chose Munpia as our data source. Moreover, as shown in Sosul Network (https://sosul.network/), a website for crowd-sourcing reviews for web novels, popular web novels on Munpia are also distributed through NAVER SERIES and Kakao Page. Thus, we assume that web novels collected from Munpia can represent Korean web novels.

We used only the first 20 episodes in each web novel since the main subjects of web novels have already appeared while describing their protagonists' motivations. Also, despite an enormous number of web novels distributed through Munpia (39,005 on October 22nd, 2020), we conducted experiments with only 200 web novels due to the following four reasons. First, the proposed methods build the genre genealogy tree by analyzing life cycles of subculture genres. Therefore, our web novel corpus should have genre diversity and be evenly distributed over the years. Second, as we discussed in Section 1, not all web novels are authored by full-time writers. Many web novels in Munpia are published by enthusiastic consumers, and some of these web novels contain only a few episodes. The proposed methods use introductions written by authors (or editors) and comments for the episodes with web novels' main texts. Thus, web novels in our corpus should have a meaningful number of comments (5.31 comments per episode). Finally, ground-truth data for our experiment was composed by human evaluators due to the absence of benchmark datasets, and we could not conduct the experiment with an excessive number of web novels. Therefore, among the available web novels, we chose web novels with three criteria, the number of episodes, genres, publication years, and the number of comments. As a general text corpus, we used a Korean Wikipedia corpus.

The web novel texts were segmented into sentences, and we conducted POS (Part-Of-Speech) tagging, stemming, and removing stop words. We restricted the search space into neologisms, nouns/noun phrases, and emotional words/phrases for all the proposed methods. However, we conducted representation learning of all the words, since adjacency between postpositional particles and other words affects the nouns' meanings in Korean. Word2Vec has various hyper-parameters, and we empirically tuned the parameters to distinguish words with the semantic difference from words without the difference well. We made $\mathcal{D}(w_n, G, \mathcal{N}_a)$ as close to 1 or 0 as possible. Thus, we found parameters that minimize $-\sum_{\forall w_n} \sum_{\forall \mathcal{N}_a} \log 2|\mathcal{D}(w_n, G, \mathcal{N}_a) - 1/2|$ using a grid search on: the number of epochs ϵ (40 to 200 with a step size +20), the learning rate ρ (0.00025 to 0.25 with a step size ×10), the number of dimensions δ (32 to

256 with a step size ×2), the number of negative samples k (5 to 15 with a step size $+2$), and the weighting factor for the noise distribution α (0.00 to 1.00 with a step size $+0.25$). We determined the hyper-parameters as: $\epsilon = 140$, $\rho = 0.0025$, $\delta = 128$, $k = 7$, and $\alpha = 0.00$.

To evaluate the proposed methods, we collected ground truth datasets by conducting a questionnaire survey. First, we composed a group of human evaluators that consists of 37 Korean web novel consumers. User communities give names to subculture genres, and there have not been official taxonomies and names of the genres. Therefore, the evaluators first annotated genre labels on web novels in our corpus that they have read. We used genre labels only that are annotated by the majority of the evaluators. We obtained 635 annotations for 147 web novels (on average 4.32 labels on each novel) with 36 genres; 53 novels had not read by any evaluator. Table 1 presents a list of the genres. Other survey questions related to experimental procedures will be described in each section.

Table 1. List of 36 genres of Korean web novels and their taxonomy acquired from one of the major Korean wiki websites (https://namu.wiki/). Genre diversity of Korean web novels is getting increased over time.

Generation		Genres		
1st Generation (1990s)		전통판타지물 (Traditional Fantasy Fiction), 도시판타지물 (Urban Fantasy Fiction), 신무협물 (Neo-Martial Arts Fiction), SF물 (Science Fiction), 밀리터리물 (Military Fiction), 대체역사물 (Alternative History Fiction), 로맨스물 (Romance Fiction)		
2nd–3rd Generation (2000s)		차원이동물 (Dimension-Shifting Fiction), 환생물 (Rebirth Fiction), 영지물 (Feud Fiction), 기갑물 (Mecha Fiction), 드래곤물 (Dragon Fiction), 게임물 (Game Fiction), 판협물 (Fantasy-Martial Arts Fiction), 라이트노벨 (Light Novel)		
4th Generation (2010s)	Fantasy World based on Different World	이세계물 (Different World Fiction), 던전운영물 (Dungeon Management Fiction), 차원유랑물 (Dimension-Traveling Fiction), 스페이스오페라물 (Space Opera Fiction)	탑등반물 (Tower Climbing Fiction)	학원물 (Academy Fiction), 빙의물 (Possession Fiction), 정치물 (Politics Fiction), 귀환물 (Homecoming Fiction), 회귀물 (Return Fiction), 육아물 (Rearing Fiction)
	Fantasy World based on Earth	헌터물 (Hunter Fiction), 레이드물 (Raid Fiction), 성좌물 (Constellation Fiction), 아포칼립스물 (Apocalypse Fiction)		
	Close to Real World	전문가물 (Expert Fiction), 기업물 (Business Fiction), 스포츠물 (Sports Fiction), 스트리머물 (Streamer Fiction), 연예계물 (Entertainment Fiction)		

4.2. Accuracy of Extracting Genre Keywords

We assess the accuracy of the proposed method for extracting genre keywords to evaluate the proposed method's usefulness and validate the effectiveness of word semantic difference for discovering imaginary subjects from literary texts (RQ 1). The accuracy was assessed by comparing the results of the proposed method with user survey results. The evaluators corrected the keyword sets by adding missing keywords and deleting incorrect ones. We took changes only agreed by the majority of the evaluators. To measure the accuracy, we employed the precision, recall, and F_1 score. When $\mathcal{K}^*(\mathcal{G}_i)$ is the edited keyword set of $\mathcal{G}_i$, the accuracy metrics are calculated as:

$$p^{\mathcal{K}}(\mathcal{G}_i) = \frac{|\mathcal{K}(\mathcal{G}_i) \cap \mathcal{K}^*(\mathcal{G}_i)|}{|\mathcal{K}(\mathcal{G}_i)|}, \;\; r^{\mathcal{K}}(\mathcal{G}_i) = \frac{|\mathcal{K}(\mathcal{G}_i) \cap \mathcal{K}^*(\mathcal{G}_i)|}{|\mathcal{K}^*(\mathcal{G}_i)|}, \;\; F_1^{\mathcal{K}}(\mathcal{G}_i) = 2 \times \frac{p^{\mathcal{K}}(\mathcal{G}_i) \times r^{\mathcal{K}}(\mathcal{G}_i)}{p^{\mathcal{K}}(\mathcal{G}_i) + r^{\mathcal{K}}(\mathcal{G}_i)}, \tag{11}$$

where $p^{\mathcal{K}}(\mathcal{G}_i)$, $r^{\mathcal{K}}(\mathcal{G}_i)$, and $F_1^{\mathcal{K}}(\mathcal{G}_i)$ indicate the precision, recall, and F_1 score for $\mathcal{G}_i$, respectively.

Word2Vec [24] and a three-layered fully-connected neural network were used as a baseline method. Word vectors for each genre have already composed in Section 3.1.3. We classified words into keywords or non-keywords using the neural network; input layer: 128 nodes, first hidden layer: 128 nodes, second hidden layer: 64 nodes, third hidden layer: 64 nodes, and output layer: 1 node. Activation functions of all the layers were the sigmoid. 80% and 20% of the words were used as training and testing data with 5-fold cross-validation. The accuracy was assessed for each genre and aggregated using the arithmetic mean.

The second column of Table 2 shows the experimental results of the keyword extraction. Due to the lack of spaces, we present results for ten genres that are ancestors of the possession fiction, which is the most popular recently. Although both precision and recall of the proposed method were reasonable (Table 3), a few genres had mostly the same keyword sets (e.g., raid fiction and hunter fiction). Also, we obtained too small keyword sets from several genres (e.g., apocalypse fiction). These results show that some subjects are difficult to be stated as a few terms. Main subjects of a few genres were close to relationships among characters (e.g., 로맨틱판타지물 (romantic fantasy fiction) and 육아물 (rearing fiction)). Significance of the subjects was also not considered by the proposed method. The raid fiction and hunter fiction had similar keywords, but a few terms (e.g., 보스 (boss monster) and 패턴 (pattern)) had distinctively higher frequencies in the raid fiction. In further research, we have to extend the concept of subjects over mere terms and find other genre characteristics more than imaginary subjects. The second column of Table 3 presents the accuracy of the proposed method.

Table 2. Results of the keyword extraction and social issue discovery. Due to the limited space, this table presents only relatives of the 'Possession fiction.' Korean texts are parts of discovered keywords for imaginary and real world-originated subjects. Texts in round brackets are English translations of the keywords. Keywords presented as the results of social issue discovery are labels of word clusters. Also, bold texts are true positives, and the others are false positives.

Genre	**Keywords**	**Social Issues**
SF물 (Science Fiction)	**워프 (warp)**, 게이트 (gate), **초광속 (faster-than-light)**, 드라이브 (drive), 빔 (beam), **인공지능 (artificial intelligence)**, 괴수 (monster), 사이킥 (psychic), 전함 (battle ship), 성계 (stellar system)	개척 (pioneer), **대기업 (mega-corporation)**, 콜로니 (colony), 성계 (stellar system), 정부 (government), 전쟁 (war), **테러 (terror)**, 연구소 (laboratory), **인공지능 (artificial intelligence)**, 자동 (automatic)
고전판타지물 (Classical Fantasy Fiction)	**오러 (aura)**, **소드마스터 (sword-master)**, **서클 (circle)**, **오크 (orc)**, **몬스터 (monster)**, **마법 (magic)**, **마나 (mana)**, **기사 (knight)**, **사제 (priest)**, **용병 (mercenary)**	성 (castle), 여관 (inn), 전쟁 (war), 노예 (slave), 로브 (robe), 귀족 (noble), 충성 (royalty), 왕 (king), 기사 (knight), **귀족파 (aristocratic faction)**
신무협물 (Neo-Martial Arts Fiction)	**무공 (martial arts)**, **내공 (aura)**, **초식 (technique)**, **오대세가 (top-5 powerful families)**, **십대고수 (top-10 masters of martial arts)**, **고수 (master)**, **마교 (demonic cult)**, **후기지수 (next generation)**, **구파일방 (nine clans and one sect)**, **비무 (sparring)**	**위선 (hypocrisy)**, **명문 (noble family)**, 은원 (favor and spite), 스승 (teacher), 수련 (training), **자질 (talent)**, 암운 (ominous clouds), **기연 (strange chance)**, 천하 (world), 기루 (brothel)
아포칼립스물 (Apocalypse Fiction)	**종말 (apocalypse)**, **부수다 (destroy)**, **마지막 (last)**, **좀비 (zombie)**, 괴수 (monster), 배신 (betrayal), **생존 (survival)**, 침략 (invasion), **뒷통수 (backstabbing)**, 외계인 (alien)	**생존 (survival)**, 편의점 (convenient store), 구역 (district), **혼자 (alone)**, 사회 (society), **그룹 (group)**, 경찰 (police), **부장 (head of department)**, 살인 (murder), **알바 (part-time job)**
차원이동물 (Dimension-Shifting Fiction)	**차원이동 (dimension-shifting)**, **집 (home)**, 가족 (family), 마나 (mana), 오러 (aura), 내공 (aura), 마을 (village), 마법 (magic), 드래곤 (dragon), 신 (god)	**백수 (free-timer)**, 학교 (school), 부모님 (parents), 귀족 (noble), 성 (castle), 차원 (dimension), 돌아가다 (return), 수련 (training), 치킨 (fried chicken), 왕 (king)
게임물 (Game Fiction)	**레벨 (level)**, **스킬 (skill)**, **퀘스트 (quest)**, **아이템 (item)**, **클래스 (class)**, **히든 (hidden)**, 사냥 (hunting), **던전 (dungeon)**, 몬스터 (monster), **랭커 (ranker)**	**가상현실 (virtual reality)**, 운영 (managing), 대회 (contest), 경매 (auction), **인공지능 (artificial intelligence)**, **재능 (talent)**, **랭커 (ranker)**, **십대길드 (top-10 guilds)**, 운빨 (luck), 백수 (free-timer)

Table 2. *Cont.*

Genre	Keywords	Social Issues
회귀물 (Return Fiction)	**회귀 (return)**, 반복 (repetition), 배신 (betrayal), 회차 (–th round), 몬스터 (monster), **후회 (regret)**, **복수 (revenge)**, **과거 (past)**, 실수 (mistake), 신 (god)	복수 (revenge), 가족 (family), **재능 (talent)**, **배신 (betrayal)**, **기회 (opportunity)**, 기억 (memory), 계획 (plan), **경험 (experience)**, **버리다 (abandon)**, 날짜 (date)
레이드물 (Raid Fiction)	상태창 (status), **레이드 (raid)**, 몬스터 (monster), 스킬 (skill), **각성자 (psychic)**, **게이트 (gate)**, **등급 (rank)**, 길드 (guild), **패턴 (pattern)**, **공략 (tactics)**	**재능 (talent)**, 시스템 (system), 스카우터 (headhunter), **정부 (government)**, **대우 (treatment)**, **협회 (association)**, **재벌 (chaebol)**, **기자 (journalist)**, **통제 (control)**, **차별 (discrimination)**
헌터물 (Hunter Fiction)	상태창 (status), **훈련소 (training center)**, 던전 (dungeon), 스킬 (skill), **각성자 (psychic)**, **게이트 (gate)**, **등급 (rank)**, 길드 (guild), **브레이크 (burst)**, 몬스터 (monster)	**루키 (rookie)**, **정부 (government)**, 테스트 (test), **재벌 (chaebol)**, 시스템 (system), 허가 (permission), 구역 (district), **협회 (association)**, **재능 (talent)**, 계약 (contract)
빙의물 (Possession Fiction)	**설정 (features of imaginary worlds)**, 게이트 (gate), 상태창 (status), 특성 (nature), 스킬 (skill), **작가 (author)**, **원작 (original work)**, 사이다 (cider), 아카데미 (academy), **빙의 (possession)**	**엑스트라 (extra)**, 주인공 (protagonist), 작가 (author), 전개 (story development), 기억 (memory), **계획 (plan)**, **사이다 (cider)**, **예지 (foresight)**, **테러 (terrorism)**, **살아남다 (survive)**

Table 3. Accuracy of the proposed and baseline methods for the genre keyword extraction, genre classification, genre genealogy tree, and social issue detection. The genealogy tree was assessed by the edit distance, and the three remaining methods were evaluated by precision (p), recall (r), and F_1 measure (F_1). Numbers in round brackets present the standard deviations.

	Keyword Extraction		Genre Classification		Genealogy Tree	Social Issue Detection	
Proposed	p	0.80 (0.07)	p	0.75 (0.15)	10 (5.51)	p	0.65 (0.22)
	r	0.78 (0.09)	r	0.85 (0.11)		r	0.72 (0.16)
	F_1	0.79 (0.08)	F_1	0.80 (0.12)		F_1	0.68 (0.18)
Baseline	p	0.71 (0.15)	p	0.72 (0.11)	21 (5.07)	p	0.60 (0.14)
	r	0.75 (0.12)	r	0.78 (0.12)		r	0.70 (0.11)
	F_1	0.73 (0.14)	F_1	0.75 (0.11)		F_1	0.65 (0.13)

The fore-mentioned problem did not affect the accuracy of keyword extraction. Nevertheless, we found another problem: Wikipedia cannot cover colloquial languages. For example, a Korean word, 사이다 (cider), is also used in the meaning of 'inexorable' or 'feisty,' not only in a kind of beverage. This word frequently occurred in several genres and was used in the same meaning. However, 사이다 (cider) was selected as keywords, since our general text corpus (Wikipedia) is written in literary style. According to Robert McKee [61], among movie scripts, play scripts, and novels, play scripts are the closest to the literary languages, movie scripts are the most similar to the colloquial languages, and novels are in the middle. As a kind of snack culture, web novels look closer to everyday languages than ordinary novels. In further research, the general text corpus should cover the colloquial languages, and we have to add methods for handling polysemy and homonymy issues to the proposed methods.

The proposed method outperformed the baseline method in terms of both accuracy and variance. Performance improvement was more distinctive in precision than in recall. Word vectors used by both methods contain information for co-occurrences of words. Imaginary subjects occur more frequently than other words. Thus, terms for the subjects (keywords) could have distinctively different locations between embedding spaces for subculture works and general texts. By comparing the two embedding spaces, the proposed method can get the difference. Since the subjects also co-occur with each other frequently, the baseline could find locations where the subjects were densely located. However, it is difficult to distinguish ordinary words from the dense area without comparing the two embedding spaces, as shown in the low precision of the baseline method.

4.3. Accuracy of Classifying Subculture Works into Genres

This section evaluates the proposed multi-label classification method for subculture genres. Since the classification is based on the genre keywords, classification performance also supports RQ 1. We used only genre annotations collected from the web novel platform to compose the keyword sets and membership functions of genres. Then, we compared the automated labeling results with the annotations acquired form the evaluators. The accuracy was also measured by the precision, recall, and F_1 score. Methods for calculating the metrics are the same as the previous experiment.

As a comparison group, we used Doc2Vec [62] that is widely applied to classify documents. After conducting Doc2Vec for each web novel, we classified the web novels into genres according to their document vectors using a three-layered fully-connected neural network. This network had a similar structure to the one in the previous experiment; we extended its output layer to 36 nodes (as with the number of genres). Training and validation methods were also the same. The third column of Table 3 shows accuracy of the genre classification.

The proposed method outperformed the baseline method, whereas precision was not significantly improved, especially in terms of variance. As discussed in the previous section, keywords of some genres were not distinctive enough. The indistinctiveness did not hinder the keyword extraction much but affected the accuracy of the genre classification. This problem had two aspects. First, the proposed method showed reasonable recall but low precision for genres with a few keywords. There might be other features determining memberships for the genres. For example, apocalypse fiction has a few keywords, such as 종말 (apocalypse), 부수다 (destroy), 마지막 (last), and so on. The core keywords enabled us to discover novels included in the genre. However, for some web novels, including these keywords, the evaluators did not label them as apocalypse fiction. To find its reasons, we examined frequencies of emotional words and phrases in both groups (true positives and false positives). We found that the false positives had much less negative emotions than the true positives.

Second, when plural genres had duplicated keywords, and the keywords occurred in a novel frequently, the proposed method assigned the novel on all the genres. Thus, the duplicated keywords made these genres exhibit low precision and high recall. Although most of the keywords are duplicated, occurrence frequencies of keywords can be a clue for classifying these kinds of genres. For example, the raid fiction and hunter fiction shared most of their keyword sets. However, the frequency of some keywords, such as 공략 (tactics), 레이드 (raid), and 패턴 (pattern), was much higher in the raid fiction.

Our classification method was affected by this problem because we used independent classifiers, which consist of a membership function and a threshold for each genre. Also, the membership functions (Equation (7)) did not consider correlations between keywords. Although the baseline method did not consider keywords, its lower variance indicate that machine learning (ML)-based classifiers can be an efficient solution for considering the correlations among genres and between keywords. Our further research will be focused on integrating the proposed heuristics with ML techniques.

4.4. Accuracy of Building Genealogy Trees of Genres

This section evaluates the proposed method's performance for building genre genealogy trees and validates correlations of subject inheritance between genres with differentiation of the genres (RQ 2). To assess the generated genre genealogy tree, we asked the evaluators to edit the tree by adding or deleting nodes/edges. Among the editions, we accepted only editions made by the majority of the evaluators.

Then, the tree's accuracy was measured by the minimum number of editions required to transform the original into the edited version. As a baseline method, Doc2Vec and the hierarchical agglomerative clustering were employed. We composed document vectors for each genre (not for each web novel) and clustered them according to the cosine similarity between the document vectors. First, we discuss the results of the proposed method regarding the content of web novel genres. Since we could not find credible publications for Korean web novel genres, we referred one of the biggest Korean wiki websites (https://namu.wiki/w/%EC%9E%A5%EB%A5%B4%EB%AC%B8%ED%95%99). Then, we evaluate the accuracy of the proposed method by comparing it with the baseline method.

Figure 2 presents a part of the genre genealogy tree built by the proposed methods. The tree is focused on the possession fiction (빙의물), which is the most popular recently, and its ancestors. The possession fiction indicates works that fictional characters of novels or video games are suddenly possessed by protagonists (who lived in the real world). Common points of these works are (i) omniscient protagonists, (ii) shifting into imaginary worlds, (iii) fighting for survival, and (iv) mechanical equality. We traced where these points come from by using the genealogy tree.

The first characteristic is a heritage of the return fiction (회귀물). The return fiction describes protagonists going back to the past and fixing all the trials and errors. Experienced protagonists of this genre know future events and have been in the events. Although the protagonists have archenemies that made them return, they also have conflicts with characters who cannot recognize and acknowledge their ability and capability (mainly privileged groups, such as 귀족 (noble), 공무원 (public servant), and 재벌 (chaebol)). Major keywords of this genre were '회귀 (return),' '후회 (regret),' '복수 (revenge),' '과거 (past),' '실수 (mistake),' and so on.

The return fiction also comes from the dimension-shifting fiction (차원이동물). In the classical genre fiction, all the fictional characters are residents of the imaginary world. Thus, authors do not have restrictions for designing characters and events. At the same time, readers feel difficulties in sympathizing with the characters. This genre uses protagonists who lived in the real world and suddenly moved into the imaginary world. The protagonists solve problems in the imaginary world to go back home. Primary keywords of this genre were '차원이동 (dimension-shifting),' '집 (home),' '가족 (family),' and other keywords inherited from the classical fantasy/neo-martial arts fictions.

Game fiction (게임물) commonly uses video games as their backgrounds. In their narrative world, everything is operated according to game systems. Every player gets fair rewards for their achievements and efforts. The main rhetoric in this genre is that protagonists are treated unfairly because of social pressures, despite their ability or talent. Under the game systems, which guarantee mechanical equality, the protagonists achieve everything that they deserve. Therefore, keywords of this genre included '시스템 (system),' '실력 (ability),' '재능 (talent),' '운 (luck),' and other keywords related to the game systems (e.g., monsters, levels, and skills).

Raid fiction (레이드물) and hunter fiction (헌터물) focus on 'raids' and 'dungeons' in Video games. At the beginning of these genres, the raid fiction imitated the famous online game, 'World Of Warcraft.' A difference from the game fiction is that their backgrounds are based on 'gamified real world.' In these genres, gigantic monsters or dungeons suddenly appeared in the real world, and heroes with superpowers hunt the monsters to save the world. These superpowers and ways to grow the superpowers follow the game systems. The backgrounds, which are a reflection of the real world, make conflicts between the protagonists and the ancien regime. The protagonists overthrow social problems radically and aggressively. As culprits causing the problems, these genres mainly use 공무원 (public servant), 재벌 (chaebol), 국회의원 (a member of national assembly) and so on. Excluding keywords inherited from the game fiction, main keywords of these genres were '게이트 (gate),' '헌터 (hunter),' '초능력 (superpower),' and so on.

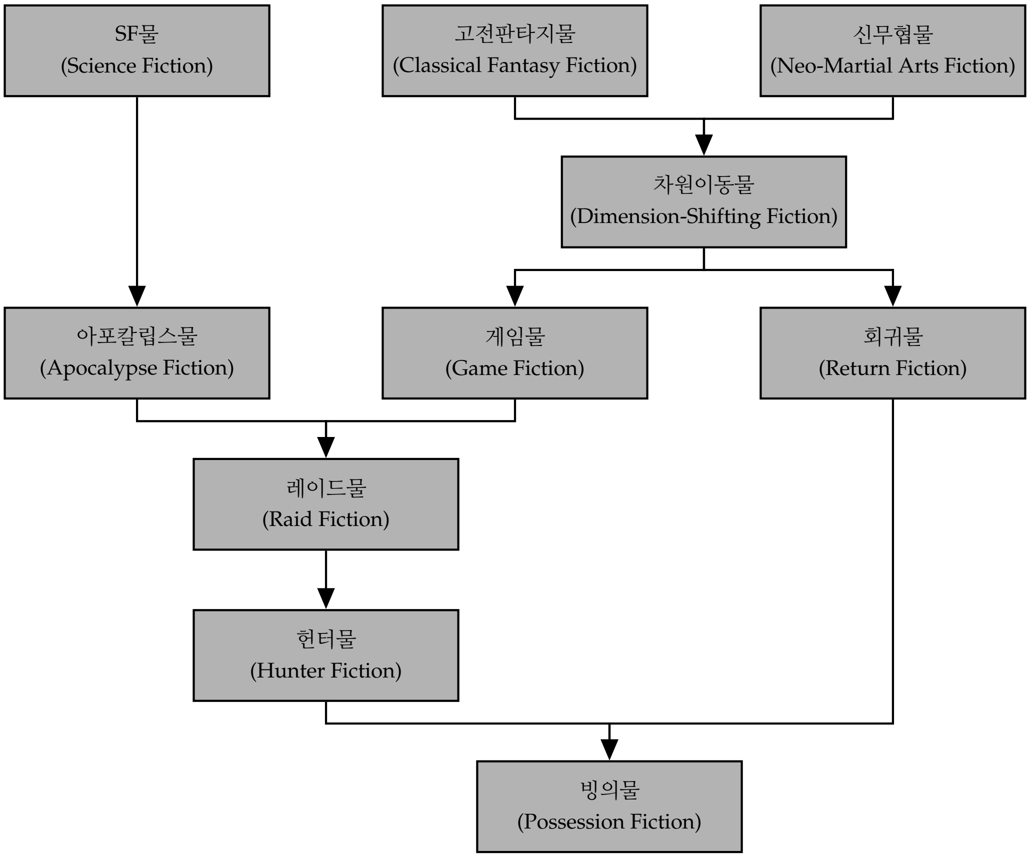


Figure 2. A part of the genre genealogy tree of Korean web novels that is constructed by the proposed method. This figure presents only the roots of the 'Possession fiction.' Each node indicates a genre. Heads and tails of edges indicate child and parent genres, respectively. Korean genre names (in the form of '–물') are labels collected from the evaluators. Texts in the round brackets are English translations labeled in this study.

Apocalypse fiction describes struggles for survival after or during the apocalypse. Although causes of the apocalypse are different in each novel (e.g., zombies, aliens, and gods), irresistible violence destroys human society and enforces the jungle's law. Therefore, keywords of this genre included '종말 (apocalypse),' '부수다 (destroy),' '배신 (betrayal),' '생존 (survival),' and so on.

The genre genealogy tree and explanations for the genres on the tree show inheritance of genre characteristics and how they are inherited. We can also see what kinds of desires are reflected by each genre (e.g., desires for recognition and fair opportunities). Excluding the running examples, there are various genres; e.g., 스트리머물 (streamer fiction), 연예계물 (entertainment fiction), and 대체역사물 (alternative history fiction). These genres are closely connected not only with the serious social problems (e.g., inequality, authoritarianism, and imperialism) but also with cultural shifts or trends (e.g., popular video games and the advent of video streaming services).

The fourth column of Table 3 shows the accuracy of the genre genealogy tree. Since the hierarchical clustering does not allow multiple parents, it showed significantly lower performance than the proposed method. This result is enough to show that genres are affected by multiple existing genres, and we need a method designed to trace the differentiation of subculture genres. We measured variances by measuring the edit distance for each evaluator's answer, and both methods exhibited high variances. Understandings for web novel content would be subjective. In further research, we should look for better representation than the genealogy tree or a more accurate method for building the tree, to find a genre differentiation model that can convince the majority. However, the results of the proposed method and the evaluators also had lots of differences. Since the genealogy tree of all the collected genres is too vast, we explain based on the possession fiction and its ancestors (Figure 2).

First, on the evaluators' answer, the apocalypse fiction had no connection. Since the proposed method does not consider a standalone genre, a few shared keywords made connections between the apocalypse fiction and science fiction (e.g., 외계인 (alien) and 괴수 (monster)) and between the apocalypse fiction and raid fiction (e.g., 뒷통수 (backstabbing)). The unnecessary connections also appeared between the game fiction and the streamer fiction and between the tower climbing fiction (탑등반물) and the hunter fiction. The proposed heuristics concentrate on only tracing the inheritance of genre keywords. We made connections between all genres that have shared keywords to increase the edge density of the tree. However, the experimental result showed that we have to consider degrees of influence between genres. Although Equation (9) measures semantic differences between genres, we used it only to filter edges roughly before applying the heuristics.

The evaluators also annotated the raid fiction and hunter fiction as siblings, and they had only one parent, the game fiction. Also, they answered that the two genres are on the same generation. The proposed method assigned generations of genres by using temporal distributions of the number of published subculture works in the genres with a range $\mu(\mathcal{G}_i) \pm \sigma(\mathcal{G}_i)$. We should extend the length of the range. Moreover, the proposed method attempts to increase the genealogy tree's depth. We found other genres met this problem (e.g., 탑등반물 (tower climbing fiction) and 이세계물 (Different world fiction)), and it indicates that the genealogy tree is not as deep as our expectation. In further research, we should find a way how we can recognize siblings that share most of the keywords.

4.5. Accuracy of Detecting Correlations of Genres to Social Issues

We evaluated the usefulness of the proposed social issue discovery method and validated the semantic consistency's effectiveness for recognizing real world-oriented subjects in subculture multimedia (RQ 3). We set top-10 word clusters according to Equation (10) as social issues detected by the proposed method. The evaluators corrected the results for each genre by adding or deleting clusters. The correction was conducted in a range of word clusters discovered in Section 3.2. Similar to the previous experiments, we aggregated corrections agreed by the majority of the evaluators and made ground truth. By comparing the original results with the corrected ones, we calculated precision, recall, and F_1 score for each genre and merge them using the arithmetic mean.

This experiment also employed a combination of word embedding models and simple neural networks as a baseline method. We composed vector representations of each word cluster by averaging vectors of words included in the cluster. A three-layered fully-connected neural network, which had the same structure with the network in Section 4.2, was trained to predict whether a word cluster corresponds to social issues. Training and validation methods were the same as the methods in Section 4.2. The fifth column of Table 3 presents the accuracy of the social issue detection.

Due to the limited space, Table 2 presents 10 of the 36 genres, which are the ancestors of the possession fiction. Tables 1 and 2 showed that the proposed methods had more difficulties in discovering social issues from old genres than from recent ones. For example, issues discovered from 고전판타지물 (classical fantasy fiction) and 신무협물 (neo-martial arts fiction) were far from problems in modern society. The evaluators also annotated much less social issues on the old genres than on the others. Considering the genre generations in Table 1, the first generation had 3.71 social issues on average, and the fourth generation got 12.50 annotations on average. However, even in the same generation, genres had a high variance in the number of annotations. Although 도시판타지물 (urban fantasy fiction) and 대체역사물

(alternative history fiction) were long-lived genres, they included various social issues as many as the latest genres. We could find 육아물 (rearing fiction) as an opposite case.

This problem made the proposed methods perform low average precision and high variance in the precision, as shown in Table 3. Also, since the proposed method fixed the number of social issues (top-10) in each genre, the baseline method performed much less variance than the proposed method, and the gap was more significant in the precision. Although the proposed method outperformed the baseline, there was a problem in our assumption that all the genres are correlated to particular social issues. The problem will be solved by using the proposed features as inputs of ML techniques.

The experimental result indicates that subculture genres have different sensitivity in reflecting social issues. Although they come from our deficiency, not all of them will be thinly veiled, and some of them will focus on telling comforts and hopes. Also, despite the high variance, subculture genres have become more connected to social issues over time. Subculture platforms have also moved from printed books via websites to mobile applications. This move might affect the changes in the social reflection of the genres. Thus, the trials for observing our society through subculture multimedia will become more significant.

5. Conclusions

This study aims at discovering social issues from subculture narrative multimedia. Consumption of narrative artworks exposes our inner desires directly without self-censorship. The proposed methods are based on the assumption that subjects of subculture genres are closely connected to their readers' vicarious satisfaction. Also, the imaginary subjects will have different meanings between the imaginary world and the real world, and otherwise, social issue-related subjects will be semantically consistent. Therefore, the proposed methods analyze each subculture genre using the semantic difference of imaginary subjects and discover related social issues by detecting semantic consistency of subjects.

We evaluated the proposed methods by applying to Korean web novels. Due to the absence of the existing methods, we compared the proposed methods with word/document embedding methods widely used for the keyword extraction or document classification. Although the proposed methods exhibited reasonable accuracy, they also had a few limitations. Our future studies will be focused on applying the proposed methods to other kinds of subculture multimedia and improving the proposed methods' limitations, as follows;

- Difference between general word usages and Wikipedia: The proposed methods labeled some words that should be in social issues as keywords. Although human evaluators thought that the miss-classified words were used in the same meanings on both subculture works and general texts, the words had a high semantic difference. This situation was more vivid in daily vocabularies, and Wikipedia could not cover colloquial languages. In future studies, we will extend the general text corpus by incorporating colloquial conversations (e.g., social media posts).
- Subject distinctiveness between subculture genres: In several cases, closely connected genres did not have distinctive subjects from each other as much as we expected. Thus, the inclusion of subjects was not effective for discovering boundaries between genres, and the indistinctness suffered the proposed methods, mainly the genre classification. The most significant feature of the apocalypse fiction was gloomy and hopeless ambiances. The raid fiction was distinguished from the hunter fiction by using relative frequencies of their subjects. In further research, we will search for other features of subculture genres that can be used with the subjects.

Author Contributions: Conceptualization, O.-J.L., E.-S.Y. and J.-T.K.; Data curation, H.H. and E.-S.Y.; Formal analysis, O.-J.L.; Funding acquisition, J.-T.K.; Investigation, O.-J.L.; Methodology, O.-J.L.; Project administration, E.-S.Y. and J.-T.K.; Resources, H.H. and E.-S.Y.; Supervision, E.-S.Y. and J.-T.K.; Validation, O.-J.L. and H.H.; Visualization, H.H.; Writing—original draft, O.-J.L. and H.H.; Writing—review & editing, O.-J.L., H.H., E.-S.Y. and J.-T.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICT Consilience Creative program (IITP-2019-2011-1-00783) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hashem, I.A.T.; Chang, V.; Anuar, N.B.; Adewole, K.; Yaqoob, I.; Gani, A.; Ahmed, E.; Chiroma, H. The role of big data in smart city. *Int. J. Inf. Manag.* **2016**, *36*, 748–758. [CrossRef]
2. Lau, B.P.L.; Marakkalage, S.H.; Zhou, Y.; Hassan, N.U.; Yuen, C.; Zhang, M.; Tan, U.X. A survey of data fusion in smart city applications. *Inf. Fusion* **2019**, *52*, 357–374. [CrossRef]
3. Yigitcanlar, T.; Kamruzzaman, M. Does smart city policy lead to sustainability of cities? *Land Use Policy* **2018**, *73*, 49–58. [CrossRef]
4. Angelidou, M.; Psaltoglou, A.; Komninos, N.; Kakderi, C.; Tsarchopoulos, P.; Panori, A. Enhancing sustainable urban development through smart city applications. *J. Sci. Technol. Policy Manag.* **2018**, *9*, 146–169. [CrossRef]
5. Costa, D.G.; Vasques, F.; Portugal, P.; Aguiar, A. A Distributed Multi-Tier Emergency Alerting System Exploiting Sensors-Based Event Detection to Support Smart City Applications. *Sensors* **2019**, *20*, 170. [CrossRef]
6. Suma, S.; Mehmood, R.; Albeshri, A. Automatic Event Detection in Smart Cities Using Big Data Analytics. In Proceedings of the 1st International Conference on Smart Cities, Infrastructure, Technologies and Applications (SCITA 2017), Jeddah, Saudi Arabia, 27–29 November 2017; Mehmood, R., Bhaduri, B., Katib, I., Chlamtac, I., Eds.; Springer: Jeddah, Saudi Arabia, 2018; Volume 224, pp. 111–122. [CrossRef]
7. Al-Turjman, F.; Malekloo, A. Smart parking in IoT-enabled cities: A survey. *Sustain. Cities Soc.* **2019**, *49*, 101608. [CrossRef]
8. Allam, Z.; Newman, P. Redefining the Smart City: Culture, Metabolism and Governance. *Smart Cities* **2018**, *1*, 4–25. [CrossRef]
9. Zhou, H.; Yin, H.; Zheng, H.; Li, Y. A survey on multi-modal social event detection. *Knowl.-Based Syst.* **2020**, *195*, 105695. [CrossRef]
10. Meel, P.; Vishwakarma, D.K. Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Syst. Appl.* **2020**, *153*, 112986. [CrossRef]
11. Das, S.; Kramer, A.D.I. Self-Censorship on Facebook. In Proceedings of the 7th International Conference on Weblogs and Social Media (ICWSM 2013), Cambridge, MA, USA, 8–11 July 2013.; Kiciman, E., Ellison, N.B., Hogan, B., Resnick, P., Soboroff, I., Eds.; The AAAI Press: Cambridge, MA, USA, 2013; pp. 120–127.
12. Sleeper, M.; Balebako, R.; Das, S.; McConahy, A.L.; Wiese, J.; Cranor, L.F. The post that wasn't: Exploring self-censorship on facebook. In Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW 2013), San Antonio, TX, USA, 23–27 February 2013; Bruckman, A.S., Counts, S., Lampe, C., Terveen, L.G., Eds.; ACM Press: San Antonio, TX, USA, 2013; pp. 793–802. [CrossRef]
13. Michel, J.B.; Shen, Y.K.; Aiden, A.P.; Veres, A.; Gray, M.K.; Pickett, J.P.; Hoiberg, D.; Clancy, D.; Norvig, P.; Orwant, J.; et al. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* **2010**, *331*, 176–182. [CrossRef] [PubMed]
14. Grayson, S.; Mulvany, M.; Wade, K.; Meaney, G.; Greene, D. Exploring the Role of Gender in 19th Century Fiction Through the Lens of Word Embeddings. In Proceedings of the 1st First International Conference on Language, Data, and Knowledge (LDK 2017), Galway, Ireland, 19–20 June 2017; Gracia, J., Bond, F., McCrae, J.P., Buitelaar, P., Chiarcos, C., Hellmann, S., Eds.; Springer: Galway, Ireland, 2017; Volume 10318, pp. 358–364. [CrossRef]
15. Chen, J.; Cui, M. Analysing Gender Bias in IMDB Films Based on Social Networks. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *806*, 012022. [CrossRef]
16. Baudrillard, J.; Baudrillard, J. *La Société de Consommation*; Gallimard Education: Paris, France, 1986; Volume 35. (In French)
17. Yecies, B.; Shim, A.; Yang, J.J.; Zhong, P.Y. Global transcreators and the extension of the Korean webtoon IP-engine. *Media Cult. Soc.* **2019**, *42*, 40–57. [CrossRef]

18. Shim, A.; Yecies, B.; Ren, X.T.; Wang, D. Cultural intermediation and the basis of trust among webtoon and webnovel communities. *Inf. Commun. Soc.* **2020**, *23*, 833–848. [CrossRef]
19. Yoesoef, M. Cyber Literature: Wattpad and Webnovel as Generation Z Reading in the Digital World. In Proceedings of the International University Symposium on Humanities and Arts (INUSHARTS 2019), Depok, Indonesia, 23–25 July 2019; Atlantis Press: Depok, Indonesia, 2019; Volume 453, pp. 128–131. [CrossRef]
20. Kim, J.H. A study on the genre related concepts of web-novel. *J. Korean Fict. Res.* **2019**, *74*, 107–137. (In Korean) [CrossRef]
21. Jun Rho, H. A study on the genre aspects of Korean web novels. *Comp. Study World Lit.* **2018**, *64*, 409–428. (In Korean)
22. Hebdige, D. Subculture: The Meaning of Style. *Crit. Q.* **1995**, *37*, 120–124. [CrossRef]
23. Jo, S.; Oh, H. Identifying Reader's Internal Needs and Characteristics Using Keywords from Korean Web Novels. *J. Korea Inst. Inf. Commun. Eng.* **2020**, *24*, 158–165. (In Korean)
24. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of the 27th Annual Conference on Neural Information Processing Systems (NIPS 2013), Lake Tahoe, NV, USA, 5–8 December 2013; Burges, C.J.C., Bottou, L., Ghahramani, Z., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Lake Tahoe, NV, USA, 2013; pp. 3111–3119.
25. Su, Q. A Study of Code-Switching in Chinese Web Novels. In Proceedings of the 2018 International Conference on Asian Language Processing (IALP 2018), Bandung, Indonesia, 15–17 November 2018; Dong, M., Bijaksana, M.A., Sujaini, H., Romadhony, A., Ruskanda, F.Z., Nurfadhilah, E., Aini, L.R., Eds.; IEEE: Bandung, Indonesia, 2018; pp. 123–128. [CrossRef]
26. Lin, Y.J.; Hsieh, S.K. The Secret to Popular Chinese Web Novels: A Corpus-Driven Study. In Proceedings of the 2nd Conference on Language, Data and Knowledge (LDK 2019), Leipzig, Germany, 20–23 May 2019; Eskevich, M., de Melo, G., Fäth, C., McCrae, J.P., Buitelaar, P., Chiarcos, C., Klimek, B., Dojchinovski, M., Eds.; Schloss Dagstuhl—Leibniz-Zentrum für Informatik: Leipzig, Germany, 2019; Volume 70, pp. 24:1–24:8. [CrossRef]
27. Lee, O.J.; Jung, J.J. Story embedding: Learning distributed representations of stories based on character networks. *Artif. Intell.* **2020**, *281*, 103235. [CrossRef]
28. Grayson, S.; Wade, K.; Meaney, G.; Greene, D. The Sense and Sensibility of Different Sliding Windows in Constructing Co-occurrence Networks from Literature. In Proceedings of the 2nd IFIP WG 12.7 International Workshop on Computational History and Data-Driven Humanities (CHDDH 2016), Dublin, Ireland, 25 May 2016; Bozic, B., Mendel-Gleason, G., Debruyne, C., O'Sullivan, D., Eds.; Springer: Cham, Switzerland, 2016; Volume 482, pp. 65–77. [CrossRef]
29. Reagan, A.J.; Mitchell, L.; Kiley, D.; Danforth, C.M.; Dodds, P.S. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Sci.* **2016**, *5*. [CrossRef]
30. Lee, O.J.; Jung, J.J. Character Network Embedding-based Plot Structure Discovery in Narrative Multimedia. In Proceedings of the 9th International Conference on Web Intelligence, Mining and Semantics (WIMS 2019), Seoul, Korea, 26–18 June 2019; Akerkar, R., Jung, J.J., Eds.; ACM: Seoul, Korea, 2019; pp. 15:1–15:9. [CrossRef]
31. Elsner, M. Character-based kernels for novelistic plot structure. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012), Avignon, France, 23–27 April 2012; Daelemans, W., Lapata, M., Màrquez, L., Eds.; The Association for Computer Linguistics: Avignon, France, 2012; pp. 634–644.
32. Tran, Q.D.; Jung, J.E. CoCharNet: Extracting Social Networks using Character Co-occurrence in Movies. *J. Univers. Comput. Sci.* **2015**, *21*, 796–815. [CrossRef]
33. Weng, C.; Chu, W.; Wu, J. RoleNet: Movie Analysis from the Perspective of Social Networks. *IEEE Trans. Multimed.* **2009**, *11*, 256–271. [CrossRef]
34. Lee, O.J.; Jung, J.J. Modeling affective character network for story analytics. *Future Gener. Comput. Syst.* **2019**, *92*, 458–478. [CrossRef]
35. Liu, C.; Last, M.; Shmilovici, A. Identifying turning points in animated cartoons. *Expert Syst. Appl.* **2019**, *123*, 246–255. [CrossRef]

36. Liu, C.; Shmilovici, A.; Last, M. Towards story-based classification of movie scenes. *PLoS ONE* **2020**, *15*, e0228579. [CrossRef] [PubMed]
37. Holanda, A.J.; Matias, M.; Ferreira, S.M.S.P.; Benevides, G.M.L.; Kinouchi, O. Character networks and book genre classification. *Int. J. Mod. Phys. C* **2019**, *30*, 1950058. [CrossRef]
38. Hettinger, L.; Becker, M.; Reger, I.; Jannidis, F.; Hotho, A. Genre Classification on German Novels. In Proceedings of the 26th International Workshop on Database and Expert Systems Applications (DEXA 2015), Valencia, Spain, 1–4 September 2015; Spies, M., Wagner, R.R., Tjoa, A.M., Eds.; IEEE Computer Society: Valencia, Spain, 2015; pp. 249–253. [CrossRef]
39. Chattoo, C.B. Oscars So White: Gender, Racial, and Ethnic Diversity and Social Issues in U.S. Documentary Films (2008–2017). *Mass Commun. Soc.* **2018**, *21*, 368–394. [CrossRef]
40. Chae, G.; Park, J.; Park, J.; Yeo, W.S.; Shi, C. Linking and clustering artworks using social tags: Revitalizing crowd-sourced information on cultural collections. *J. Assoc. Inf. Sci. Technol.* **2015**, *67*, 885–899. [CrossRef]
41. Park, D.; Nam, J.; Park, J. Novelty and influence of creative works, and quantifying patterns of advances based on probabilistic references networks. *EPJ Data Sci.* **2020**, *9*. [CrossRef]
42. Jung, J.E.; Lee, O.J.; You, E.S.; Nam, M.H. A computational model of transmedia ecosystem for story-based contents. *Multimed. Tools Appl.* **2017**, *76*, 10371–10388. [CrossRef]
43. Grayson, S.; Mulvany, M.; Wade, K.; Meaney, G.; Greene, D. Novel2Vec: Characterising 19th Century Fiction via Word Embeddings. In Proceedings of the 24th Irish Conference on Artificial Intelligence and Cognitive Science (AICS 2016), Dublin, Ireland, 20–21 September 2016; Greene, D., Namee, B.M., Ross, R.J., Eds.; CEUR-WS.org: Dublin, Ireland, 2016; Volume 1751, pp. 68–79.
44. Peng, C.; Jung, J.J. Interpretation of metaphors in Chinese poetry: Where did Li Bai place his emotions? *Digit. Scholarsh. Humanit.* **2020**, fqaa016. [CrossRef]
45. Elsner, M. *Abstract Representations of Plot Structure*; Linguistic Issues in Language Technology; CSLI Publications: Stanford, CA, USA, 2015; Volume 12.
46. Labatut, V.; Bost, X. Extraction and Analysis of Fictional Character Networks: A Survey. *ACM Comput. Surv.* **2019**, *52*, 89. [CrossRef]
47. Grayson, S.; Wade, K.; Meaney, G.; Rothwell, J.; Mulvany, M.; Greene, D. Discovering structure in social networks of 19th century fiction. In Proceedings of the 8th ACM Conference on Web Science (WebSci 2016), Hannover, Germany, 22–25 May 2016; Nejdl, W., Hall, W., Parigi, P., Staab, S., Eds.; ACM: Hannover, Germany, 2016; pp. 325–326. [CrossRef]
48. Chaturvedi, S.; Iyyer, M.; Daumé, H., III. Unsupervised Learning of Evolving Relationships Between Literary Characters. In Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI 2017), San Francisco, CA, USA, 4–9 February 2017; Singh, A.P., Markovitch, S., Eds.; AAAI Press: San Francisco, CA, USA, 2017; pp. 3159–3165.
49. Tran, Q.D.; Hwang, D.; Lee, O.J.; Jung, J.J. A Novel Method for Extracting Dynamic Character Network from Movie. In *Big Data Technologies and Applications*; Jung, J.J., Kim, P., Eds.; Springer: Seoul, Korea, 2017; Volume 194, pp. 48–53. [CrossRef]
50. Tran, Q.D.; Hwang, D.; Lee, O.J.; Jung, J.E. Exploiting Character Networks for Movie Summarization. *Multimed. Tools Appl.* **2017**, *76*, 10357–10369. [CrossRef]
51. Lee, O.J.; Jung, J.J. Integrating Character Networks for Extracting Narratives from Multimodal Data. *Inf. Process. Manag.* **2019**, *56*, 1894–1923. [CrossRef]
52. Bost, X.; Labatut, V.; Gueye, S.; Linarès, G. Extraction and Analysis of Dynamic Conversational Networks from TV Series. In *Social Network Based Big Data Analysis and Applications*; Kaya, M., Kawash, J., Khoury, S., Day, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; pp. 55–84. [CrossRef]
53. Lee, O.J.; Jung, J.J. Affective Character Network for Understanding Plots of Narrative Multimedia Contents. In Proceedings of the Workshop on Affective Computing and Context Awareness in Ambient Intelligence (AfCAI 2016), Murcia, Spain, 24–25 November 2016; Ezquerro, M.T.H., Nalepa, G.J., Mendez, J.T.P., Eds.; CEUR-WS.org: Murcia, Spain, 2016; Volume 1794.

54. Lee, O.J.; Kim, J.T. Measuring Narrative Fluency by Analyzing Dynamic Interaction Networks in Textual Narratives. In Proceedings of the 3rd Workshop on Narrative Extraction From Texts (Text2Story 2020), co-located with the 42nd European Conference on Information Retrieval (ECIR 2020), Lisbon, Portugal, 14 April 2020; Campos, R., Jorge, A.M., Jatowt, A., Bhatia, S., Eds.; CEUR-WS.org: Lisbon, Portugal, 2020; Volume 2593, pp. 15–22.
55. Bost, X.; Gueye, S.; Labatut, V.; Larson, M.; Linarès, G.; Malinas, D.; Roth, R. Remembering winter was coming. *Multimed. Tools Appl.* **2019**. [CrossRef]
56. Tsai, C.; Kang, L.; Lin, C.; Lin, W. Scene-Based Movie Summarization Via Role-Community Networks. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23*, 1927–1940. [CrossRef]
57. Lee, O.J.; Jo, N.; Jung, J.J. Measuring Character-based Story Similarity by Analyzing Movie Scripts. In Proceedings of the 1st Workshop on Narrative Extraction From Text (Text2Story 2018), Co-Located with the 40th European Conference on Information Retrieval (ECIR 2018), Grenoble, France, 26 March 2018; Jorge, A.M., Campos, R., Jatowt, A., Nunes, S., Eds.; CEUR-WS.org: Grenoble, France, 2018; Volume 2077, pp. 41–45.
58. Lee, O.J.; Jung, J.J. Explainable Movie Recommendation Systems by using Story-based Similarity. In Proceedings of the ACM IUI 2018 Workshops Co-Located with the 23rd ACM Conference on Intelligent User Interfaces (ACM IUI 2018), Tokyo, Japan, 7–11 March 2018; Said, A., Komatsu, T., Eds.; CEUR-WS.org: Tokyo, Japan, 2018; Volume 2068.
59. Lee, O.J.; Jung, J.J.; Kim, J.T. Learning Hierarchical Representations of Stories by Using Multi-layered Structures in Narrative Multimedia. *Sensors* **2020**, *20*, 1978. [CrossRef]
60. Lee, O.J.; Jung, J.J. Story Embedding: Learning Distributed Representations of Stories based on Character Networks (Extended Abstract). In Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI 2020), Yokohama, Japan, 11–17 June 2020; Bessiere, C., Ed.; International Joint Conferences on Artificial Intelligence Organization: Yokohama, Japan, 2020; pp. 5070–5074. [CrossRef]
61. McKee, R. *Story: Substance, Structure, Style and the Principles of Screenwriting*; HarperCollins: New York, NY, USA, 1997.
62. Le, Q.V.; Mikolov, T. Distributed Representations of Sentences and Documents. In Proceedings of the 31th International Conference on Machine Learning (ICML 2014), Beijing, China, 21–26 June 2014; Xing, E.P., Jebara, T., Eds.; JMLR.org: Beijing, China, 2014; Volume 32, pp. 1188–1196.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.