



Social Network Analysis in Russian Literary Studies

Frank Fischer and Daniil Skorinkin

29.1 INTRODUCTION

Network analysis has come to be an essential method in the Digital Humanities. A network can be described, in brief, as “a collection of points joined together in pairs by lines.” Terminologically, “a point is referred to as a node or vertex and a line is referred to as an edge” (Newman 2018, 1). If you can meaningfully describe a dataset with such nodes and edges, it is network data you deal with. Nodes can be entities like airports, cities, or devices connected to the Internet, linked to each other (or not) via edges. In the case of social networks, nodes represent people or, more generally, social entities, which easily extends to fictional characters. The edges between them describe their relations. While these relations can be of many types, literary network analysis at this stage is usually looking into communicative relations: Who is talking to whom and to what extent? This formal approach usually neglects the content of these interactions but can reveal larger structural patterns that would otherwise stay invisible as we will see in the use cases presented below. Network analysis is meant to complement other quantitative and qualitative approaches when it comes to interpreting literary texts.

Once we established a set of network data, the broad range of algorithms and methods developed within network theory becomes available to make the material “speak” in different ways. The visualization of network data often

The original version of this chapter was revised. A correction to this chapter can be found at https://doi.org/10.1007/978-3-030-42855-6_33

F. Fischer (✉) • D. Skorinkin
Higher School of Economics (HSE University), Moscow, Russia
e-mail: ffischer@hse.ru

comes first but is usually only the starting point of a more precise analysis, because the underlying data can be interpreted more meaningfully by help of literally hundreds of different algorithms. The nature of questions around network data can roughly be divided into graph- and node-related questions. The former allow for an analysis of the structural evolution of texts, while the latter allow for new ways of categorizing character types.

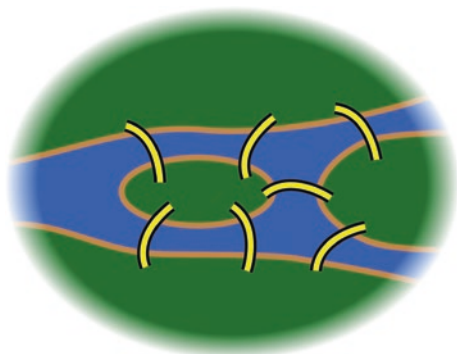
This chapter is structured as follows. A short look into the origins of (social) network analysis in general and literary network analysis in particular will be followed by a methodology section which will explain how to extract and formalize network data before introducing basic graph- and node-related measures. We then present exemplary use cases for literary network analysis, for both drama and novels.

The data for the subsection on drama originates from the Russian Drama Corpus (RusDraCor, see <https://dracor.org/rus>), a Text Encoding Initiative (TEI)-encoded collection of Russian drama from 1747 to the 1940s (Skorinkin et al. 2018). In the words of the Text Encoding Initiative, TEI is “a standard for the representation of texts in digital form” (<http://tei-c.org>). It is usually “expressed using a very widely-used formal encoding language called XML” (Burnard 2014). The data for the subsection on the novel consists in an annotated version of Tolstoy’s *War and Peace* (for more on other corpora, see Chap. 17).

29.2 THE ORIGINS OF SOCIAL NETWORK ANALYSIS

When talking about methods and tools, it is always insightful to look at their historicity, that is, the circumstances which led to their invention. In the case of graph theory, we have to go back to the year 1736 and Swiss mathematician Leonhard Euler. He was confronted with the seven bridges of the back then Prussian city of Königsberg and a question: Is it possible to cross all seven bridges reaching across river Pregel one after another without crossing a bridge twice? By finding an abstraction of the problem, Euler was able to proof that this, in fact, is impossible. He understood the four involved landmasses as nodes and the bridges as edges (see Fig. 29.1). The number of bridges and their endpoints were key for the solution of the problem: all four landmasses

Fig. 29.1 The seven bridges of Königsberg.
Wikimedia Commons,
https://commons.wikimedia.org/wiki/File:7_bridges.svg, licence:
CC BY-SA 3.0



are reached by an odd number of bridges, but for it to work there should be a maximum of two landmasses (nodes) with an odd number of bridges (edges); these two landmasses could then serve as starting and end point, whereas the other two would have to feature an even number of bridges leading to them.

From this historical anecdote, we only take with us the idea of abstracting interconnected entities as graphs and jump two centuries ahead on the timeline, to April 3, 1933. On that very day, an article appeared in *The New York Times* reporting about a new method called “psychological geography” (later renamed to “sociometry”), which was developed by psychosociologist Jacob Levy Moreno. This method promised to visualize attraction and repulsion between individuals within communities showing “the strange human currents that flow in all directions from each individual in the group toward other individuals” (McCulloh et al. 2013). Moreno was one of the first to use network visualizations to describe social phenomena.

Another jump on the timeline and we are in the 1960s at Harvard, where scholars such as Harrison White achieved the so-called “Harvard Breakthrough,” which through methodological innovations “firmly established social network analysis as a method of structural analysis” (Scott 2000, 33). Looking at these developments, Linton Freeman lists “four defining properties” of social network analysis:

1. It involves the intuition that links among social actors are important.
2. It is based on the collection and analysis of data that record social relations that link actors.
3. It draws heavily on graphic imagery to reveal and display the patterning of those links.
4. It develops mathematical and computational models to describe and explain those patterns (Freeman 2011, 26).

We will find all these properties in literary network analysis, too. So, when did the literary studies start to become interested in network analysis? At first, this was not driven by inherent research questions, but by the mere fact that literature is an entertaining use case for social network analysis. Computer scientist Donald Knuth, author of *The Art of Computer Programming* and creator of the TeX typesetting system, needed example data for the Stanford GraphBase, a program and dataset collection for the generation and manipulation of graphs and networks (Knuth 1993). The list of datasets featured character interactions in the chapters of *Anna Karenina*, *David Copperfield*, and *Les Misérables* (<https://people.sc.fsu.edu/~jburkardt/datasets/sgb/sgb.html>)

The files for these three novels contain data on the co-occurrence of literary characters per chapter, which makes for genuine network data. Interestingly, anyone who has ever opened an example file in the number-one network analysis tool in the Humanities, Gephi, will have seen the very network graph of *Les Misérables*, because it is very prominently provided as a Gephi example file (Bastian et al. 2009).

After some more individual approaches to the network analysis of novels (Schweizer and Schnegg 1998 on the post-1989 novel *Simple Stories* by East-German author Ingo Schulze), the network analysis of dramatic texts started out with Shakespeare (Stiller et al. 2003; Stiller and Hudson 2005). Yet these first incentives did not come from literary scholars, and it took some more years until that eventually happened with the studies of Franco Moretti in 2011 and Peer Trilcke in 2013.

These two papers were the starting signal for a broad application of the network paradigm in digital literary studies, leading to several dozen papers in this field within the following five years. The main focus was on dramatic texts, as under normal circumstances they are easier to segment than novels, given their clear division into acts, scenes, and speech acts. While earlier works revolved around the network analysis of just a few individual texts, now hundreds or thousands of texts were examined, following the “Distant Reading” paradigm, which sets out to complement the close reading of texts. In the practice of Distant Reading, digital methods are used to analyze a number of texts that can be orders of magnitude larger than what an individual can read.

One result of this development was the “Distant-Reading Showcase” (Fischer et al. 2016), which put 465 German-language plays on a poster in chronological order, visually illustrating the structural transformation of German drama between 1730 and 1930. Using the same method, we can plot the extracted social networks of the 144 plays contained in the Russian Drama Corpus to date (Fig. 29.2). This unusual view from the digital stratosphere can reveal macrostructures: what is visible from such a distance are general shifts from simple network structures to more complex ones throughout the two centuries between Sumarokov’s tragedy “Horev” (1747) and Mayakovsky’s and Bulgakov’s plays of the 1920s and 1930s.

29.3 METHODOLOGY

29.3.1 *Formalizing Literary Network Data: The “Digital Spectator”*

In order to extract network data from fictional texts, we have to define a consistent way to formalize character interaction. A relation between two characters as we define it is established if both characters are performing a speech act in a given segment of a play, usually a scene. Following this definition, if character A and character B are speaking in the same scene, they are linked to each other.

This formalization is inspired by Romanian mathematician Solomon Marcus who in his book *Mathematical Poetics* (1973) suggested a formalization of a theater play undertaken by an “unusual spectator,” one who is only capable to observe the entrances and exits of the actors and monitor their co-occurrences on stage without listening to what they say. In the digital age, it is very simple

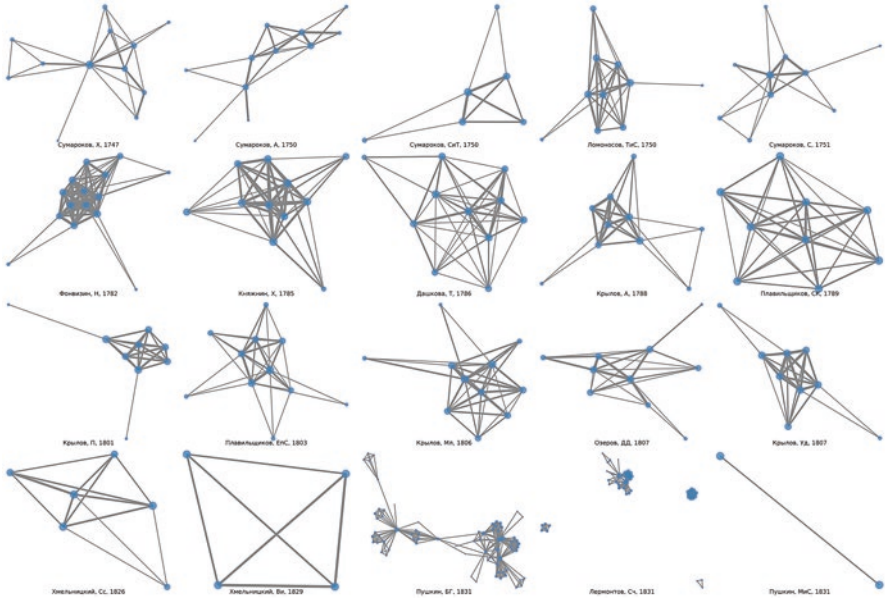


Fig. 29.2 Extracted social networks of 20 Russian plays. Excerpt (left-upper corner) from a larger poster displaying 144 plays in chronological order (1747–1940s). Version in full resolution: <https://doi.org/10.6084/m9.figshare.12058179>

Table 29.1 Number of co-occurrences of characters in A. Ostrovsky’s *Groza* (abbreviated)

Source	Target	Weight
Varvara	Katerina	12
Kabanova	Kabanov	10
Kabanov	Katerina	7
Boris	Varvara	6
Kabanov	Varvara	5
Kudrâš	Boris	5
Kabanova	Varvara	5
Kabanova	Katerina	5
...

to operationalize such formalization on a large scale, which is why we could rename the concept and call this method “the digital spectator.” Put in action, the digital spectator extracts the co-occurrences of speaking characters. Let us take Ostrovsky’s play “*Groza*” (“The Storm,” 1859) as an example, one of the pivotal Russian plays of the nineteenth century, which caused a scandal with its clear implication of adultery. The number of co-occurrences between characters looks as shown in Table 29.1.

This (abbreviated) table simply collects the number of co-occurrences between all characters of the play (in the “Weight” column). The table headers

“Source” and “Target” are interchangeable in our example since we are not collecting the direction of information flows.

This is already everything we need for a network analysis of “Groza.” A visualization of this simple formalization is shown in Fig. 29.3. It comprises all characters of the play (including side characters of acts 4 and 5 lacking proper names), and we can clearly see the core of the network, the Kabanov family with mother (Kabanova) and daughter (Varvara), son (Kabanov) plus wife (Katerina). Without involving one line of the actual text, we arguably found the protagonists of the play just by looking at their position in the network. It is important to note that the “epistemic thing” of our analysis is different than that of traditional textual analyses of literary texts (Trilcke and Fischer 2018). We are not analyzing the actual text of the play, but a strict formalization of it. There, it cannot hurt to stress once more that a formal approach like network analysis does not set out to replace more traditional approaches, but to complement them.

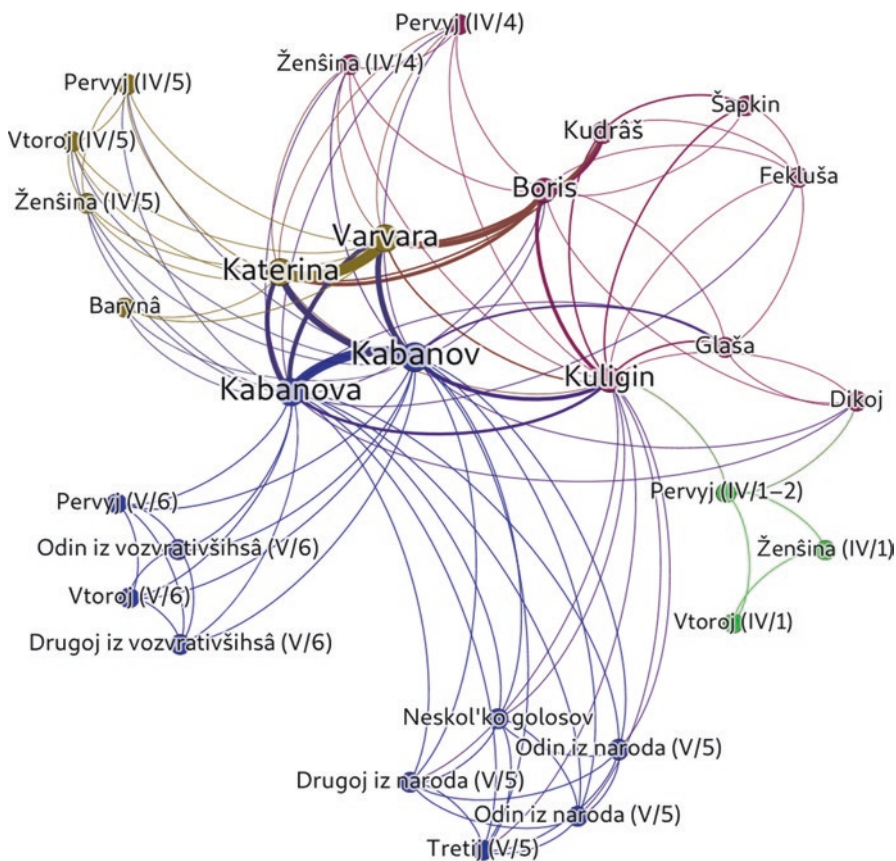


Fig. 29.3 Network graph for Ostrovsky’s *Groza*

Since the formalization step is so crucial, we developed an easy-entry tool to acquaint literary scholars with the problem and enable them to extract literary network data by hand. The tool Easy Linavis (ezlinavis)—an abbreviation for “Easy Literary Network Visualisation”—is available at <https://ezlinavis.dracor.org/>. The network data is generated live while entering speaking entities scene by scene:

```
# Act I
## Scene I
Kuligin
Kudrăș
Șapkin
## Scene II
Dikoj
Boris
## Scene III
Kuligin
Boris
Kudrăș
Șapkin
Feklușa
...
```

As its output, ezlinavis generates a CSV file which can be downloaded and opened in a network analysis tool like the aforementioned Gephi.

Our take on formalizing character interaction has some advantages (it can be easily automatized and, thus, scaled up), but also some shortcomings. It is important to not forget the rationale behind a formalization and to be consistent after a formalization method has been established. Following our operationalization, characters that do not speak are invisible to our “digital spectator.” For example, the blind old man playing the violin in the first scene of Pushkin’s “Mozart and Salieri” (1831) does not raise his voice, so he doesn’t appear in our formalization (in an admittedly not very interesting network with only two characters, i.e., Mozart and Salieri). While we might lose some information and dimensions of the literary piece, we accept this limitation in order to gain something, namely scale. By being able to automatize the extraction of character relations, we can look at a larger number of texts and distill patterns that would otherwise remain invisible.

Since we already introduced Gephi as one of the most popular tools for analysis, we should take the opportunity to mention alternative software. Other Graphical User Interface (GUI)-driven programs like Pajek, Cytoscape, and NodeXL are complemented by the two programming libraries NetworkX and igraph, which are usually used from within higher programming languages such as Python or R. These libraries have in common that most of the established network algorithms are already implemented and well documented so that they can directly be put to use.

29.3.2 *Graph-Related Measures*

From the abundance of graph-related measures that can be used to describe the properties of a network, we want to introduce six basic ones:

- *Network size*: The number of nodes of a network; in our case, the number of (speaking) characters in a play.
- *Network diameter*: The highest value among all shortest distances between two nodes. For example, the shortest distance between two directly connected nodes is 1. If node A is connected to nodes B and C, but B and C are not directly connected, then the shortest distance between them (through node A) is 2, and so forth.
- *Network density*: A value between 0 and 1 indicating the ratio between all realized to all possible connections. In average, comedies are denser than tragedies (one reason for this is that, at the end of comedies, the majority of the cast often appears on stage to witness the resolution of the comic conflict, thereby establishing relationships between characters that are reflected in a higher network density).
- *Clustering coefficient*: Another value between 0 and 1, measuring the number of transitive relations: if node A is related to node B and B is related to node C, then A is also related to C. The value is determined by the number of such closed triplets over the total number of triplets.
- *Average path length*: For each pair of nodes in a connected network, there is a shortest path length. The average path length is thus the average of all shortest path lengths.
- *Maximum degree*: The degree is the number of relations of a node to other nodes. The maximum degree shows the character with the most relations (i.e., the plurality of interactions), thus playing a central role in the play.

29.3.3 *Node-Related Measures*

Graph-related measures are complemented by node-related ones, which allow us to zoom in on single networks and talk about individual nodes. There are literally hundreds of node-related measures, among which are these three basic ones:

- *Degree*: Like stated above, the degree is the number of relations of a node to other nodes.
- *PageRank*: A recursive algorithm, different from degree insofar that it counts not only the number of relations to other nodes, but also depends on the PageRank of these other nodes. According to PageRank, the importance of a node depends on the importance of other nodes linking to it.

- *Betweenness centrality*: A measure of centrality in a graph based on shortest paths. The betweenness centrality of a node does not value the mere number of direct connections to other nodes, but the number of shortest paths between other nodes leading through that node.

Now that we have introduced some basic terminology and measures, let us look at the network properties of some literary works.

29.4 USE CASES

29.4.1 Drama

Graph-related values for five selected plays from our Russian Drama Corpus look as shown in Table 29.2.

Just by looking at the network metrics, it becomes apparent how much the two plays by Sumarokov and Pushkin differ structurally, although they are basically revolving around the same storyline (the story of False Dimitrij during the Time of Troubles around 1600). A diameter of 6 and a network size of 79 shows how Pushkin stretches the plot in a very Shakespearean manner. This strong influence is confirmed by a letter that Pushkin wrote (in French) to his friend Raevsky, dating from July, 1825, around the time he finished “Boris Godunov” (spelling follows the original):

mais quel homme que ce Schakespeare! je n’en reviens pas. ... Voyez Schakespeare. Lisez Schakespeare (now what a man is this Shakespeare! I can’t believe it ... Look at Shakespeare. Read Shakespeare). Pushkin (1962, 178)

Table 29.2 Graph-related values for five selected Russian plays

<i>Play</i>	<i>Network size</i>	<i>Network diameter</i>	<i>Network density</i>	<i>Clustering coefficient</i>	<i>Average path length</i>	<i>Maximum degree</i>
Sumarokov: <i>Dimitrij Samozvanec</i> (<i>Dimitrij the Impostor</i> , 1771)	6	2	0.73	0.77	1.27	5 (Dimitrij)
Pushkin: <i>Boris Godunov</i> (1825)	79	6	0.11	0.89	3.03	29 (Boris)
Griboedov: <i>Gore ot uma</i> (<i>Woe from Wit</i> , 1825)	31	3	0.44	0.8	1.57	23 (Čackij)
Gogol’: <i>Revizor</i> (<i>The Government Inspector</i> , 1836)	31	3	0.49	0.82	1.52	26 (Gorodnichij)
Ostrovsky: <i>Groza</i> (<i>The Storm</i> , 1859)	29	4	0.28	0.83	1.93	23 (Kabanova)

The revolutionary change that Pushkin brought to Russian drama can be shown when put into context. Figure 29.4 shows the network sizes of 144 Russian plays in chronological order. Until 1825, the network size of plays stays well below 25, but with Pushkin’s “Boris Godunov,” the network size suddenly explodes: 79 speaking entities are counted, and the diagram also shows that after Pushkin there is a broader variety of different network sizes, a changed landscape of how character networks are crafted in Russian drama after Pushkin’s initial ignition.

Without trying to overinterpret these very basic metrics, it is interesting to note that Pushkin’s play exhibits the lowest density of all plays present in the table above, but at the same time shows the highest clustering coefficient. A comparatively high clustering coefficient in a larger network with several distinguishable communities means that the individual nodes of these communities are tightly connected among each other, a property known from real-world networks, which also applies to “Boris Godunov” (cf. Fig. 29.5 below). Such real-life social networks have been called “small worlds,” building on the idea that every citizen of the world knows every other citizen over only six edges.

After looking at entire networks, let us throw a glance at node-related values and how we can use them to study literary characters. Distance and centrality measures can be used to describe and interpret the position of a node in the network. It has been suggested to use the average distance as an indicator for

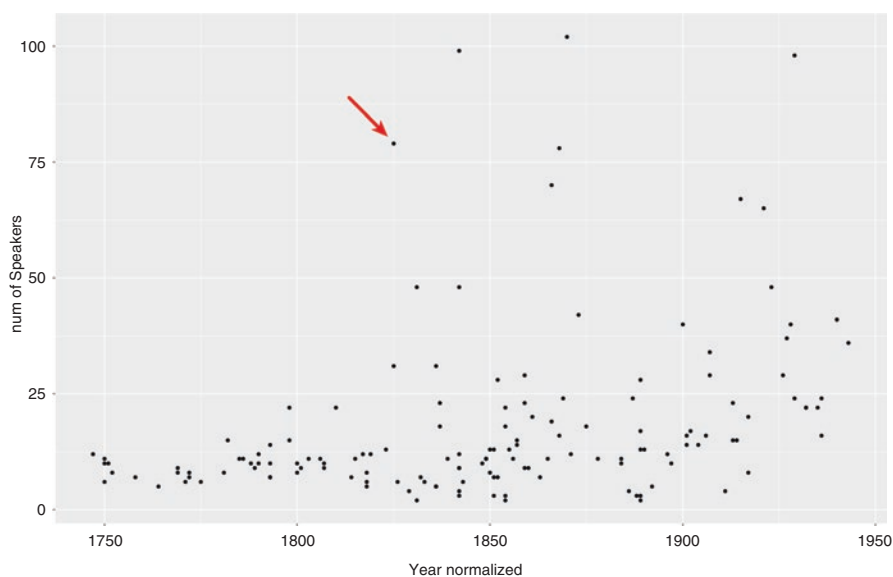


Fig. 29.4 Network sizes of 144 Russian plays in chronological order, x-axis: (normalized) year of publication, y-axis: number of speaking entities per play. Arrow indicates Pushkin’s “Boris Godunov.” Russian Drama Corpus (<https://dracor.org/rus>)

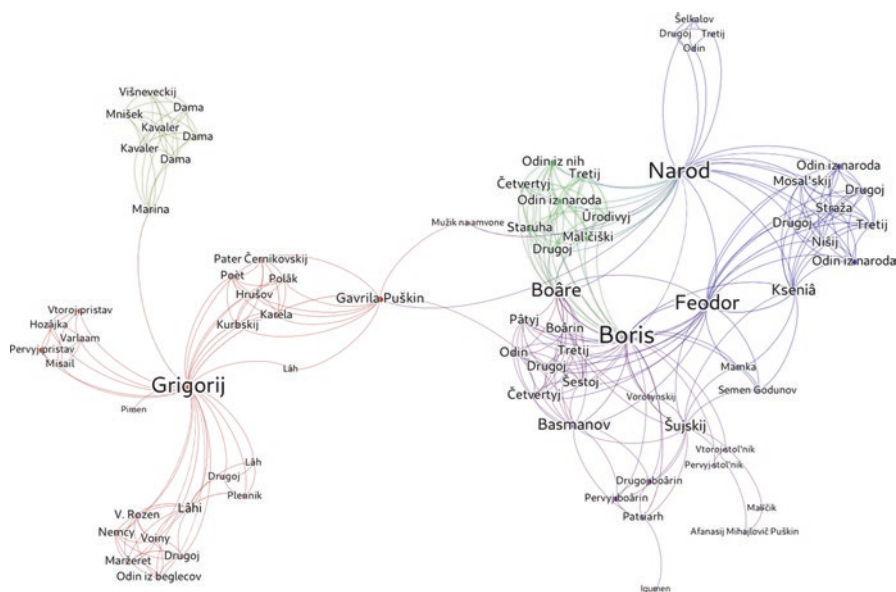


Fig. 29.5 Network visualization of A. Pushkin's *Boris Godunov*. Russian Drama Corpus (<https://dracor.org/rus>)

detecting the protagonist of a play. The character minimizing the distance to all other characters should be a candidate, Moretti argues in his above-mentioned paper. In his formalization of “Hamlet,” Hamlet has an average distance (from all other characters) of 1.45, while the average distance of Claudius is 1.62 and that of Horatio 1.69. Recent research has shown that it is not very fruitful to suggest such simple measures for very complex concepts such as “protagonist.” Instead, multidimensional approaches to describe character types have been proposed since (Algee-Hewitt 2017; Fischer et al. 2018).

Truth be told, literary networks are usually small compared to real-life social networks. Analyzing a single network of two nodes (like in the “Mozart and Salieri” example mentioned above), or even five or ten nodes, is close to being pointless. However, analyzing bigger plays can be insightful, which we demonstrate once more with Pushkin's “Boris Godunov.” This example also serves as demonstration as to how to combine a visual and a numeric analysis. To address the former, let us look at a Gephi visualization of Pushkin's play (Fig. 29.5).

We easily recognize two larger clusters on the left and right side: the forces assembled by False Dimitrij to the left, and the broader Muscovite community around the tsar, Boris Godunov, to the right. Visualizations like this make use of the so-called spring-embedding algorithms which try to assemble nodes and edges in a way that makes it easy to identify larger structures (in our case, we used “Force Atlas 2,” which comes built-in with Gephi). Next to the two major opposing parties, our attention is drawn to the one and only character that

connects the two larger clusters, Gavriła Puškin. While his degree is quite low, he occupies a strategically important position. He, in fact, acts as a messenger and mediator. He is sent from Poland to Moscow to convey to Boris the terms of False Dimitrij and later convinces Boris's military chief Basmanov to change sides, which eventually helps Dimitrij win the throne. Gavriła Puškin, as a follower of Dimitrij, also announces the decrees of the new tsar to the People ("Narod"), thus becoming the only character connecting all larger clusters of the network.

A visual interpretation of this play may be fruitful already, but pinning interpretations on actual numbers adds more precision, so let us come back to the node-related metrics. We chose five characters of the play and listed some network-analytical values in the table below, contrasted by the number of words uttered by these characters (Table 29.3).

A network-based interpretation would first ascertain that Boris has connections to more characters than his opponent Grigorij. At second glance, his position is weaker, since Grigorij is connected to more nodes completely dependent on him, strengthening his position for the eventual usurpation. And last but not least, Gavriła Puškin. Like seen above, he does not excel in the mere number of connections, but he is the bottleneck through which the important information has to pass, yielding in a very high value for betweenness centrality. We can assume that the crucial role of a side character like Gavriła Puškin is not accidental. The idea that Pushkin's noble ancestors played an active part in Russian history can be pursued up to poems like "Moã rodoslovnaã" (1830).

The fact that some of the above metrics contradict each other again strengthens the importance of a multidimensional approach when it comes to the quantitative analysis of characters and character types in literary texts.

29.4.2 Novels

The social network analysis of novels has developed a tad slower. Unlike in the case of drama, there are usually no speaker names in front of a speech act, which is why the automated extraction of communication networks is far more

Table 29.3 Selected network metrics for five characters in A. Pushkin's *Boris Godunov*

<i>Character</i>	<i>Number of spoken words (without stage directions)</i>	<i>Degree</i>	<i>PageRank</i>	<i>Betweenness centrality</i>
Boris	1660	29	0.038	405
Grigorij	1967	26	0.044	1501
Basmanov	303	15	0.020	629
Šujskij	770	13	0.021	197
Gavriła Puškin	424	12	0.018	1482

complicated here. The simpler approaches rely on named-entity recognition to extract character names before choosing a text window to relate characters to each other. This can happen on sentence, paragraph or chapter level and yields very different results, depending on the method chosen. Since characters are often mentioned indirectly via pronouns or other referring expressions, a lot of work has to go into coreference resolution. But despite the more difficult task, the network analysis of novels has yielded first promising results (Grayson et al. 2016; Jannidis 2017).

We made our own little foray into the network analysis of novels—with Leo Tolstoy’s *War and Peace*. With help of named-entity recognition tools and a pinch of manual markup, we identified character mentions throughout the novel. Though by no means comprehensive, our markup contains 25,600 unambiguously identified appearances of individual characters across the text of *War and Peace*. We used the markup to automatically extract the social network of the novel. Each time two characters were mentioned within one sentence, they were assumed to be interacting in some way. Figure 29.6 contains the visualization of the resulting network of 119 nodes.

Let us turn to numbers and compare character centralities using the multi-dimensional approach described above. The table below ranks the most central characters according to three different centrality measures (Table 29.4).

Overall, Pierre seems to be the most central character—hardly a surprise to anyone familiar with the novel. Differences between centrality measures are also telling. Betweenness centrality obviously assigns more importance to the historical/military characters. If we examine the military subnetwork of Tolstoy’s novel, we can see that it is less dense—and more hierarchical. Political and military figures in the novel do not have as much interaction as the main nonhistorical characters of *War and Peace*, who are constantly thrown into all sorts of social groups and circumstances. But when Kutuzov or Napoleon or Aleksandr I do get involved, they mostly interact with their inferiors (marshals, generals), who in turn convey the message down the command chain. Some actual examples from the novel include the scene in which Kutuzov, Russian commander-in-chief, talks to a regimental commander (interaction), who in turn talks extensively to his subordinate battalion commander (interaction). Yet, there is no direct conversation between Kutuzov and the battalion commander. The reader hardly ever notices this fact, but the structure of the network seems to highlight this setting-dependent difference in communication patterns.

Whether Tolstoy, himself a retired artillery officer with war experience, purposefully attempted to create an opposition of “War interaction” versus “Peace interaction” in his novel, remains an open question. But the difference in the social network structure in *War and Peace* clearly correlates with the settings. To show this, we produced separate networks for each of the 15 books (parts of volumes in the canonical Russian four-volume edition) and the epilogue of *War and Peace*. Figures 29.7, 29.8, and 29.9 present three sample networks for

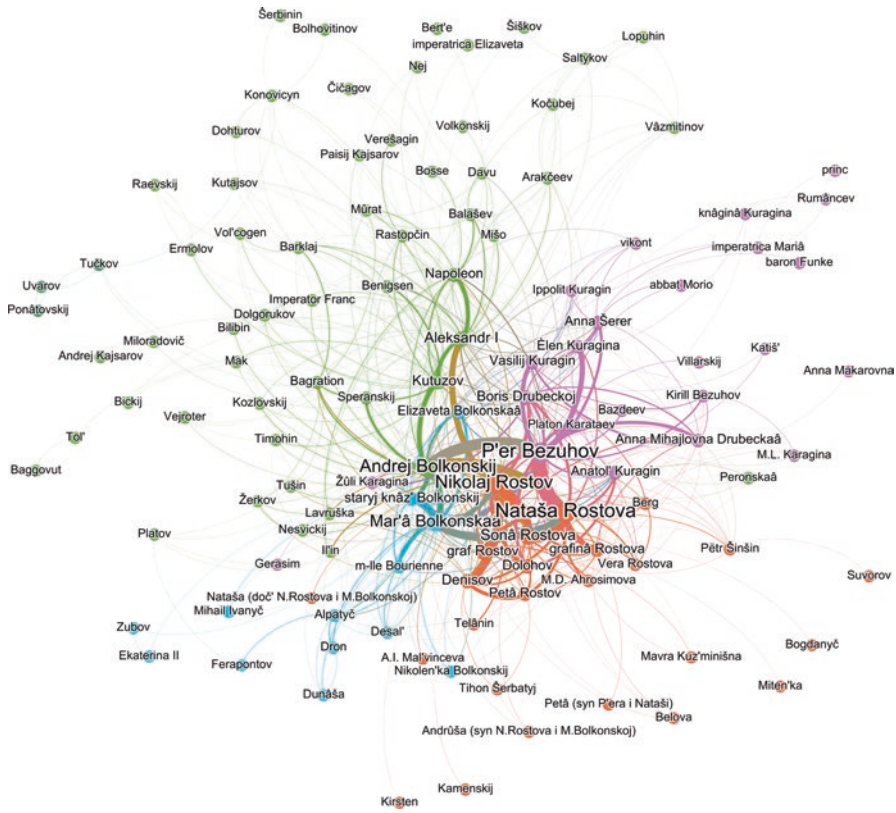


Fig. 29.6 Network visualization of L. Tolstoy’s *War and Peace*

Table 29.4 Central characters in L. Tolstoy’s *War and Peace* ranked according to three different centrality measures

<i>Degree (weighted)</i>	<i>PageRank</i>	<i>Betweenness centrality</i>
P'er Bezuhov	P'er Bezuhov	Kutuzov
Nataša Rostova	Nataša Rostova	P'er Bezuhov
Nikolaj Rostov	Nikolaj Rostov	Aleksandr I
Andrej Bolkonskij	Andrej Bolkonskij	Napoleon
Mar'ja Bolkonskaá	Mar'ja Bolkonskaá	Andrej Bolkonskij
Soná Rostova	Aleksandr I	Nikolaj Rostov
Aleksandr I	Kutuzov	graf Rostov
Kutuzov	Soná Rostova	staryj knáz' Bolkonskij
Denisov	Napoleon	Nataša Rostova
Boris Drubeckoj	Denisov	Vasilij Kuragin

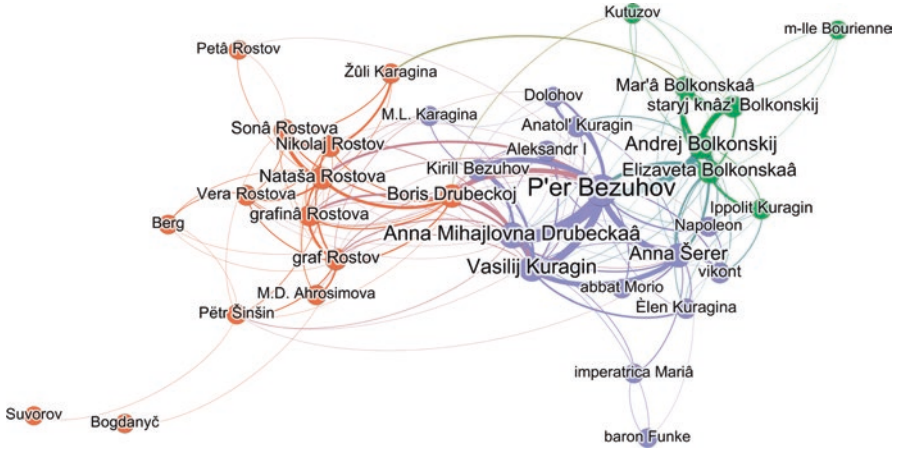


Fig. 29.7 Network visualization of L. Tolstoy's *War and Peace*, book 1 (first part of the first volume)



Fig. 29.8 Network visualization of L. Tolstoy's *War and Peace*, book 10 (second part of the third volume)

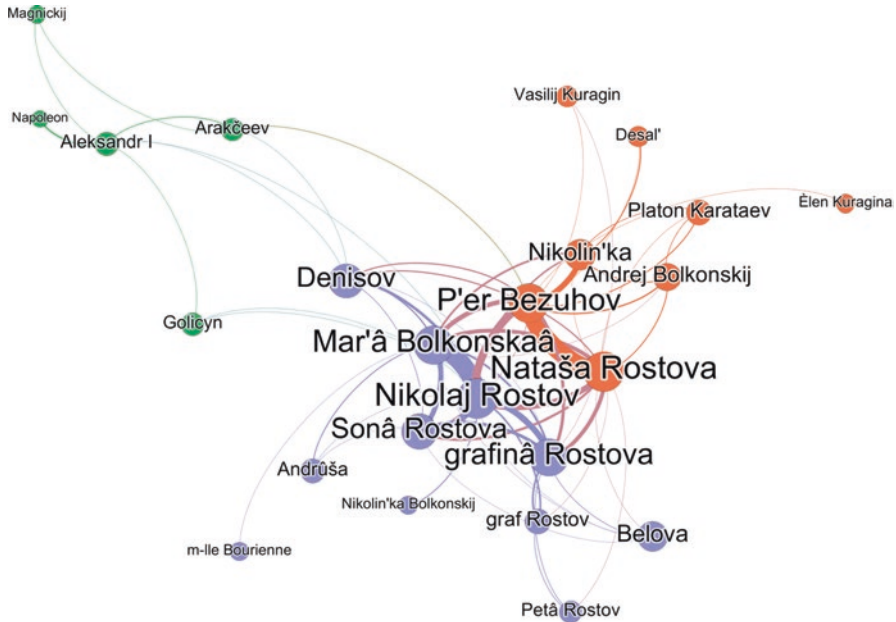


Fig. 29.9 Network visualization of L. Tolstoy's *War and Peace*, epilogue

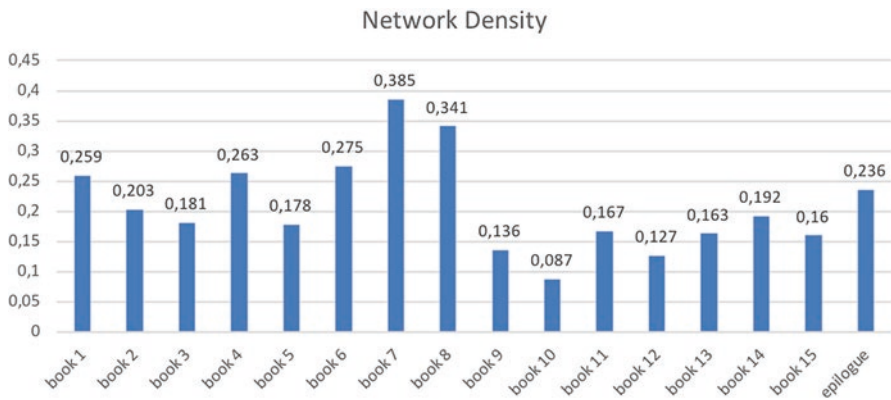


Fig. 29.10 Network densities of separate books (parts) of *War and Peace*

separate books: book 1, starting the novel; book 10, in which the Borodino battle occurs; and the epilogue that wraps up the novel.

The network in Fig. 29.8 represents Book 10 (second part of the third volume). This is one of the most battle-torn parts of *War and Peace*, as it describes the preparation and events of the Borodino battle. This network exhibits the lowest density in the whole novel—one could speculate that war and military settings disrupt human interaction.

Figure 29.10 shows the density dynamics throughout the whole novel, which can be interpreted in terms of war/peace cycles of *War and Peace*. The novel begins in book 1 with peaceful events and higher-than-average density of the character network. This is interrupted by the war of the third coalition, ending with the disastrous Austerlitz battle (books 2 and 3)—and lower-than-average density. Book 4 brings us back to the peaceful life by describing the life of the Rostov family with Nikolai Rostov on vacation from his regiment. In book 5, Nikolai, having lost a small fortune in a card game, goes back to service, the war gains momentum, Pierre breaks up with his wife completely and goes on his spiritual search—peaceful life is disrupted everywhere, and network density drops along with it. However, this time the war ends quickly in book 6 with the Treaties of Tilsit, Prince Andrej falls in love with Nataša—and the reader enters the high-density zone of peaceful life again. Book 7, the densest of all in the novel, describes the idyllic life of the Rostov family in their Otradnoe estate. The events of book 8 take place in Moscow, and this is where peace ends with Anatol's attempt to steal Nataša away. Next comes book 9—Napoleon invades Russia, not only disrupting peace, but also the social network of the novel. Then comes the Borodino battle—the watermark moment of the whole novel, and the lowest density point. The war and sorrows continue, and the density remains below the average until the very end. Only in the epilogue, which wraps up the events of the novel proper, the network density reaches the same above-average value that it had at the beginning of *War and Peace*.

29.5 CONCLUSION

The network analysis of literary texts has developed into a prolific subdiscipline of the Digital Humanities, a formal approach revealing hitherto invisible structures and structural changes in literary history.

The extraction and formalization of network-analytical data is the first step to gaining workable network data. It can be done manually or automatically, depending on the scale of the research question and the data available. Following data formalization, the visualization step oftentimes is a first indicator for the quality of the extracted network data. A visualization can be used for interpretation, too, but the real power of network analysis resides in the underlying numbers and available algorithms as we have demonstrated with a few examples in this chapter.

Further development will depend on whether it will be possible to establish versatile and stable infrastructures for the general digital analysis of literary texts, based on reliable text corpora and technical interfaces, like Application Programming Interfaces (APIs) or other endpoints that make it easier to access structural data. The DraCor platform (<https://dracor.org/>) is one such attempt addressing the digital research on drama (Fischer et al. 2019). By offering an interface for TEI-encoded drama corpora, it can open a comparative angle to the digital literary studies, and also help to position Russian drama within the context of other national literatures. A glance at the richness of

existing TEI-encoded drama corpora will help to understand these opportunities:

- Théâtre Classique: 1290 French plays from seventeenth and eighteenth centuries
- Shakespeare His Contemporaries: 853 English plays written between 1550 and 1700
- German Drama Corpus: 474 German-language plays from between 1730 and 1930
- Russian Drama Corpus: 144 Russian plays published between 1747 and the 1940s
- Letteratura teatrale nella Biblioteca italiana: 139 Italian plays
- Dramawebben: 68 Swedish plays
- Shakespeare Folger Library: all (37) Shakespeare plays
- Ludvig Holbergs skrifter: 36 comedies
- Biblioteca Electrónica Textual del Teatro en Español de 1868–1936 (BETTE): 25 Spanish plays
- Emothe: The Classics of Early Modern European Theatre: 113 plays including translations (Italian, English, French, Spanish)

Since all these corpora are encoded in TEI, they are comparable, although being written in different languages and stemming from different epochs. The comparative aspect is well within reach and complements similar efforts in the field of the analysis of the European novel (Schöch et al. 2018).

Beyond the added methodology for the study of literary texts, the knowledge of network metrics also sharpens the senses for the functions of other kinds of networks we are surrounded by in everyday life, be they online communities, metro lines, or highways. They are all based on the same assumptions and can be examined and understood using the same methods. The successful import of network analysis into the humanities thus leads to a broader understanding of realities beyond one's own discipline and to new opportunities for interdisciplinary cooperation.

REFERENCES

- Algee-Hewitt, Mark. 2017. Distributed Character: Quantitative Models of the English Stage, 1550–1900. *New Literary History* 48 (4): 751–782.
- Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *International AAAI Conference on Weblogs and Social Media*.
- Burnard, Lou. 2014. The TEI and XML. In *What is the Text Encoding Initiative? How to Add Intelligent Markup to Digital Resources*. Marseille: OpenEdition Press. <http://books.openedition.org/oep/680>.
- Fischer, Frank, Christopher Kittel, Peer Trilcke, Mathias Göbel, Andreas Vogel, Hanna-Lena Meiners, and Dario Kampkaspar. 2016. Distant-Reading Showcase. 200 Years

- of German-Language Drama at a Glance. In *Conference Proceedings of DHd 2016*, Leipzig. <https://doi.org/10.6084/m9.figshare.3101203.v2>.
- Fischer, Frank, Peer Trilcke, Christopher Kittel, Carsten Milling, and Daniil Skorinkin. 2018. To Catch a Protagonist: Quantitative Dominance Relations in German-Language Drama (1730–1930). In *Conference Proceedings of DH 2018*, 193–200. Ciudad de México. <https://dh2018.adho.org/?p=9362>.
- Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtel, Christopher Kittel, Carsten Milling, and Peer Trilcke. 2019. Programmable Corpora—Die digitale Literaturwissenschaft zwischen Forschung und Infrastruktur am Beispiel von DraCor. In *Conference Proceedings of DHd 2019*, 194–197. Mainz and Frankfurt. <https://doi.org/10.5281/zenodo.2596094>.
- Freeman, Linton C. 2011. The Development of Social Network Analysis—with an Emphasis on Recent Events. In *The SAGE Handbook of Social Network Analysis*, ed. John Scott and Peter J. Carrington, 26–39. London: SAGE Publications.
- Grayson, Siobhán, Karen Wade, Gerardine Meaney, and Derek Greene. 2016. The Sense and Sensibility of Different Sliding Windows in Constructing Co-Occurrence Networks from Literature. In *2nd IFIP International Workshop on Computational History and Data-Driven Humanities*. Trinity College: Dublin.
- Jannidis, Fotis. 2017. Netzwerke. In *Digital Humanities: Eine Einführung*, ed. Fotis Jannidis, Hubertus Kohle, and Malte Rehbein, 147–161. Stuttgart: Metzler.
- Knuth, Donald. 1993. *The Stanford Graph Base: A Platform for Combinatorial Computing*. ACM Press.
- Marcus, Solomon. 1973. *Mathematische Poetik*, 287–370. Frankfurt/M: Athenäum Verlag.
- McCulloh, Ian, Helen Armstrong, and Anthony Johnson. 2013. *Social Network Analysis with Applications*. New York, NY: Wiley.
- Moretti, Franco. 2011. *Network Theory, Plot Analysis*. Stanford Literary Lab Pamphlets 2. <http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf>.
- Newman, Mark. 2018. *Networks*. 2nd ed. Oxford University Press.
- Pushkin, Aleksandr. 1962. *Sobranie sočinenij v 10 tomah* [Collected Works in 10 Volumes]. Vol. 9: Pis'ma 1815–1830 [Letters 1815–1830]. Moscow 1959–1962.
- Schöch, Christof, et al. 2018. Distant Reading for European Literary History. A COST Action. In *Proceedings of DH2018*, Ciudad de México. <https://dh2018.adho.org/en/?p=11345>.
- Schweizer, Thomas, and Michael Schnegg. 1998. Die soziale Struktur der ‘Simple Stories’: Eine Netzwerkanalyse. Accessed April 30, 2019. <https://www.ethnologic.uni-hamburg.de/pdfs-de/michael-schnegg/simple-stories-publication-michael-schnegg.pdf>.
- Scott, John. 2000. *Social Network Analysis: A Handbook*. Thousand Oaks, CA: Sage Publications.
- Skorinkin, Daniil, Frank Fischer, and German Palchikov. 2018. Building a Corpus for the Quantitative Research of Russian Drama: Composition, Structure, Case Studies. In *Proceedings of the International Conference “Dialogue 2018,”* 662–682. <http://www.dialog-21.ru/media/4332/skorinkind.pdf>.
- Stiller, James, and Matthew Hudson. 2005. Weak Links and Scene Cliques Within the Small World of Shakespeare. *Journal of Cultural and Evolutionary Psychology* 3 (1): 57–73.
- Stiller, James, Daniel Nettle, and Robin I.M. Dunbar. 2003. The Small World of Shakespeare’s Plays. *Human Nature* 14 (4): 397–408.

- Trilcke, Peer. 2013. Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft. In *Empirie in der Literaturwissenschaft*, ed. Philip Ajouri, Katja Mellmann, and Christoph Rauen, 201–247. Münster: Paderborn mentis Verlag.
- Trilcke, Peer, and Frank Fischer. 2018. Literaturwissenschaft als Hackathon: Zur Praxeologie der Digital Literary Studies und ihren epistemischen Dingen. In *Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden*, ed. Martin Huber, and Sybille Krämer. https://doi.org/10.17175/sb003_003.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

