Network Analysis in Digital Humanities: Application Framework Postprint

Authors: Wang Linxu, Wang Jun, Wang Jun

Date: 2022-11-26T00:00:00+00:00

Abstract

Network analysis, as an effective analytical method and visualization approach, represents one of the most widely adopted directions in the digital humanities domain. A systematic review and synthesis of network analysis applications in digital humanities can assist researchers in rapidly clarifying the capabilities and limitations of network analysis, thereby enabling deeper research practices. Employing content analysis, this study reviews and summarizes articles published in the most influential journals and at international digital humanities conferences in the field over the past five years, organizing them across four dimensions: research questions, datasets, network characteristics, and network analysis metrics, ultimately distilling an application framework for network analysis methodologies in digital humanities. This framework encompasses three data scales (single text, parallel texts, and corpus), five application scenarios (character analysis network, figure association network, discourse space network, text association network, cultural theme network), two metric scales (global metrics and local metrics), and five metric types (composition, density, centrality, factions, and structure).

Full Text

Abstract

[**Purpose/Significance**] Network analysis, as an effective analytical and visualization method, represents one of the most widely applied directions in digital humanities. A systematic summary and synthesis of network analysis applications in digital humanities can help researchers quickly clarify the capabilities and limitations of this method to facilitate deeper research practice. [**Method/Process**] This study employs content analysis to review and summarize articles published in the past five years in the most influential international digital humanities journals and conferences. By examining four dimensions—research questions, datasets, network characteristics, and network analysis metrics

—we distill an application framework for network analysis methods in digital humanities. [**Results/Conclusions**] This framework includes three data scales (single text, parallel texts, and corpus), five application scenarios (character analysis networks, character relationship networks, discourse space networks, text correlation networks, and cultural theme networks), two indicator scales (global and local indicators), and five metric types (composition, density, centrality, cliques, and structure).

Keywords: network analysis; digital humanities; social network analysis; content analysis
Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2022.20.006

Since the late 20th century, digital humanities—evolving from humanities computing—has brought about a global transformation in knowledge production paradigms. After 2016, digital humanities entered an accelerated development phase in mainland China, with the boundary between network infrastructure construction and research-driven digital humanities studies becoming increasingly clear, and a "methodological community" gradually emerging [1]. Originating from network theory, network analysis not only possesses powerful analytical capabilities but also serves as an effective data visualization approach. In recent years, it has been widely applied in digital humanities, becoming one of the directions with the most substantive progress.

Cognitive, action, flow, distance, and co-occurrence relationships are common attributes. Applying social network analysis in digital humanities provides precise quantitative analysis of various social relationships, offering social examples for theoretical model construction and proposition verification, while uncovering implicit social relationships and trends. The focus of analysis and modeling using social network analysis techniques on characteristic resources lies in the acquisition and processing of nodes and the determination of relationships and relationship patterns. Conceptually distinct from traditional statistical analysis and data processing methods, social network analysis represents a new tool for exploring humanities questions in digitized historical materials. As social network analysis gradually develops in digital humanities research, characters are no longer the only nodes in networks; networks composed of texts, words, and other elements have begun to attract scholarly attention and use. Therefore, the network discussed in this paper is not limited to social networks composed solely of characters.

To explore the usage of network analysis in digital humanities research and clarify the application objectives, capabilities, and rationales of the method, this paper uses content analysis to review relevant literature and proposes a structural framework for SNA applications in digital humanities to facilitate methodological understanding and expand the boundaries of network analysis. Based on this purpose, the core questions explored in this paper are: (1) What research can be conducted using network analysis on humanities materials? (2) What nodes and connections constitute these networks? (3) Which metrics are selected during network analysis?

2. Research Design

2.1 Literature Acquisition

This study selected *Digital Scholarship in the Humanities* (DSH)—the most academically influential journal in digital humanities—and the Digital Humanities Conference (DH) organized by the Alliance of Digital Humanities Organizations as primary data sources. We collected papers from 2018 to 2022 (five years) and filtered out studies that used network analysis to deconstruct, analyze, and mine texts. Additionally, considering that DSH is a linguistics journal focusing primarily on language and literature, we also included papers from the historical journal *Historical Methods: A Journal of Quantitative and Interdisciplinary History* (2021 impact factor: 1.647, Q1 in History; 1.299, Q2 in Linguistics). Ultimately, 32 articles were selected as research objects (see Table 1).

In recent years, domestic digital humanities 学术交流 and publishing have flourished. *Digital Humanities* (launched by Tsinghua University and Zhonghua Book Company) and *Digital Humanities Research* (launched by the School of Information Resource Management at Renmin University of China) are both professional journals in the field (see Table 1). However, since both journals have been established for less than five years, they were not included in the coding discussion. Nevertheless, relevant studies in these journals provide important reference value for material processing and methodological application in the Chinese context. Europe' s *Journal of Historical Network Research* is a rare humanities and social sciences journal focusing on network analysis applications in the field. Since its inception, all five issues have published high-level applied practices of social network analysis methods in history, society, and political science. Recently, its special issue "Beyond Guanxi" was officially published, aiming to introduce recent Chinese historical network scholarship to European and American historians and providing research paradigms for network analysis oriented toward Chinese historical materials. Therefore, this journal was also considered in the discussion.

Content analysis is built upon in-depth interpretation of text content, and the coding workload is substantial. The aforementioned relevant literature collectively achieves the research purpose of this paper.

2.2 Research Methods

Content analysis is a method combining qualitative and quantitative approaches. It starts with qualitative problem assumptions and, through quantitative analysis of literature content, identifies features that reflect the essence of the literature and are easily quantifiable [3]. Particularly for the interdisciplinary digital humanities field, scholars often have different writing habits and varying levels of detail in explaining research data and methods. Simply reading to understand the current state of methodology use is not easy. Content analysis can transform linguistically represented literature into statistically depicted data and draw qualitative conclusions from statistical data, thereby helping this paper achieve a deeper and more precise understanding of relevant literature.

For academic literature, content analysis objects typically include bibliographic information such as titles, abstracts, keywords, and references. However, these are insufficient for fully understanding research questions and grasping research methods holistically. Therefore, to ensure the completeness of the coding scope, this paper uses full texts as analysis objects.

First, based on understanding of digital humanities research, we developed five primary coding categories: research question, research data, network data, network type, and network analysis metrics. Second, based on article content, we conducted secondary coding on these 32 articles (see Table 2), with coding results summarized in Section 6. Since most coding objects are English papers, secondary coding names use generalized Chinese terms after summarizing identical original descriptions as much as possible.

3. Common Source Texts and Research Questions

3.1 Source Texts

Network analysis in digital humanities typically begins with defining specific text corpora—that is, starting with processing analytical materials. Corpus construction is key to drawing meaningful conclusions [4], and by exploring the correlation between source text analysis dimensions and research questions and defining corpora, network analysis can quantify and visualize qualitative research questions [5].

Digital humanities research has inherent interdisciplinary characteristics. In addition to humanities, arts, and social sciences being highly relevant to digital humanities, computational linguistics and applied computer science research itself is also closely connected to digital humanities [6]. These research overlaps make it difficult to classify digital humanities studies by disciplinary preference. Although research paradigms vary, explorations on similar research data often complement each other, with single text analysis, parallel text analysis, and corpus-based analysis being three common approaches across different disciplines on similar datasets.

The main materials for digital humanities network analysis research include: literary works (drama, novels, poetry, prose), documentary archives (letters, historical archives), explanatory documents (dictionaries, encyclopedias), and religious and philosophical classics. Additionally, social media usage data can be used to explore memory inheritance and cultural transmission. For example, I.I.M.A. Rhodes used network analysis to analyze likes and comments on YouTube videos related to civil rights activist Paul Robeson to reveal how Robeson is remembered [7].

3.2 Research Questions

3.2.1 Character Function Analysis in Literary Works Network analysis of literary works is the most important topic in digital humanities research on such materials. Franco Moretti, a professor in the Department of Comparative Literature at Stanford University, proposed "Distant Reading" based on large-scale text data collection and statistical analysis, which has had a profound influence on world literature studies and provides theoretical foundations and preliminary ideas for social network analysis of novels. "Plot Analysis," built upon identifying key nodes (i.e., core characters) in character networks and their dialogues [8], focuses on several aspects closely related to novel narrative research: space and time in texts, division of network communities, discovery of core characters, and deconstructing character dialogues and relationships based on network theory to identify key characters with narrative potential from large-scale, multi-character texts.

As more specific quantitative practices unfold, defining "relationships" between characters has become a research difficulty and hotspot, leading to different definitions of "relationships" based on different literary genres. For example, dialogue behavior in drama almost represents the entirety of dramatic action. L. Evalyn et al. used character co-occurrence in dialogue as connections between nodes in Shakespeare' s plays, with the number of times characters speak as edge weights, to explore narrative patterns in comedies, tragedies, and historical plays [9]. However, in novels, dialogue is not the entirety of narrative; a certain amount of narration can also reveal the closeness of character relationships. Zhao Wei analyzed the narrative functions of key characters in "The Great Wave Trilogy" by distinguishing direct and indirect speech and their proportions and dialogue contexts [10]. Whether in novels or drama, both are essentially linguistic arts. When dialogue language is used as the main analytical variable, it can intuitively clarify the distribution of discourse power in works. Furthermore, combining it more precisely and deeply with traditional close reading, and discussing the narratological functions of relevant characters or texts in conjunction with important statistical concepts in character and plot research, can further excavate the unique value of works. Additionally, applying network analysis to literary works helps study the influence of specific texts or authors on other works over a period [11] and understand the evolution of relationships between different characters [12-13].

3.2.2 Character Relationship and Clustering Analysis Character function analysis is based on grasping character relationships and dialogues in literary works. However, in other materials, character relationships often require accumulation of materials covering a certain time range, particularly evident in historical studies that also focus on characters. Since M.C. Alexander et al. investigated the rise of the Medici family in the early 15th century [14], archaeologists and historians have shown increasing interest in network analysis methods [15]. Research on documentary archives such as letters and historical archives often spans long time dimensions, such as tracing communication networks from the Renaissance period [16] and revealing witness affiliation networks of Scottish elites co-witnessing charter drafting in medieval times [17], both exploring character relationships within hundred-year time windows.

Thanks to the increasing availability of semi-structured historical biographical databases such as the China Biographical Database Project (CBDB), large-scale character relationship analysis has become possible. Yan Chengxi et al. constructed political networks of Song Dynasty figures and elaborated on relationship evolution patterns [18]. Zhang Liyuan et al. analyzed ideological transformations of Song Dynasty scholar-officials by comparing academic and political networks between the Northern and Southern Song periods [19]. As more figures appear in relationship networks, character aggregation and dispersion become more valuable for interpretation, and individual positions within groups have become a major research theme in network analysis. For example, J.K. Ochab et al. analyzed clustering in 13th-century Czech noble co-witnessing networks to examine noble gatherings before and after a specific time node, i.e., noble interactions before and after a rebellion [20].

3.2.3 Text Narrative Pattern and Content Analysis Tracing the development of social network analysis reveals that it is deeply rooted in human cultural value symbol systems, with 20th-century structuralist linguistics and semiotics both providing fertile ground for social network research. Therefore, extracting interpretable symbols from texts, establishing relationship networks of fictional images, and discovering potential narrative intentions behind texts have become another feasible analytical paradigm. Qiu Weiyun et al.' s recent newspaper research depicted important ideological history transformations by constructing vocabulary concept networks [21]. Other scholars constructed co-occurrence networks of 30 terms related to nation-building themes in the writings of Liang Qichao and Chen Duxiu to examine differences and ideological transformations in their works [22].

When nodes in a network are words (or characters), phrases, or sentences, their relationships constitute narrative space, and such networks are often directed. Discussions of narrative space include differences in narrative across genres, transitivity and similarity of narrative units. However, common metrics for relationships between narrative units are relatively singular, often using co-occurrence frequency and mutual information scores—the probability of two narrative units appearing together within a specified character window. Ignoring positional information makes it difficult to conduct in-depth content analysis. Y. Yang et al. introduced resistance distances to describe lexical relationships in character co-occurrence networks [23]. Additionally, connecting narrative units with neighboring units can further explore narrative content.

3.2.4 Character Association Discovery and Database Usability Analysis Character association discovery research primarily benefits from various structured thematic databases and typically lacks specific research question orientation, representing very typical data-driven research that often serves as application demonstrations for database construction studies. Network analysis and visualization have also become important means for dataset evaluation and research potential demonstration. M. Levine demonstrated the effectiveness of the Chinese Biographical Database (CBD) through two historical network analysis cases after introducing it [24]. P.K. Bol studied blood and academic relationships among Wuzhou people in Zhejiang through network analysis, further illustrating CBD' s in-depth description of character relationships [25]. Apart from relying on existing databases, most digital humanities scholars still need to build their own datasets for research. Therefore, the focus of many digital humanities studies is no longer analytical methods and conclusions but dataset construction, with network analysis naturally becoming attempts based on partial data. J. Waxman constructed a Talmud dataset based on debate content among thousands of Babylonian scholars, including academic relationships (primarily student-colleague relationships) and 13 types of scholar interactions (collaboration, disagreement, refutation, etc.), and analyzed part of the content using network analysis and visualization [26]. Similarly, M.R. Zambrano et al. discussed modern Ecuadorian architecture based on historical archives [27].

4. Network Construction

After clarifying the data foundation and research questions, defining nodes and edges in the network and constructing the network is the next step. By examining network types and node/edge selection in relevant articles (see Table 4), we can identify common "networks" in digital humanities: character dialogue networks, character co-occurrence networks, character-location networks, character relationship networks, character affiliation networks, discourse space networks, text correlation networks, and cultural theme networks. Network types are closely associated with network data and research questions, but this section focuses solely on node and edge content to categorize network types.

4.1 Character Analysis Networks

Network types corresponding to character function analysis research goals are character dialogue networks, character co-occurrence networks, and character-location networks, which can be summarized as character analysis networks for analyzing fictional characters in texts.

Character dialogue space is the most significant manifestation of character function in such networks. In character dialogue networks, nodes are typically characters, and relationships represent dialogues between two characters, weighted by the number of dialogue occurrences. I. Pikkanen used characters from four historical plays published in Sweden and Finland between 1837 and 1869 as nodes, with character exchanges as edges, to reveal core characters and narrative spaces in dramas through metrics and visualization of four periods' exchange networks [28]. M. Kubis constructed character dialogue networks for 19th- and

20th-century novels separately, distinguishing between speakers and characters mentioned in dialogue to discover systematic differences in novels across the two centuries [29].

Character co-occurrence networks refer to characters participating in the same event. P. Jayannavar et al. studied character interaction networks in which characters participated in the same event and observation networks where mutually acquainted characters observed each other [30]. Xu Chao represented characters and event entities from the *Zuo Zhuan* annotated corpus through co-occurrence networks, discovering the small-world property of Spring and Autumn period networks [31].

Character-location networks connect characters who have visited the same location. J.S.Y. Lee et al. associated characters appearing in the same location [32]. Guo Jiaxin et al. constructed not only a "character-location" network for *The Longest Day in Chang' an* to analyze character groups' activity ranges and venues but also a "location-location" network to analyze location transitions and associations, where edge weights relate to connection tightness between nodes. Additionally, co-occurrence in the same scene is included in character-location networks [33], where locations can be considered fictional "places."

4.2 Character Relationship Networks

Character relationship networks are based on factual relationships between characters from multiple reference materials, mainly including two types: direct relationship networks describing character-scene co-occurrence, and networks describing relationships between characters and their affiliated communities (or factions).

Character relationships include kinship, academic relationships, and political relationships. W. Shang used characters in *Shishuo Xinyu* as network nodes, with character co-occurrence in each story as edges, and divided relationships into direct positive relationships (factual level) and lighter positive interactions (identification, respect, travel, dialogue, academic exchange, colleague relationships, courtesy, comfort, etc.), assigning different weights to study the social network of Eastern Jin aristocrats [34]. Even without explicit character relationships, character networks remain applicable. C. Armand marked Republican-era figures as nodes, adding a directed edge from A to B when B appeared in A' s biography, and tested whether figures formed groups based on shared attributes (provincial origin, education, etc.) [35].

With large-scale data support, relationships generated through character interactions can also reveal unique phenomena. M.J. Hill et al. constructed book transaction networks of authors based on bibliographic databases to understand authors' historical concepts and contexts [36]. Li Hui et al. connected Zeng Guofan with his correspondents, analyzing Zeng' s character relationships across different periods from personal letter networks [37].

Character affiliation networks often explore new character group divisions through existing affiliations. J.J. Yoo discovered cross-faction poet interactions and poetry community aggregation and overlap through "poet-poetry community networks" and "poet-poetry faction networks" in the Joseon Dynasty [38]. Additionally, direct connections between characters and other content provide important references for determining "character groups," such as H.L. Xiong' s association of characters with official positions to depict promotion paths of Song Dynasty elites under the "academy talent cultivation" policy [39].

4.3 Discourse Space Networks

Discourse space networks are networks of words (or characters) and terms, including term co-occurrence networks in scenes (or paragraphs), term collocation networks, and term derivation networks.

Analyzing term co-occurrence networks in the same paragraph or sentence is one of the most common and ideal ways to explore the overall semantic structure of a given text collection. Co-occurrence refers to "relationships between pairs of terms appearing within a constant-size context window" [40], which can be any text length—a certain number of characters or words, a sentence, a paragraph, or an entire document. Window size often depends on how many topics are expected in a document. Novels are typically segmented into smaller modules to capture fleeting themes as stories unfold. Although poetry uses concise language, each word has considerable independent semantic expression, making each sentence an effective window. News and other documentary short texts always center on one theme, making whole-document term co-occurrence analysis more appropriate. D. GamerMann et al. used words in *The Little Prince* as nodes, connecting two words if they co-occurred in a sentence and shared a third word, to analyze word usage in different language translations [41]. C. Lee selected 100 high-frequency words from 404 news reports and connected high-frequency words appearing in the same document [42].

Term collocation networks are networks of terms and their adjacent terms; term derivation networks are networks of words with common components (such as roots in English), both being derivatives of corpus linguistics methodology. Corpus linguistics establishes natural language processing rules through analysis and statistics of large amounts of real language material, mainly applied in teaching, translation, vocabulary, semantics, dictionaries, and grammar, focusing on both macro (overall corpus linguistic features and text types) and micro (specific lexical and grammatical phenomena) aspects [43], which helps interpret historical texts and discourse [44]. Network analysis supports research at both macro and micro levels. H. Bonin analyzed British parliamentary debate content, forming collocation networks of "people" and "democracy" (or "demo-") to analyze concept transformations [45]. Domestically, Qiu Weiyun also associated character concepts related to "politics" in *New Youth* [46]. In term derivation networks, J. Longhi compiled tweets from official accounts of 11 presidential candidates in 2017, connecting "islam" and its derivatives appearing in the same tweet to emphasize discourse analysis' s importance in analyzing fake news [47].

4.4 Text Correlation Networks

Text correlation networks focus on paragraphs and chapters, with connections being similarities between different texts, where similarity magnitude determines edge weights. Constructing such networks often requires prior text processing. G. Rotari et al. used each Grimm Brothers' letter as a node, connecting letters with common word co-occurrence, and determined similarity (edge weights) using cosine increments applied to z-scores of the 50-150 most common words to analyze similarities and differences in the Grimm Brothers' writing styles and their roles in professional or personal development [48].

Text consensus networks also use texts or text samples as nodes, with connections established in two ways: first, by calculating text distances between samples to create the strongest link to the nearest node and weaker links to the two next-nearest nodes; second, by repeating the process for different ranges of common words multiple times and combining each iteration' s network into a single consensus network similar to consensus tree analysis. C. Lee constructed author and translator consensus networks based on English and Korean novels to identify different authors' and translators' textual styles [49].

4.5 Cultural Theme Networks

Cultural theme networks differ from the above networks by emphasizing that social interactions and relationships can somehow mediate the structure and agency of various social phenomena. In 1990, the New York School distinguished four studies on the connection between networks and culture: networks as cultural channels, networks as culture-shaping, networks as cultural forms, and networks as cultural interaction [50]. However, in highly practical digital humanities research, detailed divisions in sociology have not yet been involved. Here, culture is often deconstructed into concepts, trends, or externalized narrative attributes on objects, with clustering used to mine aggregation and themes of these elements. L. Giagnolini et al. constructed a "historian-concerned theme" network and a "historian-historian" network of co-workers in the same institution based on rich biographical archives of art historians to discover relationships in art history communities and further identify cultural flows [51]. S. Milonia et al. observed how musical imitation reveals the interconnected nature of medieval romantic lyric poetry through a "song-song" network [52].

5. Network Analysis Metrics and Tools

Network analysis metrics can be classified differently based on various criteria. According to the division between whole networks and local networks, metrics can be divided into global and local indicators. However, in practice, both are often used simultaneously. Therefore, based on descriptive purposes, these metrics can be divided into five descriptive categories.

5.1 Global and Local Attributes

The history of network centrality measurement dates back more than half a century, but unlike more objective and reliable visual interpretation, metric results require contextual interpretation [53]. As shown in Table 5, statistical analysis of global attributes can compare multiple network dimensions simultaneously [54]. For example, comparing node and edge counts across networks of the same type [55] is a directly interpretable ranking tool. Additionally, network density (edge count related to nodes) and average path length are highly effective for analyzing character networks.

Local metrics include degree (number of adjacent nodes), the simplest centrality indicator that was the only systematically used metric from the late 1950s to early 1970s before more diverse measures developed [56]. Its simplicity makes interpretation clear—for instance, in literary networks, degree can calculate how many times a character speaks to another. Betweenness centrality breaks the concept of what network "centers" might consist of and is popular for revealing abilities to connect different subnetworks. Closely related to circulation concepts, it detects "bridges" or "key passages" that can open or lock parts of networks by calculating shortest paths. Thus, depending on application, these positions are both the most powerful and most vulnerable in networks. Closeness centrality allows nodes to distribute relatively evenly in networks with certain density that are not yet divided into cliques and can transform between "center" and "periphery" concepts. Eigenvector centrality is more difficult to interpret, largely depending on structures around nodes, but is particularly useful for analyzing hierarchical structures in graphs [57] and forms the basis of the PageRank algorithm.

5.2 Five Categories of Network Analysis Metrics

When describing network features, we can describe each point' s features or view the network as a whole to describe its overall characteristics. By coding relevant articles, we divided these metrics into five levels: composition, density, centrality, cliques, and structure. The usage frequency of these metrics in relevant papers is shown in Table 6.

(1) Composition. Points, edges, and size are statistics of how many nodes a network contains and how many other nodes a node connects to, serving as references for node classification.

(2) Density. In the coded papers, nearly 47% of studies analyzed lines, neighborhoods, and density—further describing network space size. We use "density" to summarize these three metrics. Network density, network diameter, average path length, and node connectivity are core analytical metrics.

(3) Centrality. Half of the studies explored centrality, peripherality, and centralization. When describing each point' s features, calculating how many directly connected points exist is an important analytical perspective—this value is "centrality." When describing whole network features, "centralization" reflects the extent to which a network revolves around a point.

(4) Cliques. Components, cores, and cliques study subgroups in networks. Attempts to discover how many categories a network can be divided into—clusters, components, cores, circles, etc.—all mine subgraphs based on node relevance and relationship tightness. These subgraphs often provide important references for researchers.

(5) Structure. Position, set, and cluster analysis include structural equivalence, regular equivalence, clustering, block models, etc. Discussions of node positions often go beyond graph theory principles to analyze set structures. When applied to digital humanities, they often enable case studies to extend to similar situation determinations.

5.3 Network Analysis Tools

Many network analysis tools are widely recognized and used in academia and industry. Tools such as Neo4j, Cytoscape, Gephi, Pajek, and JUNG all perform well in knowledge discovery, information fusion, scalability, and visualization [58]. Gephi—an open-source network analysis and visualization software package written in Java on the NetBeans platform—is widely used by digital humanities scholars for its excellent visualization capabilities. However, due to Gephi's weak support for network metric calculations, it is often paired with social network analysis software like UCINET or programming languages like R and Python.

As more researchers with humanities backgrounds enter digital humanities, many digital humanities platforms also provide network visualization and analysis tools for scholars without programming habits, such as Stanford University' s online network visualization tool Palladio [59], which allows users to easily upload data for network analysis, and the DocuSky digital humanities research platform, which provides four analytical approaches for Chinese text style network analysis [60].

6. Literature Classification

After understanding common research questions, text data, network types, and analysis metrics, this paper derives an application framework for digital humanities research using network analysis. By mapping relevant articles to core codes—that is, providing specific classification of important literature—we can both facilitate more intuitive understanding of literature details and verify the applicability and coverage of the discussed content. Although source text characteristics are the fundamental basis for quantitative analysis, their variety and differences do not affect methodological discussion, while data scale determines the upper limit of quantitative method use. Therefore, this paper classifies relevant literature into three research types: single text analysis, parallel text analysis, and corpus-based analysis (see Table 7).

6.1 Single Text Analysis

When analyzing single texts, interpreting constructed networks is crucial and often requires comprehensive judgment combining existing research and material backgrounds. With limited data, networks have relatively few nodes and edges. Descriptive analysis from points, edges, and size, and from lines, neighborhoods, density, centrality, peripherality, and centralization is usually sufficient. Additionally, combining multiple analytical methods for in-depth mining of single texts is necessary. V.H. Masías et al. used K-means clustering on weighted networks to analyze characters in *Romeo and Juliet*, selecting only closeness centrality as the key metric for weighted networks [61].

6.2 Parallel Text Analysis

Two similar texts or different language translations of the same work can be considered parallel text analysis. As network nodes and edges increase, clique (or cohesive subgroup) generation and description become important references for comparative analysis. Y. Fang et al. analyzed discourse space networks in three Chinese translations of *Alice's Adventures in Wonderland* through cohesive subgroup analysis to explore translator styles [62].

6.3 Corpus Analysis (Including Bilingual Corpus Comparative Analysis)

Whether authors build their own corpora or conduct research based on existing databases, sufficient data allows network analysis methods greater application space. Comparative analysis remains a common method. T.Y. Lim et al. compared discourse space networks built from 224 poems by Plath and 280 poems by Sexton, observing personal pronoun and emotional word networks through various metrics including important nodes, node connectivity, betweenness centrality, eigenvector centrality, and cohesive subgroups, additionally using structural topic models to identify word groups representing each theme [63].

7. Future Challenges

In recent years, with the advancement of "New Liberal Arts" construction, digital humanities has developed rapidly in China, bringing material and topic expansion and injecting new vitality into humanities and social sciences. This paper focuses on specific methodological applications and finds that although digital humanities research has no boundaries, it is full of challenges and often time-consuming, making it difficult to achieve overnight. Even when researchers have many considerations during practical exploration—such as theoretical and standard discussions, research "validity" and "reliability"—there are still specific issues worth considering that serve as current research difficulties and challenges, providing concrete guidance for future research.

7.1 Data Reuse

In recent years, digital humanities research using network analysis has increased, but methodological practice at the paradigmatic level has not gradually deepened, merely applying it to different materials and texts to solve similar problems. Even in methodological practice studies, researchers still focus considerable energy on dataset construction, as seen in paper structures. Therefore, how to utilize and share existing datasets to promote exploration based on more materials and deeper research questions, and to expand network analysis paradigms, has become a future challenge for digital humanities and will represent another expansion of digital humanities research boundaries. Fortunately, the opening of more semi-structured or structured thematic datasets enables digital humanities scholars to complete research topics spanning different periods single-handedly. M. Levine at Central Washington University used principal component analysis, geographic information visualization, and one-mode network structure analysis to interpret experiences of Chinese revolutionary leaders with support from the Chinese Biographical Database (CBD) data [74].

7.2 Multimodal Networks

Networks where all nodes are the same type are called "one-mode networks" ; networks with two node types are called "bipartite networks" or "two-mode networks" ; networks with more node categories are called "multimodal networks." Due to network analysis tool limitations, many scholars have to split two-mode networks into multiple one-mode networks for analysis and draw conclusions through comparison; analysis oriented toward multimodal networks is even rarer. For relatively common components, cores, and cliques, multimodal network clique discovery mainly includes topic model-based methods, ranking and clustering combination methods, data reconstruction methods, and dimensionality reduction methods [75]. All these methods pose more requirements for digital humanities research but also bring more possibilities.

References

[1] Zhao Wei. Digital Era Humanities Research Transformation and Transcendence–Digital Humanities in China[J]. Exploration and Free Views, 2021(6): 191-206, 232-233.

[2] WETHERELL C. Historical social network analysis[J]. International review of social history, 1998, 43(S6): 125-144.

[3] Shao Zuoyun, Li Xiuxia. Review of Literature Knowledge Discovery Methods Combining Citation Analysis and Content Analysis[J]. Information Studies: Theory & Application, 2020, 43(3): 153-159.

[4] RANDI R. Building a corpus: what are the key considerations?[M]// The Routledge handbook of corpus linguistics. London: Routledge, 2010.

[5] PAINTER D T, DANIELS B C, JOST J. Network analysis for the digital humanities: principles, problems, extensions[J]. Isis, 2019, 110(3): 538-554.

[6] LUHMANN J, BURGHARDT M. Digital humanities—a discipline in its own right? an analysis of the role and position of digital humanities in the academic landscape[J]. Journal of the Association for Information Science and Technology, 2022, 73(2): 148-171.

[7] RHODES I I M A. Paul Robeson's place in YouTube: a social spatial network analysis of digital heritage[J]. Digital scholarship in the humanities, 2019, 34(1): 174-188.

[8] MORETTI F. Network theory, plot analysis[M]. London: Verso Books, 2013.

[9] EVALYN L, GAUCH S, SHUKLA M. Analyzing social networks of xml plays: exploring shakespeare's genres[C/OL]. [2022-08-11]. https://dh2018.adho.org/analyzing-social-networks-of-xml-plays-exploring-shakespeares-genres/.

[10] Zhao Wei. Social Network Analysis and Character Functions in "The Great Wave Trilogy" [J]. Shandong Social Sciences, 2018(9): 50-64.

[11] SOR J, LONG H. Network analysis and the sociology of modernism[J]. Boundary 2, 2013, 40(2): 147-182.

[12] AGARWAL A, CORVALAN A, JENSEN J, et al. Social network analysis of alice in wonderland[C]// Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature. Montreal: Association for Computational Linguistics, 2012: 88-96.

[13] MACCARON P, KENNA R. Universal properties of mythological networks[J]. Europhysics letters, 2012, 99(2): 28002.

[14] ALEXANDER M C, DANOWSKI J A. Analysis of an ancient network: personal communication and the study of social structure in a past society[J]. Social networks, 1990, 12(4): 313-335.

[15] LEMERCIER C, ZALC C. Quantitative methods in the humanities: an introduction[M]// GOLDHAMMER A. Charlottesville: University of Virginia Press, 2019.

[16] AHNER R, AHNER S E. Metadata, surveillance and the Tudor state[J/OL]. Digital scholarship in the humanities, 2017, 32(2): 336-343.

[17] JACKSON C. Using social network analysis to reveal unseen relationships in medieval Scotland[J]. Digital scholarship in the humanities, 2017, 32(2): 336-343.

[18] Yan Chengxi, Wang Jun. Digital Humanities Perspective: Visualization Research of Song Dynasty Political Networks Based on Symbolic Analysis[J]. Journal of Library Science in China, 2018, 44(5): 87-103.

[19] Zhang Liyuan, Wang Jun. Comparative Analysis of Academic and Political Systems in the Northern and Southern Song Dynasties Based on Social Network

Dynamics[J]. Technology Intelligence Engineering, 2020, 6(1): 34-49.

[20] OCHAB J K, ŚKVRŇÁK J, ŚKVRŇÁK M. Detecting Ottokar II' s 1248-1249 uprising and its instigators in co-witnessing networks[J/OL]. Historical methods: a journal of quantitative and interdisciplinary history, 2022: 1-20[2022-08-22]. https://doi.org/10.1080/01615440.2022.2065397.

[21] Qiu Weiyun. Vocabulary, Concepts, Discourse: Construction and Representation of "Beauty" in Modern China from a Digital Humanities Perspective[J]. Journal of Art Theory and Art History, 2019(1): 182-223.

[22] CHAO A S, LIU Z, LI Q. Network of words: a co-occurrence analysis of nation-building terms in the writings of liang qichao and chen duxiu[J/OL]. Journal of historical network research, 2022, 5(1)[2022-08-22]. https://doi.org/10.25517/jhnr.v5i1.122.

[23] YANG Y, SONG Y. Exploring the similarity between Han' s and non-Han' s Yuan poetry: resistance distance metrics over character co-occurrence networks[J]. Digital scholarship in the humanities, 2022, 37(3): 880-893.

[24] LEVINE M. Biography for historical analysis: a chinese biographical database[J/OL]. Journal of historical network research, 2021, 5(1)[2022-08-22]. http://jhnr.uni.lu/index.php/jhnr/article/view/124.

[25] BOL P K. From kinship to collegiality: changing literati networks, 1100-1400[J/OL]. Journal of historical network research, 2021, 5(1)[2022-08-22]. http://jhnr.uni.lu/index.php/jhnr/article/view/121.

[26] WAXMAN J. A graph database of scholarly relationships in the Babylonian Talmud[J]. Digital scholarship in the humanities, 2021, 36(S2): ii277-ii289.

[27] ROSA ZAMBRANO M, GONZALEZ M, MEDINA A, et al. Modern architecture in the professional discourse: analysis of the Architectural Biennial of Quito' s 1976-92 archive using bipartite networks[J]. Digital scholarship in the humanities, 2022, 37(3): 894-909.

[28] PIKKANEN I. The metrics and poetics of historical drama: the dramatis personae of a premodern revolt in early nineteenth-century Finland[J]. Orbis litterarum, 2019, 74(5): 311-339.

[29] KUBIS M. Quantitative analysis of character networks in Polish xix and xx century novels[J]. Digital scholarship in the humanities, 2021, 36(S2): ii175-ii181.

[30] JAYANNAVAR P, AGARWAL A, JU M, et al. Validating literary theories using automatic social network extraction[C]// Proceedings of the fourth workshop on computational linguistics for literature. Denver: Association for Computational Linguistics, 2015.

[31] Xu Chao. Research on Language Networks and Social Networks in Zuo Zhuan[D]. Nanjing: Nanjing Normal University, 2014.

[32] LEE J S Y, YEUNG C Y. Extracting networks of people and places from literary texts[C]// Proceedings of the 26th Pacific Asia conference on language, information, and computation. Bali: Faculty of Computer Science, Universitas Indonesia, 2012.

[33] SANTA MARÍA M T, CALVO TELLEZ J, JIMÉNEZ C M. ¿Existe correlación entre importancia y centralidad? evaluación de personajes con redes sociales en obras teatrales de la Edad de Plata[J]. Digital scholarship in the humanities, 2021, 36(S1): i81-i88.

[34] SHANG W, SANG Z. Solidity in turbulent flow: the social network of aristocratic families in the Eastern Jin Dynasty (317-420 CE)[J/OL]. Journal of historical network research, 2022, 5(1)[2022-08-22]. http://jhnr.uni.lu/index.php/jhnr/article/view/126.

[35] ARMAND C, HENRIOT C. From textual to historical networks: social relations in the biographical dictionary of republican China[J/OL]. Journal of historical network research, 2021, 5(1)[2022-08-22]. http://jhnr.uni.lu/index.php/jhnr/article/view/117.

[36] HILL M J, SÓLYOM T. Patterns of early modern authorship: using metadata as historical record[C/OL]. Utrecht: DH, 2019[2022-08-22]. https://researchportal.helsinki.fi/en/activities/patterns-of-early-modern-authorship-using-metadata-as-historical-.

[37] Li Hui, Hou Junming, Chen Tao, et al. Containing Pearls and Fluorescence: Research on the Evolution of Ancient Letter Networks[J]. Digital Humanities, 2020(1): 109-117.

[38] YOO J J. Poetry in action: networks of literary communication and the cultural leverage in eighteenth-century Seoul[J]. Digital scholarship in the humanities, 2021, 36(3): 766-781.

[39] XIONG H L. Path toward the top leadership: a network analysis of the civil service system in the early southern song (1131-1164)[J/OL]. Journal of historical network research, 2022, 5(1)[2022-08-22]. http://jhnr.uni.lu/index.php/jhnr/article/view/98.

[40] GRAYSON S, WADE K, MEANEY G, et al. The sense and sensibility of different sliding windows in constructing co-occurrence networks from literature[C]// International workshop on computational history and data-driven humanities. Cham: Springer, 2016: 65-77.

[41] GAMERMANN D, MORET-TATAY C, NAVARRO-PARDO E, et al. The small-world of 'Le Petit Prince' : revisiting the word frequency distribution[J]. Digital scholarship in the humanities, 2017, 32(2): 301-311.

[42] LEE C. How are 'immigrant workers' represented in Korean news reporting? - a text mining approach to critical discourse analysis[J]. Digital scholarship in the humanities, 2019, 34(1): 82-99.

[43] Xu Tongyang, Wang Xia. Data-driven Digital Humanities Research from the Perspective of Corpus Linguistics[J]. Library Tribune, 2021, 41(10): 90-99.

[44] MCENERY A, BAKER H. Corpus linguistics and 17th-century prostitution: computational linguistics and history[M/OL]. [2022-09-01]. https://www.bloomsburycollections.com/book/corpus-linguistics-and-17th-century-prostitution-computation-linguistics-and-history.

[45] BONIN H. From antagonist to protagonist: 'democracy' and 'people' in British parliamentary debates, 1775-1885[J]. Digital scholarship in the humanities, 2020, 35(4): 759-775.

[46] Qiu Weiyun. On the Meaning and Value of Visualized Data in Digital Humanities Research—Taking Digital Conceptual History Research as an Example[J]. Theory and Criticism of Literature and Art, 2020(2): 23-29.

[47] LONGHI J. Mapping information and identifying disinformation based on digital humanities methods: from accuracy to plasticity[J]. Digital scholarship in the humanities, 2021, 36(4): 980-998.

[48] ROTARI G, JANDER M, RYBICKI J. The Grimm Brothers: a stylometric network analysis[J]. Digital scholarship in the humanities, 2021, 36(1): 172-186.

[49] LEE C. Do language combinations affect translators' stylistic visibility in translated texts?[J]. Digital scholarship in the humanities, 2018, 33(3): 592-603.

[50] Liu Jun. Social Network Analysis Handbook (Volume 1)[M]. Chongqing: Chongqing University Press, 2018.

[51] GIAGNOLINI L, DAQUINO M, MAMBELLI F, et al. Exploratory methods for relation discovery in archival data[J/OL]. Digital scholarship in the humanities, 2022[2022-08-18]. https://doi.org/10.1093/llc/fqac036.

[52] MILONIA S, MAZZA MURRO M. Temporal networks of 'Contrafacta'in the first three troubadour generations[J/OL]. Digital scholarship in the humanities, 2022[2022-08-18]. https://doi.org/10.1093/llc/fqac018.

[53] GRANDJEAN M, JACOMY M. Translating networks: assessing correspondence between network visualisation and analytics[C/OL]. [2022-08-18]. https://halshs.archives-ouvertes.fr/halshs-02179024/.

[54] TANG M C, CHEN K. A cross-language comparison of co-word networks in digital library and museum of Buddhist studies[C/OL]. Montreal: DH. 2017[2022-08-18]. http://buddhism.lib.ntu.edu.tw/FULLTEXT/JR-OTHERS/others575593.pdf.

[55] TRILCKE P, FISCHER F, GOBEL M, et al. Theatre plays as 'small worlds' ? network data on the history and typology of german drama, 1730-1930[C/OL]. Jagiellonian University & Pedagogical University, Krakow: DH. 2016[2022-08-18]. https://publications.hse.ru/en/chapters/196225767.

[56] FREEMAN L C. Centrality in social networks conceptual clarification[J]. Social networks, 1978, 1(3): 215-239.

[57] PIPER A, ALGEE-HEWITT M, SINHA K, et al. Studying literary characters and character networks[C/OL]. [2022-08-18]. https://dh2017.adho.org/abstracts/103/103.pdf.

[58] CAMACHO D, PANIZO-LLEDOT A, BELLO-ORTEGA G, et al. The four dimensions of social network analysis: an overview of research methods, applications, and software tools[J]. Information fusion, 2020, 63: 88-120.

[59] PALLADIO. Visualize complex historical data with ease[EB/OL]. [2022-08-18]. http://hdlab.stanford.edu/palladio/.

[60] TU H C, HSIANG J, HUNG I M, et al. DocuSky, a personal digital humanities platform for scholars[J]. Journal of Chinese history, 2020, 4(2): 564-580.

[61] MASÍAS V H, BALDWIN P, LAENGLE L S, et al. Exploring the prominence of Romeo and Juliet' s characters using weighted centrality measures[J]. Digital scholarship in the humanities, 2017, 32(4): 837-858.

[62] FANG Y, LIU H. Seeing various adventures through a mirror: detecting translator' s stylistic visibility in Chinese translations of Alice' s adventures in wonderland[J/OL]. Digital scholarship in the humanities, 2022[2022-08-18]. https://doi.org/10.1093/llc/fqac024.

[63] LIM T Y, LEE S H, TAK J Y, et al. Revisiting sylvia Plath' s and Anne sexton' s confessional poetry: analyzing stylistic differences and evolution of poetic voice(s) through computational text analysis[J]. Digital scholarship in the humanities, 2021, 36(4): 950-970.

[64] ALGEE-HEWITT M A. The hidden dictionary: text mining eighteenth-century knowledge networks[C/OL]. Mexico City: DH. 2018[2022-08-18]. https://dh2018.adho.org/en/the-hidden-dictionary-text-mining-eighteenth-century-knowledge-networks/.

[65] ZHITOMIRSKY-GEFET M, PREBOR G. Sagebook: toward a cross-generational social network for the Jewish sages' prosopography[J]. Digital scholarship in the humanities, 2019, 34(3): 676-695.

[66] RUEGG C, LEE J J. Epic social networks and eve' s centrality in Milton' s paradise lost[J]. Digital scholarship in the humanities, 2020, 35(1): 146-159.

[67] KUBIS M, MICKIEWICZ A. Quantitative analysis of character networks in Polish xix and xx century novels[C/OL]. Utrecht: DH, 2019[2022-08-18]. https://mkubis.home.amu.edu.pl/dh2019.pdf.

[68] LEE C. Do language combinations affect translators' stylistic visibility in translated texts?[J]. Digital scholarship in the humanities, 2018, 33(3): 592-603.

[69] LI W. Visualizing networks of artistic ideas in history paintings in seventeenth-century Netherlands[C/OL]. Utrecht: DH, 2019[2022-08-18]. https://dataverse.nl/dataset.xhtml?persistentId=doi:10.34894/YVFEZO.

[70] HAIDER T, VAN DYCK-HEMMING A. Extracting a social network of musicologists[C]// Digital Humanities 2020 Book of Abstracts. Ottawa: DH,

2020: 143.

[71] LAPARRA-PEREZ A, MUÑOZ F F, FERNANDEZ-DE-PINEDO E N. Econhist: a relational database for analyzing the evolution of economic history (1980-2019)[J]. Historical methods: a journal of quantitative and interdisciplinary history, 2022, 55(1): 45-61.

[72] BROWN D M, SOTO-COROMINAS A, SUÁREZ J L. The preliminaries project: geography, networks, and publication in the Spanish golden age[J]. Digital scholarship in the humanities, 2017, 32(4): 709-732.

[73] LEDOLTER J, VANDERVELDE L. A case study in text mining: textual analysis of the territorial papers[J]. Digital scholarship in the humanities, 2020, 35(1): 101-126.

[74] LEVINE M. Post-WWI Chinese revolutionary leaders in Europe[J/OL]. Journal of historical network research, 2021, 5(1)[2022-08-18]. http://jhnr.uni.lu/index.php/jhnr/article/view/1

[75] Yang Yu, Guo Yong, Li Hailong, et al. Research Progress on Heterogeneous Network Community Discovery[J]. Application Research of Computers, 2018, 35(10): 2881-2887.

Author Contributions: Wang Linxu: Research material collection and analysis, paper writing.
Wang Jun: Research design, paper revision.

Note: Figure translations are in progress. See original paper for figures.