‘Literary studies must be a science’ (Yury Lotman):

Representing and Interpreting Character Relationships in Russian Drama by Means of Quantitative Methods

Irina Pavlova

Wolfson College

Submission for the Degree of Doctor of Philosophy

Medieval and Modern Languages

Hilary Term, 2022

Table of Contents

Short abstract 11

Long abstract 13

Introduction 21

Goals 24

Materials 25

Methodology background 27

Methods 30

A case-study of the quantitative representation of 'key protagonists' in three Russian plays 40

Conclusion 67

A comparative analysis of *A Month in the Country* (Turgenev) and *Uncle Vanya* (Chekhov) 69

Basic network analysis 75

'Configuration change' mark-up 82

Network analysis of *A Month in the Country* 88

Network graph and centralities 88

Edges weight and 'dialogues' 97

Network analysis of *Uncle Vanya* 104

Network graph and centralities 104

Edges weight and 'dialogues' 115

Comparison and conclusions 122

Women in Russian Drama 127

Ostrovsky and women 138

The Storm 142

Without a Dowry 177

Conclusion 196

New Drama 197

The later plays in the RusDraCor 197

Within and beyond Socialist Realism 199

Perestroika 215

The 21st Century 223

Documentary Theatre 235

Fiction and mockumentary plays 243

Conclusion 247

Conclusion 249

Bibliography 256

Table of Figures

Figure 1. Example of TEI encoding....26

Figure 2. Example of TEI encoding....27

Figure 3. Network graph of *Much Ado About Nothing* by Shakespeare....32

Figure 4. A network graph of *Boris Godunov* by Pushkin....43

Figure 5. The network graph of *The Government Inspector* by Gogol....52

Figure 6. Network graph of *Masquerade* by Lermontov....60

Figure 7. Love triangles in *A Month in the Country*....70

Figure 8. Love triangles in *Uncle Vanya*....70

Figure 9. Network graph of *A Month in the Country*....76

Figure 10. Network graph of *Uncle Vanya*....79

Figure 11. Configuration change examples in *A Month in the Country*....83

Figure 12. Configuration change examples in *A Month in the Country*....83

Figure 13. Extended version of the example in Figure 11....84

Figure 14. Problematic case for configuration change mark-up in *A Month in the Country*...85

Figure 15. Problematic case for configuration change mark-up in *A Month in the Country*...87

Figure 16. Network graph of *A Month in the Country* based on configuration changes....88

Figure 17. Network graph of *Uncle Vanya* based on configuration changes....105

Figure 18. Network graph of *Uncle Vanya* based on configuration changes, without the workman and the watchman....109

Figure 19. Network graph of *Uncle Vanya* based on configuration changes, without the workman, the watchman's voice, Elena Andreevna and Marina....111

Figure 20. Number of plays written by the authors of the 19th-century subcorpus....137

Figure 21. Average percentage of female characters among the authors of the 19th-century subcorpus....138

Figure 22. Network of *The Storm* with all characters 146

Figure 23. Network of *The Storm* with 12 'core characters' 146

Figure 24. Network graph of *Without a Dowry* 180

Figure 25. Extract from the first act of *Farewell in June* 202

Figure 26. Extract from the first act of *Farewell in June* 202

Figure 27. Extract from the first act of *Farewell in June* 202

Figure 28. Network graph of *Farewell in June* by Vampilov 204

Figure 29. Network graph of *The Elder Son* by Vampilov 205

Figure 30. Network graph of *Duck Hunting* by Vampilov 207

Figure 31. Extract from *Last Summer in Chulimsk* 208

Figure 32. Extract from *Last Summer in Chulimsk* 208

Figure 33. Network graph of *Last Summer in Chulimsk* by Vampilov 209

Figure 34. Network graph of *Three Girls in Blue* by Petrushevskaya 211

Figure 35. Network graph of *The Moscow Choir* by Petrushevskaya 213

Figure 36. Network graph of *Cinzano* by Petrushevskaya 214

Figure 37. Network graph of *The Isolation Ward* by Petrushevskaya 215

Figure 38. Network graph of *Vladimir Square* by Razumovskaya 217

Figure 39. Network graph of *The Life of Yura Kurochkin* by Razumovskaya 219

Figure 40. Network graph of *Without a Dowry* by Razumovskaya 220

Figure 41. Network graph of *Halloween for Russians* by Razumovskaya 222

Figure 42. Network graph of *Terrorism* by the Presnyakov brothers, version 1 225

Figure 43. Network graph of *Terrorism* by the Presnyakov brothers, version 2 226

Figure 44. Network graph of *Playing the Victim* by the Presnyakov brothers, version 1 228

Figure 45. Network graph of *Playing the Victim* by the Presnyakov brothers, version 2 229

Figure 46. Network graph of *Oxygen* by Vyrypaev, version 1 231

Figure 47. Network graph of *Oxygen* by Vyrypaev, version 2 232

Figure 48. Network graph of *Oxygen* by Vyrypaev, version 3 232

Figure 49.Network graph of *July* by Vyrypaev, version 1 233

Figure 50. Network graph of *July* by Vyrypaev, version 2 234

Figure 51. Network graph of *July* by Vyrypaev, version 3 234

Figure 52. Network graph of *My First Man* by Isaeva 236

Figure 53. Network graph of *Doc.Tor* by Isaeva 237

Figure 54. Network graph of *One Hour and 18 Minutes* by Gremina 241

Figure 55. Network graph of *The Two in Your House* by Gremina 243

Figure 56. Network graph of *Pagans* by Yablonskaya 245

Figure 57. Network graph of *A Man from Podol'sk* by Danilov 246

Figure 58. Network graph of *Choose Three* by Danilov 247

Figure 59. 'Superposter' of Russian-language plays 254

Table of Tables

Table 1. Count-based metrics for Lzhedmitry, Tsar Boris and Gavrila Pushkin 46

Table 2. Most frequent words of Lzhedmitry, Tsar Boris and Gavrila Pushkin 48

Table 3. Ratio of 1st-, 2nd- and 2rd-person verbforms in the speech of Lzhedmitry, Tsar Boris and Gavrila Pushkin 49

Table 4. Most frequent lemmas of Lzhedmitry, Tsar Boris and Gavrila Pushkin 50

Table 5. Count-based metrics for Khlestakov and Gorodnichii 54

Table 6. Most frequent words of Khlestakov and Gorodnichii 56

Table 7. Ratio of 1st-, 2nd- and 3rd-person verbs in the speech of Khlestakov and Gorodnichii 56

Table 8. Most frequent lemmas of Khlestakov and Gorodnichii 57

Table 9. Count-based metrics for Arbenin, Knyaz' Zvezdich and Nina 61

Table 10. Most frequent words of Arbenin, Knyaz' Zvezdich and Nina 62

Table 11. Ratio of 1st-, 2nd and 3rd-person verbs in the speech of Arbenin, Knyaz' Zvezdich and Nina 63

Table 12. Most frequent lemmas of Arbenin, Knyaz' Zvezdich and Nina 63

Table 13. Weighted degree of characters in *A Month in the Country* 78

Table 14. Degree of characters in *A Month in the Country*, based on configuration changes mark-up 90

Table 15. Betweenness centrality of characters in *A Month in the Country*, based on configuration changes mark-up 90

Table 16. Closeness centrality of characters in *A Month in the Country*, based on configuration changes mark-up 91

Table 17. Eigenvector centrality of characters in *A Month in the Country*, based on configuration changes mark-up 92

Table 18. Weighted degree of characters in *A Month in the Country*, based on configuration changes mark-up 92
Table 19. Weight of edges between characters in *A Month in the Country*, based on configuration changes mark-up.....99
Table 20. Weight of edges between characters in *A Month in the Country*, based on configuration changes mark-up, top 14 pairs.....100
Table 21. Dialogue-based connection between characters in *A Month in the Country* 103
Table 22. Degree of characters in *Uncle Vanya*, based on configuration changes mark-up .106
Table 23. Betweenness centrality of characters in *Uncle Vanya*, based on configuration changes mark-up 107
Table 24. Closeness centrality of characters in *Uncle Vanya*, based on configuration changes mark-up 107
Table 25. Eigenvector centrality of characters in *Uncle Vanya*, based on configuration changes mark-up 108
Table 26. Centrality values of characters in *Uncle Vanya*, based on configuration changes mark-up, without the workman and the watchman's voice 110
Table 27. Weighted degree of characters in *Uncle Vanya*, based on configuration changes mark-up 112
Table 28. Weight of edges between characters in *Uncle Vanya*, based on configuration changes mark-up 116
Table 29. Dialogue-based connection between characters in *Uncle Vanya* 118
Table 30. Number of speech acts 125
Table 31. Number of words 125
Table 32. Plays with a high proportion of group-speech-acts 134
Table 33. Ostrovsky's plays with 50% and more of female characters 141

Table 34. Ostrovsky's plays with 50% and more of female speech-acts 141
Table 35. Ostrovsky's plays with 50% and more of female words.....142
Table 36. Network values of the “core characters” of *The Storm* 147
Table 37. Number of scenes for the ‘core’ characters of *The Storm* 149
Table 38. Number of speech-acts for the ‘core’ characters of *The Storm* 149
Table 39. Number of words for the ‘core’ characters of *The Storm*.....150
Table 40. Number of stage directions per speech-act for the ‘core’ characters of *The Storm* 151
Table 41. Average length of speech-acts of the ‘core’ characters of *The Storm* 153
Table 42. Proportion of exclamation marks in the characters’ speech in *The Storm*.....156
Table 43. Proportion of ellipses in the characters’ speech in *The Storm*.....159
Table 44. Distribution of parts of speech in speech-acts of the ‘core’ female and male characters of *The Storm* 161
Table 45. TTR of the characters’ speech in *The Storm*.....165
Table 46. Lexical diversity measures for characters of *The Storm* with > 50 lemmas. 167
Table 47. MSTTR for characters of *The Storm* with 50 > lemmas.....170
Table 48. Most frequent words of the ‘core’ characters of *The Storm* 172
Table 49. Strength of the dialogue-based connection between the ‘core’ characters of *The Storm*.....175
Table 50. Degree and betweenness centrality in *Without a Dowry*.....180
Table 51. Weighted degree in *Without a Dowry*.....181
Table 52. Number of scenes for the characters of *Without a Dowry*.....182
Table 53. Number of speech-acts for the characters of *Without a Dowry*.....183
Table 54. Number of words for the characters of *Without a Dowry*.....183
Table 55. Average number of words per response for the characters of *Without a Dowry*...184

Table 56. Proportion of stage directions related to the characters of *Without a Dowry*.....185

Table 57. Proportion of exclamation marks in the speech of the characters of *Without a Dowry*.....186

Table 58. Proportion of ellipses in the speech of the characters of *Without a Dowry*.....188

Table 59. Proportion of different parts-of-speech in the speech of the characters of *Without a Dowry*.....190

Table 60. Number of lemmas in the speech-acts of the characters of *Without a Dowry*.....191

Table 61. Lexical diversity measure of the speech of the characters of *Without a Dowry*....192

Table 62. Weight of the edges between the characters of *Without a Dowry*.....194

Table 63. Strength of the dialogue-based connections between the characters of *Without a Dowry*.....195

Short abstract

This thesis explores and demonstrates the possibility of using quantitative and digital methods to analyse Russian drama texts. The methodology under discussion lies within the fields of Digital Humanities and Digital Literary Studies, and includes Social Network Analysis (SNA) and Natural Language Processing (NLP) techniques. The project is structured as several independent case-studies that are united by a discussion of the advantages and limitations of the methodology.

The first chapter of this thesis introduces the reader to the encoding standards of the texts (TEI XML) necessary for carrying out the quantitative analysis of the plays, as well as the basics of the methodology, including definitions and formulas for calculations. These methods are then illustrated in detail by presenting a first case-study, showing how automatic methods can be used to establish the hierarchy of the protagonists in three 19th-century Russian plays.

This thesis aims to reflect the exploratory and innovative nature of the application of digital methods to literary data. Hence, each chapter introduces new methods in addition to the ones discussed in the Introduction. The next case-study is devoted to the comparison of two plays with similar plots but different network structures, and includes a more detailed analysis of the speech of the characters in the two plays.

The third case-study focuses on the representation of women in 19th-century Russian drama and attempts to combine 'distant' and 'close' reading methods, to explore the full capacity of the digital approach. It is followed by a final section discussing the applicability of the methods to contemporary plays from the Soviet period and the beginning of the 21st century, with a greater focus on the limitations of the methodology.

The work is rounded off with a concluding section, in which an overview is provided of the general structural evolution of Russian-language drama, alongside a discussion of the potential for further research in this area.

Long abstract

In recent years approaches based on Digital Humanities have become popular for analysing literary texts. They offer a fruitful opportunity for scholars to perform large-scale analysis in a short amount of time. Besides that, they provide a new angle for looking at familiar texts, which sometimes leads to the reinterpretation of the texts and contributes to more traditional literary scholarship. One of the main principles that lies behind Digital Literary Studies is the quantitative reliability of such analysis, as opposed to the subjective interpretation of literature by critics.

In order to carry out an automatic, semi-automatic and quantitative analysis of literary works, it is necessary to have them encoded in a machine-readable format. This task is the most time-consuming step in any digital analysis. Thus, there are many projects devoted solely to creating collections of texts that are digitised and encoded into a specific format. Because of the time and effort required, such individual projects are concentrated on certain types or categories of texts, be that diary entries, the works of a single author, or drama texts, as discussed in this project.

While it is equally interesting to apply digital methods to different kinds of literary data, drama texts appear to be the most rewarding candidates for such analysis. The structured composition of drama texts makes it more feasible to (semi-)automatically encode the plays in a machine-readable format, and afterwards to select and extract particular elements from a large number of plays.

One of the encoding formats frequently used in Digital Humanities is the TEI XML encoding scheme. It provides users with a unified set of tags to encode a text in any language, which then can be exchanged within the international community for research purposes. There are several projects working on compiling different corpora in TEI, including drama corpora.

The DraCor (Drama Corpora) project started out by offering two large corpora of German and Russian plays. The project currently continues to encode plays in other languages, as well as collaborating with the creators of other TEI drama corpora, to include these into its database. Thus, at the moment there are more than 12 corpora in different languages that are available via the DraCor website and its API (Application Programming Interface). The DraCor API makes it easy for its users to interact with the project's data. For instance, via the API that works with the TEI-encoded plays, it is easy automatically and quickly to extract all the characters in a play, their speech-acts, and the play's stage directions. Besides that, all the data is available for downloading, and for performing independent analysis on it.

For this thesis, I have benefited greatly from the DraCor data and worked primarily with the plays included in the RusDraCor (Russian Drama Corpus) branch of the project. The potential for quantitative analysis of the texts offered by the RusDraCor is striking. Inspired by the research projects of the DraCor team, of which I was a member, I have continued to work with SNA (Social Network Analysis) and NLP (Natural Language Processing) methods to analyse Russian-language plays.

When working with the SNA method, the plays are represented as social network graphs, where nodes represent the characters of the plays and the edges between the nodes represent the connections between the characters. The social network graphs of the plays can be analysed as mathematical objects. Furthermore, deriving from Social Network Theory, the mathematical features of the plays' networks and the nodes of their characters can be interpreted in terms of network density and the influence and importance of its nodes in this network.

The basic assumption of such analysis, as offered by the DraCor team, is that the nodes of two characters are connected if they appear together on stage. Thus, even if the characters do not address each other directly, but are present within one scene, act, or other kind of segment of a play, they are considered to be connected. Such an approach can be criticised as too simplistic,

but it provides the basis for creating larger corpora of texts, where such simplifications are necessary to adopt a unified encoding scheme.

The NLP methods used in this study include lexical and morphological analysis of the characters' speech. For instance, I have discussed the proportion of different parts-of-speech in the speech-acts, the most frequent words and lemmas used by the characters, the proportion of the verb forms, and measures of lexical complexity. In addition, I have considered simple quantitative measures of the speaking activity or presence of characters by calculating the number of speech-acts, words and sentences that the characters utter.

The main research question of this thesis is the extent of the applicability of these methods and the possibility of providing significant contributions to traditional literary studies by analysing drama texts from a quantitative and digital perspective. Despite the ambitious claims of Digital Literary Studies, there is still controversy regarding the necessity and accuracy of quantitative methods in literature. Even though a digital humanities scholar is able automatically to process literary texts and produce some calculations, these then need to be verified and interpreted, which requires expertise in literary studies. In this project I have thus included several case-studies that demonstrate the usage of primarily SNA and NLP analysis for Russian-language plays, and discussed in detail the limitations of the methodology together with possible solutions.

The first chapter of this thesis fills in the background of Digital Humanities methodologies. It covers the connection of Formalist ideas with the quantitative analysis of literature. In addition to that, in the first chapter I introduce the majority of the methods that are used in the work, together with an illustrative case-study. This case-study is devoted to using SNA and linguistic analysis to identify the main protagonists in three plays – *Boris Godunov* by Pushkin, *The Government Inspector* by Gogol and *Masquerade* by Lermontov. The analysis of *Boris Godunov* benefits greatly from the SNA approach. We are able to see that Tsar Boris and

Lzhedmitry are competing as the two main protagonists of the play. However, the SNA metrics also highlight the *narod*'s importance in the play's network. Besides that, the graph of *Boris Godunov* clearly shows the vital role of Gavrila Pushkin in the play, who connects the two clusters (the Tsar Boris group and Lzhedmitry's accomplices).

On the other hand, the insights gained from the SNA performed on *The Government Inspector* and *Masquerade* are not as informative. The reason behind this might be the smaller scale of the plays compared to *Boris Godunov*. But with the SNA and the quantitative and linguistic analysis we are still able to conclude that while structurally Khlestakov in *The Government Inspector* acts as an outsider, he is nevertheless the most active character in the play, and his speech is more self-centred and engaging than the speech of Gorodnichii.

The SNA analysis of *Masquerade* suggests the competition over the network influence between Knyaz Zvezdich and Arbenin. Judging from the network metrics and their speaking activity, Arbenin appears to be more influential than Zvezdich. Nina, being the third most 'important' character of the play, is left far behind.

Building on the results of the introductory case-study and the possibility of highlighting the structural importance of certain characters in the plays, in the second chapter, I explore how the discussed methodology is suitable for the comparison of similar plays. The two plays I discuss are *A Month in the Country* by Turgenev and *Uncle Vanya* by Chekhov. They have an identical *fabula* and their basic network graphs look the same. However, with the help of quantitative methods it is possible to highlight the fundamental differences between the two plays. While in *A Month in the Country* we are able to identify the three most 'important' characters – Rakitin, Shpigelsky and Natalya Petrovna, the network graph of *Uncle Vanya* reflects the homogeneity and decentralisation of the play.

However, in order to arrive at these conclusions it was necessary to modify the original available TEI versions of the two plays. According to the original segmentation of the plays' texts into acts and scenes, *A Month in the Country* and *Uncle Vanya* have only five and four segments respectively, and the network graphs produced from such segmentation are similar and not informative. I therefore implemented the segmentation of the plays into 'configuration changes' according to the stage directions, indicating when someone enters or leaves the stage. This is an example of the 'close reading'-interventions that still need to be performed, despite the availability of computer-readable TEI-encoded texts.

To provide a fuller analysis, in the second chapter I also introduce additional methods. For instance, I discussed the strength of connection between the comparable pairs of characters in *A Month in the Country* and *Uncle Vanya*, both from a network point of view and using quantitative lexical analysis, by measuring the number and length of dialogues between the characters.

In general, this thesis is structured as a journey of exploration of the possibilities of quantitative analysis of literature, where I aim to emphasise the pioneering nature of such work that includes a lot of trial and error. Thus, with every chapter I offer new additional methods for the analysis, and the selection of such methods needs to be tailored to a specific research question and plays. Moreover, the methods that can be useful for answering these research questions are not limited to the ones discussed in this work, and there is a great potential for expanding the set of techniques for the analysis of the plays in question.

The next chapter offers both a large-scale overview of the 19th-century Russian drama available in the RusDraCor, and the representation of women in it, as well as a detailed analysis of two plays by Ostrovsky. Ostrovsky's plays were chosen after a survey of the 19th-century plays concluded that he is the most 'egalitarian' author in the corpus, based on the number of female characters and their speech-acts in his plays. The detailed analysis of *The Storm* shows

that Ostrovsky, indeed, focused a great deal on the destiny of women and on their communications with each other. Half of the most influential and active characters in *The Storm* are female. The strongest connections in the play also occur between female characters. In *Without a Dowry*, by contrast, a greater focus is still placed on the communication between males, and their merchant deals, and the women of the play are treated as unimportant items. For this chapter, again, I had to reconsider the original network of *The Storm* and disregard the 'crowd' character, in order to perform a detailed analysis of the 'core' characters in the play.

Having seen that the proposed methodology, with certain alterations, is quite useful for the analysis of 19th-century conventionally structured plays, it was then important to consider more experimental and modern plays. In the last case-study I therefore focussed on plays written in the second half of the 20th century and in the New Drama movement. The last chapter discusses the social, political, and cultural changes that influenced late Soviet and post-Soviet Russian-language drama, while demonstrating the applicability of the SNA methods to contemporary plays. 23 plays by nine authors were chosen to be encoded in the TEI for this last case-study, based on the popularity of the playwrights and the importance of the plays for the development of Russian-language drama.

As mentioned earlier, during the whole thesis careful attention is paid to the limitations of the methodology and the necessity of close analysis of the material despite the availability of computer-readable versions of the texts. The analysis of the network representation of these modern plays, in accordance with the previously defined TEI and SNA logic, brings into focus the shortcomings of the methodology. Thus, the majority of the discussed plays required 'close' reading to be accurately represented as network graphs. And the last chapter is devoted mostly to raising questions regarding the methodology, rather than answering them. It becomes clear that, in order to include contemporary plays in any quantitative and digital analysis of Russian drama, it becomes necessary to reconsider basic assumptions about the nature of drama within and

beyond the digital approach. For instance, in several plays, including *Duck Hunting* by Vampilov and the more recent *Oxygen* by Vyrypaev, we can see that the definition of a play's character in 20th- and 21st-century drama is transformed from simply being a protagonist (*герой*) to also serving as a narrator. Drama itself is being transformed in the 21st century, and becomes more focussed on the text. This applies to some plays that rely heavily on the monologues of the characters instead of their interactions with each other, as in *My First Man* by Isaeva or *July* by Vyrypaev. Some contemporary playwrights have also taken on the role of theatre directors, and they create the composition of their plays depending on how many and what kind of actors may perform the characters in the plays (this can be seen in *Doc.Tor* by Isaeva and in *Terrorism* by the Presnyakov brothers).

All these experimental features that are currently extremely important for the development of Russian-language drama throw up issues for the proposed methodology. Thus, in the last chapter, I offer several suggestions for possible adjustments of the network representation of contemporary plays to be considered for further research, such as modifying the way some characters are displayed on the network graphs or using dynamic network graphs.

The conclusion of this thesis provides a large poster including the network graphs of 235 Russian plays written from 1747 till 2020. Despite the limitations of the methodology, especially when it comes to contemporary drama, this poster demonstrates to us the great capacity of digital methods. We can see the evolution of Russian drama and trace the trend towards greater structural simplicity in Russian-language plays, which otherwise would have taken much more time and effort to establish objectively and with numerical ground, if analysed manually. Besides the general tendency towards simplicity, Russian contemporary drama, at the same time, appears to be more complex in terms of the nature of its characters and the connections between them, and its focus on performance and on the involvement of the audience. Thus, this thesis concludes that the proposed digital methods can be useful to provide insights for traditional literary

scholarship, but emphasises that these methods need to be applied to Russian-language plays carefully, with a great ‘close’ focus on individual texts.

Introduction

Humanities and, in particular, literary studies are frequently thought to be separate from the exact sciences, even though researchers have started working with combinations of both disciplines a long time ago. The idea of using a quantitative approach for 'authorship attribution' of a text arose already in 1851, when a British mathematician Augustus De Morgan suggested calculating the length of the words in some of the Letters of Paul to establish their authorship (Ogata and Taisuke 2016). Mendenhall (1887, 1903) developed this idea and used word frequency to distinguish different writing styles in Shakespeare's texts. Such works gave way to the field of stylometry which uses linguistic and quantitative approaches to identify the style of a text.

In parallel with applied research, scholars were developing the theoretical basis for the quantitative analysis of fictional texts. In 1893 the *Analytics of Literature* by Sherman was published. A decade or so later, Shklovsky and his associates created the Russian critical school of formalism, and the works of Shklovsky, Eikhenbaum and Tynyanov still inspire literary scholars to incorporate scientific approaches into the field.

Building on these ideas, and drawing upon engineering technologies, a new area of studies has been formed – Digital Humanities (DH). These involve a variety of interdisciplinary studies that use quantitative and/or computational (digital) approaches for exploring and modeling humanities data. There are two scholars who are considered to be the pioneers of DH. In 1957 a concordance of poems by John Dryden was published by the poet and literary critic Josephine Miles. (Montgomery 1957) This concordance was created using card-reading computers. Even before the Dryden concordance was published, Miles had applied a quantitative approach to study poetry by manually counting the frequency of words in poems (Miles 1946).

Even though some consider Miles to be the pioneer of distant reading,[1] and of DH in general, another scholar's work is always referred to – the groundbreaking project "Index Thomisticus" (Busa 1980). Together with his team, Roberto Busa, an Italian priest, created a concordance of the writings of Thomas Aquinas. The project was sponsored and assisted by IBM, and the Index was 'Automatically Compiled and Printed by IBM Punched Card Machines'. "Index Thomisticus" made available an 11-million-word corpus of Latin and still continues to contribute to Latin studies (see, for example, 'Syntactic discontinuities in Latin — A treebank-based study' by Dag Haug (Haug 2017)).

As can be seen, a great part of DH studies is represented by work with texts. Even though there is research on using digital approaches to images and maps, text-based research seems to be the most frequent focus for DH. Moreover, a great part of this attention falls on to fictional literature. With time, the DH community came up with a name for this subfield – Digital Literary Studies. It includes a plethora of ideas, which can be traced back to the 'authorship attribution' work mentioned above and, also, to the formalists' desire to understand how fictional literature works. Eikhenbaum, following Shklovsky, worked on developing the notions of *fabula* and *siuzhet* to describe plot structure, and published a very important paper, *Kak sdelana «Shinel'» Gogolya* (Eikhenbaum 1919). In his work, Eikhenbaum explains the scheme of *Shinel'*s *siuzhet* as well as focussing on the nature of Gogol's comic literary devices (including '*сказ*'), that he considers the most important in *Shinel* and other works by Gogol.

Propp was thinking in the same direction and created a general paradigm of a fairytale, formalising the 'functions' or events of folklore tales, as well as establishing seven universal types of their characters (Propp 1928). Propp's work was a 'giant' step for folklore studies and literary studies in general (Propp 1968, Introduction to the Second Edition by Alan Dundes).

[1] The concept of distant reading is discussed on page 22.

This kind of research established the concept that literary features can be calculated, and that such complex components of fictional texts as style, plot, mood and rhythm can be transformed into numbers. And these numbers are mostly analysed using statistics rather than human reasoning. For that and other reasons, some of the relevant works, inevitably and reasonably, were criticised for following the structuralist approach as an end in itself. I will address this issue in the following chapters by discussing the limitations of quantitative methods for each specific case-study.

Contemporary digital literary studies is a combination of different frameworks and branches of science. Apart from the formalist and structuralist background, digital literary studies borrow their methodology from various disciplines. For example, computational linguistics is helpful when performing quantitative analysis of texts. Biological concepts are used when discussing the evolution of literature. Mathematical theories, such as graph theory, provide new ways of text representation.

Among more recent digital literary research, there are a couple of canonical works. First of all, there is *Distant Reading* by Franco Moretti (Moretti 2013). Moretti formulates the notion of distant reading as the opposite to close reading. He suggests not reading novels as such, but zooming out and processing the various statistical features of large amounts of texts. Distant reading is based on the fact that a person cannot read more than 5,000 books during a lifetime, while an "average" US reader will only get through about 700 books (Temple 2017). This idea has gained ground in the field and was taken forward in another iconic book on digital literary studies – *Macroanalysis* by Matthew Jockers (Jockers 2013). He did multiple experiments and computationally explored the style and national features of texts, as well as modelling the topics of more than 3,000 novels.

However, digital literary studies are not only about distant reading. The advantage of using quantitative approaches on literature is that they allow us to look at familiar data from a

different angle. They do not necessarily provide new discoveries, but they always do let us incorporate new notions into the field, and to formulate common phenomena within a new framework.

Despite the groundbreaking opportunity of distant reading, it is necessary to carefully apply it to fictional texts and drama in particular in order to not overlook crucial details of the texts. This thesis will focus on the combination of the distant and close reading approaches to avoid the simplified interpretation of the results of the quantitative analysis. Such approach proves to be the most appropriate when applied to dramatic texts, which will be discussed in further chapters.

Goals

This thesis will address the following research question – how can we use quantitative and digital methods to study Russian drama, and drama in general? To do this, I will introduce several case-studies focusing on a general analysis of the development of Russian drama, as well as detailed analysis of particular plays. I will discuss how the chosen methodology enhances our understanding of these plays and present the results of my findings together with their interpretation, taking into account the knowledge, already available to us from traditional literary criticism. Besides providing insights on Russian drama by using this new methodology, I consider it very important to acknowledge the limitations of the methodology, which I will discuss in detail in every chapter.

This thesis consists of five chapters. The Introduction chapter presents the methodology together with an introductory case-study. The second chapter provides a detailed analysis of *A Month in the Country* and *Uncle Vanya* as two plays with very similar plots, but different network graphs. The third chapter provides an overview of the representation of women in Russian 19th-century drama, and specifically in Ostrovsky's plays. The final chapter of this thesis

is dedicated to the review and analysis of recent 'New Drama', with a greater focus on the limitations of the methodology. The Conclusion chapter includes the summary of my findings and arguments, and a discussion of further research opportunities.

Materials

As already mentioned, I am working with Russian drama texts. My choice of materials does not imply that drama is more interesting to study than prose. This choice is limited by functional complications. A play has a basic structure which allows us to identify characters, their presence, their speech and actions. While there are works devoted to building models for processing prose texts and extracting characters and their connections from them, it is still an ongoing process. This task is itself a stand-alone study.

I have concentrated on working with Russian plays that are available in the Russian Drama Corpus (RusDraCor)[2], part of the dracor.org projects. The corpus uses the TEI (Text Encoding Initiative)[3] xml-scheme to encode the texts of plays. The TEI provides tags for encoding almost any kind of documents, such as medieval texts, translations (keeping the "layers" of the original and translated text) and dictionaries, poetry and prose. The importance of the TEI is its interdisciplinary nature and international recognition, which enables researchers to work with a range of documents and corpora without even knowing the language of the texts.

The RusDraCor project was started by a research group in the Higher School of Economics, Moscow. The project focuses on both creating a corpus of TEI-marked Russian plays and performing digital research with the available data. The corpus started from having around 30 plays in March 2017. It has now reached 212 plays, written from 1747 to 1947 (accessed on 6 January 2022), and is constantly growing. The TEI-scheme allows us to mark up the metadata of plays, such as the author, title, genre, writing, publishing and first staging dates,

[2] https://dracor.org/rus
[3] http://www.tei-c.org

lists of characters (with gender indication); and, as mentioned, the structure and content, including stage directions. Here is an example of a TEI-encoded piece of the characters' dialogue from the first scene of the first act in Chekhov's *Ivanov*:

```
<div type="act">
  <head>Действие первое</head>
  <stage>Сад в имении Иванова. Слева фасад дома с террасой. Одно окно открыто. Перед террасой
    широкая полукруглая площадка, от которой в сад, прямо и вправо, идут аллеи. На правой
    стороне садовые диванчики и столики. На одном из последних горит лампа. Вечереет. При
    поднятии занавеса слышно, как в доме разучивают дуэт на рояле и виолончели.</stage>
  <div type="scene">
    <head>I</head>
    <stage>Иванов и Боркин.</stage>
    <stage>Иванов сидит за столом и читает книгу. Боркин в больших сапогах, с ружьем,
      показывается в глубине сада; он навеселе; увидев Иванова, на цыпочках идет к нему и,
      поравнявшись с ним, прицеливается в его лицо.</stage>
    <sp who="#Ivanov">
      <speaker>Иванов</speaker>
      <stage>(увидев Боркина, вздрагивает и вскакивает).</stage>
      <p>Миша, бог знает что... вы меня испугали... Я и так расстроен, а вы еще с глупыми
        шутками...</p>
      <stage>(Садится).</stage>
      <p>Испугал и радуется...</p>
    </sp>
    <sp who="#Borkin">
      <speaker>Боркин</speaker>
      <stage>(хохочет).</stage>
      <p>Ну, ну... виноват, виноват.</p>
      <stage>(Садится рядом.)</stage>
      <p>Не буду больше, не буду...</p>
      <stage>(Снимает фуражку.)</stage>
      <p>Жарко. Верите ли, душа моя, в какие-нибудь три часа семнадцать верст отмахал...
        замучился... Пощупайте-ка, как у меня сердце бьется...</p>
    </sp>
```

Figure 1. Example of TEI encoding

For verse drama, the TEI allows us to note the separate 'parts' of a single fragmented line of verse. For instance, in Blok's *Balaganchik*:

```
<sp who="#pervyj_mistik">
  <speaker>Первый мистик</speaker>
  <l part="I">Ты слушаешь?</l>
</sp>
<sp who="#vtoroj_mistik">
  <speaker>Второй мистик</speaker>
  <l part="M">Да.</l>
</sp>
<sp who="#tretij_mistik">
  <speaker>Третий мистик</speaker>
  <l part="F">Наступит событие.</l>
</sp>
```

Первый мистик

Ты слушаешь?

Второй мистик

Да.

Третий мистик

Наступит событие.

Figure 2. Example of TEI encoding

It is important to note that the RusDraCor is not a digital edition. It focuses on producing the TEI representation of Russian drama for the purpose of digital literary studies. The RusDraCor team tries to maximise the automatisation of the encoding process and concentrates on developing converters of existing HTML mark-up into the TEI format. In this way, the project does not (yet) aim to digitise texts that are not available online, which is the main limitation of its representativeness. Despite that, it is the first and only corpus of Russian plays digitised in a machine-readable format, and it provides a great scope of opportunities for research.

Methodology background

In the same way as digital literary studies, the idea of using quantitative approaches for researching plays is not new either. Already in the 1930s Boris Yarkho (Yarkho 2013) was working on the distribution of speech in five-act tragedies in order to draw a formal line between classicism and romanticism. He studied around 200 tragedies, which he broke down into four style periods, and established four parameters that he attributed to each of the style periods. The parameters were the number of speakers in an act, standard deviation from the average number of speakers in an act, the number of scenes, and the number of characters. It is astonishing that

Yarkho performed this research by going through each play and manually calculating all these parameters.

Apart from formalising certain text parameters, using graph (or network) theory approach for studying plays is gaining momentum. Even though scholars and developers have used graphs for the representation of fictional texts since 1993 (Knuth 1994), it was Franco Moretti who paved the way for this method by providing the literary grounding for it. In his book *Network Theory, Plot Analysis* (Moretti 2011) Moretti showed how the presentation of a play as a network of interactions between characters allows us to make observations that are unobtainable or not obvious during an act of ordinary reading.

Quantitative, text-based and network methods are being used to explore different literary research questions. There are papers that are specifically dedicated to studying certain characters' "profiles" and identifying "main" characters in a play. For instance, Jonathan Culpeper investigates the vocabulary of the six most talkative characters of *Romeo and Juliet* by comparing their most frequent words and keywords in his work 'Keywords and Characterization: An Analysis of Six Characters in *Romeo and Juliet*' (Hoover, Culpeper, O'Halloran 2014). The conclusions he draws from simply comparing the distribution of personal pronouns in characters' speeches are impressive. For instance, according to Culpeper, Juliet uses a lot of *I* because she "spends much time in the play baring her soul", while Romeo has a higher frequency of *me* and *mine* since he "has a much more external consciousness of his own" and "has an egotistical streak" (Hoover, Culpeper, O'Halloran 2014). Even though findings of such kind are subject to interpretation, the quantitative data that comes out of this research provides new ground for other scholars and is, therefore, valuable in its own right.

Another interesting work that also incorporates the Social Network Analysis is the paper 'To Catch a Protagonist' (Fischer, Trilcke, Kittel, Milling, Skorinkin 2018). This study uses the data of the German Drama Corpus – part of the DraCor project – and discusses various metrics for identifying the protagonists in 465 German plays. The metrics that are used are split into two categories: count-based (same as quantitative) and network-based (based on graph theory). Count-based measures include word count and number of speech-acts, while network measures consist of different graph-based characteristics that will be discussed later. Even though the idea of the protagonist is subjective and there might be no correct answer, the combination of all the measures often points at the title character. The highlights of the work are the significance of "middle-level" characters for the "social fabric of a play" (Fischer, Trilcke, Kittel, Milling, Skorinkin 2018) and, mainly, the importance of multi-dimensional approaches to the research question.

In 'Opening the Stage: A Quantitative Look at Stage Direction in German Drama' the "epification" of stage directions is investigated in 384 plays from the German Drama Corpus (Trilcke et al. 2020). The authors conclude that with time the stage directions become longer, more emotional and include more verbs, which brings them closer to the spoken text of the plays. The material of Russian Drama Corpus is used by Fischer and Skorinkin in 'Social Network Analysis in Russian Literary Studies' to introduce the SNA approach on five Russian plays, including *Boris Godunov* by Pushkin and *The Storm* by Ostrovsky (Fischer, Skorinkin 2021). The authors demonstrate the network graphs of the plays and briefly discuss the results by interpreting the network-values of selected characters or pairs of characters in the plays.

While mentioned studies showcase the usage of DH methodology when analysing drama and Russian drama in particular, current thesis aims to provide an extensive overview of the applicability of quantitative approach by focussing not only on the methods but on the detailed interpretation of the results, which takes into account close reading of a large number of plays

and existing literary criticism. Moreover, in this project I incorporate multiple quantitative approaches without limiting myself to either SNA or linguistic analysis of the plays. For most of the plays discussed in this work, I analyse their network metrics and the strength of connections between the characters, the linguistic characteristics of characters' speech and their stage directions. Besides that, I investigate the possibility of using SNA when working with contemporary plays by producing and analysing a subcorpus of texts from the late Soviet period and the New Drama movement. Lastly, a great focus of this thesis is the discussion of the limitations of the methodology which offers possible solutions on how to make the methods more productive for the analysis of Russian drama.

Methods

The approaches mentioned above (quantitative, linguistic and network analysis) will be used in this work to explore several research questions about the application of digital methods to studying drama. I will explain below the most important concepts of the methodology. Through the study, I will also be introducing additional methods within these three major approaches, depending on the results of each case-study. I will elaborate on those methods in the corresponding chapters.

During the data extraction process, I use the dracor API (Application Programming Interface). This is a service, built on the dracor.org corpora, that allows users to request certain data from the corpus. With this API it is possible to receive information such as a list of characters, network data, and all the spoken text and stage directions of a play in one click. I send these requests, process the data, and, finally, calculate the measures using scripts that I wrote in Python[4].

[4] Python is an interpreted high-level general-purpose programming language (https://www.python.org/doc/essays/blurb/).

Network-based measures

Based on the TEI annotation of the plays, it is possible to build networks of characters' interactions and visualise them as graphs. See, for example, a graph of *Much Ado About Nothing*[5]:

[5] The graph was produced using the TEI file from the Shakespeare Drama Corpus of the DraCor (https://dracor.org/shake) derived from the Folger Shakespeare Library.

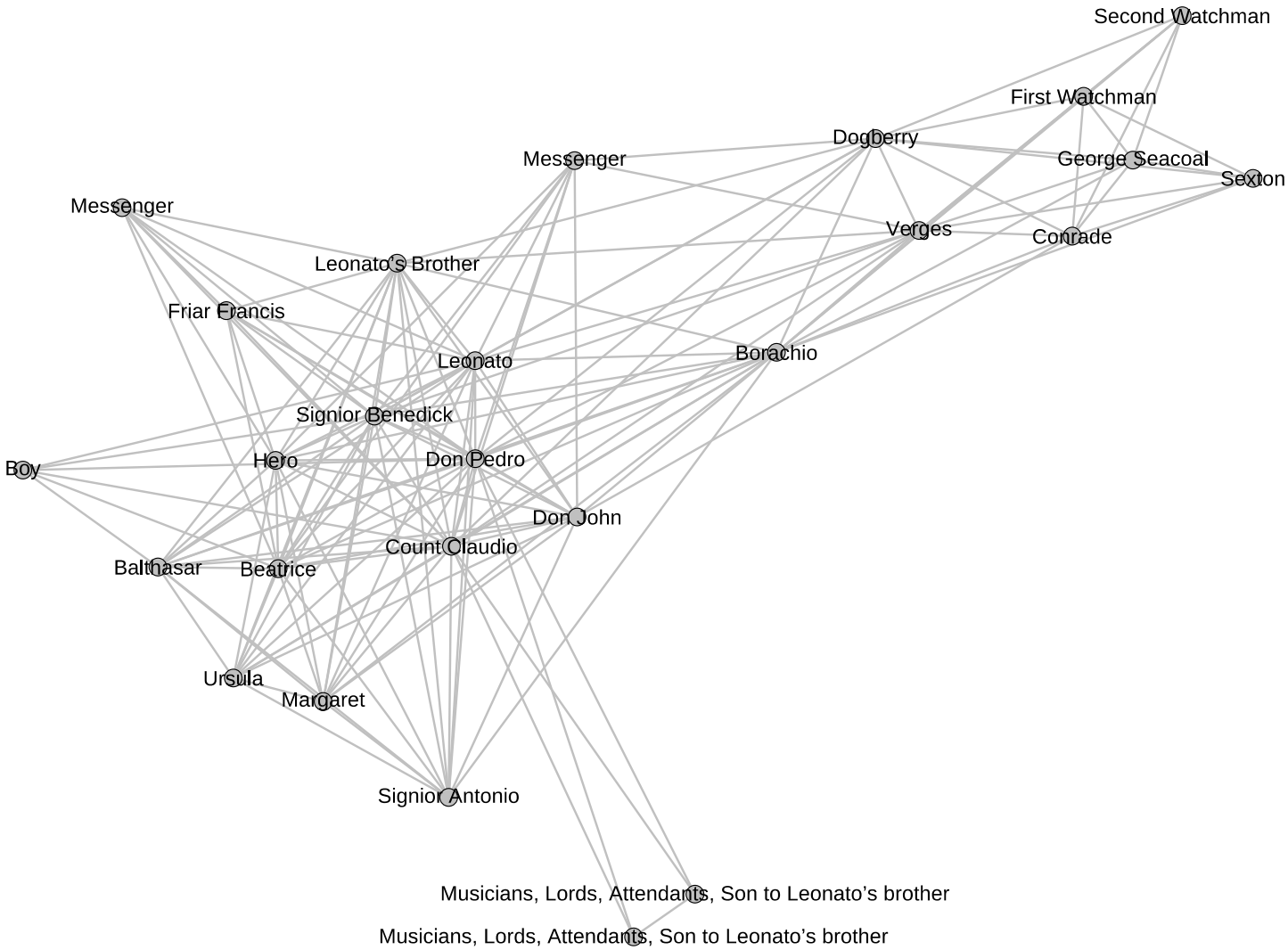


Figure 3. Network graph of *Much Ado About Nothing* by Shakespeare

The DraCor annotation implies that two characters are connected if they both speak during one scene or act. Thus, the TEI encoding in the DraCor does not account for characters leaving and entering the stage during one scene or act. Consider this TEI extract from the second scene of the third act of *The Government Inspector* (*Revizor*) by Gogol':

```
<sp who="#dobchinskij">
<speaker>Добчинский</speaker>
<stage>(бежит и кричит в дверь).</stage>
<p>Мишка! Мишка! Мишка!</p>
</sp>
<stage>Мишка входит.</stage>
```

We can see that the stage directions are marked with the <stage> tag only. There is no indication that *Dobchinsky* moves and speaks, as well as that *Mishka* enters the room. In this scene *Mishka* never speaks. Thus, even though he is present on stage together with *Dobchinsky*, *Mishka* will not be connected to him in the network representation.

It is a challenge to mark up this information in all 212 plays. Firstly, it cannot be done automatically, because stage directions often indicate characters using personal pronouns and not proper names. Coreference[6] resolution is not yet a solved problem and has been a long-running task for computational linguistics. Moreover, annotating characters' stage presence can result in inconsistency, since authors sometimes forget to include this information in stage directions. Thus, for this case-study I will stick with the network based on the act of speaking, though I do not exclude the possibility of building networks based on characters' presence for further research.

Another feature of the existing mark-up is non-directed speech-acts. The TEI files do not contain information on who addresses whom when speaking. This is, again, a task that cannot be automatised and, at the same time, will produce inconsistent results. It is not always clear from

[6] The relation between noun phrases which have the same reference. (Matthews P.H. 2007. *The Concise Oxford Dictionary of Linguistics* (2 ed.). (Oxford University Press)).

the text what characters are being addressed and which of them hears or does not hear the speaker. As a result of these limitations, when using the DraCor annotation, we can create undirected graphs where characters *A* and *B* are connected if they both speak during the same scene (or other smallest segment of a play). Using the graph theory terminology, I will be calling characters 'nodes' or 'vertices' and their connections – 'edges'. The RusDraCor-based graph will include information about the weight of edges. For example, if characters *A* and *B* both speak during the same three scenes, their connection's weight is 3.

Apart from weight, there are other graph metrics that can be calculated after we create a graph. I am going to elaborate on the ones that I am using for this research. These metrics are called 'centralities', and they are used to indicate the most important nodes. There are plenty of definitions of "important" and, in this way, there are various 'centrality' measures that indicate the different kinds of importance of nodes.

The simplest is '**degree centrality**'. It is calculated as the number of "neighbours" that a node has. Two nodes are considered neighbours if they are connected with one edge, or, in other words, the length of the shortest path between these two nodes equals 1. Nodes with the highest 'degree centrality' will be characters who "meet" or speak to the highest proportion of a play's community.

'**Betweenness centrality**' of a node reflects how often the node lies on the shortest path between two other nodes. To calculate this measure for node *A*, we need to:

1) take all pairs of nodes, that are not node *A*, in a graph;
2) count the number of possible 'shortest paths' between each pair of nodes;
3) count how many times node A lies on the 'shortest paths' between each pair of nodes;
4) for each pair, divide the value in 3) by value in 2);

5) sum up these fractions.

This can be also described by the formula:

$$g(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where *v* is the node in question, σ_{st} is the number of 'shortest paths' between *s* and *t* and $\sigma_{st}(v)$ is the number of those paths with the node *v* involved. Nodes with the highest 'betweenness centrality' can be thought of as characters that control the information's flow. But more interestingly, removing this kind of node from a network might result in disconnecting a graph, because a lot of connections flow through these nodes. "A character may speak only four times, but if each of those speeches is to a different group whose members do not speak to any other character, then the character who provides that crucial bridge among factions has the highest BC [betweenness centrality]" (Algee-Hewitt 2017).

'**Closeness centrality**' of a node represents how close the node is to all the other nodes in a graph. To calculate this measure for node *A*, we need to:

1) sum up the lengths of the 'shortest paths' between node *A* and all the other nodes in the graph;

2) take a reciprocal of this value.

The 'closeness centrality' of node *x* can also formalised as:

$$C(x) = \frac{1}{\sum_y d(y,x)}.$$

where *d(y, x)* is the length of the shortest path between the nodes *x* and *y*. Characters with high 'closeness centrality' can be more influential as the information that they are passing on reaches others faster.

‘**Eigenvector centrality**’ of a node reflects the importance of this node by the importance of the nodes it is connected to. “In plays […] it suggests information on how characters relate to both the core and the periphery. Characters who interact with many characters on the periphery of the play have a lower EC [eigenvector centrality] than characters who talk to fewer core characters” (Algee-Hewitt 2017).

Performing network-based analysis would normally be the first step when working with a play. I would then continue to examine the most interesting or important characters, according to the network metrics or in some other way, depending on the research question.

Count-based measures

The count-based metrics I used include:

1) number of scenes during which a character speaks;
2) number of speech-acts of a character;
3) number of words of a character;
4) mean sentence length of a character’s speech-acts;
5) mean word length of a character’s speech-acts;
6) number of stage directions that relate to a character.

1, 2, 3, and 6 should contribute to the level of presence of a character or his/her speaking activity. Mean sentence length would describe the speech’s tempo. Shorter sentences support the pace of the text. The speech-acts of characters using shorter sentences might be easier to read and to perceive. Lastly, mean word length tends to reflect the greater complexity of a character’s speech.

Text-based analysis

Finally, the third major approach I apply in this work is text- or semantics-based. It is connected to the count-based approach, because the results are also produced with a quantitative method. But the analysis itself focuses on semantics rather than pure numbers, as it does with network-based and count-based measures. I explored the most frequent words in characters' speech-acts, the most "valuable" lemmas in their speech-acts, and the most frequent words and bigrams in their stage directions. While the most frequent wordforms, including pronouns and function words, should give us insights into such characteristics of characters' speech as attitude and emotionality, the analysis of the most frequent lemmas refocuses our attention on the subject of their speeches.

When working with characters' speech-acts, I extracted all the speech-acts of every character of each play as single documents. In this way, I would have 10 documents for a play with 10 characters. I am going to refer to this array of speech-acts of a play's characters as a corpus of speech-acts.

To produce the lists of the most frequent words, I calculated the frequency of all words in the corpus of speech-acts for each character. Then I calculated the relative frequencies of words. Relative frequency (RF) reflects the frequency of a word in relation to the size of a corpus or sub-corpus in question. When I was analysing character *A*, I would calculate the RFs of words for its speech-acts and then the RFs of words for the rest of the characters in total, excluding character *A*'s speech-acts. It was necessary to exclude character *A*'s speech-acts from the larger corpus, so that they did not influence the RFs of the words in the larger corpus. By doing this, we can see how character *A*'s speech is different from the general picture in a play. I compared the RFs of the character *A*'s speech-acts to the RFs of the larger corpus. If the relative frequency of a word in character *A*'s speech-acts was higher than in the larger corpus, the word was included in

the list of the most frequent words for this character. Thereafter, I analysed the most frequent words for the selected characters.

At this stage, I performed only basic preprocessing. I lowercased all the words in the texts, so that the computer recognises the same words of different register as one word. It is necessary to note that proper names in the texts were also lowercased, and this is why they will appear so in my text-based analysis. After lowercasing I performed tokenization, which is a process of breaking up a text into separate words – tokens. I did not include any other preprocessing steps, such as removing words of low semantic value, in order to obtain rough information including the most frequent function words that the characters use. I expected these words to be mostly function words and pronouns. This should not be attributed to the fact that characters' speeches are specifically primitive, because these kind of words should be the most frequent ones in any text in any language. See, for example, the frequency list of lemmas in Russian by Lyashevskaya and Sharov[7], or the most frequent lemmas of the Corpus of Contemporary American English[8]. According to these sources, the most frequent lemmas for Russian are: *и*, *в*, *не*, *на*, *я*, *быть*, *он*, *с*, *что*, *а*, *по*, *это*, *она*, *этот*, *к*, and for American English are: the, be, and, of, a, in, to (as an infinitive marker), have, to (as a preposition), it. My goal was to look at the low-level lexicon to explore the behavioral patterns of characters' speeches.

As mentioned previously, analysing the most frequent words should lead us to clues about the nature of characters' speech such as talking a lot about themselves or using a lot of negation. If we want to look at the substance of characters' speech, or the topics, for example, we need to look at the lemmas and exclude any uninformative words. For the most frequent

[7] http://dict.ruslang.ru/freq.php
[8] https://www.wordfrequency.info/free.asp?s=y

lemmas, I used the tf-idf (term frequency-inverse document frequency) measure. The tf-idf of a word reflects the value of this word for a certain document and is calculated in several steps:

1) tf (term frequency) is the frequency of a word in a given document or, in our case, the character's speech-acts;

2) idf (inverse document frequency) is "an inverse function of the number of documents in which it occurs"[9]. Idf is high if a word is common for all the documents, or all the characters;

3) tf-idf = tf * idf.

In order to eliminate any noise (unwanted information from the data), I needed to perform more advanced preprocessing on the texts. The noise might be the stop-words (function words, words with low semantic value) and very frequent words. In the tf-idf calculation I considered only words that appear in no more than 80% of the documents. Thus, if the word was used by more than 80% of the characters, it was excluded from the analysis, because it might be a too common word. I removed all the stop-words from the characters' speech-acts. Then I lemmatised the texts using a morphological analyser for the Russian language (Mystem)[10]. After doing this, I produced and analysed the 10 lemmas with the highest tf-idf for the selected characters.

It is necessary to acknowledge that the accuracy of automatic lemmatization tools and morphological analysers is below 100%. In their study 'Automatic Morphological Analysis for Russian: A Comparative Study' the authors demonstrate that the accuracy of ten open-source POS-taggers might vary from 20.07% to 96.94% (Dereza, Kayutenko, Fenogenova 2021). The Mystem tagger, when tested on the Russian National Disambiguated Corpus, performs the task

[9] Sparck Jones, K. 'A Statistical Interpretation of Term Specificity and its Application in Retrieval', *Journal of Documentation*, 1972, vol. 28 no. 1, pp. 11-21.
[10] https://tech.yandex.ru/mystem/

of POS-tagging with the accuracy of 96.43%. Thus, even though it is quite reliable in most cases, some errors might occur.

I will use the methods discussed above extensively in this work. I will start off by demonstrating all of them in the following case-study. As I go along, I will pay attention to the advantages and disadvantages of certain methods and, where needed, I will introduce additional methods to provide a deeper analysis of the plays. I will be discussing these additional approaches as I introduce them.

A case-study of the quantitative representation of 'key protagonists' in three Russian plays

The goal of this case-study is to investigate how different quantitative and computational approaches reflect characters in fictional texts and help to identify the 'main protagonists' or important characters. When reading a novel or a play, it might seem easy to identify the most important characters, based solely on the plot of the text. My assumption is that there is more to this than the plot, and that it is possible to use non-traditional methods to explore the different perspectives of addressing this task. There must also be certain functions of some characters, and peculiarities of their speech and actions, that make them special or significant.

My methods include studying the 'communities structure' of characters within texts. By 'communities' I mean the groups of characters that are densely connected with each other. The network structure also helps us to identify the most valuable actors in the network of characters. Apart from the network analysis, I will concentrate on quantitative measures of the characters' activity and speech. I want to explore whether a reader's impression of protagonists is supported by numerical analysis.

For this case-study, I chose three plays of different genres, written and published in the 19th century: the historical tragedy *Boris Godunov* by Pushkin, the comedy *The Government Inspector* by Gogol, and the drama *Masquerade* (*Maskarad*) by Lermontov. *Drama* can be seen

as the third macro genre in Russian playwriting. However, there is no consensus as to whether *drama* is an independent genre – the medium between tragedy and comedy – or a debased tragedy with more commonplace settings (Gerasimov, Lotman, Priima, 1982). These plays are written in the early 19th century by authors who had a neo-classical background, but who were moving away from it.

To study these plays, I used the same approach for each of them, took into account my intuition as an ordinary reader, and compared the results to it.

Boris Godunov by Pushkin

The first play to be discussed is *Boris Godunov* by Pushkin. It was written in 1825 and published in 1831. It is a groundbreaking play for its time. Pushkin, following Shakespeare, created a great tragedy with many more characters than was usual in neo-classical drama, and violated the unities of action, place and time. The tragedy is based on historical events that took place at the end of the 16th-beginning of the 17th centuries. The major plot of the play is the struggle for power between Tsar Boris and Grigory Otrepjev (Lzhedmitry – the False Dmitry I). In his tragedy, Pushkin uses different names for Grigory, which are Grigory, Samozvanets (the impostor), Dimitry and Lzhedmitry. The names sometimes change even within one scene. For these names to be considered one entity by the computer, the RusDraCor team manually attributed them to the one id – grigorij_dimitrij_lzhedimitrij_samozvanets. I will be switching between these names in my work. Tsar Boris is also labeled differently in the play. Pushkin calls him Boris before he accepts the throne, and Tsar after. For the same reason, as with Grigory's labels, the character of Tsar Boris has a single id – tsar_boris.

Tsar Boris and Grigory are normally considered to be the 'main protagonists' of *Boris Godunov*. I will discuss how this maps on to the quantitative analysis of the play. But first of all, let us take a look at the network's graph illustration of *Boris Godunov*:

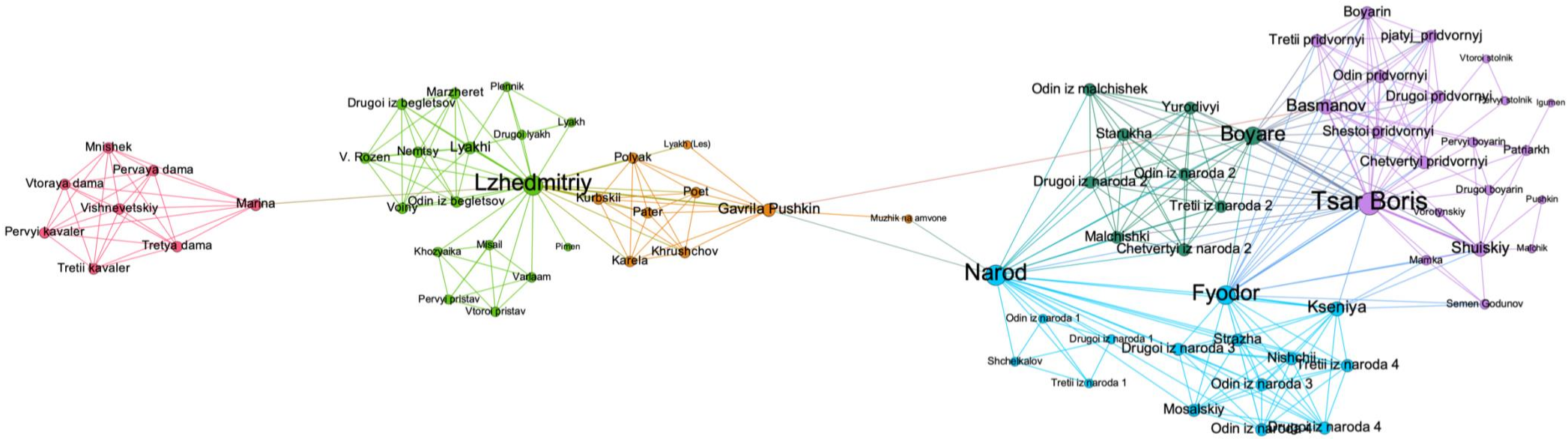


Figure 4. A network graph of *Boris Godunov* by Pushkin

This graph was visualised using Gephi which is the most popular software for visualising and exploring networks. Gephi also allows us to calculate different metrics, including 'centrality' metrics, and they can be used for visualisation. The colours of the graph represent the clusters or communities within the play. When running the "Modularity" function in Gephi, which is a community detection algorithm (Blondel, Guillaume, Lambiotte, Lefebvre 2008), each node gets assigned a cluster number. The clusters or communities are groups of nodes that have much closer ties or, in other words, higher density of connections between each other. These communities usually have bridges to other clusters. For example, on the graph above, we can see the community of Grigory, and the community of Marina and Mnishek. They are connected by a bridge, an edge, between Grigory and Marina.

It is necessary to understand that network theory methods do not provide universal results. Some metrics have different ways of implementation, and there are also multiple clustering algorithms. It is a not a big limitation for this work since the networks are not as large and the analysis of the results still involves close reading. Moreover, we will see that clustering results tend to be persuasive when comparing them to the perception of the play. Clusterisation can, as well, simply be useful for visualisation.

The nodes' size on the graph corresponds to the 'degree centrality' of the characters and here are the three characters with the highest degree in *Boris Godunov*:

1. Tsar Boris with degree of 29
2. Narod (the people) with degree of 26
3. Grigory - Lzhedmitry with degree of 25

Boris Godunov is a large network for a play, and it includes 79 characters. We can see that Tsar Boris and Lzhedmitry are quite close in the degree, but Boris has the highest number of direct contacts with others. Lzhedmitry, in turn, has the highest 'betweenness centrality', and we

can see from the graph that he connects the "Marina – Mnishek" cluster with the rest of the communities. Tsar Boris, again, has the higher 'eigenvector centrality', which means that he is an important member of the community, and, as we know, is connected to other important characters. It is quite common that one node would be the first in all the 'centrality' metrics, and, from reading the play, we could assume it would be either Tsar Boris or Lzhedmitry. But it also comes as no surprise that Boris and Lzhedmitry would be in competition in terms of the influence and position they hold in the network. The graph and its metrics reflect quite accurately the main plot of the play.

At the same time, it was unexpected that the character with the highest 'closeness centrality' would turn out to be Gavrila Pushkin. He is the conspirator who is trying to sabotage the position of Tsar Boris by persuading the main governor to betray the Tsar'. We can see that he has a privileged point and should have better access to information or connections, because he has short communication paths to other characters. Moreover, all the 'centrality' metrics of the nodes, proceeding from their formulas, are connected to the number of direct connections – the 'degree centrality'. Consequently, it makes sense not only to look at the nodes with the highest 'centralities' as such, but to take into account the degree of the characters. Then, Gavrila Pushkin also has the highest 'betweenness centrality' in consideration of his degree. In fact, if we removed Gavrila's node from the network, the whole plot of the play would be disrupted as Boris and Lzhedmitry would be unaware of each other. It is a discovery that we were able to make with the network theory, and this "profile" of Gavrila Pushkin is not obvious from the close reading of the play. Therefore, I looked closely at these three characters of *Boris Godunov* (Lzhedmitry, Boris and Gavrila) since they have high network metrics and can be called the key figures of the play.

The people, in this case, is the collective character that is labelled by Pushkin as *народ*. It does not include occasional characters such as *один из народа* (one of the people) or *другой*

(another one). It is interesting to note that *narod* has the second highest degree in the play. Some literary scholars have said that Pushkin depicted the *narod* as the foundation for government and historical changes (Litvinenko 1974, Bondi 1975). Lotman wrote: "Все писавшие о «Борисе Годунове» Пушкина отмечали, что народ выступает в трагедии как независимая историческая сила, что он присутствует в «афише», списке действующих лиц, и что это отражает его «суверенное» значение в трагедии" (Lotman 1996). Now we can have the actual statistics of how active and important the *narod* is in *Boris Godunov*.

It would also be interesting to analyse the network features of such characters as *один* and *другой,* together with and separately from the *narod* entity. Should they be combined with the *narod*, the people's 'centrality degree' would then become the highest in the play.

In terms of the count-based metrics, Lzhedmitry is the leader. He speaks in 8 out of 22 scenes, which is the highest number among all characters. He also has the highest number of speech-acts and, unsurprisingly, the greatest number of words. Tsar Boris is the second by the number of scenes, speaking during 6 scenes. He has only 38 speech-acts in comparison to the 73 of Lzhedmitry's. Boris is in third place for the speech-acts' results and he is outstripped by Shuisky. If we compare the number of scenes, speech-acts and words for these three characters, we can again see how outstanding are Gavrila Pushkin's high 'network centralities' in accordance with his count-based "activity" metrics:

Character / Metric	Lzhedmitry	Tsar Boris	Gavrila Pushkin
Number of scenes	8 (1st)	6 (2nd)	4 (4th)
Number of speech-acts	73 (1st)	38 (3rd)	15 (5th)
Number of words	1967 (1st)	1660 (2nd)	424 (7th)

Table 1. Count-based metrics for Lzhedmitry, Tsar Boris and Gavrila Pushkin

Moving on to consider the next count-based metric, let us look at the mean length of sentences in the character's speech-acts. Lzhedmitry's mean sentences' length is 8.9 words, Tsar Boris' is 10.6 words, and Gavrila Pushkin's is 11.2 words. To compare these with the whole corpus, I calculated the mean lengths of sentences for all the characters in *Boris Godunov*. It resulted into having an array of 79 numbers – 79 mean lengths of sentences. Then I looked at the maximum and minimum numbers in this array, its median and its mean. The minimum of mean sentences' lengths is 1 word, the maximum is 40, the median and mean are 5.8 and 7.4 respectively. The mean of this array is the average mean sentences' length for all the characters of *Boris Godunov*, the median is its "middle" value.

It could be more typical practice to simply compare Lzhedmitry's, Tsar Boris' and Gavrila Pushkin's values to the mean length of sentences in the whole play. However, the goal was to compare characters against each other, rather than to a general characteristic of a play. In this way, I decided to take into consideration the array of mean values for all the characters, and not one mean value for the whole corpus.

Lzhedmitry, Boris and Gavrila all have longer sentences than all the characters in general, because the mean lengths of their sentences are higher than the mean of mean lengths for all the characters ('the overall mean'). But among the three of them, Lzhedmitry's sentences are the shortest. Thus, we can assume that Lzhedmitry's speech is livelier and has a higher tempo, which makes his speech closer to natural speech.

To support this statement, I have also calculated the proportion of finite verbs to all verbs in their speech. Mystem provides a detailed morphological analysis which allowed me to automatically separate the finite verbs from infinitive verbforms and participles. The mean ratio of finite verbs in the sentences of Lzhedmitry, Boris and Gavrila is 0.8, 0.78 and 0.74 respectively. A higher proportion of finite verbs contributes to the energy and pace of the narration and confirms that Lzhedmitry's speech appears to be more engaging.

The word length analysis did not reveal any specific distinctiveness. The three characters use words of similar length, and it does not differ much from the general corpus statistics.

Regarding the numbers of stage directions, Lzhedmitry has 38 stage directions, relevant to his speech-acts, Tsar Boris has 16 and Gavrila Pushkin only 6. It is around 0.4 stage directions per speech-act for each of them, which is also not a significant insight.

Let us now turn to the semantic analysis of these three characters by looking at their most frequent words and lemmas. The 15 most frequent words in Lzhedmitry's, Tsar Boris' and Gavrila Pushkin's speech-acts, obtained as described in the Methods sections, are:

Lzhedmitry	Tsar Boris	Gavrila Pushkin
я	я	и
не	ты	он
в	мне	да
ты	но	но
что	так	вы
на	о	а
мне	сын	с
как	мой	ему
меня	тебя	басманов
но	меня	димитрий
ли	теперь	ли
о	ж	сам
нет	со	ль
ни	чтоб	когда
мой	всё	лишь

Table 2. Most frequent words of Lzhedmitry, Tsar Boris and Gavrila Pushkin

Lzhedmitry and Boris both use many first-person pronouns, as they talk a lot about themselves and are self-oriented characters. Gavrila's most frequent words, by contrast, do not include any first-person pronouns. He seems to be often speaking about others (*он*, *вы*, *ему*, *басманов*, *димитрий*) which underlines his connecting function. This is quite a simplistic approach on which to base such conclusions, and to elaborate on that I have also taken into account the person agreement of verbs in their speech. Again, appealing to the morphological analysis, I have calculated the ratio of first-, second- and third-person verbs to all verbs in the responses of Lzhedmitry, Boris and Gavrila.

Character	Ratio of 1st-person verbs	Ratio of 2nd-person verbs	Ratio of 3rd-person verbs
Lzhedmitry	0.37	0.27	0.37
Boris	0.26	0.43	0.31
Gavrila	0.14	0.4	0.46

Table 3. Ratio of 1st-, 2nd- and 2rd-person verbforms in the speech of Lzhedmitry, Tsar Boris and Gavrila Pushkin

As we can see, Lzhedmitry has the highest proportion of first-person verbs in his speech, while Boris turns out to address other characters more frequently. Gavrila, indeed, most often talks about other characters and frequently addresses others, while talking very little about himself and his own actions.

Going through the 10 most 'valuable' lemmas (with the highest tf-idf) for Lzhedmitry, Boris and Gavrila gives us insights into the nature of their speeches and thoughts:

Lzhedmitry	Tsar Boris	Gavrila Pushkin
любовь	сын	димитрий
марина	власть	басманов
видеть	шуйский	законный
судьба	народ	московский

мир	душа	войско
знать	гроб	становиться
курбский	царь	гражданин
бедный	казнь	прах
борис	князь	поздно
вести	нужный	идти

Table 4. Most frequent lemmas of Lzhedmitry, Tsar Boris and Gavrila Pushkin

These are the lemmas that are specifically important for these three characters and are distinctive for their speech-acts. Tsar Boris often gives grandiose speeches, talking about strong concepts and events: *власть*, *душа*, *гроб*, *казнь*. He is discussing how he became the ruler, and how power can be a burden. He also talks about baring his soul and being unhappy.

Lzhedmitry does not, in fact, give such speeches. Surprisingly, he does not speak or think a lot about power, even though he competes for power with Tsar Boris. Lzhedmitry is reflecting about his fate, his relationships with the world, and seems to be passive ("*я* ***бедный***", "*назначила слепая мне* ***судьба***"). Pushkin lets us get to know Lzhedmitry from a personal perspective, and his distinctive lemmas also depict his affair with Marina and conversations about others.

Gavrila Pushkin's most valuable lemmas, again, represent the fact that he talks a lot about the mechanisms of power struggle (*законный*, *войско*, *гражданин*) and other actors (*димитрий*, *басманов*). He seems to be a fascinated observer of this competition, keenly awaiting its results.

Moving to the content of the stage directions for these three characters, we can say that they are purely functional. The most frequent words used in the stage directions for Lzhedmitry and Tsar Boris are *входит* and *уходит*. We will see later that different authors use stage directions in quite different ways.

The investigation of *Boris Godunov* has shown that the network approach is really useful for this play. It provided significant insights into the play's structure and the functions of its characters. We can see ways in which Boris and Lzhedmitry are connected and the kind of influence they are fighting for.

The graph metrics reflect the contribution that each character makes to the network. The 'degree centralities' echo the view that *Boris Godunov* is a play about the people. Gavrila Pushkin turns out to be a critically valuable character of the play. We may need to reconsider the notion of 'main protagonists'. Perhaps, they are not only the ones who form the plot, but also the ones who keep the structure together and support the plot. It is also interesting how the functions of the characters are then upheld by their vocabulary.

The Government Inspector by Gogol

The Government Inspector was written in 1835 and published in 1836. Gogol's aim was "to collect all the imperfections and injustices in Russia and to laugh at all of it at once" (Gogol 1855). *The Government Inspector* is considered to be a new drama with a new aesthetics for its time. The play was not well received by the audience and the critics because it was not artistically skilled enough according to the rules of neoclassicism, and not comic enough either. The plot of the play is built on the illusory government inspector intrigue, because there is in fact no real inspector. Nevertheless, I would assume that Khlestakov, who is thought to be the inspector, is the main character. I will describe how I explored this hypothesis below.

The Government Inspector's network graph looks as follows:

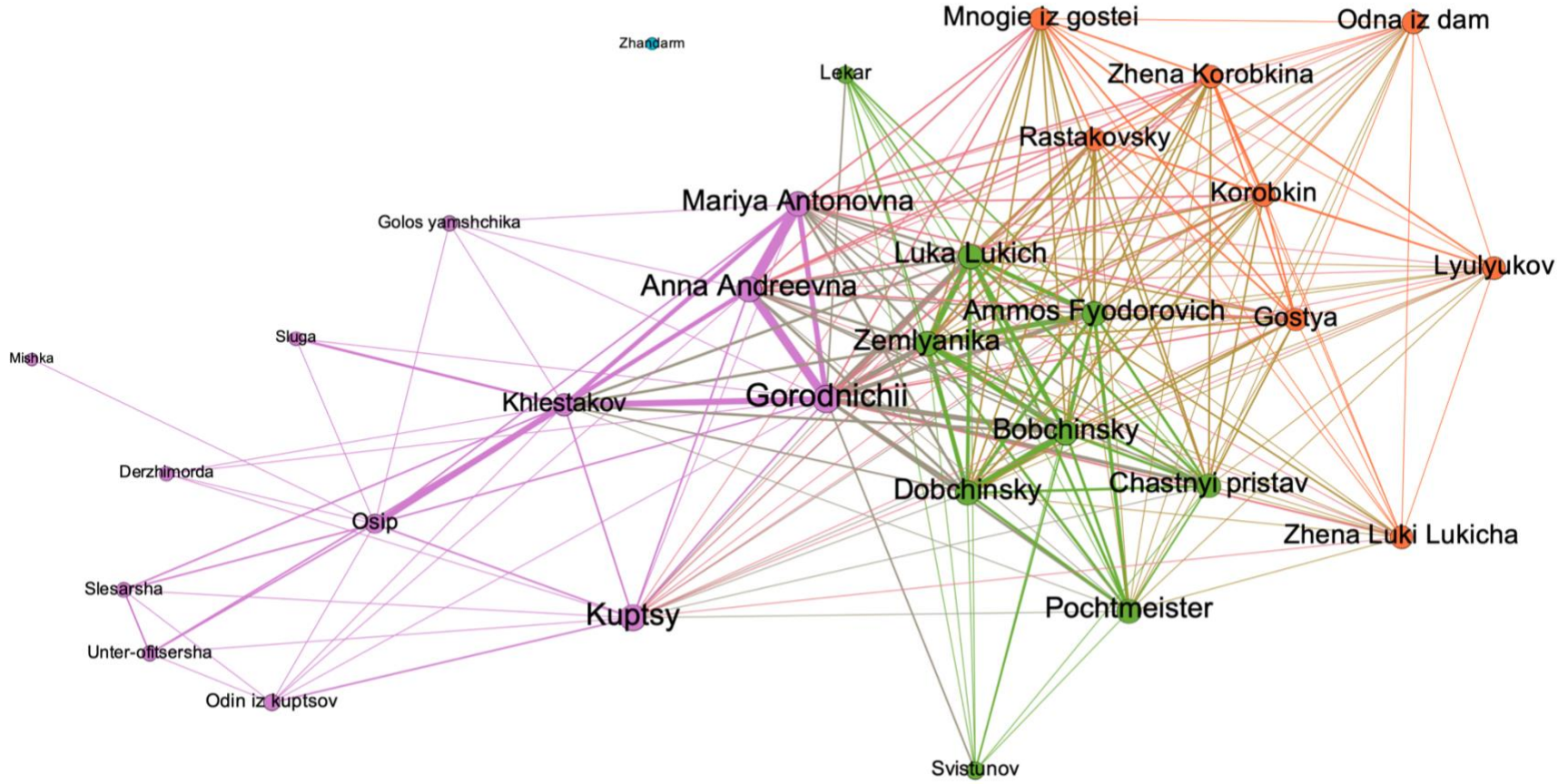


Figure 5. The network graph of *The Government Inspector* by Gogol

The nodes' size, again, depends on the 'degree centrality' of the characters. Surprisingly, Khlestakov does not have as many direct connections as expected. Intuitively, it would be normal to assume that he could be called the main protagonist of the play. Khlestakov is the key figure in the *revizor* (the government inspector) intrigue – "the *revizor* is sitting in everyone's head" as Gogol said (Mann 2007). However, the network metrics that we will discuss later do not support this assumption.

We can see that there are many characters with similar degrees, such as the Gorodnichii (the Mayor) and those around him. Indeed, if we look at the numbers, we can say that, with a slight lead by Gorodnichii, the play has no "main" character:

1. Gorodnichii with a degree of 26
2. Kuptsy (the shopkeepers) with a degree of 24
3. Anna Andreevna with a degree of 22
4. Ammos Lyapkin-Tyapkin, Artemy Zemlyanika, Luka Lukich, Pochtmeister, Bobchinsky, Dobchinsky and Maria Antonovna with a degree of 21
5. Chastnyi Pristav (the chief of police) with a degree of 20
6. Rastakovsky, Korobkin, Zhena Korobkina (Korobkin's wife), Mnogie iz gostei (many of the guests), Gost'ia (a guest) and Odna iz dam (one of the ladies) with a degree of 18
7. Khlestakov with a degree of 17
8. Osip with a degree of 12

Gorodnichii has the highest 'degree centrality' because he is the leader of the city community. He is connected to the majority of the actors and gives orders. But others from his

cluster also do not lag far behind, as they communicate in a lot of shared scenes. It can be said that these numbers support the ambition of Gogol to generalise and summarise the history of Russia in the concept of a town. In this way, there are no outstanding protagonists, and, as Mann writes, the production of the play could only be successful if **all** the actors performed successfully (Mann 2007).

Khlestakov's lower 'degree centrality' emphasises the fact that he is an outsider. The Gorodnichii has the highest 'betweenness', 'closeness' and 'eigenvector centralities'. This is unsurprising due to his degree and his actions that we know from the plot of *The Government Inspector*. It means that the Gorodnichii is the most involved character, the connector and influencer.

Interestingly, if we perform the same action as with *Boris Godunov*, it turns out that Osip has the highest 'closeness centrality' when his degree is taken into account. This is for the reason that Osip is the only one in the network, who is connected with Mishka. Mishka appears in multiple scenes together with the Gorodnichii and others, but he never speaks to any of them. He just takes orders. So Osip is the only one who has a conversation with Mishka and they form a pair of servants in the graph.

As regards count-based metrics, Khlestakov is the leading figure. He is not the most involved character within the network. But he seems to be the most active, contending with Gorodnichii:

Character / Metric	Khlestakov	Gorodnichii
Number of scenes	26 (1st)	21 (2nd)
Number of speech-acts	229 (1st)	172 (2nd)
Number of words	3850 (2nd)	4919 (1st)

Table 5. Count-based metrics for Khlestakov and Gorodnichii

Thus, Khlestakov is present in the greatest number of scenes. He also speaks more frequently than anybody else in the play. The greater number of words in Gorodnichii's speech-acts may be explained with the average length of the sentences he uses. Khlestakov's mean length of sentences is 7 words. Gorodnichii's mean is 8.5 words. The overall corpus statistics are: 2.6 words as the minimum of means, 12.2 words as the maximum, 6.8 words as the median, and 6.7 words as the mean of means. So the Gorodnichii generally has longer sentences than Khlestakov, and the mean length of his sentences is higher than for other characters. As a result, the Gorodnichii's conversations can become less lively than Khlestakov's. The latter contributes to keeping up the fast rhythm of the play's dialogues. This agrees with the high proportion of finite verbs in Khlestakov's sentences – 0.84, on average, versus 0.80 in the Gorodnichii's sentences. However, the distinction is not a major one.

The word length, as in *Boris Godunov*, does not differ much for both characters. Likewise, it is not untypical of the whole corpus.

I will now concentrate on the lexicon of Khlestakov and the Gorodnichii, as the two characters who are competing for the most statistically significant character of *The Government Inspector*. They have a higher significance in the structure because they are the most active. Let us discuss the 15 words with the highest relative frequency in their speech-acts:

Gorodnichii	Khlestakov
и	я
что	не
на	что
так	в
с	а
ну	как
ты	мне

бы	за
вам	это
то	вы
или	нет
теперь	же
если	меня
то	у
же	ну

Table 6. Most frequent words of Khlestakov and Gorodnichii

Khlestakov has three first-person pronouns in the list while the Gorodnichii has none, and the most frequent pronoun he uses is ты. These lists show that Khlestakov speaks much more about himself, which can attract the readers' and audience's attention to his personality, and because of that, makes him look like (or sound like) a central character.

Analysing the person-agreement of verbs in their speech, however, suggests that Khlestakov also addresses other characters frequently:

Character	Ratio of 1st-person verbs	Ratio of 2nd-person verbs	Ratio of 3rd-person verbs
Khlestakov	0.35	0.38	0.26
Gorodnichii	0.19	0.44	0.38

Table 7. Ratio of 1st-, 2nd- and 3rd-person verbs in the speech of Khlestakov and Gorodnichii

Nevertheless, the proportion of first-person verbs in Khlestakov's speech is significantly larger than those in the Gorodnichii's speech, who turns out to talk mostly to and about other characters.

Moving on to the most distinctive lemmas:

Gorodnichii	Khlestakov
знать	любовь
черт	сударыня
дело	признаваться
взять	любить
говорить	взаймы
друг	говорить
всякий	давать
сделать	знать
изволить	видеть
проклятый	просить

Table 8. Most frequent lemmas of Khlestakov and Gorodnichii

They reflect the fact that Khlestakov is the only character in the play who has a (sort of) love intrigue. He also uses more abstract verbs belonging to the semantic classes of speech (*признаваться*, *говорить*, *просить)*, emotions (*любить*), and mental processes (*знать*, *видеть*). While the Gorodnichii also uses some of these (*знать*, *говорить*), he is big on the action verbs *взять* and *сделать*. These particularities convey Khlestakov's dreamy character, with not much action taken in his life. At the same time, the Gorodnichii is quite action-oriented, because he is trying to solve the *revizor* problem without wasting time.

Khlestakov's stage directions are as well more diverse than the Gorodnichii's. Khlestakov has 108 stage directions, while the Gorodnichii has 90. The most frequent words in the Gorodnichii's stage directions are *в* (19), *и* (12), *сторону* (11), *вслух* (7), к (7), *осипу* (6), *с* (6), *на* (5), *за* (5), *уходит* (5), *дверь* (3), *бобчинский* (3), *бобчинскому* (3), *добчинскому* (3), *говорит* (3).

The most frequent bigrams for him are *в сторону* (11), *к осипу* (4), *за ним* (2), *вместо шляпы* (2), *всем телом* (2), *обращаясь к* (2), *за голову* (2). So the Gorodnichii's stage directions are rather uniform and practical. The fact that *в сторону* is the most common phrase for him shows his insincerity when communicating with others, especially Khlestakov. Unlike the Gorodnichii, Khlestakov's stage directions include *в сторону* only once. As Mann (Mann 2007) noticed, it indicates that Khlestakov says what he thinks directly, and it characterises him as a simple person.

Khlestakov's most frequent stage directions words are *и* (25), *в* (14), *не* (9), *к* (7), *на* (6), *с* (6), *из* (5), *ему* (5), *пишет* (5), *голосом* (5), *осип* (4), *один* (4), *ее* (4), *ты* (3), *окно* (3), *продолжает* (3), *лукич* (3), *решительным* (3), *говорит* (3), *ест* (3), *добчинскому* (3), *по* (3), *громким* (3), *дверей* (3), *деньги* (3), *городничий* (3), *писать* (3), *лука* (3), *окну* (3), *но* (3), *рисуется* (3). It looks as though Khlestakov performs more different actions than Gorodnichii: he eats, writes, and poses.

Khlestakov's most frequent stage directions bigrams include *к окну* (3), *лука лукич* (3), *в окно* (3), *решительным голосом* (2), *бобчинский выглядывает* (2), *придвигая стул* (2), *продолжает писать* (2), *городничий и* (2), *слышен голос* (2), *целует ее* (2), *на колени* (2), *подает ему* (2), *подходит к* (2), *и прочие* (2), *из дверей* (2). They often feature other characters and are more varied, since there is no such common bigram for Khlestakov as *в сторону* is for the Gorodnichii.

It is important to note that the ratios of stage directions per speech-acts for Khlestakov and Gorodnichii are higher than for the characters of *Boris Godunov* that we discussed earlier. The content is also very different. It shows how Pushkin was only using stage directions for purely functional tasks. Gogol, on the other side, being very familiar with theatre, had a clear and detailed image of how he wanted the play to be staged. The nature of the stage directions in *The*

Government Inspector resonates with the exceptionally thorough instructions for actors he provided at the beginning of the play.

To summarise the analysis of *The Government Inspector*: the network approach, unlike in *Boris Godunov*, does not reveal any unexpected protagonists. It also does not expose any rivalry between Khlestakov and Gorodnichii. In contrast, it reveals a group of highly connected, almost equally influential and important characters. It does show us the concept behind the play. Gogol wanted to picture the town as a homogeneous entity. We can see that because the same group of characters has a similarly high 'degree centrality'. They often appear and talk on stage together. Still, other statistics uncover that there is a reason why Khlestakov seems to be a 'key protagonist' of the play, despite figuring in only three of the play's five acts. He is more active in terms of talking. His character is self-centered and his speech is more dynamic than the Gorodnichii's.

Masquerade by Lermontov

We will now discuss the last play for this case-study. *Masquerade* was written in 1835, the same year as *The Government Inspector*, but published in 1842. Lermontov called it a drama; however, it is also said to be a social tragedy (Lomunov 1941). In this play Lermontov wanted to expose the vices of high society in Saint Petersburg. The plot of the play focuses on the personal drama of the gambler who has given up cards – Arbenin. There can be no doubt that Arbenin is the 'main protagonist' of the drama. In 1836 Lermontov even rewrote *Masquerade*, and called its new version *Arbenin*.

Let us look at the graph of *Masquerade*:[11]

[11] The numbers in the labels of the nodes correspond to the segments of the play. For instance, 'Pervyi igrok 1 1 2' refers to *Первый игрок* in the second entrance of the first scene of the first act.

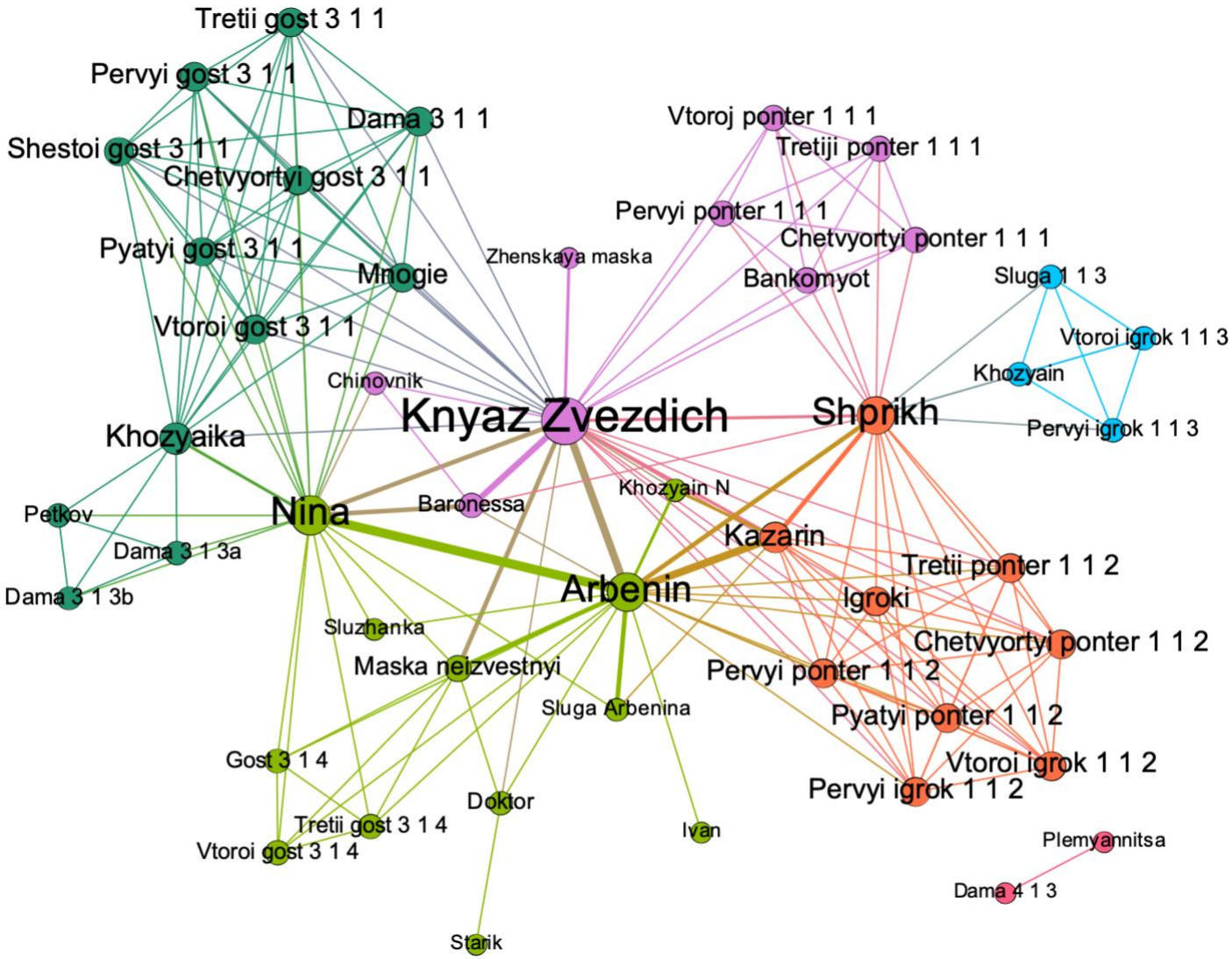


Figure 6. Network graph of *Masquerade* by Lermontov

The character with the highest ‘degree centrality’ is Knyaz’ Zvezdich (Prince Zvezdich). Zvezdich also has the highest ‘betweenness’, and ‘closeness centralities’. This is, to some extent, unexpected, as with a reader’s intuition one could assume that Arbenin should be the leader in the network metrics as he seems to be the ‘main protagonist’. Thus, according to the network metrics, Zvezdich has the highest number of direct connections and is the most influential and connective. But despite Zvezdich’s greatest degree, Arbenin has the highest ‘eigenvector centrality’. This means that Zvezdich is often connected to the peripheral characters while Arbenin is connected to the core ones. Indeed, from the graph we can see, that Arbenin is missing out on connections with the card players and some guests at the ball. From that we can say that Zvezdich is more socially active, while Arbenin is trying to distance himself from people in society, because he detests them. In the same way, Nina has a slightly higher degree than Arbenin, which is due to the fact that she is more socially active at the balls:

1. Knyaz’ Zvezdich with a degree of 31
2. Nina with a degree of 22
3. Arbenin with a degree of 21

Even though he is the third by degree, Arbenin is the leader in count-based metrics:

Character / Metric	Arbenin	Knyaz’ Zvezdich	Nina
Number of scenes	25 (1st)	16 (2nd)	11 (3rd)
Number of speech-acts	214 (1st)	143 (2nd)	100 (3rd)
Number of words	5002 (1st)	1350 (2nd)	1213 (4th)

Table 9. Count-based metrics for Arbenin, Knyaz’ Zvezdich and Nina

I will be looking closer at these three characters, because they are the three most involved or present according to the network- and count-based metrics. They also, allegedly, form a love triangle in the play. The words of the highest relative frequency for all of them are quite similar:

Knyaz' Zvezdich	Arbenin	Nina
я	я	я
не	и	не
но	вы	ты
вы	а	в
что	все	но
мне	как	мне
о	на	меня
а	нет	о
вас	так	как
он	был	тебя
нет	да	нет
вам	то	ой
это	вам	ж
да	же	так
если	ни	конечно

Table 10. Most frequent words of Arbenin, Knyaz' Zvezdich and Nina

They all use a lot of *я* and second-person pronouns. But Arbenin seems to be less self-involved than Zvezdich and Nina, because they also have *мне* and *меня* among their most frequent words. The proportion of different person-agreement in the verbs that these characters use provides additional evidence for that:

Character	Ratio of 1st-person verbs	Ratio of 2nd-person verbs	Ratio of 3rd-person verbs
Knyaz' Zvezdich	0.41	0.3	0.3
Arbenin	0.31	0.32	0.37
Nina	0.36	0.42	0.22

Table 11. Ratio of 1st-, 2nd and 3rd-person verbs in the speech of Arbenin, Knyaz' Zvezdich and Nina

Zvezdich turns out to be the most self-centered among them, while Nina concentrates on addressing other characters and Arbenin's usage of verbs is almost equally distributed between the three persons.

All of them use a lot of negation: *не*, *нет*, and *ни*. Other than that, the analysis of the most frequent words does not reveal any differences or peculiarities for these characters.

The analysis of the lemmas with the highest tf-idf gives us more information:

Knyaz' Zvezdich	Arbenin	Nina
браслет	любить	евгений
узнавать	знать	любить
слово	душа	знать
знать	счастие	умирать
сделать	умирать	смерть
честь	грудь	бледный
видеть	любовь	сюда
совет	давать	странный
счастие	сердце	браслет
цель	честь	говорить

Table 12. Most frequent lemmas of Arbenin, Knyaz' Zvezdich and Nina

Zvezdich's top tf-idf word is *браслет*, which is explained by his desire to find out who the unknown woman is that he meets at the ball. Both Arbenin and Zvezdich talk about честь. Zvezdich talks about his dignity as a player, Arbenin's dignity, and his own dignity once Arbenin finds out about Zvezdich flirting with Nina. Arbenin mostly talks about Zvezdich's dignity and how he is disappointed in Zvezdich. Arbenin's most valuable lemmas are more emotion-related than Zvezdich's: *любить*, *душа*, *умирать*, *любовь*, *сердце*. It characterises him as a thoughtful person who reflects a lot about himself and his relationships with others. He is in constant search of *счастие*. Whereas Arbenin is desperate to find his, Zvezdich is full of hope after he meets the mysterious lady.

Nina's lemmas are associated with anxiety and death: *бледный*, *странный*, *умирать*, *смерть*. Despite Arbenin's accusations of *Nina* neglecting him, the fact that *евгений* has the highest tf-idf in Nina's speech-acts means that she is constantly thinking and talking about him. Moreover, she seems to be the only one, who calls him by his first name.

It is worth noting that Arbenin's mean sentence length is 9.1 words, which is higher than Nina's 8.5 and Zvezdich's 7.3. The median for the means of all characters is 6 words and the mean is 6.6. In this way, Arbenin's sentences are, on average, about three words longer than the others'.

The difference in the means of Arbenin, Nina and Zvezdich can be interpreted in the following way. Arbenin loves words and, sometimes, talks without thinking. We saw that in his speeches he is concentrating on abstract notions. Zvezdich, by contrast, is a man of action and in his speeches he sticks to the point more than Arbenin.

The fact that Nina's average sentence length is a bit lower than Arbenin's but higher than Zvezdich's may have a tricky explanation. It is a hackneyed common assumption that women talk more than men. Accordingly, the number of speech-acts by women should be

higher and the length of women's speeches should be greater. Surprisingly, it is not true for the RusDraCor, and in the majority of plays women talk less than men.[12] This may be because of the patriarchal society in Russia at the time when the plays were written, or the expectations for gender roles on the stage. Likewise, even sentimental Nina speaks and thinks in shorter sentences than Arbenin. This reflects, at the same time, his personality and Nina's position in the play.

In terms of the speech tempo, the three characters have a high proportion of finite verbs in their sentences, on average. But Nina seems to be more concentrated on the events and the mean ratio of finite verbs in her sentences is the highest – 0.82 versus 0.77 and 0.8 in the sentences of Zvezdich and Arbenin, respectively.

Let us now look at the last bit of the play's analysis – the content of stage directions. The stage directions reflect the fact that *Masquerade* is indeed a "*социальная трагедия*" (Lomunov 1941). The most important bigram for all the three characters is *в сторону*. They all hide their real thoughts and pretend.

Arbenin's stage directions most frequently include *в* (19), *на* (18), *и* (13), *уходит* (11), *сторону* (9), *руки* (5), *вдруг* (5), *ей* (5), *не* (5), *подходит* (5), *к* (5), *себя* (4), *слуга* (4), *молчание* (4), *слуге* (4), *встает* (4), *один* (4), *берет* (4), *улыбаясь* (4). The most frequent bigrams are *в сторону* (9), *про себя* (3), *подходит к* (3), *на колени* (2), *сложив руки* (2), *по комнате* (2), *на него* (2), *в размышлении* (2), *ходит по* (2), *смотрит на* (2), *него пристально* (2), *падает на* (2), *на нее* (2), *с диким* (2), *слуга уходит* (2), *переменив тон* (2), *за руку* (2). Apart from *в сторону*, Arbenin also speaks *про себя*. He seems to be the

[12] "Approaching the Evolution of Russian Drama by Means of Network Analysis" by Irina Pavlova, unpublished first year Master's paper presented at the "Digitizing the Stage" conference in Oxford in July 2017.

most secretive about his thoughts among the three of them. There are a lot of phrases that characterise how Arbenin talks or reflects emotively, walking from side to side, wondering: *ходит по*, *по комнате*, *в размышлении*, *падает на*, *на колени*, *с диким*, *переменив тон*.

Zvezdich and Nina have far fewer stage directions and they are mostly action-related. Zvezdich's most frequent stage directions words are: *в* (6), *сторону* (5), *на* (4), *и* (4), *берет* (3), *руку* (3), *за* (3), *ее* (3). Apart from в сторону, his most frequent bigrams are *ее за* (2) and *за руку* (3). These are the traces of his interactions with Baronessa (The Baroness).

Nina's most frequent stage directions words are *в* (8), *и* (5), *к* (3), *упадает* (3), *на* (3), and the bigrams are *упадает на* (2), *в сторону* (2).

This analysis of *Masquerade* using the suggested methods is quite productive. First of all, network metrics reveal that Zvezdich often communicates with peripheral characters and Arbenin is the core of the community of *Masquerade*. Arbenin is also more present in terms of his speaking activity.

Most importantly for the general description of the play, stage directions analysis reveals the stealthy nature of the interactions within the community. Another general note about the play is that, sadly, among the three of them, Nina is the least present and active, as she appears in the fewest scenes and speaks seldom. I attribute this to women's position in Russian society of that time, as well as her role in the play. It resonates well with the Baronessa saying: "*Что ныне женщина? Создание без воли*, | *игрушка для страстей иль прихотей других!*".[13]

[13] *Masquerade*, Act II, Scene 1.

Conclusion

In this case-study I have aimed to demonstrate the new perspectives on drama texts that can be gained with quantitative methods. Literature is subjective and is not meant to be understood by everyone in the same way. But authors create certain structures in their texts. Consciously or subconsciously, they include some patterns and regularities in their works. There are, of course, certain limitations. As we have seen, the methods I use are not universally applicable to any play. They need to be adjusted, and some may produce little or no results. However, by concentrating on different approaches while working with each play, I have either statistically contributed to the existing literary knowledge, or produced new findings. Despite that, I want to point out that this work is not about replacing a traditional approach to literary studies. Regrettably, and perhaps luckily, computers cannot yet interpret the results of this kind of analysis. Thus, it is necessary to turn back to the texts and to close reading to understand how the numbers are derived, which I will do in each chapter.

As mentioned earlier, I am not going to limit my research to the methods used in this case-study. For instance, apart from various 'centrality' measures, it is possible to look at the characters who have the strongest connection with each other – the connection of the greatest weight. Such characters talk on the same stage together more often than others. It might provide information about the important actors who support or influence the 'key protagonists'. For example, in *Boris Godunov* such pairs are Tsar Boris and Shuisky, and Tsar Boris and the Boyare (The Boyars). In *The Government Inspector* the pair of characters with the strongest connection is Anna Andreevna and Maria Antonovna (the wife and daughter of Gorodnichii). In my analysis I overlooked these characters since they did not have outstanding network metrics. However, the fact that they have the strongest connection suggests that they can be specifically valuable in the network structure of the play. In

Masquerade, the ‘strongest’ pair is Arbenin and Nina, which reflects the fact that the plot of the play concentrates on this couple.

Furthermore, it is necessary to pay more attention to general observations about the plays. For example, in the next chapters, I am going to consider various overall graph metrics, such as the density and the clustering coefficient. The graph’s density is a value that reflects how many of all possible connections are actually established in the network – the level of the network’s connectedness. The clustering coefficient shows how well characters group into tied communities. These measures might contribute essential information about the structure of the plays.

Moreover, it would be interesting to move beyond the simple measures of speech analysis that have been used in this case-study. For instance, the average length of words was meant to represent the complexity of the speech of a character. But it turned out to be relatively similar for all the characters I studied in all the three plays. In his quantitative analysis of *War and Peace* Skorinkin discovered that Bolkonsky has the most complicated speech while Natasha’s speech is the simplest in terms of distinctive words, punctuation and other non-lexical parameters of speech (Skorinkin 2018). I believe that the speech texts of the ‘main protagonists’ may have different ‘readability’ values than those of the rest of the characters, and I will continue to explore the methods of measuring that in the following chapters.

Having demonstrated the foundations of my methodology in general, I will now move on to the next chapter, where I will focus closely on two texts – *A Month in the Country* by Turgenev and *Uncle Vanya* by Chekhov – and bring in some additional methods for their analysis

A comparative analysis of *A Month in the Country* (Turgenev) and *Uncle Vanya* (Chekhov)

It is not unusual for authors to get inspired by each other's works and to borrow and adapt certain characters' profiles or plot lines. There have been a few critics (Berdnikov 1953) who have considered Turgenev's drama to be mostly the product of his adaptation of European plays. And Turgenev's play *A Month in the Country* is frequently associated with *The Stepmother* (*La Marâtre*) by Balzac (Grossman 1924, Vishnevskaya 1989). The two plays, indeed, have much in common as regards plot and dramatis personae. *The Stepmother* describes the life of a small household of people who are unsatisfied, in love or heartbroken. The main theme of the play is the rivalry between two women who are competing for the love of a man. The women are the young stepmother, who is unhappily married to an old general, and his daughter. A similar plot-line is one of the main subjects of *A Month in the Country*. It tells the story of a conflict between an unhappy married woman and her foster-daughter. They both fall in love with a young student, and the older woman, driven by jealousy, ruins the friendship between the student and her foster-daughter.

The plays are thus apparently similar, and this has provoked a discussion around the importance of Turgenev's borrowings, and the degree of their influence on the originality of his theatre works. One of the conclusions of this scholarship is that, despite their blatant unoriginality, some of Turgenev's plays, including *A Month in the Country* (his most famous play), nevertheless introduced a new wave of Russian drama – the psychological drama. This finds its fulfilment in Chekhov's psychological realism 50 years later.

One of the greatest of Chekhov's plays, according to the critics, is *Uncle Vanya*. Interestingly, both its manner and its plot greatly resemble *A Month in the Country*. Briggs (1994) points us to Isaiah Berlin's observation when comparing *A Month in the Country* with

The Stepmother: "[The] parallels can scarcely be accidental". "[This] applies with equal rigour to that between Turgenev and Chekhov", says Briggs. Below are very basic graphs that illustrate the similarity of the plot-lines in *A Month in the Country* and *Uncle Vanya*. These graphs were created manually and are based simply on the connections and affections between the key characters that we recognise while reading the plays.

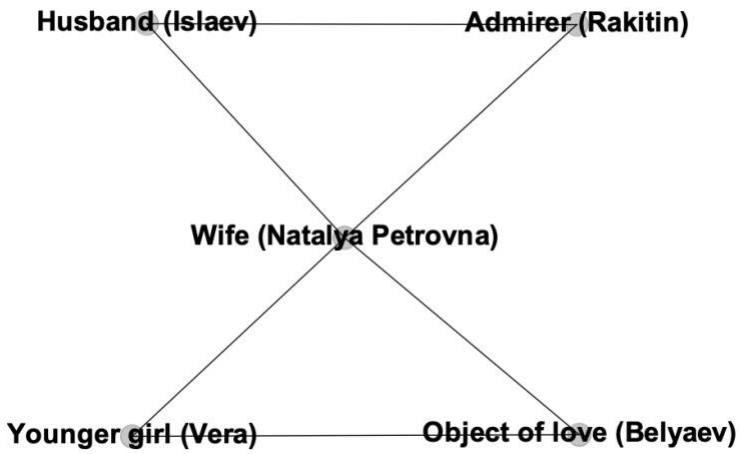


Figure 7. Love triangles in *A Month in the Country*

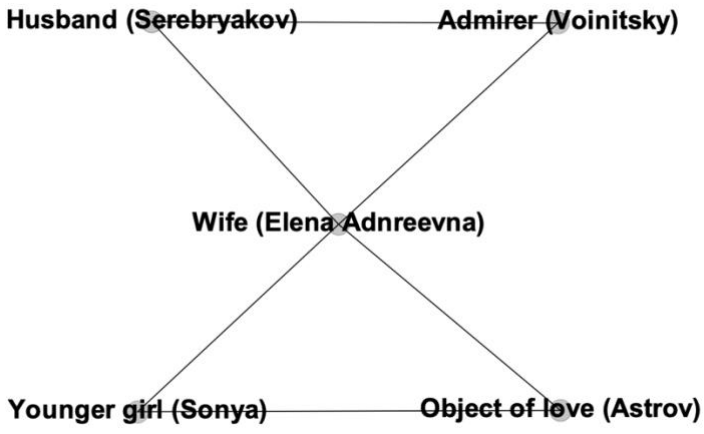


Figure 8. Love triangles in *Uncle Vanya*

We see two structurally identical sets of love triangles, formed by comparable characters. Looking at these graphs and taking into account the fact that they feature the key

characters in the plays, we might ask ourselves: are there at least any structural differences between the plots of *A Month in the Country* and *Uncle Vanya*? In this chapter I will explore how quantitative methods can help enhance the analysis of these two plays with very similar, at first sight, network structures and plots. I will use several approaches to evaluate the input of the characters for each play, in an attempt to highlight the differences in the overall structures of the interactions between the characters. But let us first explore further in what ways *A Month in the Country* and *Uncle Vanya* seem so similar.

A Month in the Country was written by 1850, but was first published only in 1855 due to censorship issues. It is a five-act comedy with thirteen characters. However, Turgenev later called it a short novel in dramatic form, and insisted that *A Month in the Country* was not suited for the stage.

Set in the 1840s in the private estate of a rich landowner, Islaev, the plot of the play focuses on multiple love triangles. The parties involved are Natalya Petrovna – Islaev's wife, Vera – their foster-daughter, Aleksey Belyaev, and Rakitin – the family friend. Belyaev is a young teacher of Russian who has come from Moscow to teach the Islaevs' son Kolya. Both Natalya Petrovna and Vera fall in love with Belyaev, and Rakitin has been a long-time admirer of Natalya Petrovna. Knowing about both Rakitin's and Vera's feelings, Natalya Petrovna nevertheless cannot resist approaching Belyaev. Natalya Petrovna reveals Vera's secret to Belyaev in order to establish that he is not in love with Vera, and then she confesses her own love to Belyaev. Confused and disturbed by Natalya Petrovna's confession, Belyaev eventually leaves the estate, and so does Rakitin. Vera and Natalya Petrovna are left heartbroken and unsettled. We will now see how similar this plot is to *Uncle Vanya*.

Uncle Vanya was written in 1896 and published in 1897. Chekhov called his play "scenes from country life". It has four acts and nine characters. Its plot revolves around the

family of Serebryakov, a retired professor. We see, again, several love triangles, formed by Elena Andreevna – Serebryakov's young wife, Sonya – his daughter, Voinitsky (Vanya) – his brother-in-law on his late wife's side, and Voinitsky's friend, the town's doctor – Astrov. Voinitsky is in love with Elena Andreevna, Sonya is in love with Astrov, and Astrov also has feelings for Elena Andreevna. Elena Andreevna tries to help Sonya to connect with Astrov, but, instead, almost starts a love affair with Astrov herself. Together with this, the atmosphere becomes heated because of Voinitsky's conflict with Serebryakov, and the professor and Elena Andreevna eventually leave the estate.

Both plays were considered unusual for the theatre. *A Month in the Country* was often called a "play for reading", while *Uncle Vanya* was criticised for "not going anywhere" (Tolstoy about *Uncle Vanya*, Semyonov 1913). The ways the plays have been approached by the critics are also similar.

Soviet literary scholars (Berdnikov 1953, Vishnevskaya 1989) chose to read *A Month in the Country* as a manifestation of the decaying character of the gentry and its purposeless lives. They also focused their interpretation of the play on the conflict between two social groups – the gentry and the classless people (raznochintsy). In the introduction to his translation of the play, Isaiah Berlin (Turgenev 1981) recognises that Soviet critics were obliged to follow this interpretation. The strong Russian reactions to the European revolutions of 1848, when Turgenev was working on the play, would have made it impossible for the audience to miss the harsh social criticism in *A Month in the Country*, says Berlin. Like Berdnikov and Vishnevskaya, Berlin also notices Belyaev's resemblance to Vissarion Belinsky through his name, interests and background: Belinsky was a Russian literary critic and a revolutionary thinker, who supported the ideas of socialism and democratic revolution. Turgenev befriended him in 1843, not long before Belinsky's death in

1848. Turgenev was very fond of Belinsky and, to quote Berlin, he "remained devoted [to Belinsky, both as a man and a critic,] all of his life". The assumption that Belinsky was somehow a prototype of the character of Belyaev imparts a very strong political tone to the play.

Consequently, Berlin acknowledges the great importance of two particular ways of interpreting *A Month in the Country*, including the socio-political interpretation. He writes that the play "may be approached in many ways, but if either the psychological realism of the 'comedy' […] or the faith of the intelligentsia incorporated in it are ignored, the point of the play, the purpose of the writer will be missed". However, he also says that the political interpretation of *A Month in the Country* "has been driven too far" and that "there had been no mention […] of these pronounced social and political features of the play […] before the Revolution" (referring to the Russian Revolution of 1917). The 'traditional', as Berlin calls it, interpretation, proposed by both Russian and Western scholars, focuses on the psychological realism of the play that uncovers the personal tragedies of the characters. Richard Freeborn (1985) says that the social element of the play is important for its understanding, but it is secondary to the emotional relationships between the characters. This is also how Stanislavsky saw *A Month in the Country*, and he was the one who managed to direct the first truly successful production of the play in 1909.

Uncle Vanya repeats, in a sense, the destiny of *A Month in the Country*: Soviet critics laid particular emphasis on the characters' involvement with the world of work. The professor and his wife are contrasted to Voinitsky, Astrov and Sonya. The latter are hard-working people who dedicate themselves to others, while the former are lazy, exploitative, useless and fake (Ermilov 1946). However, Pitcher argues that we should not reduce *Uncle Vanya*'s meaning to moralisation. Moreover, Tolstoy criticised the play for not

actually having any moralising element, and said that Voinitsky and Astrov were mediocre and their ideas were weak (Semyonov 1913).

The theme of work is constantly present in the play, though it does not necessarily provide a 'right or wrong' distinction. According to Charles Meister, Chekhov "goes beyond romance to examine *the philosophy of work* as his unifying theme". And Pitcher (1985) points out that the play focuses on the contrast between the takers and the givers, i.e. between people who see work as a contribution to humanity and people whose work is "aimed at trying to satisfy a thirst for glory and self-advertisement" (p. 90). He says that work became an essential aspect of the play in order to give Voinitsky, Astrov and Sonya an extra emotional dimension, and not to demonstrate the moral contrast between them and the professor.

As we can see, *Uncle Vanya*'s plot is apparently similar to the plot of *A Month in the Country*. Both Belyaev and Astrov are not part of the families, they disturb Natalya Petrovna and Elena Andreevna respectively, and get involved in relationships with them. Elena Andreevna, like Natalya Petrovna, reveals the younger girl's feelings to Astrov, who then invites Elena Andreevna to have a love affair with him instead.

The differences between the parallel characters are also obvious for the most part. Belyaev, however, is drastically different from Astrov. The former is a young, naïve student, while the latter is a not-so-young doctor, already disappointed in life. Rakitin and Voinitsky are as well completely dissimilar in their mode of life. Voinitsky is a hardworking man and Rakitin is a bored intellectual. Natalya Petrovna is an active actor, who is pursuing her feelings for Belyaev and *consciously* ruining his friendship with Vera. Elena Andreevna, by contrast, is a passive, "episodic character", as she describes herself. She *almost unknowingly* interferes with the lives of Voinitsky, Sonya and Astrov. Moreover, in *A Month in the*

Country Belyaev is an outsider, whereas in *Uncle Vanya* the outsiders seem to be Serebryakov and Elena Andreevna. And while Belyaev, by the logic of comparison, should correlate with Astrov, it is Elena Andreevna who he seems to be equivalent to, because he is both an outsider and a catalyst (Freeborn 1985) of the play's action.

These are some of the differences in the personalities of the characters, that are complemented by other structural and plot-related differences between the plays. As Briggs (1994) wrote, "No-one would wish to say […] that *Uncle Vanya* is a rewriting of *A Month in the Country*. […] many resemblances between these two plays do not indicate equivalence of story […]" (p. 30).

I am going to explore which similarities and, most importantly, which differences between the two systems of characters, with very similar dramatis personae and patterns of interactions, can be reflected with quantitative methods. There are certain features that have made *A Month in the Country* and *Uncle Vanya* so different that many critics would not care to juxtapose them. The question is, which of these features can be shown with exact methods, and how.

Basic network analysis

Let us now examine the most basic values of the network graphs of *A Month in the Country* and *Uncle Vanya*. These graphs will provide us with a sense of the extent to which the characters are connected with each other within the plays. We consider two characters to be connected if they both speak within any single act. Thus, on these network graphs we will see the overall structure of the character systems and the volume of communication in the plays.

The characters are represented by the nodes, where the size of a node corresponds to the degree of that node. If a character is connected to three other characters, its degree will be 3 and its node will be larger than that of the character with a degree of 2. The connections between the characters, or the nodes, are visualised as the edges of the graphs. The thickness of the edges corresponds to the weight of the connection between the characters. If two characters both speak within the same two acts, their connection will have a weight of 2, and the edge between their nodes will be two times thicker than between the nodes of the characters who only speak together within one act.

Below is a network graph of *A Month in the Country* constructed on this basis.

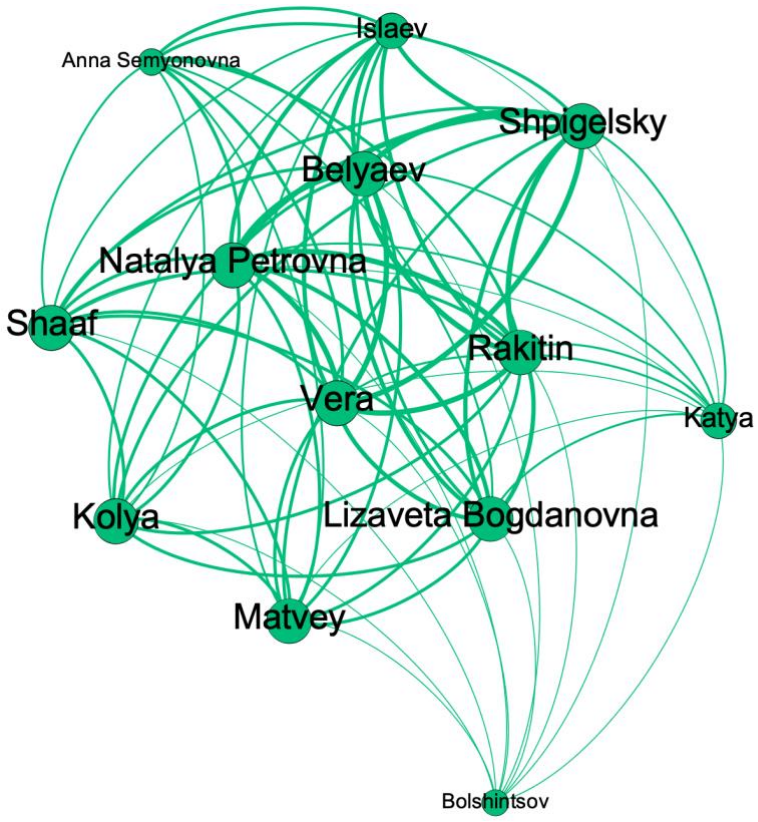


Figure 9. Network graph of *A Month in the Country*

All the nodes are well-connected to each other and most of them have the same degree. We cannot spot any particularly significant protagonists at the moment.

When writing a first version of the play, Turgenev called it ‘The Student’. Later he renamed it as ‘Two Women’. It seems reasonable, then, to expect that the two major plot-focusses in the play would be the student himself and the rivalry between the two women, as Belyaeva (2005) suggests. This would make us assume that either Belyaev, or Natalya Petrovna and Vera are supposed to be the ‘main protagonists’. However, the title *A Month in the Country* suggests refocusing the plot away from certain characters, to all the interactions happening within this month in the country. As Vishnevskaya (1989) points out, in *A Month in the Country* Turgenev is talking about a ‘nest of gentry’ (referring to the title of one of Turgenev’s later novels), as a whole social phenomenon.

This is exactly what we see from the graph. Its density is 0.96, and nine out of thirteen characters have a degree of 12, which is the maximum possible degree in this play. These nine characters are Islaev’s family (Natalya Petrovna, their son Kolya, and their foster-daughter, Vera), their staff (Shaaf, Lizaveta Bogdanovna and Matvey), plus Rakitin, Belyaev and the doctor (Shpigel’sky). Islaev himself is not included in this tightly connected community because he is almost always off the social scene, busy working in the estate.

Consequently, this group of nine characters with the highest degree also shares the highest betweenness, closeness and Eigenvector centralities, which all serve to contrast them with the other four characters.

If we look at the weighted degree of the nodes, we will see a slightly different picture. Weighted degree takes into account the weight of the edges by which a node is connected to other nodes. To have a high weighted degree, it is not enough for a character to ‘meet’ everybody else just once. They also need to interact with (some of) them more often than with other characters, meaning that the number of acts that they ‘share’ (speak within together) with other characters should be higher. This allows us slightly to overcome the

illusion of almost all the characters being equally active and ‘important’. This illusion comes from the fact that there are only five segments (acts) in the play. It is quite easy for a character to ‘meet’ all the other characters at least once, even if this character speaks only once during the play. Moreover, weighted degree releases us from comparing the act-based presence of characters, since the weight of an edge between the characters is determined by the number of acts where they both speak.

Character	Weighted Degree
Belyaev	43
Natalya Petrovna	43
Rakitin	43
Shpigelsky	43
Vera	43
Lizaveta Bogdanovna	37
Islaev	33
Kolya	30
Matvey	30
Shaaf	30
Anna Semyonovna	26
Katya	17
Bolshintsov	10

Table 13. Weighted degree of characters in *A Month in the Country*

The characters with the highest weighted degree are Belyaev, Natalya Petrovna, Rakitin, Shpigelsky and Vera. Then we see Lizaveta and Islaev, followed by Kolya, Shaaf and Matvey, and then by Anna Semyonovna and Katya. Kolya, Shaaf, Lizaveta Bogdanovna and Matvey are thereby now excluded from the group of most active characters. And the

character with the lowest weighted degree is the neighbour, Bolshintsov. Even though it is not surprising that Bolshintsov is the one with the lowest weighted degree, this is not entirely due to the fact that he has, as well, the lowest unweighted degree. He and Anna Semyonovna both have the lowest unweighted degree of 10, but Bolshintsov's weighted degree is much lower than Anna Semyonovna's. This supports his profile of an 'outsider', who not only has the lowest number of connections but also appears on the stage only once. We can say therefore say that the weighted degree allows us to make a more precise judgement of the characters' involvement in interactions with other characters. It already refocuses our attention on the most essential characters of *A Month in the Country*. We will see later how a different network representation of the play divides the characters even more, according to their importance and activity.

Now, let us compare the basic network structure of *A Month in the Country* to the *Uncle Vanya* network graph:

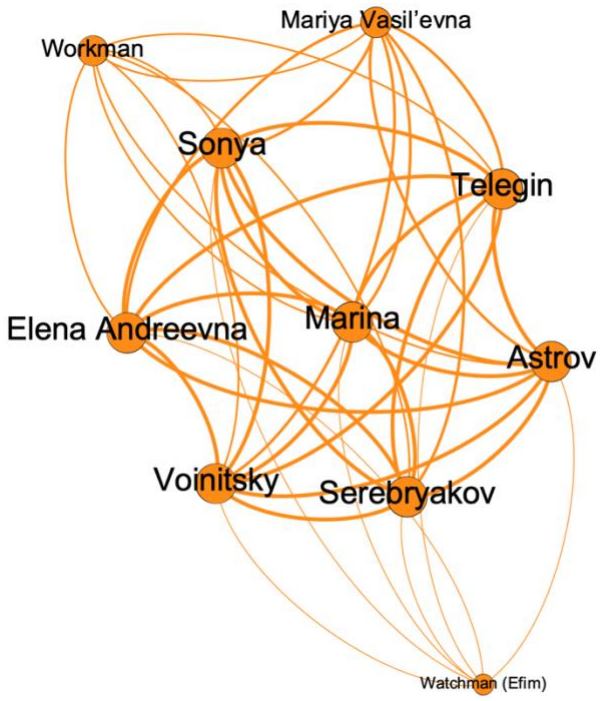


Figure 10. Network graph of *Uncle Vanya*

Uncle Vanya is known to be a transformation of Chekhov's earlier play *The Wood Demon* (Rayfield 1975, Pitcher 1985). In comparison to *The Wood Demon*, *Uncle Vanya* is shorter and some of the characters of *The Wood Demon* were removed from it. Rayfield (1975) says that, as a result of this transformation, the focus of the play shifted from the doctor to the uncle. Thus, we might expect Voinitsky to occupy the central position within the structure of the play's characters.

However, we can see a graph picture almost identical to *A Month in the Country*'s. The graph's density is also 0.96 (0.956). If we include Efim, the watchman, in the list of the characters, the play will have ten characters. And seven out of ten characters have a degree of 9, which is the maximum for this play. The well-connected characters with the highest degree are Serebryakov's family, including the 'head of the family' (the professor), Astrov, the family friend Telegin and their servant Marina. The only person from the family who has a slightly lower degree (of 8) is Mariya Vasil'evna, the mother of Serebryakov's late wife and of Voinitsky. This is because during the second act, where Efim talks to Elena Andreevna, Mariya Vasil'evna is partly present on stage, but she does not speak. So technically, she also should be included in the tight 'family and friends' circle of equally 'important' or active characters.

Everybody except Mariya Vasil'evna, the workman and Efim (the watchman), shares the same and the highest betweenness, closeness and Eigenvector centralities, as well as **the same weighted degree**. In this way, we can say there are no 'outsiders' or peripheral actors in *Uncle Vanya*, such as Bolshintsov in *A Month in the Country*.

Again, as in *A Month in the Country*, the characters with the highest degree have the same and the highest betweenness, closeness and Eigenvector centralities. The plot seems to

be constantly focusing on the interactions between all the 'significant' characters (excluding the staff).

The network graph of *Uncle Vanya*, where the majority of characters are connected with all the other characters, might reflect Chekhov's ideas about moving beyond the concept of having a distinct hero in a play. Pitcher (1985) writes that "one of Chekhov's most important innovations was to decentralize his cast. The traditional idea of a hero and heroine has been discarded. No single character is allowed to stand out as more central than any other […] every member of the group knows everyone else" (p. 14). Pitcher explains that such a structure was necessary for Chekhov to create "psychological intimacy", in order to explore the emotions within "a group of people whose lives are already connected". Rayfield (1975) also noted that Chekhov's plays of 1890s and 1900s have a polyphonic structure, which means that they are not focused around individual characters. We will see later if other ways of quantitative analysis of the play highlight the 'polyphonic' nature of *Uncle Vanya*.

The main concern of this chapter is to investigate whether there is a way to distinguish *A Month in the Country* from *Uncle Vanya* within the proposed framework, which includes simple quantitative methods and network theory. For the moment, comparing the graphs in Figures 7 and 8 and Figures 9 and 10, the plays look structurally identical.

As mentioned above, the fact that all the characters in both plays are very well-connected to each other might be due to Turgenev's and Chekhov's ideas about prioritising the community over individuals. However, I believe that it is also connected to the number of segments in the play. *A Month in the Country* and *Uncle Vanya* have five and four segments, namely acts, respectively. While in some other plays, for example in *The Government Inspector*, there are also five acts, these are split into 52 scenes. Having only five large segments in the play makes it hard for an author, even if they wished, to distance characters

and to allow some smaller communities to form in such a way that we would see it on a network graph. Moreover, acts and scenes in a play might often represent a structural function for a text, but they do not necessarily offer guidance for a theatre production. These segments of a play are not static in their settings: characters who talk during the same act or scene do not need always to be present on stage all together at the same time. A more precise way of building speech-based networks, such as the ones that we have been discussing so far, might be on the basis of stage directions that indicate the entrances and exits of characters. For these reasons, I have introduced 'configuration changes' to the TEI mark-up of the plays. This might help us to see more accurate speech-based networks in the plays, as well as to distinguish more active characters, around whom the plots revolve for most of the time.

'Configuration change' mark-up

So far I have been using two types of tags (indicating division in to segments) in the plays' mark-up – act or scene. I manually added tags to the TEI files of *A Month in the Country* and *Uncle Vanya*. The new configuration 'opens' when any character either enters or leaves the stage, according to the stage directions. This has allowed me to split the plays into smaller segments and to build networks based on these segments. The mark-up was not always as straightforward as it may seem. While most of the time it was quite clear how to split the text into configuration s, there might be questionable cases. Let us look at typical examples of from *A Month in the Country*. These two extracts come from the first act of the play. The first one introduces a conversation between Rakitin and Natalya Petrovna. They start talking about the changes which he has spotted in her face and her manners, when Kolya interrupts them.

```
<sp who="#rakitin">
    <speaker>Ракитин.</speaker>
    <p> А вот погодите...
    </p>
</sp>
</div>
<div type='configuration'>
    <stage> (Коля вдруг с шумом вбегает из залы прямо к Анне
        Семеновне.) </stage>
<sp who="#kolja">
    <speaker>Коля.</speaker>
    <p> Бабушка, бабушка! посмотри-ка, что у меня! <stage>(Показывает ей лук и
            стрелы.)</stage> Посмотри-ка! </p>
</sp>
```

Figure 11. Configuration change examples in *A Month in the Country*

The next extract shows how Shaaf, the teacher, makes Kolya leave the room to go to his lessons, and Natalya Petrovna starts questioning Vera about what she, Kolya and Belyaev were busy with in the garden.

```
<sp who="#kolja">
    <speaker>Коля</speaker>
    <stage>(уходя, шёпотом Ракитину).</stage>
    <p> А ты всё-таки клей прикажи... <stage>(Ракитин кивает.)</stage>
    </p>
</sp>
<sp who="#shaaf">
    <speaker>Шааф</speaker>
    <stage>(дергая Колю).</stage>
    <p> Kommen Sie, mein Herr... <note place="foot">Идемте, сударь... (Нем.)</note>
        <stage> (Уходит с ним в залу. Ракитин уходит вслед за ними.) </stage>
    </p>
</sp>
</div>
<div type='configuration'>
<sp who="#natalja_petrovna">
    <speaker>Наталья Петровна</speaker>
    <stage>(Вере).</stage>
    <p> Сядь... ты, должно быть, устала... <stage>(Садится сама.)</stage>
    </p>
</sp>
```

Figure 12. Configuration change examples in *A Month in the Country*

In Figure 11, the new configuration starts when Kolya enters the stage. In Figure 12, it starts after Shaaf, Kolya and Rakitin leave the stage.

It is necessary to note that a network that uses configuration changes as segments still has some of the limitations of speech-based networks. It allows us to eliminate most of the connections between characters who do not really communicate with each other but just happen to speak during the same scene or act. However, it still does not account for characters who are present in a certain segment (configuration setting) and hear what other characters say, but do not speak. Here is an extended version of the example in Figure 11:

```
<sp who="#natalja_petrovna">
    <speaker>Наталья Петровна.</speaker>
    <p> Да?.. В таком случае сделайте одолженье... Вы ведь меня знаете – <pb n="290"/>
        <pb n="291"/> угадайте, в чем состоит эта перемена, что во мне такое
        произошло – а? </p>
</sp>
<sp who="#rakitin">
    <speaker>Ракитин.</speaker>
    <p> А вот погодите...
    </p>
</sp>
</div>
<div type='configuration'>
    <stage> (Коля вдруг с шумом вбегает из залы прямо к Анне
        Семеновне.) </stage>
<sp who="#kolja">
    <speaker>Коля.</speaker>
    <p> Бабушка, бабушка! посмотри-ка, что у меня! <stage>(Показывает ей лук и
            стрелы.)</stage> Посмотри-ка! </p>
</sp>
<sp who="#anna_semenovna">
    <speaker>Анна Семеновна.</speaker>
    <p> Покажи, душа моя... Ах, какой славный лук! кто тебе его сделал? </p>
</sp>
<sp who="#kolja">
    <speaker>Коля.</speaker>
    <p> Вот он... он... <stage> (Указывает на Беляева, который остановился у двери
            залы.) </stage>
    </p>
</sp>
</div>
```

Figure 13. Extended version of the example in Figure 11

Kolya enters the stage when Natalya Petrovna and Rakitin are talking to each other. He only speaks to Anna Semyonovna, and then a new configuration segment begins when Belyaev enters the stage. While Natalya Petrovna and Rakitin hear everything that Kolya says to Anna Semyonovna, there will not be a connection (an edge) between them and Kolya in this segment, because they do not say anything in this configuration segment. For all that,

this limitation also exists in the act- or scene-segmented type of networks that we discussed. One could argue that such a feature, in this case, strengthens the approach of focusing on connections between the characters who *really communicate* with each other.

Another important issue in splitting the play into configuration changes is ambiguous mark-up cases. Let us take a look at the following examples from *A Month in the Country*. In Figure 14 we see an extract from the first act of the play, when Shpigelsky, the doctor, arrives at the estate and Matvey, the servant, introduces him. They find Anna Semyonovna and Shaaf playing cards.

```
<div type='configuration'>
<sp who="#matvej">
    <speaker>Матвей</speaker>
    <stage>(входит из залы и докладывает).</stage>
    <p>Игнатий Ильич приехали-с.</p>
</sp>
<sp who="#shpigelskij">
    <speaker>Шпигельский</speaker>
    <stage>(входя по его следам).</stage>
    <p> Об докторах не докладывают. <stage>(Матвей уходит.)</stage> Нижайшее мое
        почтенье всему семейству. <stage>(Подходит к Анне Семеновне к
            ручке.)</stage> Здравствуйте, барыня. Чай, в выигрыше? </p>
</sp>
</div>
<div type='configuration'>
<sp who="#anna_semenovna">
    <speaker>Анна Семеновна.</speaker>
    <p> Какое в выигрыше! Насилу отыгралась... И то слава богу! Всё вот этот злодей.
            <stage>(Указывая на Шаафа.)</stage>
    </p>
</sp>
<sp who="#shpigelskij">
    <speaker>Шпигельский</speaker>
    <stage>(Шаафу).</stage>
    <p> Адам Иваныч, с дамами-то! это нехорошо... Я вас не узнаю. </p>
</sp>
```

Figure 14. Problematic case for configuration change mark-up in *A Month in the Country*

First Matvey enters the stage and announces that Shpigelsky has arrived. Then Shpigelsky enters the stage and addresses Matvey. Then Matvey leaves and Shpigelsky starts talking to everybody else who was on stage before he arrived. Because the stage direction that indicates that Matvey leaves the stage is in the middle of Shpigelsky's speech-act, it is

technically not possible to start a new configuration segment right after Matvey leaves the stage. The XML document structure must be hierarchical and there must not be any cross-over tags, therefore tags must enframe <sp> tags and cannot be opened within <sp> elements. Thus, there are two options for how to mark up this piece of *A Month in the Country*. The first one is to create a short configuration segment for Matvey's speech-act, and open a new configuration when Shpigelsky enters after Matvey. Another option is to create a short configuration segment for Matvey's and Shpigelsky's speech-acts, and open a new configuration after Matvey leaves the stage, i.e. after Shpigelsky's speech-act that starts with "Об докторах не докладывают". I decided to choose the second option, since otherwise it would look like Matvey is talking to himself and nobody hears him, while, in fact, everybody hears him, and Shpigelsky even addresses him in the beginning of his speech-act. This step would not create a connection between Matvey and Anna Semyonovna or Matvey and Shaaf. But it would at least preserve the connection between Matvey and Shpigelsky.

Figure 15 shows us an extract from the first act with the conversation between Islaev and Rakitin, when Belyaev enters the room, interrupting them.

```
<sp who="#rakitin">
    <speaker>Ракитин.</speaker>
    <p>Кто такое Беляев?</p>
</sp>
</div>
<div type='configuration'>
<sp who="#islaev">
    <speaker>Ислаев.</speaker>
    <p> А новый наш учитель, русский. Дичок еще порядочный; ну, да привыкнет. Малый
        неглупый. Я его попросил сегодня посмотреть, что постройка... <stage>(Входит
            Беляев.)</stage> А, да вот и он! Ну, что? Как там? Ничего не делают
        небось? А? </p>
</sp>
<pb n="303"/>
<sp who="#beljaev">
    <speaker>Беляев.</speaker>
    <p>Нет-с; работают.</p>
</sp>
```

Figure 15. Problematic case for configuration change mark-up in *A Month in the Country*

There is a similar issue here as in Figure 14. Belyaev enters the stage in the middle of Islaev's speech-act. Because of the reasons discussed above, it is not possible to split Islaev's speech-act, so that its first part would be in the segment with Rakitin and its second part in the segment with Belyaev. I assume that Belyaev does not hear Rakitin's response "Кто такое Беляев?", but hears and replies to Islaev's question. Moreover, Rakitin does not speak till the next configuration change, when Natalya Petrovna enters the stage. Therefore, I decided to split Rakitin's and Islaev's dialogue with a configuration change and included Islaev's response into the same configuration with Belyaev's first response.

Whereas most of the configuration changes mark-up were quite easy to decide, there were debatable cases from time to time. This required me to make certain decisions based on my personal judgement. I tried to stick to the idea of preserving connection between the characters who heard each other and then spoke to each other, where possible. I believe there are not any ultimately *right* choices when doing such work, and therefore this mark-up might vary when done by different annotators. Nevertheless, such cases are not very frequent in the plays and should not significantly affect their network metrics.

Network analysis of A Month in the Country

Network graph and centralities

Below is the new network graph for *A Month in the Country*. It is based on the configuration changes mark-up. The connection between two characters is now established if these characters both speak within a single configuration segment. Each segment now corresponds to a single stage setting, meaning that there is a static group of characters appearing on stage at the same time. On such a network graph we would expect to see characters connected in a more rational way, rather than on a graph based on connections within acts. We should not see any more close connections between the characters who appear on stage rarely and the ones who are active during the whole play.

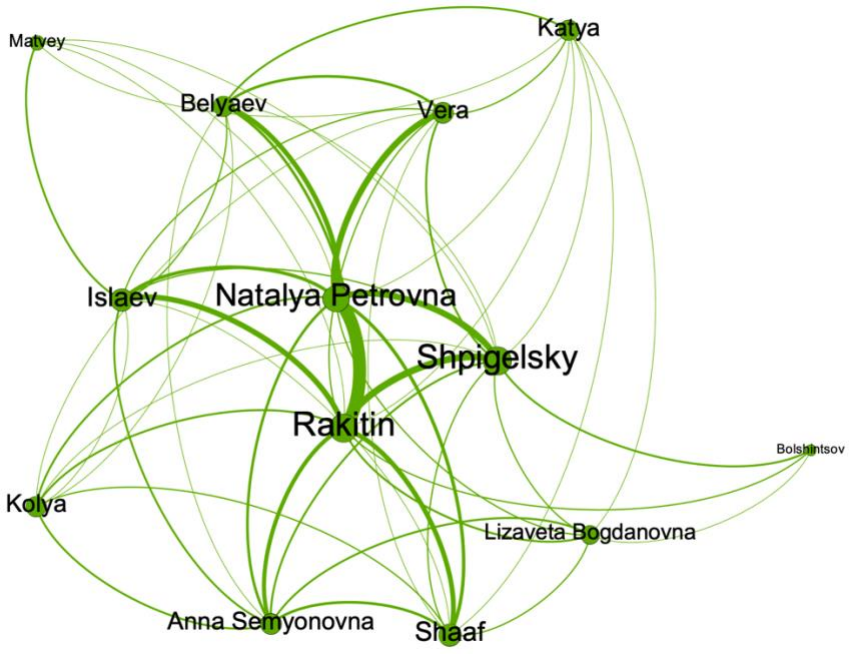


Figure 16. Network graph of *A Month in the Country* based on configuration changes

And indeed we see that here many characters are no longer part of the tight and equal community — the one that we saw on the graph in Figure 9. The size of the nodes and the thickness of the edges between them in this new network reflects the fact that there are some

characters who are apparently more important and influential network-wise, and also that there are pairs of characters who have a stronger connection than others. We will now explore closely what the new network values are that the characters possess, and which relationships are focused on more than others within *A Month in the Country*.

In *Turgenev: a reading of his fiction*, Seeley (1991) splits the characters of *A Month in the Country* into three groups: principals, subsidiary roles and the doctor, who is *in between* the two groups. Seeley considers Natalya Petrovna to be the key protagonist of the play, but he says "she does not dwarf or eclipse the other personages: the cast includes at least four other fully drawn characters". He also talks about the percentage of stage time these four characters have together with Natalya Petrovna. According to his calculation, Natalya Petrovna is on stage during 64% of performance time and Rakitin – during 57%. Belyaev, Vera and Islaev occupy 34, 27 and 10% of the performance time respectively. We will soon see what kind of distribution of presence and activity among the characters is drawn by the network metrics, and how Seeley's interpretation is reflected by the new network.

The density of the configuration-based network is 0.69, compared to 0.96 of the acts-based network. Because in this new network characters are not that densely connected to each other, we can spot specific descriptors for some of them when interpreting the characters' centralities values.

Character	Degree
Shpigelsky	12
Rakitin	12
Natalya Petrovna	11
Islaev	9
Shaaf	9

Kolya	8
Anna Semyonovna	8
Vera	8
Belyaev	8
Katya	8
Lizaveta Bogdanovna	7
Matvey	5
Bolshintsov	3

Table 14. Degree of characters in *A Month in the Country*, based on configuration changes mark-up

Character	Betweenness centrality
Shpigelsky	6.73
Rakitin	6.73
Natalya Petrovna	2.89
Lizaveta Bogdanovna	1.50
Katya	1.38
Islaev	1.37
Shaaf	1.22
Anna Semyonovna	0.73
Belyaev	0.62
Vera	0.43
Kolya	0.27
Matvey	0.14
Bolshintsov	0.00

Table 15. Betweenness centrality of characters in *A Month in the Country*, based on configuration changes mark-up

Character	Closeness centrality
Shpigelsky	1.00
Rakitin	1.00
Natalya Petrovna	0.92
Islaev	0.80
Shaaf	0.80
Katya	0.75
Anna Semyonovna	0.75
Belyaev	0.75
Vera	0.75
Kolya	0.75
Lizaveta Bogdanovna	0.71
Matvey	0.63
Bolshintsov	0.57

Table 16. Closeness centrality of characters in *A Month in the Country*, based on configuration changes mark-up

Character	Eigenvector centrality
Shpigelsky	1.00
Rakitin	1.00
Natalya Petrovna	0.97
Shaaf	0.84
Islaev	0.83
Kolya	0.78
Vera	0.78
Belyaev	0.77
Anna Semyonovna	0.77

Katya	0.73
Lizaveta Bogdanovna	0.63
Matvey	0.51
Bolshintsov	0.30

Table 17. Eigenvector centrality of characters in *A Month in the Country*, based on configuration changes mark-up

Character	Weighted degree
Rakitin	66
Natalya Petrovna	65
Shpigelsky	38
Islaev	27
Anna Semyonovna	27
Shaaf	25
Vera	24
Belyaev	23
Kolya	15
Lizaveta Bogdanovna	14
Katya	11
Matvey	7
Bolshintsov	6

Table 18. Weighted degree of characters in *A Month in the Country*, based on configuration changes mark-up

There are now only two characters with the maximum degree of 12 – Rakitin and Shpigelsky. They also have the highest betweenness, closeness and Eigenvector centralities. This means that they are the 'bridges' for the information flow between the characters of the

play, and they are well-connected to the 'important' or network-wise active characters. All of this makes them network-wise the most influential characters in *A Month in the Country*.

It is interesting to see Shpigelsky stepping forward in the network structure. His character is described by Berdnikov (1953) as an intelligent, sharp-sighted person, who hates and despises the entire noble circle with which he is forced to mix. This hatred comes from his awareness of a social inequality to which he cannot be reconciled. However, Shpigelsky not only 'meets' everyone, but stands between some of the characters' interactions. He is fighting his way to the top by establishing connections with everyone, so that he can later cut deals with them. As Grossman (1924) says, Shpigelsky is an organiser of family affairs. Shpigelsky is indeed situated in between the two groups of the characters, as Seeley points out. Freeborn (1985) even calls him a *go-between*. All these observations are supported by Shpigelsky's centralities values, especially by the fact that he has the highest betweenness centrality.

Rakitin has the highest centralities, probably for different reasons than Shpigelsky. Rakitin is always present on the social scene of the overall community as he is a well-established member of it. As a long-time admirer of Natalya Petrovna, he keeps in contact with everyone in the estate in order to learn about any news and about Natalya Petrovna's moods, when he is unable to find this out from her directly. Rakitin is also simply a good old friend of Islaev's. This, and the fact, that a major part of the play's action focuses on his relationship with Natalya Petrovna, provides Rakitin with such high network values.

Natalya Petrovna follows Rakitin and Shpigelsky with a degree of 11 and the second highest betweenness, closeness and Eigenvector centralities. The only difference between her and Rakitin and Shpigelsky is that she has no connection with Bolshintsov. She gets to know

about his plans from Shpigelsky, because of his most evident ‘deal’ in the play. Otherwise, she is as influential and important as her admirer and the doctor.

It is curious that Belyaev and Vera occupy the ninth and tenth place, when ranked by betweenness centrality. They are situated in that ranking below everybody else, except Kolya, Matvey and Bolshintsov. This reflects the fact that Vera and Belyaev do not participate in connecting others or transmitting messages. I would rather say that they are dragged into some interactions by the more active characters, namely, Natalya Petrovna, Rakitin, Shpigelsky and Islaev. We will see that Belyaev and Vera occupy a middle position in the network and, were it not for Natalya Petrovna’s affection for Belyaev, they would probably have even lower network values.

As for the weighted degree, which corresponds both to the number and to the volume of interactions that each character has, Rakitin and Natalya Petrovna occupy the highest position with the weighted degree of 66 and 65, respectively. This tells us that they are the most active characters in terms of being present on stage and talking to others. It also provides more evidence for the contention that the main focus of *A Month in the Country* is Natalya Petrovna’s life.

Shpigelsky takes the third position in the weighted degree rank, but with a much lower value of 38. He does ‘meet’ everyone in the play and acts as a connecting bridge between the characters, but he is not communicating with other characters as frequently as Rakitin and Natalya Petrovna are. Perhaps Shpigelsky does not want or does not need to have such tight connections to others. Moreover, he does not live in the estate, which leaves him fewer opportunities for more frequent communication with others.

Anna Semyonovna, Islaev, Shaaf, Vera and Belyaev have a degree of 9 or 8 and a weighted degree in the range of 27-23. They take up the middle tier of the weighted degree

rank. The difference between the weighted degree of these five characters and the weighted degree of Shpigelsky is not that drastic, even though they have a much lower unweighted degree than him. This emphasises even more the outstandingly high weighted degree of Rakitin and Natalya Petrovna, considering that their other centrality values are either equal to or lower than Shpigelsky's. We will talk about other reasons for this a bit later.

It is interesting to compare Anna Semyonovna's and Lizaveta Bogdanovna's metrics. Lizaveta Bogdanovna's unweighted degree is only 1 point lower than the degree of Anna Semyonovna. Their closeness and Eigenvector centralities are quite close to each other. But Anna Semyonovna has a weighted degree almost twice as large as Lizaveta Bogdanovna. The former, in her turn, has twice as large a betweenness centrality as Anna Semyonovna. I would say that these measures characterise Lizaveta Bogdanovna as a member of the household staff, who needs to be here and there, connecting people, letting them know where somebody is or how somebody feels. The higher weighted degree of Anna Semyonovna may correspond to her being a family member who needs to keep in touch with her family and the residents of the estate, without intentionally participating in all the plot intrigues.

There is another case that demonstrates how different social position, occupation and behavioral patterns are reflected in the network values. Kolya and Katya do not have a high weighted degree. They both have the same unweighted degree and closeness centrality. They also have quite close Eigenvector centralities. However, it is peculiar that Kolya's betweenness centrality is only 0.27 compared to Katya's betweenness centrality of 1.38. Katya's metrics look very similar to Lizaveta Bogdanovna's. This informs us, again, that servants and staff members who are not greatly involved in communication tend to have a higher betweenness centrality than more active characters. Thus, Lizaveta Bogdanovna and

Katya occupy the fourth and the fifth place in the ranking by betweenness centrality, right after Shpigelsky, Rakitin and Natalya Petrovna.

Moving on to the periphery of the play's network, Matvey and Bolshintsov have the lowest weighted degree, as expected. Bolshintsov has no claim to be included in the main plot line and social circle. But in the previous network graph in Figure 9, Matvey was one of the nine characters with the highest network values. But now, the configuration changes mark-up approach separates him from that group of characters, who are apparently more active and 'important'. In the new network his position is always the second from last, according to the centralities values.

To summarise this section, the new network of *A Month in the Country*, based on the mark-up of configuration changes, allows us to identify different categories of the characters within the play. We see that Natalya Petrovna, Rakitin and Shpigelsky occupy the highest positions regarding all of the centralities values. We would not be so aware of the evidence of their activity and influence in the act-based network which we considered before the configuration changes mark-up. At this time, we can tell that neither Belyaev nor Vera carry as important a storyline in the play as Natalya Petrovna and Rakitin. And even though Belyaev acts as a catalyst, the major theme of *A Month in the Country* is to show how other characters react to him and what changes come about in their moods and lives, because of the feelings that he evokes in them.

We have seen how the characters' communication shapes their network positions, and we have identified the most active and influential characters of the play. Let us now see how their communication is distributed between themselves – what pairs of characters get more attention from the author and the audience, and how that correlates to the plot of the play.

Edges weight and 'dialogues'

To understand where exactly the weighted degree of characters comes from, we should look at the weights of the edges between them and all the other characters. The weight of an edge between two characters is the number of segments in which both of the characters appear on stage and speak, i.e. the number of 'common' segments between two characters. By looking at the weight of edges, we can see the most connected pairs of characters and assume that they must be some of the 'most important' characters of the play, as there is a lot of focus on their interactions and communication in the play's text.

When working with the act-based mark-up of the play, it would not make much sense to take into account the edge weights, since the majority of characters are connected to each other via five or four acts. However, there are distinctive values in the weights of edges between the characters if we examine them in the network based on the configuration changes mark-up:

Character 1	Character 2	Weight
Natalya Petrovna	Rakitin	19
Rakitin	Shpigelsky	10
Natalya Petrovna	Vera	9
Natalya Petrovna	Shpigelsky	9
Islaev	Rakitin	8
Rakitin	Shaaf	7
Belyaev	Natalya Petrovna	7
Anna Semyonovna	Rakitin	6
Natalya Petrovna	Shaaf	5
Islaev	Natalya Petrovna	5
Belyaev	Vera	4
Belyaev	Rakitin	4
Anna Semyonovna	Shaaf	4
Anna Semyonovna	Natalya Petrovna	4
Shpigelsky	Vera	3
Lizaveta Bogdanovna	Rakitin	3
Kolya	Natalya Petrovna	3
Kolya	Rakitin	3
Islaev	Matvey	3
Bolshintsov	Shpigelsky	3
Belyaev	Katya	3
Anna Semyonovna	Shpigelsky	3

Character 1	Character 2	Weight
Rakitin	Vera	2
Lizaveta Bogdanovna	Natalya Petrovna	2
Lizaveta Bogdanovna	Shpigelsky	2
Lizaveta Bogdanovna	Shaaf	2
Kolya	Shaaf	2
Katya	Vera	2
Islaev	Shpigelsky	2
Islaev	Vera	2
Bolshintsov	Rakitin	2
Belyaev	Islaev	2
Shaaf	Vera	1
Matvey	Shpigelsky	1
Matvey	Rakitin	1
Matvey	Natalya Petrovna	1
Kolya	Vera	1
Kolya	Shpigelsky	1
Katya	Shpigelsky	1
Katya	Shaaf	1
Katya	Matvey	1
Katya	Lizaveta Bogdanovna	1
Katya	Natalya Petrovna	1
Katya	Rakitin	1

Anna Semyonovna	Kolya	3
Anna Semyonovna	Lizaveta Bogdanovna	3
Anna Semyonovna	Islaev	3
Shaaf	Shpigelsky	2
Islaev	Shaaf	1
Islaev	Kolya	1
Bolshintsov	Lizaveta Bogdanovna	1
Belyaev	Kolya	1
Belyaev	Shpigelsky	1
Anna Semyonovna	Belyaev	1

Table 19. Weight of edges between characters in *A Month in the Country*, based on configuration changes mark-up

The total number of configuration segments in *A Month in the Country*, according to my own mark-up, is 73. That would be the maximum edge weight between any pair of characters. Let us look closely at the most connected pairs of characters, whose edges' weight is 4 or more, which means they 'communicate' in at least 5% of the play's segments. The participants in the two love triangles of the play are represented in bold in order to highlight and compare how often they appear in the top of this ranking.

Character 1	Character 2	Weight
Natalya Petrovna	**Rakitin**	19
Rakitin	Shpigelsky	10
Natalya Petrovna	**Vera**	9
Natalya Petrovna	Shpigelsky	9
Islaev	**Rakitin**	8
Rakitin	Shaaf	7
Belyaev	**Natalya Petrovna**	7
Anna Semyonovna	**Rakitin**	6
Natalya Petrovna	Shaaf	5
Islaev	**Natalya Petrovna**	5
Belyaev	**Vera**	4
Belyaev	**Rakitin**	4
Anna Semyonovna	Shaaf	4
Anna Semyonovna	**Natalya Petrovna**	4

Table 20. Weight of edges between characters in *A Month in the Country*, based on configuration changes mark-up, top 14 pairs

All of the first ten pairs of this list include either Natalya Petrovna or Rakitin. Natalya Petrovna and Rakitin also have the highest weight of the edge between them. They 'speak to each other' in more than 25% of the play's segments. The next most connected pair, Rakitin and Shpigelsky, have an edge weight which is almost half as small. Natalya Petrovna also has strong connections with Vera and Shpigelsky. Rakitin and Islaev are well-connected, as well. They occupy the fifth place in the ranking of weighted edges, with the weight of 8. All this explains the distinctively high weighted degree of Natalya Petrovna and Rakitin, because they form all of the most actively communicating pairs. Moreover, this brings us to consider how vital they are in shaping the majority of communication in the play.

Again, as we saw with the centralities values, Vera and Belyaev do not appear on the frontline of the action of the play. Keeping in mind the two plot focuses initially proposed for *A Month in the Country* (the student and the two women), it is striking how slender the student's storyline is in the final version. We can speculate that Turgenev chose to step back from the political content when he chose not to focus on the student.

Although Vera has a strong connection with Natalya Petrovna and their pair occupies a third position when ranked by the weight of edges, she does not seem to be as active in communicating with any other character. The same can be observed about the weight of the edges between Belyaev and the rest of the characters. Moreover, the weight of the edge between Vera and Belyaev is more than twice as small as the one between her and Natalya Petrovna. This demonstrates, again, that they become reasonably active characters only because Natalya Petrovna is interested in both of them for different reasons, and pulls them into the active circle of the network.

We have now seen how actively all the characters communicate with each other. This reflects the general level of connectedness within certain pairs of characters, which is especially high between Natalya Petrovna and Rakitin, as well as the level of focus on their communication. However, it would be interesting to measure the quality of their connections. Some of the pairs mentioned above often happen to speak to each other, but in a brief way rather than having thorough verbal interactions. The weight of the edge between two characters who only say something once during their 'common' segment would be the same as for the characters who have long conversations. For these reasons, I decided to measure the value of the dialogue-based strength of the connections between the characters in some way. I will explain how this value is calculated through the example of the pair of Natalya Petrovna and Rakitin.

I defined a dialogue as a sequence of responses between the two characters, if these responses were not interrupted by the responses of any other characters. In other words, a dialogue would be a segment of text where Rakitin and Natalya Petrovna exchange responses with each other back and forth. Then I automatically extracted the 'dialogues' that take place between Natalya Petrovna and Rakitin.

I identified 25 such dialogues. The total number of Natalya Petrovna's responses in the play is 383. The number of her responses in those 25 dialogues is 184. The total number of Rakitin's responses in the play is 297. The number of his responses in those 25 dialogues is 181. Let us also assume that for each dialogue we counted in one 'accidental' response (either the first one or the last one), that did not actually relate to the dialogue, but was addressed to some other character.

The measure of the strength of their dialogues would be the division of the sum of their responses in those 25 dialogues (excluding the number of possible 'accidental' responses) by the sum of their total responses in the play: ((184 + 181) - 25) / (383 + 297) = 340 / 680 = 0,5.

It is important to note that this value is approximate, for the sake of automatising the process of extracting the dialogues. Even if two characters only speak to each other, i.e. all their responses in the play are part of the dialogues between them, the strength of their dialogue-based connection would not equal 1, since we have excluded a certain number of responses from the calculation to account for 'accidental' responses.

Let us turn to the dialogue-based strength of the connections for the nine pairs of characters who are related to the two possible major plot lines of the play:

Pair	Strength of connection

Natalya Petrovna and Rakitin	0.50
Belyaev and Vera	0.42
Natalya Petrovna and Vera	0.29
Natalya Petrovna and Belyaev	0.20
Rakitin and Islaev	0.20
Belyaev and Rakitin	0.18
Belyaev and Islaev	0.048
Natalya Petrovna and Islaev	0.042
Rakitin and Vera	0.00

Table 21. Dialogue-based connection between characters in *A Month in the Country*

Again, as with the weight of edges, Natalya Petrovna and Rakitin have the strongest connection. Their conversations are not only frequent, but also long and substantial.

Interestingly, the dialogue-based strength of connection between Belyaev and Vera contrasts with the previously discussed values of their activity. 0.42 means that Belyaev and Vera devote almost half of their total communication efforts to each other. Their connection now outstrips the connection between Vera and Natalya Petrovna. This brings our attention back to the 'Belyaev-Vera' storyline and proves that Natalya Petrovna did have a reason to worry about their relationship. There was, perhaps, not a romantic connection between Vera and Belyaev, but at least a friendship that Natalya Petrovna felt she needed to break up.

The dialogue-based measure also sheds light on the level of connection between Rakitin and Belyaev. The strength of their connection almost equals the strength of the connection between Belyaev and Natalya Petrovna. This emphasises Rakitin's interest in Belyaev, even though he does not speak to him that often. Rakitin feels his own inferiority to his "young, ambitious and easily socially mobile" rival, while Rakitin himself is "superfluous

and middle-aged" (Freeborn 1985). He becomes fascinated by Belyaev's interests and political views, when he tries to understand what attracts Natalya Petrovna to Belyaev.

In the same way as the weight of edges, these calculations display how Rakitin is much closer to Islaev than Natalya Petrovna is. Even Belyaev's connection to Islaev is slightly higher than Natalya Petrovna's. Natalya Petrovna is not engaged in communicating with her husband. While keeping her relationship with Rakitin and trying to start a new one with Belyaev, Natalya Petrovna attempts to escape her misery in being married to Islaev.

The dialogue-based measure allows us to draw interesting conclusions about the quality of connection between the key members on the estate. It still identifies the pair of Natalya Petrovna and Rakitin as the most active and, therefore, important for the play. But it also highlights the relationship of Belyaev and Vera. Belyaev's connection to Natalya Petrovna gives way to his friendship with Vera, even though Natalya Petrovna tries to create a strong bond with him.

Let us now turn to the analysis of the new network of *Uncle Vanya* to see how the configuration changes mark-up transforms the network graph of that play. We will also explore the new centralities values as well as the dialogue-based connection of the characters, to see if they bring us closer to establishing the differences between *Uncle Vanya* and *A Month in the Country*.

Network analysis of Uncle Vanya

Network graph and centralities

Below is the network representation of *Uncle Vanya* based on the configuration changes mark-up.

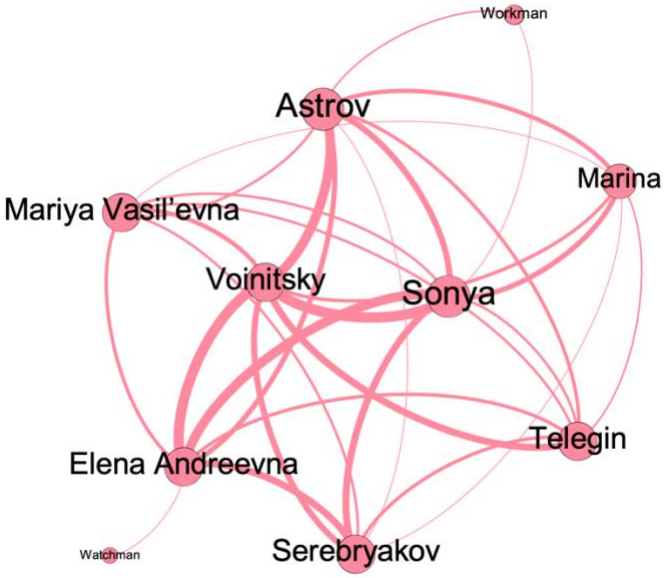


Figure 17. Network graph of *Uncle Vanya* based on configuration changes

As in *A Month in the Country*, we would expect the new graph of *Uncle Vanya* to be more sparse and some characters of the play to stand out in their network values. But we can see that most of the characters are tightly connected to each other, as they were in the previous version of the network graph. However, the density of this graph is 0.67, compared to 0.96, as it was before. Similarly to *A Month in the Country*, the configuration changes mark-up allows us to distance some household staff members from the main communicating group. By splitting the play into configuration segments, I separated off the events of Astrov and Sonya hearing the workman's speech, and of Elena Andreevna hearing the watchman's voice. In this way, the workman and the watchman became disconnected from the rest of the network, whereas previously they were connected to everyone who spoke in the fourth and second acts respectively.

Apart from that, there are still seven characters who seem to be equally influential in the network. Let us discuss the centralities of the nodes further to find if there are any significant outliers in the new network.

Character	Degree
Astrov	8
Sonya	8
Voinitsky	7
Telegin	7
Elena Andreevna	7
Mariya Vasil'evna	7
Serebryakov	7
Marina	6
Workman	2
Watchman (Efim)	1

Table 22. Degree of characters in *Uncle Vanya*, based on configuration changes mark-up

In the new network Astrov and Sonya have the maximum degree of 8, they 'meet' every other character in the play, except for the watchman. And the rest of the active characters, who form the connected group in the graph, follow them with a degree of 7, or 6 in the case of Marina. As mentioned before, the workman and the watchman are now kept outside of the main communication circle, with a degree of 2 and 1 respectively. We should consider other centralities values to see if Astrov and Sonya hold any outstanding influence in the network, since now their degree is only 1 point above everybody else's (or 2 points above, in the case of Marina).

Character	Betweenness centrality
Elena Andreevna	8.00
Astrov	3.80
Sonya	3.80
Voinitsky	0.30

Telegin	0.30
Mariya Vasil’evna	0.30
Serebryakov	0.30
Marina	0.00
Workman	0.00
Watchman (Efim)	0.00

Table 23. Betweenness centrality of characters in *Uncle Vanya*, based on configuration changes mark-up

Character	Closeness centrality
Astrov	0.90
Sonya	0.90
Elena Andreevna	0.82
Voinitsky	0.82
Telegin	0.82
Mariya Vasil’evna	0.82
Serebryakov	0.82
Marina	0.69
Workman	0.53
Watchman (Efim)	0.47

Table 24. Closeness centrality of characters in *Uncle Vanya*, based on configuration changes mark-up

Quite surprisingly, we see that Elena Andreevna has the highest betweenness centrality, and it is much higher than that of the rest of the characters. However, this is due to a small detail regarding the characters she is connected to. As mentioned earlier, Elena Andreevna is the only one who hears the watchman’s voice, which grants her such a high betweenness centrality – without this connection the graph would disconnect and disintegrate in to two parts. In the same way, Astrov and Sonya share the second place in the betweenness

centrality rank, because each of them holds half of the responsibility for providing a bridge between the workman and other characters. We will discuss later how to overcome this illusion of Elena Andreevna, Sonya and Astrov being the most influential characters in the network.

Moving on to other centralities, Astrov and Sonya have the highest closeness centrality, which means that they are situated closer to other characters than the rest of them, and can obtain information more effectively. However, Elena Andreevna, Voinitsky, Telegin, Mariya Vasil'evna, Serebryakov and Marina have closeness centrality just a bit lower than Astrov and Sonya. The higher closeness centrality of Sonya and Astrov, again, is due to their connection with the workman. They are closer to the workman than anyone else and they have the highest degree among the characters. This is what makes their closeness and Eigenvector centralities slightly higher than of the other characters:

Character	Eigenvector centrality
Astrov	1.00
Sonya	1.00
Voinitsky	0.96
Telegin	0.96
Mariya Vasil'evna	0.96
Serebryakov	0.96
Elena Andreevna	0.87
Marina	0.85
Workman	0.29
Watchman (Efim)	0.13

Table 25. Eigenvector centrality of characters in *Uncle Vanya*, based on configuration changes mark-up

So far, we can conclude that Sonya and Astrov are the most ‘popular’ characters, closest to everybody else, and are connected to other ‘important’ characters, even though the rest of the characters, except the workman and the watchman Efim, share almost the same network values. The only especially remarkable value is Elena Andreevna’s betweenness centrality. However, it does not seem fair to give her the title of the ‘connector’ of the play, only because she hears the watchman’s voice. Moreover, the watchman Efim is not even included in the cast list, as defined by Chekhov. After removing that edge between Efim and Elena Andreevna, we will see that Sonya and Astrov have the highest betweenness centrality because of their connection to the workman. To create a more consistent picture, we also should remove the edges between Sonya and the workman and Astrov and the workman.

That being done, we can see the following network graph and metrics:

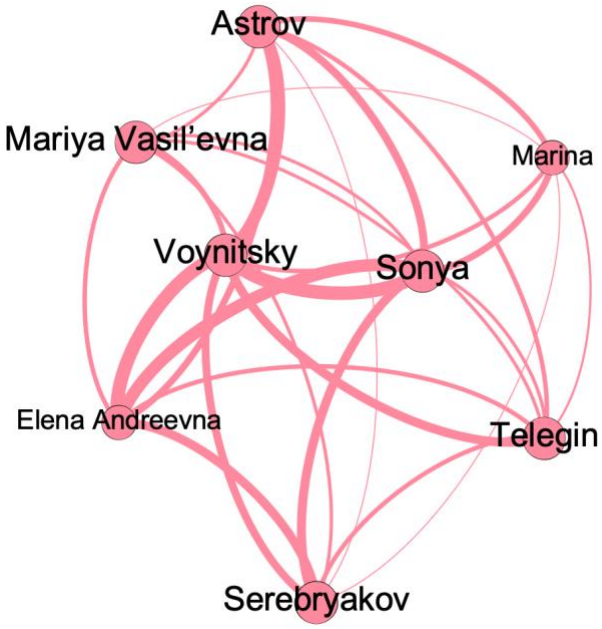


Figure 18. Network graph of *Uncle Vanya* based on configuration changes, without the workman and the watchman

Character	Degree	Eigenvector centrality	Closeness centrality	Betweenness centrality
Astrov	7.00	1.00	1.00	0.17
Voinitsky	7.00	1.00	1.00	0.17
Telegin	7.00	1.00	1.00	0.17
Mariya Vasil'evna	7.00	1.00	1.00	0.17
Serebryakov	7.00	1.00	1.00	0.17
Sonya	7.00	1.00	1.00	0.17
Marina	6.00	0.89	0.88	0.00
Elena Andreevna	6.00	0.89	0.88	0.00

Table 26. Centrality values of characters in *Uncle Vanya*, based on configuration changes mark-up, without the workman and the watchman's voice

The density of this network is 0.96. All of the characters, except Marina and Elena Andreevna, share the same metrics, apart from the weighted degree that we will discuss later. Astrov, Voinitsky, Telegin, Mariya Vasilevna, Serebryakov and Sonya are connected to everybody else in the network. Funnily enough, Elena Andreevna and Marina step back in the network values because they are not connected with each other. Their betweenness centrality is now 0. With a lower, not a maximum degree, they have no value as 'connectors' in this network. Moreover, if we exclude them from the network, we would see a dense complete graph[14] where nobody occupies a privileged position in any way:

[14] A simple graph in which each pair of distinct vertices are adjacent is a complete graph. (Wilson R. J. 1996. *Introduction to Graph Theory. Fourth edition.* (Longman Group Ltd.)).

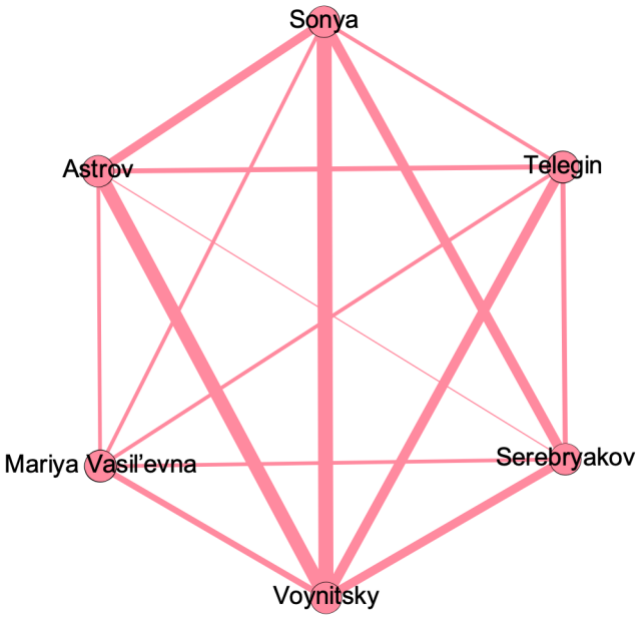


Figure 19. Network graph of *Uncle Vanya* based on configuration changes, without the workman, the watchman's voice, Elena Andreevna and Marina

The density of this network is 1. The fact, that these six nodes occupy the same position network-wise, according to their degree, betweenness, closeness and Eigenvector centralities, reflects even more clearly Chekhov's intention to move away from the 'single hero' perspective. He was aiming at revealing a character's nature through all the other characters.

Nevertheless, we can spot the difference in the thickness of the edges between the characters, equivalent to the weight of the edges. It means that some characters communicate with each other more frequently than the others. We will discuss the strength of these connections in the next section. For now, let us come back to the graph in Figure 17 and compare the weighted degree of the characters in this network. The value of a weighted degree corresponds to the weight of the edges between the characters. By looking at it we will see which characters are generally the most active in communication.

Character	Weighted degree
Voinitsky	59
Sonya	51
Elena Andreevna	46
Astrov	39
Serebryakov	32
Telegin	28
Mariya Vasil'evna	22
Marina	19
Workman	3
Watchman (Efim)	1

Table 27. Weighted degree of characters in *Uncle Vanya*, based on configuration changes mark-up

Voinitsky has the highest weighted degree, which means that he is a 'talkative', active character, who appears on stage often and has strong speech-based connections with others. After all, the play is called *Uncle Vanya*, and it is valuable to see that there is a measure that reflects his particular importance as a character.

Sonya follows Voinitsky with a high weighted degree. This has been interpreted by some critics as bringing the focus of the play on to 'simple people', who are beautiful inside and who have high morals (Ermilov 1946). We will see later that Sonya has a lot of communication with the estate's residents and visitors.

Together with Voinitsky and Sonya being the most active characters, in terms of the number of interactions with others, Astrov takes the fourth position in this ranking with a higher weighted degree than the outsider Serebryakov. Astrov is a good friend of Voinitsky, and also he is a more long-standing member of the community, whereas Serebryakov and

Elena Andreevna arrive to visit the estate for a short time, and would not have time to establish strong connections with others.

In this case, it is interesting to see that Elena Andreevna occupies the third place in the weighted degree rank. She is quite active in terms of her presence on stage and communication. Even though she technically does not hold an advantageous position in the graph, because she is not connected to Marina, she has tight communicative bonds with other characters. We will see in the next section who she is particularly close to.

Although Telegin is not a very visible character of the play, he is often around. As Magarshack (1952) writes: "[Telegin] takes no part in the external action at all […] But he has a very important part in the inner movement of the play […] the part of the chorus whose comments on the action heighten the dramatic tension" (pp. 207-208). He is also the one who often plays music in the play, providing what can be called another dimension of the mood in the estate.

Moving down the ranking, Mariya Vasil'evna and Marina have a lower weighted degree than the rest of the family and the residents. It is not surprising to see that, since they do not appear in as many conversations in the play as others.

We have already discussed the low network value of the workman and the watchman, so we would not expect them to have a high weighted degree. Although it could be the case that they would be particularly 'close' to their few connections, the network graph and metrics show that they are not.

To sum up, *Uncle Vanya*'s network metrics reflect how the majority of the characters share the same network centralities values. There are six characters who seem to communicate and act as a united nucleus. We will explore further if there is more focus on

certain storylines of the play, but for the moment it looks like almost all the characters of *Uncle Vanya* are equally important.

To specify, we are able to identify two members of the estate who disrupt this overall connectedness – Elena Andreevna and Marina. Even though they have lower network metrics than others due to their lower degree (Table 26), the difference in their network values is not drastic. And, as already noted, the only reason they have lower values is because they are not connected with each other. We could not say that they are redundant or the least influential characters. Indeed, Elena Andreevna is still a key character and acts as an accelerator of disappointment for other central characters (Ermilov 1946). She acts as a catalyst, which "brings the crisis to a head", but she also "waits for something to happen to [her] from outside before [she] can make up [her] mind about the nature of [her] own feelings" (Pitcher 1985, p. 93).

Marina, according to Rayfield (1975), represents an important theme in Chekhov – "the old servants […] [who] throw into greater relief the idleness, the vacillation and disorder of their masters" (p. 109). She might not be in the heart of the action and the love intrigues in the estate, but she backs the community up by establishing the household's routine, as well as providing support to the other characters, especially to Astrov.

Other distinctions between the characters that we have seen so far come from the weighted degree. The weighted degree refocuses us on the participants in the love triangles, and especially Voinitsky, who seems to be the most talkative and, maybe, the most present on stage. Let us now move to the next section to discuss the strength of connections between the characters to see how the weight of edges between all the characters establish the weighted degrees that we talked about. We will also explore the quality of the connections using the dialogue-based measure, as we did when discussing *A Month in the Country*.

Edges weight and 'dialogues'

To recap: the weight of the edge between two characters is the number of segments in which they both speak. We would assume that in such segments characters speak to each other. Such analysis would not be sensible using the act-based mark-up of the play, since in that case almost half of the edges would have a weight of 4.

My presumption is that the higher the weight of the edge between two characters, the more attention the play pays to this pair of characters. Thus, by analysing the weight of the edges in *Uncle Vanya*'s network, we can interpret the importance of certain storylines in the play.

Character 1	Character 2	Weight
Elena Andreevna	Voinitsky	12
Sonya	Voinitsky	12
Astrov	Voinitsky	11
Elena Andreevna	Sonya	11
Elena Andreevna	Serebryakov	8
Serebryakov	Sonya	8
Telegin	Voinitsky	8
Astrov	Sonya	7
Serebryakov	Voinitsky	7
Astrov	Elena Andreevna	6
Marina	Sonya	6
Astrov	Marina	5
Mariya Vasil'evna	Voinitsky	5
Astrov	Telegin	4

Elena Andreevna	Mariya Vasil’evna	4
Elena Andreevna	Telegin	4
Marina	Voinitsky	4
Serebryakov	Telegin	4
Astrov	Mariya Vasil’evna	3
Mariya Vasil’evna	Telegin	3
Mariya Vasil’evna	Sonya	3
Mariya Vasil’evna	Serebryakov	3
Sonya	Telegin	3
Astrov	Workman	2
Marina	Telegin	2
Astrov	Serebryakov	1
Elena Andreevna	Watchman (Efim)	1
Mariya Vasil’evna	Marina	1
Marina	Serebryakov	1
Workman	Sonya	1

Table 28. Weight of edges between characters in *Uncle Vanya*, based on configuration changes mark-up

Consequently, the characters who have the highest weighted degree form part of the pairs who occupy the top places when ranked by the weight of edges. This, again, fixes our focus on the communication within the love triangles of the play. There are two pairs with the highest weight of edges – Elena Andreevna-Voinitsky and Sonya-Voinitsky. They speak to each other in 12 segments of the play. The total number of segments in *Uncle Vanya*, according to my own configuration changes mark-up, is 47. This means that these pairs speak to each other in more than 25% of the play. The other two pairs who communicate with each other in 23% of the play are Astrov-Voinitsky and Elena Andreevna-Sonya.

From reading the play, we do not notice a lot of communication between Serebryakov and Sonya, the father and daughter. However, the strength of their connection is the same as the connection between Serebryakov and Elena Andreevna. Both pairs speak to each other in 17% of the play. We will soon see how their communication varies in its quality.

Serebryakov, generally, seems to be quite active in communication. Judging from the weight of edges, the conflict between Voinitsky and Serebryakov appears to be as important as the connection between Astrov and Sonya. Both pairs communicate during 15% of the play. As Pitcher (1985) notices, it is the clash between Voinitsky and Serebryakov that highlights the contrast between the residents and the outsiders. This is indeed a very important topic of *Uncle Vanya*, and the conflict between Voinitsky and the professor plays a critical part in exposing Voinitsky's inner struggles and insecurities. Witnessing the unravelling of the conflict, we understand Voinitsky's disillusionment in life and disappointment in his own fate.

Astrov and Elena Andreevna have a slightly lower weight of the edge between them than Voinitsky and Serebryakov. They talk to each other in 13% of the play. Thus, at the moment, the weight of the edges suggests that Sonya's connection with Astrov is stronger than his connection with Elena Andreevna. However, we will see how this reverses when we look at the dialogue-based strength of their relationships.

I will analyse the 'dialogues' between the participants of the love triangles – Sonya, Astrov, Elena Andreevna and Voinitsky, as well as Serebryakov (to take into account that the important idea behind *Uncle Vanya* is the conflict between Serebryakov and his wife and the permanent residents of the estate).

Pair	Strength of connection

Elena Andreevna and Sonya	0.32
Voinitsky and Astrov	0.32
Elena Andreevna and Astrov	0.30
Voinitsky and Serebryakov	0.24
Astrov and Sonya	0.23
Elena Andreevna and Voinitsky	0.22
Elena Andreevna and Serebryakov	0.22
Voinitsky and Sonya	0.14
Serebryakov and Sonya	0.09
Astrov and Serebryakov	0.00

Table 29. Dialogue-based connection between characters in *Uncle Vanya*

In addition to sharing the second place of the ranking by the weight of the edges, Elena Andreevna-Sonya and Voinitsky-Astrov occupy the top positions in the dialogue-based ranking. Astrov and Voinitsky have the strongest connection, dialogue-wise. They are two middle-aged friends, who share a lot of common frustrations about their lives and the future of Russia. See, for instance, this extract from the fourth act where Voynitsky begs Astrov to tell him what he should do with the rest of his life, since it is so unbearable for him to continue living, and Astrov concludes that they are both hopeless:

> **Войницкий.** Что ж, я - сумасшедший, невменяем, я имею право говорить глупости.
>
> **Астров.** Стара штука. Ты не сумасшедший, а просто чудак. Шут гороховый. Прежде и я всякого чудака считал больным, ненормальным, а теперь я такого мнения, что нормальное состояние человека - это быть чудаком. Ты вполне нормален.

Войницкий ***(закрывает лицо руками)*****.** Стыдно! Если бы ты знал, как мне стыдно! Это острое чувство стыда не может сравниться ни с какою болью. ***(С тоской.)*** Невыносимо! ***(Склоняется к столу.)*** Что мне делать? Что мне делать?

Астров. Ничего.

Войницкий. Дай мне чего-нибудь. О боже мой... Мне сорок семь лет; если, положим, я проживу до шестидесяти, то мне остается еще тринадцать. Долго! Как я проживу эти тринадцать лет? Что буду делать, чем наполню их? О, понимаешь... ***(судорожно жмет Астрову руку)*** понимаешь, если бы можно было прожить остаток жизни как-нибудь по-новому. Проснуться бы в ясное, тихое утро и почувствовать, что жить ты начал снова, что все прошлое забыто, рассеялось, как дым. ***(Плачет.)*** Начать новую жизнь... Подскажи мне, как начать... с чего начать...

Астров ***(с досадой).*** Э, ну тебя! Какая еще там новая жизнь! Наше положение, твое и мое, безнадежно.

In the same way as with the conflict between Serebryakov and Voinitsky, the friendship between Astrov and Voinitsky provides more space within the play to express and illustrate their points of view and their general desperation in life.

Elena Andreevna and Sonya are definitely not such close friends, if at all, as Voinitsky and Astrov. But demographically, so to speak, they are the closest ones to each other in the estate. They are both lonely and seeking female company. Elena Andreevna, especially, has no friends in the estate except for the old professor. She also believes that she is genuinely trying to help Sonya with her love for Astrov. Elena Andreevna wants to be kind

to Sonya, even though she fails to achieve this. Sonya, in turn, seeks advice and support from an older female figure, who seems interesting and knowledgeable and differs from the rest of the people around Sonya.

The next pair in this ranking is Astrov and Elena Andreevna. Now we see that he has a 'deeper' connection with Elena Andreevna than with Sonya, i.e. longer conversations. Pitcher (1985) speaks of 'the tragic lack of communication' between Sonya and Astrov. Sonya tries, as best as she can, to make Astrov notice her feelings towards him, but he fails to fully notice them and understand. Instead, Astrov is more absorbed in his connection with Elena Andreevna. The main reason for that is his desire to have a love affair with her.

Sonya and Voinitsky have the highest number of dialogues. However, almost all of them consist of only two responses. They do not have either elegant high society conversations, or long philosophical discussions. Critics frequently noticed the disconnectedness in Chekhov's dialogues. "[It] is not that they are all alienated from each other, but, on the contrary, they are intimate and there is no need to be so formal and explicit" (Pitcher 1985). This must be the case with Sonya and Voinitsky. They do not have exhaustive discussions because of their domestic familiarity and because they are already so close that they can understand each other better than other people, without communicating most of their feelings and thoughts.

Elena Andreevna's connection with Voinitsky is as strong as with Serebryakov. Voinitsky is trying to rescue her from Serebryakov, to lure her away to his world. He does not succeed in the end, but he is fighting for her attention towards himself and to his opinions about Serebryakov.

As expected, the relationship of Sonya and Serebryakov does not have a high dialogue-wise strength. They are around each other in 17% of the play (Table 16), but, as

with Voinitsky, their communication mostly consists of two responses. This might be, as well, due to their familiarity as a father and daughter or, more likely, due to the fact that Serebryakov is generally not interested in Sonya and her fate.

Astrov 'meets' Serebryakov in only one segment and they do not have a single dialogue. It is perhaps a bit odd that Elena Andreevna's admirer, Astrov, never speaks to her husband on stage (he does come to treat him one night though, but we do not see them speaking). As well, it is striking that Astrov is a very good friend with Voinitsky, and he represents the same kind of person, but he does not participate in the conflict with Serebryakov.

Voinitsky himself speaks to Serebryakov in 15% of the play and they devote around 25% of their mutual communication effort to each other. Thus, both of the Elena Andreevna-Astrov and Voinitsky-Serebryakov pairs have a stronger dialogue-based bond, than Sonya with Astrov or Voinitsky. We could say that the strength of their connections highlights the main theme of the play – the clash between two different social groupings. However, it is necessary to acknowledge here the limitations of these methods. Such quantitative analysis does not allow us in any way to identify the characters whose relationships are expressed implicitly. For example, Sonya and Voinitsky do no talk to each other, not so much because they do not want to, but because they do not need to.

Even though, according to the majority of the network values, the activity and communication in *Uncle Vanya* is evenly divided among the characters, we have been able to identify certain plot-lines that occupy more space in the play. However, we did not see outliers, such as were obvious in *A Month in Country*. In Turgenev's play there is one pair of characters on whom we can spot a distinctly noticeable focus. If we compare the weight of the edges and the dialogues-based connections in the two plays, we will see that in *Uncle*

Vanya these values are distributed more equally. There are explicitly important topics in the play, but it is less obvious which one is the most important.

Let us now turn to the overall comparison of the two plays by summing up the conclusions of each part of the analysis.

Comparison and conclusions

At first, when looking at the act-based network graphs and metrics of *A Month in the Country* and *Uncle Vanya*, we see quite a similar picture. Both graphs are very dense, all the characters are well-connected. We see a united community in both plays. The fact that the plays are not split into scenes gives actors more space for establishing their own views and interpretations. However, such a breakdown of the plays would not allow us to analyse effectively the text-based structure that inevitably unfolds for the readers. The mark-up based on configuration changes reveals for us the differences in the structures and consequently, the plots of the plays.

In *A Month in the Country* we see a lot of focus on Natalya Petrovna and Rakitin. Generally, we do not sense a lot of focus on either Belyaev, or his relationship with Vera. They are both powerless figures. Belyaev is rather the passive object of Natalya Petrovna's passion and Vera is disposed of by Natalya Petrovna by the end of the play. Belyaev's character is used in the play rather to unravel the personalities of Natalya Petrovna and Rakitin. This puts to rest the speculation that the main focus of *A Month in Country* is reflected by its earlier versions and titles.

One of the revelations of the analysis was Shpigelsky's crucial importance to the network in terms of bringing people together for his own benefit. But setting this apart, Rakitin and Natalya Petrovna can be called the main protagonists. Rakitin is slightly better

established within the network because he ‘meets’ Bolshintsov and Natalya Petrovna does not. However, she in any case communicates with Bolshintsov through the go-between Shpigelsky.

Even though in the dialogue-based connection we see a strong link between Belyaev and Vera, Rakitin and Natalya Petrovna take the first place in both the weight of edges and the dialogue-based connection ranks. And the difference between the weight of the edge between them and between the other characters is huge.

A Month in the Country is about human passions, sick love, incompatibility and heartbreak (Seeley, Freeborn). This is what happens with Natalya Petrovna and Rakitin. Natalya Petrovna is unhappy in her marriage with Islaev and tries to find comfort outside of their marriage. She is desperate to fulfil her desire for love and to enjoy her youth, which she is currently wasting by living as she is. But Rakitin’s love for Natalya Petrovna has never been mutual, he has been struggling and devoting all his thoughts and efforts to someone who will not love him back or even treat him with proper attention. These struggles are what constitutes *A Month in the Country*. In summary, I would say that the theme of this play is developed mostly through the two major characters of Natalya Petrovna and Rakitin. Furthermore, the focus on their relationship emphasises the autobiographical nature of the story. Turgenev once told Savina, the actress who played Vera and then later Natalya Petrovna in the productions of *A Month in the Country*: “Ракитин это я. Я всегда в своих романах неудачным любовником изображаю себя” (Turgenev, Savina 1918).

In *Uncle Vanya*, on the other hand, we see that multiple characters share the same network value, even after the configuration changes mark-up. Moreover, looking at Tables 19 and 28 we see that in *Uncle Vanya* the weight of the edges is more equally distributed

between the pairs of characters, while in *A Month in the Country* more than 30% of the pairs have an edge weight of 1.

None of the characters has a privileged position in the graph. Elena Andreevna and Marina are slightly behind everybody else on their network values. But, as discussed above, these differences are minor and do not remove them from the overall connected network. Moreover, Elena Andreevna is very active in communicating with other characters: she is present in the pairs that are top-ranked by both the weight of edges and the dialogue-based connection. As mentioned earlier, she is the catalyst of the play, activating or bringing attention to the conflict between the residents and the outsiders. Serebryakov represents the conflict too, of course. He is, by the way, much more active than Islaev, which also contributes to the more equally organised network of this play, as opposed to *A Month in the Country*.

When contrasting certain pairs of characters across the plays, we see that the Elena Andreevna-Astrov connection takes up more space in their speech, than the connection between Natalya Petrovna and Belyaev. However, Elena Andreevna is also engaged in conversations with Sonya for the same proportion of her responses, and Astrov and Voinitsky concentrate the same amount of their communication on conversations with each other.

Natalya Petrovna's almost non-existent dialogue-based connection with Islaev shows how detached she is from him. Elena Andreevna and Serebryakov, on the contrary, have quite a strong connection, though again, their dialogue-based connection is almost the same as that of several other pairs of characters (see Table 29). Elena Andreevna exists alongside Serebryakov. They arrive at and leave the estate together.

If we look at the 'speaking' activity in the plays, we can verify the structural differences between *A Month in the Country* and *Uncle Vanya*:

Rank	A Month in the Country	Uncle Vanya
1	Natalya Petrovna - 383	Elena Andreevna - 111
2	Rakitin - 297	Astrov - 109
3	Belyaev - 219	Voinitsky - 107
4	Vera - 180	Sonya - 88

Table 30. Number of speech acts

Rank	A Month in the Country	Uncle Vanya
1	Natalya Petrovna - 7374	Astrov - 2931
2	Shpigelsky - 3947	Voinitsky - 2225
3	Rakitin - 3877	Elena Andreevna - 1745
4	Belyaev - 2786	Sonya - 1490
5	Vera - 2019	Serebryakov - 1101
6	Islaev - 1582	

Table 31. Number of words

There is a drastic difference between the number of speech acts and words uttered by Natalya Petrovna and everybody else. She is followed in the number of speech acts by Rakitin. In *Uncle Vanya*, Elena Andreevna also takes the first place by the number of speech acts, but she only has 2 and 4 more speech acts than Voinitsky and Astrov, respectively. In the number of words, we again see the dominant position of Natalya Petrovna.

In *Uncle Vanya*, the 'speaking leaders' are Astrov and Voinitsky. Multiple critics (Ermilov, Magarshack) consider Astrov's personality to be one of the main topics of the play. They say that *Uncle Vanya* has two themes, that are described by Magarshack (1952) as the selfish idealist Voinitsky, who realises that he has been worshipping a false idol, and the unselfish idealist Astrov, who has high hopes about the future. Ermilov writes "…образы

Астрова и Войницкого имеют равноценное значение в пьесе и вообще может казаться, что это две разные пьесы". Indeed, we can agree with this, based on the fact that they are both part of the 'important pairs' of the play and have a high 'speaking' activity.

However, in general, we see *Uncle Vanya*'s characters as more of a solid circle, as opposed to having distinctive protagonists, as in *A Month in the Country*. If we compare Tables 21 and 29 we will see that the dialogue-based strength of connections in *Uncle Vanya* is more evenly distributed between the pairs of characters than in *A Month in the Country*.

These are the main conclusions in contrasting the two plays and their key characters that could be easily equated to each other by someone who did not immerse themselves in a more thorough study of the plays. Certainly, an attentive reader of the plays can conclude that they are not identical and that the paralleled characters do not necessarily carry the same functions or fates. My analysis shows how these characters differ in terms of their 'importance' or influence, and, in general, how the plays differ in their structure. We see that *A Month in the Country* is about a woman, a man and their failure to find mutual love. *Uncle Vanya*, in turn, appears to offer direct evidence of Chekhov's drive towards the decentralisation of the cast. These are the distinctions between the two plays that I was able to highlight with the network analysis methods and other quantitative measures.

Since the SNA methods and other quantitative measures have proved to be productive in this chapter, the next chapter will contain a similar 'close' analysis of two plays using both the methods from this chapter and the new measures that I will introduce later. Apart from this, I will also explore the possibility of the analysis of Russian drama on a larger scale, focusing on the representation of women in the Russian plays written in the 19th century.

Women in Russian Drama

As mentioned in the introductory chapter, one of the greatest advantages of such projects as the DraCor is the possibility of conducting large-scale analysis of literary texts in order to gain insight into specific tendencies and trends in literature. In this chapter, I will implement such an analysis to describe the quantitative representation of women in 19th-century Russian drama. This specific period has been chosen for analysis because it is the only century fully covered by the DraCor: there are no plays from the first half of the 18th century and from the second half of the 20th century in the Russian Drama Corpus. In addition to this, I will perform a more detailed analysis of female characters in two plays from that period. The choice of the plays will be discussed further in this chapter.

The fact that women around the world have obviously not occupied the same position as men is true for many societies, and it is especially true for Russian culture. There have not been many female writers in Russia until recently, nor famous women artists in general. This means that women have largely been portrayed in Russian culture by men, and the narrative of their lives has been created by male authors most of the time.

If we specifically consider Russian drama, among the 211 plays that were written between 1747 and 1940 and are available in the DraCor, there are only 4 plays written by female authors. Out of 57 authors represented in the corpus only two are women – Catherine the Great and countess Evdokiya Rostopchina. Such underrepresentation has two reasons: the deficiency of the data in the corpus that does not include such important female writers as Tsvetayeva or Gippius, and the general prevalence of men in Russian literature and its canon. As Barbara Heldt puts it: “There is no lack of general pronouncements about how women act or feel or think in Russian literature: these, however, have been overwhelmingly made by men” (Heldt 1987, p. 2). While other scholars have embarked on the important task of

reconstructing women's narratives and discovering hidden or forgotten female voices in Russian literature, I am dedicating this chapter to an investigation of how women were represented in 19th-century Russian drama, carried out with the help of exact methods. In her book Heldt (1987) says about famous Russian novels and plays: "If all these works were measured purely for the number of words used on the male or the female, clearly the female is considerably short-changed" (p. 13). I will explore whether female characters were indeed quantitatively underrepresented in Russian plays and if there were any significant authors who allowed more space for women than others.

At the moment, there are 127 plays of the 19th century in the Russian Drama Corpus (December 2020). I have collected various characteristics of these plays to discuss the representation of women in them. These characteristics include a play's author, title, 'normalised' year, number of speakers, percentage of female characters, percentage of female speech-acts, percentage of female spoken words, and number of female and male words per response on average. I will start by explaining what these values mean and how they were obtained.

Using the API (Application Programming Interface) of the DraCor, I retrieved the plays with a 'normalised' year greater than 1799 and less than 1900.

The 'normalised' year is a value that is calculated separately within the DraCor web application, based on the written, publication and premiere dates. Especially in Russia, with its history of censorship, it is not always easy to decide when a literary work became

available to a wider audience. The DraCor defines the "normalised" year according to this formula developed by the DLINA[15] project:

1. "Print" and "premiere" dates indicate when a play was officially released, in print or on stage.
    a. If both "print" and "premiere" dates are available, the earlier one becomes the 'normalised' date.
    b. If only one of them is available, it becomes the 'normalised' date. Otherwise, the 'normalised' date stays undefined.
2. However, the date when a play was written is also important, because even without publication a play could have been spread amongst certain circles, including literary societies. Pushkin's *Boris Godunov* and Griboyedov's *Woe from Wit* (*Gore ot uma*) were long known to the public, even before their official releases.
    a. If the only information we have is when a play was written, this becomes the 'normalised' date (this rarely occurs in my experience).
    b. If we already have a 'normalised' date plus a "written" date, we check the time gap between them.
        i. If the gap is bigger than 10 years, the "written" date becomes the 'normalised' date. The value of 10 years is chosen based on empirical evidence. According to the DLINA project the "10 years" measure limits the chance of a play being put too far away from its date of conception in the chronology.

[15] Digitally-Driven Literary Network Analysis (https://dlina.github.io).

ii. Otherwise, the ‘normalised’ date stays as established before (either “print” or “premiere”).

As a result, we obtain a list of Russian plays, available in the Russian Drama Corpus, together with their authors’ names, that were written, performed and/or published in the 19th century.

Furthermore, I extracted the number of speakers in the plays. Each play’s TEI includes the list of speakers of the play, which may be different from the cast list defined by the author. Sometimes authors do not include minor characters in a *dramatis personae* list. However, all the speech-acts are encoded in the TEI versions of plays, and thus we can easily compile the list of actual speakers, which is called ‘participation description’ in the TEI standard. This list would include all the characters speaking in a play, even if they are not part of the cast list. This principle of encoding enables us to see not only the statistics for the ‘core characters’, who are part of the *dramatis personae*, but also to get an overview of the general atmosphere and context of a play, including its background actors – the crowd or *narod*, who are often still important. Talking about Ostrovsky’s *The Storm* and referring to Pushkin’s famous concluding stage direction “народ безмолвствует” in *Boris Godunov*, Sverdlov (1995) points out that “«Народ» – это больше, чем «разные лица»: и количественно (народ – множество), и качественно (народ – сила, даже если он безмолвствует)” (p. 107). In the case of Pushkin, as it turns out, remaining silent is a form of speech act, at least in semantic terms, even though it is not directly relevant to my investigation.

The participation description also includes information about the sex or gender of each speaker. This information is encoded in the following way:

```
<listPerson>
```

```
<person xml:id="gorodnichij" sex="MALE">
        <persName>Городничий</persName>
        <persName xml:lang="ger">Polizeimeister</persName>
    </person>
                ...
</listPerson>
```

There are three categories of gender of speakers – male, female and unknown. The "unknown" category refers to groups of characters, whose composition, and therefore gender, is unknown.

Such cases do not occur very often. All in all, in the 127 plays, on average, there are 3% of unknown gender characters. The median value, as well as the minimum value of the percentage of such characters in the plays is 0%. And there is one play in the subcorpus, *The Affinity of the World Powers* (*Srodstvo mirovykh sil)* by Prutkov, where 90% of characters are of unknown gender.

In total, there are 30 plays that have unknown gender characters, and only 12 plays have 10% or more of such characters.

For example, in Ostrovsky's play *The Snow Maiden* (*Snegurochka*) there are 38% of characters of unknown – or no – gender: different choirs, voices of wood spirits, and tribes. In Chekhov's *Tatyana Repina* there are 18% of characters of unknown gender, who are also represented by choirs, characters defined as "someone", and the crowd.

Even though I will not pay much attention to the characters of unknown gender, we will take them into account when calculating, for example, the percentage of female characters or female speech acts. I will do so to obtain accurate numbers and so as not artificially to raise the statistics for male and female characters in the plays, as would happen if I excluded the unknown gender characters.

In this way, the percentage of female characters in a play is calculated by dividing the number of female characters by the number of all characters, which includes female, male and unknown characters, if there are any.

In addition to the percentage of female characters in the plays, I have also calculated the percentage of female responses, i.e. speech-acts. I wrote a script that goes through the TEI xml file of a play and finds the <sp> elements that contain a speech-act and the information about the character who utters this speech-act. The script then checks the character's identity in the speakers' list, to determine the character's gender. After that is done, it adds the text of this particular speech act to the overall array of responses of each particular gender (male, female, or unknown). In this way, for each play I create three arrays that contain all the responses of its characters, grouped by the gender of these characters.

There is another complication in classifying the speech-acts by gender. Some speech-acts are group-speech-acts. This is not to be confused with the speech-acts of group characters of unknown gender. What I define as a group-speech-act looks the following way:

```
<sp who="#Mozgovoj #Jat #Zhigalov #Aplombov #Dashenka #NastasjaTimofeevna
#Kavalery #Shafer #Zmejukina">
<speaker>Все</speaker>
        <stage>(Дымбе).</stage>
                <p>Речь! речь! Ваша очередь!</p>
</sp>
```

This is a speech-act from Chekhov's *The Wedding* (*Svad'ba*), where he gives the stage direction <speaker>Все</speaker>, which means that each character who is present on stage speaks. This is reflected in the TEI mark-up as having multiple speakers in the speech-act.

Another example can be found in Ostrovsky's *Without a Dowry* (*Bespridannitsa*):

```
<sp who="#Paratov #Vozhevatov #Robinzon">
        <speaker>Паратов, Вожеватов и Робинзон.</speaker>
        <p>Ура!</p>
 </sp>
```

Here, again, the <sp> mark-up reflects the author's stage direction indicating the speakers. We see that there are several characters participating in this speech-act. Even though we know their genders, it is impossible to classify this speech-act as either male or female.

There are also group-speech-acts involving only characters of the same gender. However, even in this case it is not clear how, for example, to calculate the number of words per response for a certain gender, if taking these speech-acts into account. Would it be reasonable to divide them by the number of speakers? Or should this count as a speech-act made by one character of a certain gender? This might lead to uncertainties in calculating necessary values for genders and characters.

Therefore, I made a decision to include all group-speech-acts in the category of the unknown gender speech-acts. In any case, these group-speech-acts seem to be largely of a certain type, most frequently exclamations, which do not usually carry significant semantic value.

Moreover, there are only a few such instances of those in the plays. In the whole subcorpus of 127 plays there are 59 plays that have group-speech-acts (i.e. not initially "unknown" gender speech acts). In 47 out of these 59 plays, there are less than 1% of such speech acts. And there are only 5 plays with more than 3% of such speech acts:

Play	Number of group-speech-acts	Per cent of group-speech-acts
Nekrasov *The Condemned* (*Zabrakovannye*)	2	4.08%

Prutkov *The Fantasy* (*Fantaziya*)	23	6.76%
Krylov *The Sleeping Powder* (*Sonnyi poroshok*)	32	7.22%
Prutkov *Skull-Reader the Phrenologist* (*Chereposlov, sirech Frenolog*)	19	16.24%
Prutkov *The Reckless Turk* (*Oprometchivyi Turka*)	6	16.67%

Table 32. Plays with a high proportion of group-speech-acts

We can see that there are only two plays whose quantitative description of gender representation would be affected by including such group-speech-acts in the category of unknown gender – Prutkov's[16] *Skull-Reader the Phrenologist* and *The Reckless Turk*. Thus, the prevalence of this type of group-speech-acts remains statistically fairly insignificant.

Finally, for calculating the percentage of female responses in a play, I calculated the number of responses in the array of female responses and divided it by the sum of the numbers of female, male and unknown gender responses.

Having such arrays of male, female and unknown gender speech-acts, it is possible to calculate not only the proportion of female speech-acts, but also the proportion of words uttered by female characters in a play. I calculated the length of each response (text within a speech-act) in each array of speech-acts. Then I added up these lengths to calculate the overall number of female, male or unknown gender words in a play. Similarly to the percentage of female speech-acts, the percentage of female words in a play is calculated by dividing the number of female words by the sum of the numbers of female, male and unknown words.

[16] This was a collective pseudonym invented and used by Aleksei Tolstoy and his cousins.

Again, having these arrays of speech-acts of female and male characters, I calculated the average number of words per female or male response in a play. I took the lengths of all female and male responses in a play and calculated the mean values of these two arrays.

Let us now discuss how, based on these calculations, women are represented in the selected 127 Russian plays of the 19th century. In these plays, women 'occupy' less space than men. Out of 127 plays, there are only 38 plays with 40% or more of female characters. The average percentage of female characters in our subcorpus is 31%. The maximum percentage of female characters in a play of this subcorpus is 80%, while there are 9 plays with zero female characters. This means that there are plays with only male characters and no plays with only female characters.

The average percentage of female speech-acts is 31%, which is exactly equal to the percentage of female characters. More or less the same can be said about the average percentage of female spoken words, which is 29%. To conclude, women are given one third of the presence and the speech in Russian drama of the 19th century.

We can also use the opportunity of being able to perform such large-scale analysis to dispel some myths about women's patterns of communication. According to a common stereotype, women talk more than men. However, if that was true, female speech-acts would be longer than male speech-acts, and, therefore, women would 'occupy' more 'word space'. Consequently, we would expect women to represent more than 29% of the spoken words in the plays even though they utter only 31% of the responses. More specifically, although women only utter 31% of speech-acts in a play on average, if female responses were longer than male responses, women would utter more than 31% of words in a play on average. However, the average number of words per response for women is lower than for men: thus the percentage of female words is even lower than the percentage of female responses. This

demonstrates that in our subcorpus, men not only talk more often (i.e. uttering more responses), but they simply talk more than women. The average length of a female response is around 16.25 words per response, with a minimum of 5.43 words and a maximum of 59.88 words. For male responses, the average length is 19.69 words: the minimum is 5.13 words and the maximum is 97.80 words.

To sum up, in 19th-century Russian drama, based on the 127 plays available in the Russian Drama Corpus, there are, in general, fewer female characters, they speak less frequently than male characters, and their responses are shorter. Male actions and voices dominated 19th-century Russian drama.

Many readers, however, have formed an impression of the strong woman praised in Russian literature, who "Коня на скаку остановит, // В горящую избу войдет!".[17] Indeed, some characteristics with which male authors invest their female characters might be understood as a celebration of women's strengths. However, Barbara Heldt (1987) makes a point that these female characters are often used in Russian literature to highlight the "inadequacies and weaknesses of some male protagonists" (p. 12). The "terrible perfection" of a Russian literary heroine is an unobtainable ideal for real Russian women. Therefore, such strong and brave heroines do not represent real female lives and issues, but simply serve the purpose of highlighting the character of a "superfluous man".

Thus, there are two separate directions in which one should look in order to understand the women's issues in 19th-century Russian drama – quantitative and qualitative analysis. As was mentioned earlier, in this chapter we are taking advantage of a relatively big corpus of Russian plays, and of the digital methods available to us in this study. Thus, the focus of this chapter shifts to the matter of the quantitative representation of women in these

[17] A famous quotation from Nekrasov's poem "Moroz, Krasny nos" written in 1863.

plays. However, it is acknowledged that merely quantitative indicators are not enough to construct a comprehensive picture about the female characters in Russian drama.

Bearing that in mind, let us now look more closely at the subcorpus of the 19th-century Russian plays, to see who amongst these authors writes plays with more or less equal quantitative representation of both men and women. The subcorpus represents 31 authors. The distribution of plays per author looks as follows (the authors are arranged in chronological order based on their dates of birth):

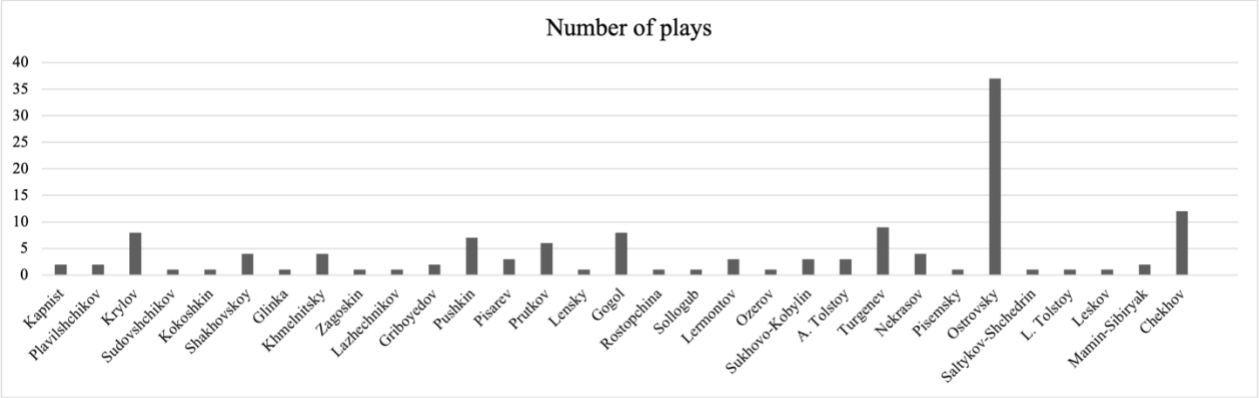


Figure 20. Number of plays written by the authors of the 19th-century subcorpus

The authors that have the greater number of plays in the subcorpus (more than 10 plays) are Ostrovsky and Chekhov. They were, in fact, the most productive dramatists of that period.

Next, let us explore the average percentage of female characters for each author. I calculated this value by taking the mean value of the percentages of female characters in all the plays by each author.

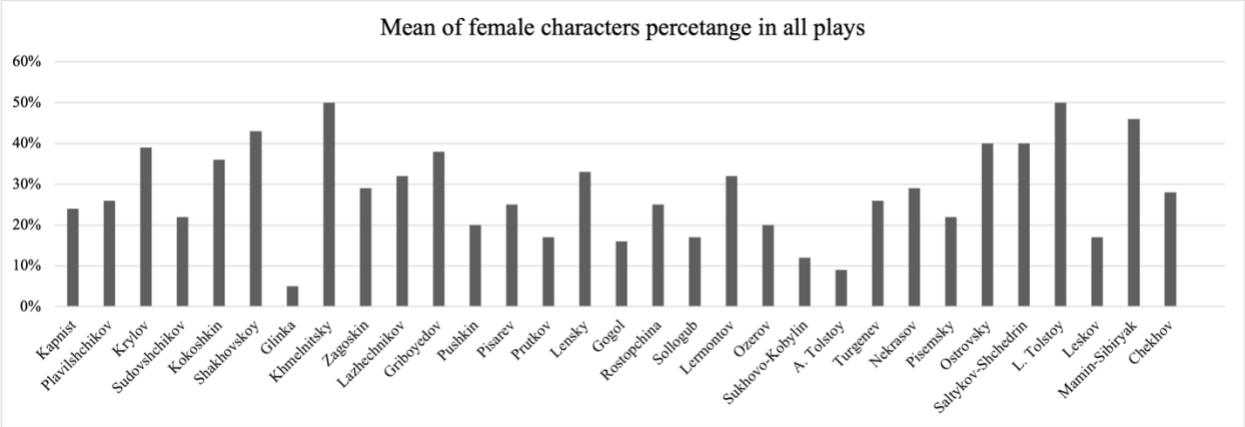


Figure 21. Average percentage of female characters among the authors of the 19th-century subcorpus

As we can see, there are only two authors in whose plays the average percentage of female characters hits 50% (Khmelnitsky and Leo Tolstoy). Out of 31 authors, 29 have less than half female characters in their plays on average.

Khmelnitsky and Leo Tolstoy have only 4 plays and 1 play in the subcorpus respectively. They are followed by Mamin-Sibiryak with 46% of female characters on average and only 2 plays, and by Shakhovskoy with 43% of female characters on average and 4 plays.

Next to them in this ranking is Ostrovsky, who has the greatest number of plays in the subcorpus with 40% of female characters on average. He is the only author in our case whom we can consider – because of the number of his plays available – as somebody who demonstrates a consistent tendency to include a more equal proportion of female characters in his plays.

Ostrovsky and women

The 37 plays by Ostrovsky included in the corpus were written between 1847 and 1884. In these plays, on average, the values of the quantitative representation of women are

higher than in the whole subcorpus in general. Ostrovsky has 40% of female characters, 42% of female speech acts, and 41% of female words in his plays on average.

Out of 37 plays in the corpus, Ostrovsky has 6 plays with 60% and more of female characters, 7 plays with 50-59% of female characters, 7 plays with 40-49% of female characters, and 7 plays with 30-39% of female characters, while the average percentage of female characters in a play for the whole corpus is 31%.

Finally, the mean length of male responses in his plays is 15.16 (compared to 19.55 in the whole corpus) and the mean length of female responses in his plays is 15.05 (compared to 16.16 in the whole corpus). Thus, the mean lengths of female and male responses in Ostrovsky's plays are almost equal.

What presumably shapes the tendency of Ostrovsky to focus on women as much as on men is the subject-matter of most of his plays. Andreev (1995) writes that it is well established that the theme of all Ostrovsky's plays, except for the historical chronicles, lies strictly within the family and *byt* (pp. 5-6). And family requires both men and women.

Ostrovsky, nevertheless, is not innocent in terms of creating what Heldt calls "terrible perfection". Several male critics, including Henley (1970) and Andreev (1995), confirm that a major preoccupation of Ostrovsky's drama is the spiritual inequality between lovers: "In Ostrovsky's plays […] the male lover is often a weak-charactered puppet" (Henley 1970, p. 231). However, Natalya Zhuravkina (1997) concludes that Ostrovsky was somebody who depicted the emancipation process in Russian society, even though it was not his stated goal: "Он изменил психологический климат женских образов, готовя тем самым совершенно новый характер героинь, которые найдут дальнейшее развитие в драмах Чехова, Горького, романах Толстого" (p. 178). Thus, although Heldt (1987) states that only Chekhov came somehow close to shifting away from the "terrible perfection" perspective in

Russian literature, he was not the most "balanced" author from the quantitative perspective, and it may have been Ostrovsky who in fact plotted the vector for a kind of equality in Russian drama. In his plays he explored the reasons why emancipation for women would be a welcome transformation of existing societal norms, even though he did not show that emancipation being achieved.

In order to scrutinise the 'most equal' dramatist within our subcorpus, I will analyse in detail two of his plays. To give an overview of Ostrovsky's plays, let us look at the lists of his plays with 50% and more of female characters, female speech-acts and female words:

Play	Percentage of female characters
A Holiday Dream Before Dinner (*Prazdnichnyi son do obeda*)	80
What You Seek, You Will Find (*Za chem poidyosh, to i naidyosh*)	78
When Your Own Dogs Are Fighting, a Strange Dog Should Not Meddle (*Svoyi sobaki gryzutsya, chuzhaya ne pristavai*)	75
A Domestic Picture (*Semeinaya kartina*)	67
The Poor Bride (*Bednaya nevesta*)	61
A Protegee of the Mistress (*Vospitannitsa*)	60
It's a Family Affair – We'll Settle It Ourselves (*Svoi lyudi, sochtyomsya*)	50
An Unexpected Event (*Neozhidannyi sluchai*)	50
Poverty is No Vice (*Bednost ne porok*)	50
Drunk at a Foreign Feast (*V chuzhom piru pokhmeliye*)	50
Hard-earned Bread (*Trudovoy khleb*)	50
Rich Brides (*Bogatye nevesty*)	50

Table 33. Ostrovsky's plays with 50% and more of female characters

Play	Percentage of female speech-acts
An Old Friend is Better Than Two New Ones (*Staryi drug luchshe novykh dvukh*)	72
When Your Own Dogs Are Fighting, a Strange Dog Should Not Meddle (*Svoyi sobaki gryzutsya, chuzhaya ne pristavai*)	70
A Holiday Dream Before Dinner (*Prazdnichnyi son do obeda*)	69
What You Seek, You Will Find (*Za chem poidyosh, to i naidyosh*)	65
A Protegee of the Mistress (*Vospitannitsa*)	62
The Poor Bride (*Bednaya nevesta*)	58
Bondwomen (*Nevol'nitsy*)	54
Live Not as You Would Like (*Ne tak zhivi, kak khochetsya*)	53
Late Love (*Pozdnyaya lyubov'*)	50

Table 34. Ostrovsky's plays with 50% and more of female speech-acts

Play	Percentage of female words
An Old Friend is Better Than Two New Ones (*Staryi drug luchshe novykh dvukh*)	71
A Protegee of the Mistress (*Vospitannitsa*)	70
When Your Own Dogs Are Fighting, a Strange Dog Should Not Meddle (*Svoyi sobaki gryzutsya, chuzhaya ne pristavai*)	66
A Holiday Dream Before Dinner (*Prazdnichnyi son do obeda*)	66
What You Seek, You Will Find (*Za chem poidyosh, to i naidyosh*)	61
The Poor Bride (*Bednaya nevesta*)	60

Bondwomen (*Nevol'nitsy*)	55
The Storm (*Groza*)	53
Late Love (*Pozdnyaya lyubov'*)	51

Table 35. Ostrovsky's plays with 50% and more of female words

It does not feel possible to leave *The Storm* out of the picture when writing about Ostrovsky and women in his works. It is his most famous play, the one that is known best both in Russia and worldwide. This play, among other things, examines the role and the nature of women in a provincial Russian society. Even though the play only has 31% of female characters, they utter 48% of speech-acts and 53% of the words in total, which means that there is plenty of material for analysing the representation of women in it.

The second play that I will discuss in this chapter is *Without a Dowry*. Although it only has 27% of female characters and female speech-acts, with 29% of female words, it is an important play for the Russian audience. It is well-known partly owing to a famous Soviet film, *A Cruel Romance*, based on the play. *Without a Dowry* introduced a significant female figure to Russian literature of the 19th century – the proud and strong Larisa – and, therefore, it provides insight into the issue of women in Russian drama. In addition, it would be interesting to compare this play to *The Storm*, which has greater 'female values'.

The Storm

Let us begin with *The Storm*, the play that first comes to mind when we think about Ostrovsky.

It was written in 1859 and published in 1860, and is a five-act drama with 41 scenes. The play tells the story of a merchant family, the Kabanovs, from Kalinov – a fictional small provincial town situated on the Volga. The young woman Katerina is unhappily married to

Tikhon and has difficult relations with her strict and powerful mother-in-law Kabanova ("Kabanikha"). When Tikhon goes away for a business trip, Katerina embarks on an affair with Boris, the visiting nephew of another merchant, Dikoy. Katerina is supported in her amours by Varvara, Tikhon's sister, who is unmarried and therefore free to spend her time with Kudryash – Dikoy's clerk.

Once Tikhon returns home, Katerina confesses to him about having an affair, unable to continue carrying the burden of her sin. The whole town finds out about it, and Kabanikha insists on Tikhon punishing his unfaithful wife. Neither society in the town nor Kabanikha, let alone Katerina herself, will pardon her, so she commits suicide to end her insufferable life.

Other characters of the play include Kuligin – an educated man with scientific interests, who gives valuable, rational comments on everything happening in Kalinov, the religious pilgrim Feklusha, Glasha – the Kabanovs' servant, and an elderly noblewoman. There are also background characters, who are defined by Ostrovsky as "town residents of both sexes".

Through the story of the tragedy of Katerina's life we learn about the society of Kalinov and its morals, and how restricted and dominated women are in this society. Let us see how this is reflected in the play in a quantitative way.

Already at this stage of giving an overview of the play, we can notice an interesting detail. As already mentioned, the DraCor versions of plays include all characters speaking on stage. In the *dramatis personae* of *The Storm* there are 13 characters, including the item that represents the various 'crowd' characters of "both sexes". This results in 29 'participating characters' (the ones from the participation description).

The DraCor team applied the following algorithm for manually encoding occasional and episodic speakers, such as *первый* (first) or *женщина* (a woman). If a character could be clearly identified throughout the text, it receives permanent id (unique identifier). If it was ever unclear whether, for example, two characters, defined as *женщина*, are the same, they each receive their own separate id[18]. If it was unclear who constitutes a group of characters, for instance, who is *все* (all), the group receives a separate id instead of specifying each participating character, as is done in the example from Chekhov's *The Wedding* on page 128.

As you can see, such a mark-up may result in producing too many individual characters, more than the author specified. But it is not always possible to understand whether two 'crowd characters' are the same person or not. Thus, the annotators of the DraCor aimed for the most objective way of encoding such cases.

As already mentioned, among the 29 'participating characters', there is 31% of female characters. They utter 48% of the responses with 53% of the words in the play. Male characters in *The Storm*, on average, have 14.82 words in a response, while female characters have 18.46 words.

However, if we look at the 12 'core characters', excluding the crowd, we will see a different picture. There is now an equal number of female and male characters, with female characters 'occupying' 50% of the speech-acts and 55% of the words in the play. The average length of male speech-acts slightly rises and equals 15.4 words, while the average length of female speech-acts among the "core" female characters stays almost the same at 18.51 words.

Thus we see that in general in his play, when creating the atmosphere and constructing the narrative about a provincial Russian town, Ostrovsky slightly favoured men

[18] Such characters would be referred to, for instance, as zhenschina_4_3 (*женщина* from the third scene of the fourth act).

over women, even though his female characters have longer speech-acts. But he was absolutely even-handed when it came to the 'core characters' of *The Storm*. This probably has to do with the fact that one of the main plot lines of the play is the confrontation between two women – young Katerina and her mother-in-law Kabanikha.

Network graphs

I will concentrate on the 12 'core characters' of the play and analyse how the female characters among them are constructed quantitatively and qualitatively. Let us begin by comparing the network metrics for the 'core characters' within the overall network of the play and the downsized version of the network formed by these 12 characters.[19]

[19] Where it was not possible to accurately identify and match the characters named as *первый*, *второй*, *женщина*, etc., who appeared several times in different segments of the play, the characters were encoded as 'Pervyi 4a' referring to one of the *первый* from the fourth act, 'Pervyi 4b' referring to the other *первый* from the fourth act, etc.

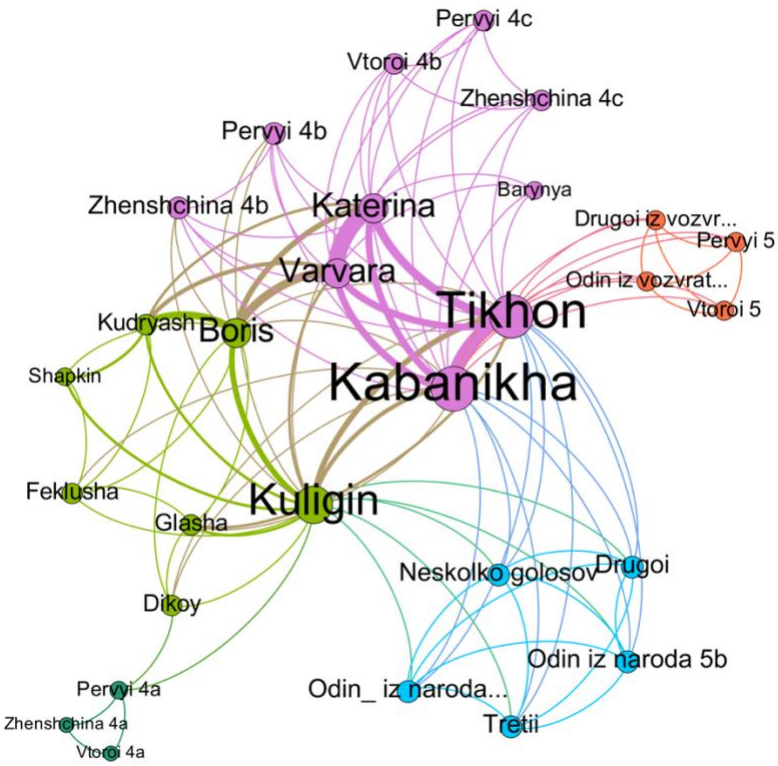


Figure 22. Network of *The Storm* with all characters

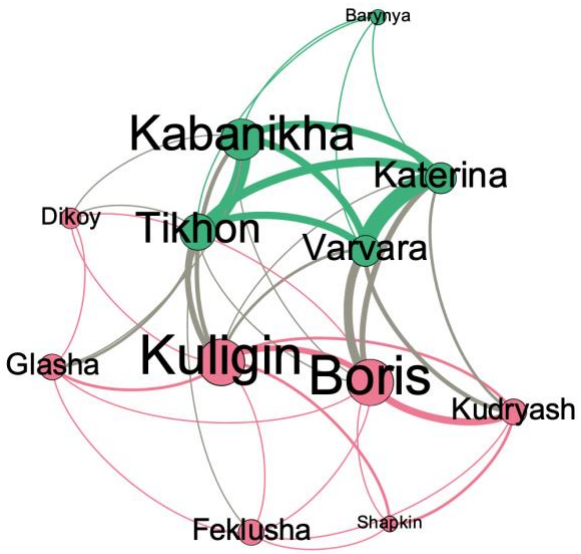


Figure 23. Network of *The Storm* with 12 'core characters'

Character	Degree in all	Degree in 12	BC in all	BC in 12
Kabanikha	23	9	94.3	4.77
Tikhon	22	8	80.52	2.68
Kuligin	18	10	92.89	5.98
Boris	12	10	12.73	5.98
Varvara	12	7	11.68	1.72
Katerina	12	7	11.68	1.72
Kudryash	6	6	1.89	1.35
Dikoy	6	5	16.0	0.0
Feklusha	6	6	2.8	1.29
Glasha	6	6	0.5	0.5
Shapkin	4	4	0.0	0.0
Barynya	4	4	0.0	0.0

Table 36. Network values of the "core characters" of *The Storm*[20]

The density of the first network is 0.28 out of 1. The most influential characters of this network listed by their degree are Kabanikha, Tikhon and Kuligin. They are followed by Boris, Katerina and Varvara with the degree of 12.

In the second network with the density of 0.62, the disparity of the centrality values of the characters is not as significant as before. The characters with the highest degree are now Kuligin and Boris, followed straightaway by Kabanikha. The next are Tikhon, Katerina and Varvara.

[20] "in all" refers to the overall network comprising 29 characters, "in 12" refers to the network consisting of 12 'core characters', "BC" means Betweenness Centrality.

Thus, there are four characters that occupy the top positions in both networks – Kabanikha, Tikhon, Kuligin and Boris. It is not surprising to see Kabanikha in this list, since she appears to be the most influential and the toughest character of *The Storm*.

These four characters are followed by Katerina and Varvara, which means that out of the six most influential and well-situated within the network characters there are three women. Let us see whether this changes when looking at other attributes of the characters' activity.

Stage presence and speaking activity

Below is the data on the number of scenes each character is present in, and the number of speech-acts and words they utter. The number of scenes per character was retrieved using the DraCor API "metadata" function; further characteristics were calculated by me, working directly with the TEI file of the play.

For calculating the total number of words of each character, I have used the same algorithm as described in the beginning of this chapter. I have collected the arrays of speech-acts for each character of *The Storm*. Then I calculated the lengths of the speech-acts for each character. The sum of these lengths is the total number of words that a character utters.

Rank	Character	Number of scenes
1	Katerina	18
2	Varvara	15
3	Kabanikha	13
4	Tikhon	13
5	Boris	12
6	Kuligin	9

7	Kudryash	7
8	Glasha	4
9	Dikoy	3
10	Feklusha	3
11	Shapkin	2
12	Barynya	2

Table 37. Number of scenes for the 'core' characters of *The Storm*

Rank	Character	Number of speech-acts
1	Katerina	139
2	Boris	106
3	Varvara	104
4	Tikhon	95
5	Kabanikha	89
6	Kuligin	61
7	Kudryash	59
8	Dikoy	39
9	Glasha	19
10	Feklusha	18
11	Shapkin	11
12	Barynya	2

Table 38. Number of speech-acts for the 'core' characters of *The Storm*

Rank	Character	Number of words
1	Katerina	3100
2	Kabanikha	1448

3	Kuligin	1425
4	Boris	1342
5	Tikhon	1242
6	Varvara	1221
7	Kudryash	917
8	Feklusha	766
9	Dikoy	724
10	Glasha	190
11	Barynya	142
12	Shapkin	62

Table 39. Number of words for the 'core' characters of *The Storm*

We see that Katerina is the central figure of the play in terms of her presence on stage and speaking activity. She has the highest number of scenes, speech-acts and words. One of the reasons Katerina is present in 18 out of 41 scenes is the four monologues (scenes when only Katerina is on stage and speaks) that she is given by Ostrovsky.

The other most active characters, who follow Katerina in the rankings according to the number of scenes, speech-acts and words are the same ones, only alternating in their order. Based on these three tables, we can identify the six characters who are given most space in the play: Katerina, Varvara, Kabanikha, Tikhon, Boris and Kuligin. There is an equal number of women and men in this list with women, in general, being more active (occupying higher positions in the three rankings).

It is worth looking at the number of stage directions that complement the characters' speech-acts. More specifically, we will consider the proportion of stage directions per speech-act for each character.

As mentioned earlier, I extracted the speech-acts for all characters. Further, I extracted all stage directions that refer to each character's speech-acts. In this way, I know both the number of speech-acts and the number of stage directions for each character. It is important to note that the number of stage directions in one speech-act might be greater than 1.

By dividing the total number of stage directions of a character by the number of their speech-acts, we obtain the mean value of stage directions per speech-act for a character, or, as I am calling it, the proportion of stage directions.

Rank	Character	Proportion of stage directions per speech-act
1	Barynya	4.0
2	Katerina	0.53
3	Varvara	0.36
4	Dikoy	0.28
5	Kabanikha	0.25
6	Tikhon	0.23
7	Kuligin	0.21
8	Boris	0.20
9	Kudryash	0.15
10	Glasha	0.10
11	Feklusha	0.10
12	Shapkin	0.09

Table 40. Number of stage directions per speech-act for the 'core' characters of *The Storm*

First of all, we should address the Barynya (noblewoman). She is a unique character in terms of her speaking statistics. Uttering only two speech-acts, she has the greatest

proportion of stage directions and the longest speech-acts (see below). Since the Barynya is an episodic character, we will not spend a lot of time discussing her, and will move on to the next positions in this ranking.

After her, Katerina has the greatest proportion of stage directions per speech-act. Even though we know that some speech-acts have more than one stage direction, the number of 0.53 tells us that, in general, there is one stage direction referring to every other speech-act of Katerina's. Varvara has a stage direction for approximately every third of her speech-acts, Kabanikha – for every fourth.

As a result, out of the top six characters in the ranking of stage directions, four characters are female. There might be, of course, a connection between the number of speech-acts and the number of stage directions too. The characters who speak more frequently may qualify for the greatest number of stage directions. However, this is not the case with the Barynya, who utters 2 speech-acts with 8 stage directions in them. It is also not true for Boris, who has a greater number of speech-acts than Varvara and Kabanikha, but a lower proportion of stage directions.

In his work *Teoriya literatury* Tomashevsky (1996) says that receiving the most vivid emotional description and, therefore, capturing the attention and the compassion of readers, is what distinguishes any one character among others as a "hero". I would argue that the proportion of stage directions per speech-act correlates with such a degree of emotional description. Characters whose actions are accompanied by stage directions have a higher chance of grasping the attention of the reader, and perhaps the audience. Via stage directions, we understand a character better by learning about their nature, manners, reactions, and affections. This, I believe, contributes to establishing the emotional connection between a reader or spectator and a character. In any production of a play, however, stage directions

may become secondary and can be reinterpreted or omitted. Therefore, it is important to note that in this work I am focusing on a reader's experience and also assume that most conventional productions will follow the author's ideas.

From what we see in Table 40, the female characters of *The Storm* are the ones who qualify for such a definition of a "hero", to a greater extent than the male characters. Female characters also occupy the first positions in the ranking by the average number of words per response.

As before, I used the arrays of speech-acts of characters and calculated the lengths of these speech-acts. The mean value of such lengths for each character is their average number of words per response.

Rank	Character	Number of words per response
1	Barynya	71.00
2	Feklusha	42.56
3	Kuligin	23.36
4	Katerina	22.30
5	Dikoy	18.56
6	Kabanikha	16.27
7	Kudryash	15.54
8	Tikhon	13.07
9	Boris	12.67
10	Varvara	11.74
11	Glasha	10.00
12	Shapkin	5.64

Table 41. Average length of speech-acts of the 'core' characters of *The Storm*

Although it is interesting to see that the Barynya and Feklusha, who speak in only 2 and 3 scenes respectively, have the highest average number of words per response, this is not entirely unexpected. The Barynya with her hectoring and Feklusha with her preaching are evidently the outliers in these statistics. The next three positions in the ranking are occupied by Kuligin, Katerina, Dikoy and Kabanikha with mean lengths of speech-acts within a more "normal" range. But, taking into account the Barynya and Feklusha, among the top six characters in this ranking there are four women.

Apart from calculating the number and lengths of speech-acts, I have also considered the monologues in the play. By monologue I mean a speech-act or speech-acts from a scene where only one character is present on stage. Having the TEI file of a play, it is possible to go through the play automatically to identify such scenes. There are seven monologues of this kind in *The Storm*:

1) In the fourth scene of the first act there is a monologue by Boris about him being in love with Katerina.

2) In the sixth scene of the second act there is a monologue by Kabanikha about her son and daughter-in-law never doing anything the right way.

3) In the eighth scene of the second act there is a monologue by Katerina about her being bored and not knowing what to do while Tikhon is away for two weeks.

4) In the tenth scene of the second act there is a monologue by Katerina about her wanting to meet with Boris.

5) In the first scene of the third act there is a song by Kudryash, that he sings while waiting for Varvara at night.

6) In the second scene of the fifth act there is a monologue by Katerina talking about her sinful affair with Boris.

7) Lastly, in the fourth scene of the fifth act there is a monologue by Katerina, right before she throws herself in the Volga river.

In this way, four out of seven monologues in the play are uttered by Katerina. It is an important fact that Ostrovsky used monologues for Katerina's character. Through monologues a character expresses their feelings and thoughts in a more direct way than through dialogues. Monologues often convey the most important messages to the audience and can be used to set the scene's atmosphere. Thus the fact that Katerina has the greatest number of monologues means that it is especially important in the play to draw the audience's attention to her emotions and thoughts, and to let her address the audience directly.

Speech analysis

Moving on to the contents of speech-acts, I have looked at certain punctuation marks to identify some patterns in the speech of the characters. First of all, I looked at the exclamation marks to see how emotional the characters are. The way I calculated it was the following. I took the speech-acts of every character and split them into sentences. Then, within each sentence, I established the number of exclamation marks. We need to note that there might be more than one exclamation mark in a sentence. For example, in the sixth scene of the fourth act Tikhon exclaims "*Не надо, не надо! не говори!*", when Katerina begins her confession to him in front of Kabanikha.

In this way, I collected a list of numbers where each number corresponds to the number of exclamation marks in a sentence. Next, I calculated the mean value in this list, i.e. the number of exclamation marks per sentence for each character.

Rank	Character	Number of exclamation marks per sentence
1	Barynya	0.68
2	Katerina	0.46
3	Tikhon	0.46
4	Dikoy	0.43
5	Kabanikha	0.41
6	Boris	0.41
7	Kuligin	0.34
8	Kudryash	0.33
9	Shapkin	0.33
10	Varvara	0.28
11	Glasha	0.27
12	Feklusha	0.27

Table 42. Proportion of exclamation marks in the characters' speech in *The Storm*

Looking at the results, we may note that the proportion of exclamation marks in the speech of characters does not necessarily reflect the level of their emotionality. For instance, the Barynya has the greatest number of exclamation marks per sentence. Even though she utters just a few speech-acts, her role in the play is quite noticeable. She represents a particular view of femininity that matches the principles of the older generation of the play. The Barynya's way of communicating these principles and judgements is loud, aggressive and firm. And her confident condemnations of Katerina serve as a prophecy of Katerina's fate.

On the other hand, we could hardly say that Varvara is not an emotional character, and yet she occupies the tenth position in this ranking. Her character, when reading the play,

seems so full of life, optimism and joy. Therefore, Varvara is unexpectedly quiet, if we are to define loudness by the presence of exclamations in speech.

Following the Barynya, Katerina and Tikhon have the next greatest number of exclamation marks per sentence. Perhaps, exclamations in *The Storm* should be associated not with emotionality, but with the need of young people to assert or fight for freedom and protest against the restrictions in their lives. They are trying to object to living according to the rules of Kabanikha and of the merchant class. Katerina falls into the category of "rebels" in Henley's (1970) subdivision of Ostrovsky's characters. Consider these responses by Katerina to Varvara about Katerina's affair and her unwillingness to obey the society's norms.

> **Катерина.** Что же мне делать! Сил моих не хватает. Куда мне деваться; я от тоски что-нибудь сделаю над собой!
>
> **Катерина**. Эх, Варя, не знаешь ты моего характеру! Конечно, не дай Бог этому случиться! А уж коли очень мне здесь опостынет, так не удержат меня никакой силой. В окно выброшусь, в Волгу кинусь. Не хочу здесь жить, так не стану, хоть ты меня режь!

Even though Tikhon, according to Henley, is a puppet, he too manages to "rise above himself" in the final scenes of the play. When Tikhon learns about Katerina's death, he finally makes a stand against his mother and voices his inconvenience with his own life:

> **Кабанов.** Маменька, вы ее погубили! вы, вы, вы...
>
> **Кабанов**. Вы ее погубили! Вы! Вы!

Кабанов. Хорошо тебе, Катя! А я-то зачем остался жить на свете да мучиться!

Varvara does not have this need, because her freedom is not yet challenged. Before writing *The Storm,* Ostrovsky spent some time exploring life in the Volga region as part of a quasi-anthropological survey commissioned by the government of Alexander II. He observes: "Девушки пользуются совершенной свободой […] Образ жизни замужних совершенно противоположен образу жизни девушек; женщины не пользуются никакой свободой и постоянно сидят дома" (Ostrovsky 1859). And he incorporates this custom in depicting the differences between Katerina's and Varvara's behaviours.

Kabanikha and Dikoy also have a reasonably high proportion of exclamation marks per response. In the play, they represent an older generation, who are rooted in the 17th century in terms of their beliefs, as Peace (1989) thinks. They are "monstrous *samodurs*" (a neologism invented by Ostrovsky) (Henley 1970), who intimidate others in order to earn their respect through fear (Peace 1989). This is why, I believe, there are frequently exclamations in their speech, reflecting imperative instructions rather than expressions of emotion.

Boris, in his turn, depends on Dikoy and has to obey his orders and follow his rules. Thus, Boris has a reason, as do Katerina and Tikhon, to protest against the will of Dikoy. Unfortunately, the protest features in his speech do not materialise in any action. Even when Dikoy, punishing Boris for his love affair with Katerina, decides to send him away, the latter leaves Kalinov for Siberia without a question. Henley (1970) defines the character of Boris in an unflattering way: "In direct confrontation with Katerina's complete self-sacrifice for Boris, it is almost impossible not to think of Boris as a coward" (p. 322).

The other punctuation mark used as a literary device which I have looked at is ellipsis. The proportion of ellipses per sentence was calculated in the same way as the proportion of exclamation marks. It is used by Ostrovsky much less than exclamations:

Rank	Character	Number of ellipses per sentence
1	Katerina	0.07
2	Tikhon	0.05
3	Glasha	0.03
4	Kuligin	0.02
5	Varvara	0.02
6	Boris	0.01
7	Kudryash	0.01
8	Dikoy	0.00
9	Kabanikha	0.00
10	Shapkin	0.00
11	Barynya	0.00
12	Feklusha	0.00

Table 43. Proportion of ellipses in the characters' speech in *The Storm*

Katerina, again, has the greatest proportion of ellipses in her speech, and Tikhon follows her in this ranking. Pauses in speech can reflect hesitation, timidity and introspective moments. Sverdlov (2003) notes the disconnectedness and hesitation in Katerina's speech and associates it with her apprehension of her death. See this extract from Katerina's monologue in the fourth scene of the fifth act of the play:

Катерина. Куда теперь? Домой идти? Нет, мне что домой, что в могилу – все равно. Да, что домой, что в могилу!.. что в могилу! В могиле лучше… Под

деревцом могилушка… как хорошо!.. Солнышко ее греет, дождичком ее мочит… весной на ней травка вырастет, мягкая такая… птицы прилетят на дерево, будут петь, детей выведут, цветочки расцветут: желтенькие, красненькие, голубенькие… всякие ***(задумывается)***, всякие…

Other characters have a lower ratio of ellipses, and almost half of them do not reach the value of one hundredth of an ellipsis per sentence.

The characteristics of speech discussed above do not prevail in either female or male speech in *The Storm*. They enable us to describe certain characters better, but do not highlight any specific differences between women and men in the play.

The next step of my analysis is exploring the usage of particular parts of speech by male and female characters. I used the Mystem morphological analyser developed by Yandex[21] for this task. These are the 11 parts of speech that I considered: adjective, adverb, conjunction, interjection, noun, particle, preposition, pronoun-adjective (*твой*, *этот*, *любой*), pronoun-adverb (*там*, *здесь*, *где*), pronoun-noun (*я*, *ты*, *вы*), and verb.

This is how I have calculated the proportion of certain parts of speech in characters' speech-acts. For each character, I took the array of their speech-acts. I processed each speech-act with the Mystem, assigning each word in a speech-act a part of speech. Knowing the length of a speech-act and how many words of this speech-act are constituted by certain parts of speech, I calculated the proportion of all parts of speech in this speech-act. In this way, for each speech-act of every character I established the proportion of nouns or verbs in it, for instance. Next, I calculated the mean value of these proportions for each character. As a result, for all the "core" characters I had the percentage of all parts of speech in their speech-

[21] https://yandex.ru/dev/mystem/.

acts on average. Further, I grouped these by the gender of the characters and calculated the mean values of the proportion of each part of speech per speech-act among female and male characters. Below is the table with these values.

Part of speech	Female characters	Male characters	Difference
Verb	0.205	0.168	0.037
Pronoun-noun	0.138	0.17	0.032
Noun	0.147	0.173	0.026
Preposition	0.077	0.055	0.022
Particle	0.137	0.157	0.02
Adjective	0.042	0.03	0.012
Pronoun-adverb	0.047	0.042	0.005
Pronoun-adjective	0.028	0.033	0.005
Conjunction	0.082	0.087	0.005
Interjection	0.02	0.015	0.005
Adverb	0.053	0.052	0.001

Table 44. Distribution of parts of speech in speech-acts of the 'core' female and male characters of *The Storm*

Although the difference in the proportion of parts of speech in female and male speech is not large, it is still important. In the introduction to his translation of *Without a Dowry*, Henley (2014) notes that "Ostrovsky was especially conscious about the language, striving above all for natural speech" and that he "was unsurpassed in using colloquial language and that he is a master of characterisation" (p. 13). Thus, Ostrovsky would pay a lot of attention to the specific words and, consequently, the parts of speech that his characters use.

Let us consider the first five parts of speech in the table, where the proportion of usage between women and men of *The Storm* differs by at least 2%. Female characters use many more verbs than men. Male characters, in their turn, use both more nouns and pronoun-nouns than women. Nouns and verbs are the key elements to the syntactic structure of a sentence.

The frequent usage of verbs contributes to the speed of narration and focuses on action rather than concepts. In this way, we may characterise the speech of the female characters of the play as lively, describing events or demanding actions.

Nouns may describe objects and concepts. But, most importantly for this analysis, using more nouns requires a more complex process of structuring a sentence. Moreover, nominalisation is more common in academic writing, which, in our case, indicates a more complex prose that is used for male speech-acts. The frequent usage of verbs and pronouns is, on the other hand, associated with a spoken discourse (Biber and Gray, 2010).

Thus, in *The Storm*, female speech is more lively and action-oriented, while male speech is more syntactically and conceptually complex. To back this up, let us look at the most educated character of the play – Kuligin. In his speech, 31% of the words are nouns and 18% verbs.

It is interesting that, even though the male characters use more nouns than female, they use fewer prepositions, that can only be used with nouns. Prepositions are mostly used in those temporal and spatial expressions that women in *The Storm* must therefore use more frequently than men.

Moreover, there is a higher proportion of particles in the female speech of the play. Particles impart certain emotional connotations to expressions. This may mean that the

female characters are more likely to colour their speech with emotions, intentions and evaluative statements. However, particles appear to be one the most complex parts of speech for Russian, since they can be split into several groups by their function: word-forming particles that act in combination with verbs (such as *бы* and *пусть*), the negation particles (*не* and *ни*) and modal particles (for instance, *-ка*, *вон*, *ж*/*же* and *ишь*), that carry the emotional colouring of the speech. Mystem recongises all of these simply as particles. Therefore, it is necessary to acknowledge such limitation when discussing the proportion of different POS in characters' speech. For further research, it would be interesting to separate the particles into different groups for a more precise comparison of female and male speech.

But, to sum up, the results of this analysis of parts of speech matches with the conclusion of Heldt (1987) that in Russian literature female culture is based on the expression of feelings, while male culture is based on the articulation of ideas.

Another feature that should highlight the difference in education and intellect between the women and men in the play is the lexical diversity of their speech. Lexical diversity reflects the range of unique words in speech. There are multiple ways of measuring this range, and the most straightforward would be to calculate the proportion of unique words in a character's speech. Such metric of lexical density is called type-token ratio (TTR) (Templin, 1957) where "type" refers to unique tokens. The TTR formula is

$$TTR = \frac{T}{N}$$

where T is the number of unique tokens (lemmas in our case) and N is the total number of tokens (lemmas).

To improve this measure, I used lemmas instead of simple tokens (words, as they appear in text), so that different word-forms of the same word would not count as unique

types. I took all the speech-acts of each character as a single text, lemmatised it and calculated the proportion of unique lemmas in this text. Below is the table with characters ranked according to such lemma-based TTR of their speech. The maximum possible value of TTR is 1.

Rank	Character	TTR	Number of all lemmas
1	Vtoroi 5	1.0	13
2	Tretiy	1.0	3
3	Drugoy	1.0	3
4	Neskol'ko golosov	1.0	10
5	Zhenshchina 4c	1.0	9
6	Zhenshchina 4b	1.0	14
7	Zhenshchina 4a	1.0	14
8	Pervyi 4b	1.0	10
9	Odin iz naroda 5b	1.0	2
10	Odin iz naroda 5a	1.0	2
11	Pervyi 5	1.0	16
12	Odin iz vozvrativshikhsya	0.88	24
13	Vtoroi 4b	0.87	30
14	Pervyi 4c	0.85	27
15	Drugoy iz vozvrativshikhsya	0.82	34
16	Shapkin	0.78	64
17	Vtoroi 4a	0.67	102
18	Glasha	0.59	194
19	The Barynya	0.57	152
20	Pervyi 4a	0.54	89

21	Feklusha	0.4	800
22	Kudryash	0.4	941
23	Kuligin	0.38	1450
24	Dikoy	0.37	741
25	Boris	0.34	1371
26	Varvara	0.34	1253
27	Kabanikha	0.31	1500
28	Tikhon	0.31	1277
29	Katerina	0.22	3152

Table 45. TTR of the characters' speech in *The Storm*

It is important to note that the number of lemmas in the characters' speech-acts might differ slightly from the number of words in their speech-acts due to the fact that the morphological analyser considers particles as separate lemmas, and, for instance, words such as *дома-то* are split into two lemmas, while in the word count they are considered as single words.

Straight away, we can see a correlation between TTR and the total number of lemmas in characters' speech. Thus, according to TTR, the most diverse lexicon in the play belongs to the episodic characters with the least number of words or lemmas. This is the main disadvantage of this measure that makes it so inaccurate. TTR is text-length dependent, and we should understand that with the increase of the length of a text, the probability of unique words or lemmas in this text reduces. At the same time, this probability increases for short texts, since if a character utters only one sentence, their speech is more likely to be constituted of unique words than the speech of a character who utters hundreds of sentences. For that reason, I will introduce three other lexical diversity measures that aim to even out the text-length disparity of characters' speech.

The first measure is called LOGTTR (Herdan, 1960). It follows the same logic as TTR, only it uses the logarithmic values of the number of unique types and tokens (lemmas in our case). This does not lower the rate of lexical diversity for characters with fewer words, but it increases this rate for characters with a higher number of words. The maximum possible value of LOGTTR is 1 and its formula is:

$$LOGTTR = \frac{Log\ T}{Log\ N}$$

The next measure that we will look at is called UBER. It is also a logarithmic value with a more complicated formula:

$$UBER = \frac{Log\ N^2}{Log\ \frac{N}{T}}$$

UBER (Dugast, 1978) is a relative measure, and thus it does not have a maximum value. For characters whose speech consists of only unique lemmas, UBER value is infinite.

Finally, the last measure we will discussed is called MSTTR (Johnson, 1994) which stands for mean segmental TTR. To calculate MSTTR for a text, we need to split the text into fixed-size segments, calculate TTR of each segment and take the mean value of these TTRs. The possible maximum value of MSTTR is 1.

For my analysis, I used 50-words segments, as has been done in a recent study on measuring lexical density and diversity (Nasseri and Thompson, 2021). Due to this, I had to disregard the first fifteen characters in Table 14, because there are fewer than 50 lemmas in total in their speech-acts.

Let us compare the ranks of characters with 50 or more lemmas in their speech-acts, according to the four measures mentioned:

Rank	TTR	TTR value	LOGTTR	LOGTTR value	UBER	UBER value	MSTTR	MSTTR value
1	Shapkin	0.78	Shapkin	0.94	Shapkin	70.07	Shapkin	0.82
2	Vtoroi 4a	0.67	Vtoroi 4a	0.91	Kuligin	54.46	Kuligin	0.81
3	Glasha	0.59	Glasha	0.90	Glasha	53.07	Glasha	0.78
4	The Barynya	0.57	The Barynya	0.89	Vtoroi 4a	52.76	Feklusha	0.78
5	Pervyi 4a	0.54	Kudryash	0.87	Kudryash	50.96	Boris	0.78
6	Feklusha	0.40	Kuligin	0.87	Feklusha	48.93	Kudryash	0.77
7	Kudryash	0.40	Pervyi 4a	0.86	Boris	48.83	Varvara	0.77
8	Kuligin	0.38	Feklusha	0.86	Varvara	47.37	Kabanikha	0.76
9	Dikoy	0.37	Dikoy	0.85	Kabanikha	45.92	Vtoroi 4a	0.75
10	Boris	0.34	Boris	0.85	The Barynya	45.23	Tikhon	0.74
11	Varvara	0.34	Varvara	0.85	Tikhon	43.88	The Barynya	0.73
12	Kabanikha	0.31	Kabanikha	0.84	Dikoy	43.57	Dikoy	0.72
13	Tikhon	0.31	Tikhon	0.84	Katerina	43.25	Katerina	0.72
14	Katerina	0.22	Katerina	0.81	Pervyi 4a	32.63	Pervyi 4a	0.70

Table 46. Lexical diversity measures for characters of *The Storm* with > 50 lemmas.

Comparing the TTR and LOGTTR results, we can see that the majority of characters have the same position in the LOGTTR rank, as they have in the TTR one. However, the LOGTTR logarithmic formula serves its purpose and the difference of lexical diversity in their speech become less significant. Katerina, who utters the greatest number of words in the play, has a TTR value of 0.22, while Shapkin with many fewer words has 0.78. In LOGTTR, Shapkin is still in the first position among the characters who utter at least 50 lemmas and Katerina still occupies the last position in this rank, but now the difference between their lexical diversity values is 0.13 and not 0.56.

Despite the fact that UBER and MSTTR metrics are more distinct from TTR and LOGTTR, there are characters that occupy the same position, or almost the same, position in all four ranks. These are Shapkin, who is in the first position of the four ranks, Glasha, who always occupies the third place, and Katerina, who is either the last or the one before last in the ranks. Other characters have less stable positions. For instance, the Barynya's lexical diversity value was quite high in TTR and LOGTTR ranks, which must be due to the fact that she utters fewer speech-acts than many other characters. This dependency levels out in UBER and MSTTR, where she is not close to the top of the rank. This highlights even more the leadership of Shapkin, who also utters fewer speech-acts than most of the characters, but the value of the lexical diversity of his speech stays high.

In general, the four metrics can be divided in two pairs judging by the level of similarity of the ranks that they produce: TTR and LOGTTR, and UBER and MSTTR. The authors of the "Lexical density and diversity in dissertation abstracts…" (Nasseri and Thompson, 2021) give an overview and evaluate twelve lexical diversity measures using the dissertation abstracts written in English by native speakers and by people who have learnt English as a second language. The results of the study suggest that UBER and MSTTR are

two of the four metrics that are proven to capture the difference in lexical diversity of abstract written by native and non-native speakers. Moreover, MSTTR is one of the two metrics that the authors concluded to be the most efficient ones. Even though in this study the lexical diversity measures are used for different purposes, it offers us some guidance on choosing the measure that would reflect the lexical diversity of the characters' speech in a more reliable way. In addition, MSTTR is a value between 0 and 1, which allows us to compare the lexical diversity of different characters in a clearer way, than is the case with UBER, where no specific possible maximum value exists. Thus, let us take a closer look at the MSTTR ranking of the lexical diversity of characters' speech.

Rank	Character	MSTTR value
1	Shapkin	0.82
2	Kuligin	0.81
3	Glasha	0.78
4	Feklusha	0.78
5	Boris	0.78
6	Kudryash	0.77
7	Varvara	0.77
8	Kabanikha	0.76
9	Vtoroi 4a	0.75
10	Tikhon	0.74
11	The Barynya	0.73
12	Dikoy	0.72
13	Katerina	0.72
14	Pervyi 4a	0.70

Table 47. MSTTR for characters of *The Storm* with 50 > lemmas

Although the difference between the MSTTR values of the speech of different characters is not very drastic, it is possible to draw some conclusions. First of all, it is surprising to see that Shapkin's speech is the most diverse, since his character is only present on stage in two scenes and is not notable in other metrics that we discussed earlier. Secondly, Kuligin has almost the same MSTTR value as Shapkin, which is as expected. This demonstrates that such lexical diversity measures are suitable for distinguishing how rich a character's speech is. If we split this ranking into two sections of seven characters, we would see that the second half of the ranking (the less diverse speech) is occupied by the characters who represent the older generation's principles – Dikoy, Kabanikha, the Barynya – and Tikhon and Katerina. On the whole, if we compare the average MSTTR values of male and

female groups of major characters, the difference is very small: male characters have MSTRR of 0.77 on average and female characters – of 0.76. Thus, there is no evidence to say that the women in *The Storm* have less "interesting" (diverse) vocabulary than men.

However, the major female characters (Katerina, Varvara and Kabanikha) tend to occupy lower positions in the lexical diversity ranks. And it is interesting that among the major characters Katerina is the one with the least diverse speech. This emphasises the uniqueness of Katerina and the strong direct image of her that Ostrovsky was portraying. Katerina is preoccupied by a single idea of freedom and is not in a situation of thinking and talking about new things. She is determined to escape, and this is the main focus of her dialogues and monologues.

Moving from the abstract mathematical representation of the characters' speech to the actual words, we will consider the most frequent words that the characters use. By means of the method described in the Introduction to Methodology chapter, I collected the lists of most frequent words for each character. This method requires calculating the frequencies of words used by each character and then comparing them to the frequencies of these words in the play in general. If a character uses a word more often than it is generally used in the whole play, it becomes their most frequent word. There are some tendencies among the female and male characters of the play in using certain function words and pronouns more frequently.

Below is the table with 10 most frequent words for the twelve characters we are discussing.

Katerina	Kuligin	Tikhon	Glasha	Kudryash	Barynya	Kabanikha	Varvara	Dikoy	Boris	Feklusha	Shapkin
Не	Сударь	Что	Не	А	В	Не	Что	Что	Не	И	Он
Я	И	Да	Так	Я	Вот	Что	Не	Ты	И	А	И
И	А	Я	Уж	Да	Все	Ты	Ты	Я	Я	В	Что
Мне	Да	Маменька	На	На	Весело	Ну	А	А	В	Милая	Уж
Меня	У	На	Батюшка	У	Ха	Так	Да	Вот	Мне	Матушка	А
На	Бы	Мне	Да	Он	Будете	Тебя	Так	Ну	С	Все	Как
Так	Вот	Так	Нас	Это	Куда	Уж	С	Меня	Меня	Так	Еще
В	Вы	Уж	Вы	Как	Хочется	С	На	Тебе	Бы	Вот	Бы
Как	Вам	Как	Вам	С	Омут	Как	Тебя	У	У	За	Тебя
Все	Ваше	Это	У	Уж	Ведет	Тебе	Вот	В	Уж	Девушка	Ею

Table 48. Most frequent words of the ‘core’ characters of *The Storm*

Katerina, Glasha, Kabanikha, Varvara and Boris use a lot of negation (*не*). Apart from Boris, this seems to be a largely female feature.

Kuligin, Glasha, Kabanikha, Varvara, Dikoy, Feklusha, Shapkin address others or talk about them a lot, using pronouns (*ты*, *тебе*, *тебя*, *вы*, *вам*), nouns and adjectives (*сударь*, *батюшка*, *матушка*, *девушка*, *милая*).

Tikhon frequently addresses Kabanikha as *матушка*. It emphasises Tikhon's personality: instead of asserting his authority as the man of the house, he is deferring to his mother in the same way as Glasha the servant and Feklusha the pilgrim do when addressing men and women that they obey.

Katerina, Tikhon, Kudryash, Dikoy and Boris talk a lot about themselves (*я*, *мне*, *меня*). Thus, while many women in the play talk to or about others, Katerina joins the men in talking about herself, as she tries to understand the direction her life should take.

Kuligin, Boris and Shapkin frequently use the subjunctive mood, realised via *бы*. We can speculate that they, being men in a patriarchal society, have more opportunities and, therefore, have a reason to contemplate a range of possible future outcomes.

Weight of edges and dialogue-based connection

The last piece of my analysis of the representation of women in *The Storm* is dedicated to the strength of connection between the characters.

First of all, let us look at the weight of edges (the number of scenes when characters are on stage at the same time).

The pairs of characters who speak together in at least 10% of the total number of 41 scenes are:

1) Katerina and Varvara with the edge's weight equal to 12.

2) Kabanikha and Tikhon with the edge's weight of 10.

3) Katerina and Tikhon have the weight of the edge between them equal to 6.

4) The edge between Varvara and Boris also equals 6.

5) Katerina and Kabanikha with the edge's weight equal to 5.

6) Kabanikha and Varvara, again, with the edge's weight of 5.

7) Varvara and Tikhon, whose edge's weight is also 5.

8) And finally, Katerina and Boris, who have the weight of the edge equal to 4.

As can be noted, among these eight pairs, there are connections both between women and women and between women and men. Katerina and Varvara have the highest weight of the edge between them and speak together in almost a third of the scenes. There are not, however, such strong edge-based connections between men and men. We will see the same picture when comparing the strength of the dialogue-based connections between the characters.

To repeat the definition of a dialogue in this study, it is an uninterrupted sequence of responses between two characters. The strength of the dialogue-based connection between two characters is calculated in the following way. Once such dialogues are identified, I calculate the number of responses of the two characters in these dialogues and divide it by the total number of responses uttered by these characters in the whole play. Below is the table with the pairs of characters who share at least one dialogue.

Pair	Dialogue-based strength of connection
Katerina and Varvara	0.56

Glasha and Feklusha	0.51
Tikhon and Kabanikha	0.47
Kudryash and Boris	0.35
Katerina and Boris	0.33
Kudryash and Shapkin	0.30
Kabanikha and Dikoy	0.30
Kuligin and Dikoy	0.29
Kuligin and Boris	0.23
Kuligin and Tikhon	0.19
Katerina and Tikhon	0.17
Kuligin and Kudryash	0.14
Kudryash and Varvara	0.13
Kabanikha and Feklusha	0.13
Varvara and Boris	0.12
Tikhon and Glasha	0.19
Tikhon and Varvara	0.07
Katerina and Kabanikha	0.07
Kabanikha and Varvara	0.04

Table 49. Strength of the dialogue-based connection between the 'core' characters of *The Storm*

The strongest connection is, again, between Katerina and Varvara. They devote half of their responses to each other. It is interesting to note that their preoccupations with love (Katerina and Tikhon, Katerina and Boris, and Varvara and Kudryash) take up much less dialogue space than the friendship between the two young women.

Thus, we can say that *The Storm* passes the Bechdel test,[22] and we can disagree with Heldt (1987) to argue that *The Storm*, in fact, does not correspond to her review of the patriarchal features of Russian literature. She writes that Russian male authors "tended to isolate female characters from one another, attending to their interactions with men under a male observing eye" (p. 62). In women's autobiographical works she notices the "complicated interactions between woman and woman, between women and family or society, and between woman and herself" and points out that women's narratives "rarely […] dwell much upon romantic relationships with men" (p. 62). In *The Storm* there is a basis for such woman-to-woman connection.

The next most well-connected pair is Glasha and Feklusha. They both have few speech-acts and they do not speak to many characters of the play.

Throughout the whole analysis we have seen how Kabanikha and her son Tikhon are almost inextricably tied to one another. The dialogue-based connection is no exception.

It is noteworthy to see that, despite a relatively strong edge connection, Kabanikha is really disconnected from both Katerina and Varvara in terms of productive communication. The only two people she has significant dialogues with are Tikhon and Dikoy. Kabanikha does make a lot of comments and gives orders, but she is not interested in hearing what others have to say in return.

Another striking insight, given to us by looking at the dialogue-base connections, is the strength of such connections between men. As we saw with the weight of edges, men do

[22] A way to measure the representation of women in fiction by establishing whether a fiction work contains at least two female characters who share a dialogue about something other than a man.

not appear on stage speaking at the same with each other very often. But, according to their dialogues-connections, once they do, their conversations are substantial.

To conclude, *The Storm* appears to be a play about women who are discussed by women. Even though the female characters of the play do not dominate in the network graph, they are the most active characters in terms of speaking activity. Judging by the edges' weights and the dialogue-based connections, as well as by the content of these conversations, women in *The Storm* talk a lot to each other about each other. Besides that, Katerina is clearly set apart from the other characters by Ostrovsky. She has the maximum number of scenes, speech-acts and words in the play, as well as the greatest number of monologues, with entire scenes dedicated exclusively to her speech. Her speech is emotional (the greatest proportion of exclamation marks after Barynya) and persistent (the least diverse), and her actions are described in the greatest detail (the greatest proportion of stage directions after the Barynya). Even though *The Storm* is a play about the tragic destiny of a young woman who is forced to take her own life because of the unbearable pressure of the merchant society, it is a story told by her. While the next play we are going to discuss – *Without a Dowry* – has an ideologically similar plot-line, we will see that the women there are represented quite differently.

Without a Dowry

Without a Dowry, like *The Storm*, tells the story of a young woman from a town in the Volga region. It was written and premiered in 1878 and first published in 1879. Ostrovsky defined its genre as drama. The play has four acts and 44 scenes. The main character of *Without a Dowry* is Larisa Ogudalova who lives in the fictional town of Bryakhimov, which shares its name with a previously existing historic town in the Volga region.

Larisa is the third daughter of a widow, Kharita Ignatievna. Their family is quite poor and the mother is desperate to marry Larisa off to a decent man. Larisa's fiancé,

Karandyshev, is a minor civil servant. Larisa does not love or respect Karandyshev, who is also not as wealthy as would have been preferred by her mother, but there are no other candidates for her. Even though she is beautiful and interesting to the men of Bryakhimov, she is without a dowry.

The action of the play happens in the course of one day. We are introduced to the small community of the friends of the Ogudalov family – two merchants, Knurov and Vozhevatov, the nobleman and shipowner Paratov, and Ilya the gypsy.

Knurov is a very wealthy and respected old man, while Vozhevatov is a young childhood friend of Larisa. Sergey Sergeyich Paratov is the former love of Larisa who, as we find out, had left the town and abandoned Larisa unexpectedly one year ago. He comes back to Bryakhimov to sell his ship and leave again to marry a well-off bride.

When Paratov finds out that Larisa is about to get married to the ordinary, somewhat despicable Karandyshev, he decides to teach him a lesson at their engagement celebration. Together with Knurov, Vozhevatov and Paratov's recent acquaintance Robinzon, he gets Karandyshev drunk and bullies him. Paratov, Knurov and Vozhevatov also persuade Larisa to go on a trip across the Volga with them. Larisa runs away from Karandyshev's house and spends the night with the men and the gypsies. In the morning she finds out that Paratov will not stay with her, and is getting married.

In the meantime, Knurov and Vozhevatov toss a coin to decide which of them will step in to try rescue Larisa from her disgrace and an unfortunate marriage with Karandyshev. Knurov (an old and married man) wins and offers Larisa to take her away to live in Paris. She refuses. At this moment, she realises that she has become a mere thing for all these men. Larisa for them is a transactional object of the same interest as Lastochka – Paratov's boat that he is selling.

Heart-broken and humiliated, Larisa is even more disgusted by Karandyshev, and so they have a heated argument once he arrives with a gun to fight the men who took her away from him. Mad with rage, Karandyshev shoots Larisa, who is happy to meet her death and thanks Karandyshev for it, while assuming the blame for it herself.

The other characters are Karandyshev's aunt Efrosinya Potapovna, the café's owner Gavrilo, and Ivan, who works as a servant in the café.

Let us turn to the quantitative representation of the play to see how Larisa's position is reflected in it, and whether she is an active or a passive participant in the events of the play.

Instead of a single *dramatis personae* list for the play, there are four different lists of the play's characters, one before each act. Thus, all the 11 actual speakers are mentioned in these cast lists. Unlike *The Storm*, *Without a Dowry* does not have other "hidden" participating characters, such as a crowd or "one woman". Ilya and a few more gypsies often accompany other characters. While only Ilya utters any speech-acts, there are several occasions when the speech of other gypsies is included in the stage directions related to his speech-acts.

As mentioned earlier, in *Without a Dowry* there are 27% of female characters, with the same proportion of female speech-acts and 29% of female words. Male responses, on average, contain 11.91 words and female responses – 12.67 words. To conclude, female characters occupy roughly a third of space in the play and the length of their responses slightly exceeds the length of male responses.

Network graphs and speaking activity

Below are the network graph and network metrics of *Without a Dowry*.

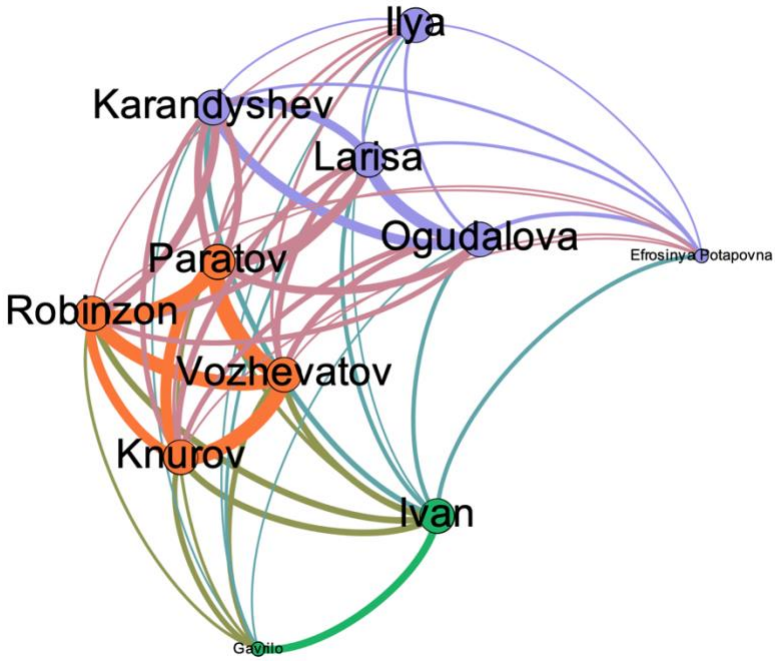


Figure 24. Network graph of *Without a Dowry*

Character	Degree	BC
Ivan	10	0.11
Vozhevatov	10	0.11
Knurov	10	0.11
Ogudalova	10	0.11
Karandyshev	10	0.11
Larisa	10	0.11
Paratov	10	0.11
Robinzon	10	0.11
Ilya	10	0.11
Gavrilo	9	0
Efrosinya Potapovna	9	0

Table 50. Degree and betweenness centrality in *Without a Dowry*

The density of this graph is 0.98, which means that almost every character is connected to every other character. Thus, all the characters, except for Gavrilo and Efrosinya Potapovna, share the same (the maximum) degree and the same betweenness centrality. Gavrilo and Efrosinya Potapovna have a degree of 9 out of 10, because they miss one connection in the network – they do not 'speak' to each other.

However, if we look at the weighted degree that reflects the volume of communication of the characters, we will see that some of them are more active and, therefore, influential than others:

Character	Weighted Degree
Vozhevatov	67
Paratov	63
Robinzon	57
Knurov	55
Larisa	49
Ogudalova	46
Karandyshev	46
Ivan	38
Gavrilo	23
Ilya	15
Efrosinya Potapovna	15

Table 51. Weighted degree in *Without a Dowry*

All the "important" men of the play occupy the highest positions when ranked by the weighted degree – Vozhevatov, Paratov and Knurov. They are accompanied in this

leadership by Robinzon, who plays the role of their jester. The picture changes when we look at the characters' presence and speaking activity:

Rank	Character	Number of scenes
1	Paratov	19
2	Vozhevatov	18
3	Larisa	18
4	Robinzon	17
5	Ogudalova	16
6	Knurov	14
7	Karandyshev	13
8	Ivan	12
9	Gavrilo	8
10	Ilya	4
11	Efrosinya Potapovna	4

Table 52. Number of scenes for the characters of *Without a Dowry*

Rank	Character	Number of speech-acts
1	Paratov	203
2	Vozhevatov	192
3	Larisa	176
4	Karandyshev	163
5	Ogudalova	148
6	Knurov	127
7	Robinzon	111
8	Ivan	50

9	Gavrilo	32
10	Ilya	23
11	Efrosinya Potapovna	14

Table 53. Number of speech-acts for the characters of *Without a Dowry*

Rank	Character	Number of words
1	Paratov	2654
2	Larisa	2437
3	Vozhevatov	2313
4	Karandyshev	2007
5	Knurov	1514
6	Ogudalova	1421
7	Robinzon	1092
8	Efrosinya Potapovna	425
9	Ivan	423
10	Gavrilo	409
11	Ilya	315

Table 54. Number of words for the characters of *Without a Dowry*

The top three characters when ranked by number of scenes, speech-acts and uttered words are Paratov, Vozhevatov and Larisa. They are also followed by Karandyshev in the ranking of the number of speech-acts and number of words. Thus, even though Larisa does not share the network influence with the men of the play, she is present and active on stage. It would be surprising if Larisa was not an active character since the title of the play, if translated literally, would be “The Woman without a Dowry”. Nevertheless, the most active character, according to the combination of weighted degree and number of scenes, speech-acts and words, is Paratov. It conforms with his agency and the authority that he uses over

other characters. Paratov makes the choice to leave or come back to Bryakhimov, to bully Karandyshev for nothing, to be and then not to be with Larisa.

We already see that the characters with the highest number of speech-acts also have the highest number of words, which seems proportionate. Let us look at the average number of words per response to confirm that.

Rank	Character	Number of words per response
1	Efrosinya Potapovna	30.36
2	Larisa	13.85
3	Ilya	13.70
4	Paratov	13.07
5	Gavrilo	12.78
6	Karandyshev	12.31
7	Vozhevatov	12.05
8	Knurov	11.92
9	Robinzon	9.84
10	Ogudalova	9.60
11	Ivan	8.46

Table 55. Average number of words per response for the characters of *Without a Dowry*

Seven out of eleven characters have relatively the same number of words per speech-act in general. However, Efrosinya Potapovna's speech-acts turn out to be the longest. It is noticeable because she occupies the last position in the ranking of number of speech-acts and the eighth position in the ranking of number of words. This reminds us of the Barynya in *The Storm*, who had the longest speech-acts in the play.

Speech analysis

As with previous plays, we will look into the proportion of stage directions per response to evaluate the scale of the characterisation of characters' speech and actions, since the characters with more detailed description are presumably more important for the author and the play.

Rank	Character	Proportion of stage directions per speech-act
1	Ilya	0.57
2	Larisa	0.31
3	Karandyshev	0.30
4	Paratov	0.23
5	Robinzon	0.23
6	Knurov	0.21
7	Efrosinya Potapovna	0.21
8	Vozhevatov	0.20
9	Ogudalova	0.16
10	Ivan	0.12
11	Gavrilo	0.09

Table 56. Proportion of stage directions related to the characters of *Without a Dowry*

Surprisingly at first, Ilya, who only utters 23 speech-acts, has the highest proportion of stage directions per speech-act. However, if we look at these stage directions, they are mostly about him coming and leaving at the request of other characters and singing or accompanying Larisa with his guitar. Thus, his stage directions do not reflect any detailed elaboration on his manners, but rather they indicate his appearance on stage and his singing.

The other two characters with the highest number of stage directions per speech-act are Larisa and Karandyshev. The stage directions related to them include the characterisation of their speech: *задумчиво*, *обидясь*, *взглянув в окно*, *холодно*, *запальчиво*, *сквозь слезы*, *сдержанно*, etc. The next five characters in this ranking all have a similar proportion of stage directions per speech-act, and Ogudalova, Ivan and especially Gavrilo have the lowest numbers. Thus, Larisa and Karandyshev are the two characters with the most detailed portrayal. The interpretation could be that they are the most sincere and sensitive characters, who are motivated by their genuine feelings and reactions, marked by the author.

Another feature that can be used for indicating the emotional colouring of speech is punctuation. Below is the table with the characters ranked by the proportion of exclamation marks in their speech.

Rank	Character	Mean number of exclamation marks per sentence
1	Ilya	0.32
2	Ogudalova	0.28
3	Efrosinya Potapovna	0.27
4	Robinzon	0.27
5	Larisa	0.26
6	Paratov	0.25
7	Karandyshev	0.21
8	Vozhevatov	0.19
9	Ivan	0.19
10	Gavrilo	0.18
11	Knurov	0.16

Table 57. Proportion of exclamation marks in the speech of the characters of *Without a Dowry*

Ilya appears to be the most emotional character of the play. He is a loud and lively person who earns his living by entertaining the richer men. Some of the exclamations in Ilya's speech-acts appear during his singing, which indicates a passionate song rather than characterising his speech specifically.

The next five characters in this ranking have approximately the same proportion of exclamations per sentence. It is interesting to note who appears at the bottom of the table – Vozhevatov, Ivan, Gavrilo and, lastly, Knurov. It indicates that Paratov is, evidently, more involved in the communication and relationships in the play. On the other hand, Paratov is, in the end, a lively and passionate character. Vozhevatov and Knurov are more indifferent and cold. Knurov's lowest position in this ranking reflects very well his practical, experienced and cynical character.

In general, the few women in the play occupy the first half of the exclamation marks ranking, while the active men of the play gravitate toward the bottom of the ranking.

Looking at the proportion of ellipses in the characters' speech, we see a similarity with *The Storm*:

Rank	Character	Mean number of ellipsis per sentence
1	Larisa	0.14
2	Ivan	0.12
3	Robinzon	0.08
4	Gavrilo	0.08
5	Karandyshev	0.05
6	Paratov	0.04
7	Knurov	0.03

8	Ilya	0.03
9	Ogudalova	0.02
10	Vozhevatov	0.01
11	Efrosinya Potapovna	0.00

Table 58. Proportion of ellipses in the speech of the characters of *Without a Dowry*

Larisa, like Katerina, has the highest proportion of ellipses, which demonstrates the deep level of self-reflection and uncertainty.

The high number of ellipses in Ivan's speech reflects his complaisance and cautiousness when talking to and about people. Consider this extract from the sixth scene of the first act, when Paratov arrives at the café and Ivan starts sweeping off the dust and dirt from Paratov's shoes and clothes:

> **Паратов *(Ивану).*** Да что ты! Я с воды, на Волге-то не пыльно.
>
> **Иван.** Все-таки, сударь, нельзя же... порядок требует. Целый год-то вас не видали, да чтобы... с приездом, сударь.

Ivan talks to Paratov in this servile way to insist on his actions. In another excerpt Ivan, again, is careful and circumspect when discussing with Robinzon the actions of the gentlemen of the town who do not bother to pay for their bills:

> **Робинзон.** Какой народ! Удивляюсь. Везде поспеют; где только можно взять, все уж взято, непочатых мест нет. Ну, не надо, не нуждаюсь я в нем. Ты ему не говори ничего, а то он подумает, что и я хочу обмануть; а я горд.
>
> **Иван.** Да-с, оно, конечно...

The ratio of ellipses in Karandyshev's speech is almost three times lower than in Larisa's. Further down the ranking we see Paratov, Knurov, Larisa's mother and Vozhevatov. In general, the picture changes here and Kharita Ignatievna, who has a high ratio of

exclamations in her sentences, now joins the men with her unhesitant and unapologetic speech.

It is understandable that Ilya and Efrosinya Potapovna, who have few speech-acts, would not necessarily be expected to have a lot of profound speech characterisation. Thus, it is notable that Vozhevatov, who utters almost as many sentences as Larisa and more speech-acts than her, has very few ellipses in them, which means that he does not hesitate about his thoughts and actions and is not as contemplative as Larisa.

On a general note, it is interesting to compare the numbers of exclamation marks and ellipses in *The Storm* and *Without a Dowry*. The characters of *Without a Dowry* appear to be more quiet and more hesitant or self-reflective than the characters of *The Storm*.

As discussed earlier in *The Storm* section, one of the ways to analyse the complexity and the constitution of the speech is to look at the different parts of speech distribution in characters' speech-acts.

Part of speech	Female characters	Male characters	Difference
Noun	0.18	0.224	0.044
Pronoun-noun	0.163	0.129	0.035
Particle	0.12	0.093	0.028
Pronoun-adjective	0.06	0.039	0.021
Verb	0.167	0.186	0.02
Preposition	0.063	0.08	0.017
Conjunction	0.087	0.073	0.014
Adjective	0.037	0.05	0.013
Interjection	0.017	0.01	0.007
Adverb	0.067	0.071	0.005

Pronoun-adverb	0.037	0.036	0.000

Table 59. Proportion of different parts-of-speech in the speech of the characters of *Without a Dowry*

As in *The Storm*, the women in *Without a Dowry* use fewer nouns in their speech than men. But, unlike in *The Storm*, men in this play use both more nouns and verbs than women.

In fact, it is Larisa who significantly lowers the proportion of nouns in female speech in the play. Nouns constitute approximately 21% of the speech-acts of both Larisa's mother and Efrosinya Potapovna and only 12% of Larisa's speech (the lowest ratio among all characters). This is compensated by a relatively high proportion of pronoun-nouns in her speech, and we see that the pronoun-nouns ratio is higher in female speech-acts.

Efrosinya Potapovna, in her turn, has the lowest proportion of verbs among all characters, therefore lowering this ratio for the whole female group of characters.

The outlier among male characters is Ilya. He contributes to a significant difference in the proportion of nouns in the male and female speech. Ilya's speech consists of 33% nouns, and this must be connected to the fact that he uses many fewer pronoun-nouns than the others.

As we can see, a small number of characters in the play can lead to the fact that such outliers can significantly influence the overall picture. A higher number of characters would allow for a more even distribution of each character's contribution to the statistics of this kind. In this way, it is important to note that for a more substantial analysis of plays with smaller networks it is worth looking at individual characters as well as at the general numbers.

On the whole, the women in the play (especially Larisa) talk less about various concepts than men because of the lower proportion of nouns in their speech. This is

compensated by the usage of pronoun-nouns, meaning that the subjects and objects of female responses are themselves or other characters that they refer to by pronouns. Moreover, women use more particles, which might contribute to more lively or more emotional responses.

When it comes to the lexical diversity of the characters' speech, the results are quite similar across all the characters, unlike in *The Storm*:

Character	Number of all lemmas
Paratov	2668
Larisa	2445
Vozhevatov	2363
Karandyshev	2023
Knurov	1533
Ogudalova	1441
Robinzon	1107
Ivan	459
Efrosinya Potapovna	436
Gavrilo	431
Ilya	316

Table 60. Number of lemmas in the speech-acts of the characters of *Without a Dowry*

Rank	TTR	TTR value	LOGTTR	LOGTTR value	UBER	UBER value	MSTTR	MSTTR value
1	Efrosinya Potapovna	0.54	Efrosinya Potapovna	0.90	Paratov	59.48	Paratov	0.83
2	Ivan	0.52	Ivan	0.89	Efrosinya Potapovna	59.35	Efrosinya Potapovna	0.83
3	Gavrilo	0.52	Gavrilo	0.89	Vozhevatov	58.20	Vozhevatov	0.83
4	Ilya	0.49	Ilya	0.88	Ivan	57.93	Knurov	0.82
5	Robinzon	0.40	Robinzon	0.87	Gavrilo	56.23	Ivan	0.79
6	Knurov	0.36	Paratov	0.87	Robinzon	53.65	Robinzon	0.79
7	Paratov	0.35	Vozhevatov	0.87	Knurov	52.3	Ogudalova	0.79
8	Vozhevatov	0.35	Knurov	0.86	Karandyshev	50.69	Gavrilo	0.78
9	Ogudalova	0.35	Ogudalova	0.85	Ogudalova	49.88	Larisa	0.76
10	Karandyshev	0.32	Karandyshev	0.85	Larisa	47.02	Ilya	0.76
11	Larisa	0.27	Larisa	0.83	Ilya	46.93	Karandyshev	0.75

Table 61. Lexical diversity measure of the speech of the characters of *Without a Dowry*

As we know from the previous section, TTR value is text-length dependent. Thus, it is important to note that even though Paratov has the greatest number of lemmas (see Table 60), his position even in TTR ranking is not close to the bottom, while Larisa occupies the last position in this ranking with rather fewer lemmas than Paratov.

If we look at the MSTTR values, Paratov, Efrosinya Potapovna and Vozhevatov are the characters with the most diverse speech. And Efrosinya Potapovna occupies either the first or the second position in all the four rankings, which indicates that her lexical diversity characteristic is the most reliable.

At the same time, other characters are not far much behind them. If we split the MSTTR ranking into two parts – characters with MSTTR greater than 0.80 and smaller than 0.80 – we can separate the three businessmen of the play, together with Efrosinya Potapovna, from the rest of the play. However, as already mentioned, the overall difference between the characters is not drastic in either of the LOGTTR, UBER or MSTTR ranks. Moreover, the average MSTTR values for women and men in the play are the same, and equal 0.79. However, similarly to *The Storm*, the main heroine of the play – Larisa – lingers toward the bottom of the lexical diversity ranks (within the 'own language' of the characters in the play).

And finally, let us discuss the strength of the relationship bonds between different characters to see how well the women, and specifically Larisa, are involved in communication. Since the network graph of *Without a Dowry* has a higher density than the network graph of *The Storm*, there are many more pairs of connected characters, as well as many more characters who share at least one "dialogue".

Below is the table with the pairs of characters who speak together in at least 15% out of 44 scenes of the play.

Source	Target	Weight

Vozhevatov	Paratov	14
Vozhevatov	Knurov	13
Vozhevatov	Robinzon	12
Paratov	Robinzon	12
Ogudalova	Larisa	10
Knurov	Paratov	9
Knurov	Robinzon	8
Ogudalova	Karandyshev	8
Larisa	Paratov	8
Karandyshev	Larisa	7
Karandyshev	Robinzon	7

Table 62. Weight of the edges between the characters of *Without a Dowry*

Paratov and Vozhevatov "talk to each other" most frequently than any other characters. The next three pairs of characters are also man-to-man connections. Even though Larisa appears in almost as many scenes as Paratov, her network connection is significantly strong only with her mother, Paratov and Karandyshev. Moreover, according to this data, only Paratov and Karandyshev have frequent interactions with the women, while overall the strongest bonds are between the men of the play.

If we look at the characters whose dialogue-based connection is at least 0.15 strong, we will also see substantial man-to-man bonds:

Pair	Dialogue-based strength of connection
Knurov and Vozhevatov	0.46
Karandyshev and Larisa	0.36
Larisa and Paratov	0.34
Gavrilo and Ivan	0.33

Robinzon and Vozhevatov	0.28
Knurov and Ogudalova	0.25
Ivan and Robinzon	0.21
Larisa and Ogudalova	0.21
Karandyshev and Paratov	0.18
Ogudalova and Paratov	0.17
Paratov and Robinzon	0.17
Karandyshev and Robinzon	0.16
Paratov and Vozhevatov	0.15

Table 63. Strength of the dialogue-based connections between the characters of *Without a Dowry*

Judging by the "dialogues", Larisa has a strong connection and long conversations with Karandyshev, her fiancé, and Paratov, her lover. Even though Paratov often appears together with other men, he does not become absorbed in profound conversations with them except, perhaps, for his argument with Karandyshev.

It is interesting to note the dialogue-based connection between Knurov and Ogudalova that we can sense from reading the play. As in *The Storm*, the older generation appears to share conversations.

However, the strength of connection between Knurov and Vozhevatov and between Gavrilo and Ivan reminds us that the play is about the men doing business together. Furthermore, this male practice of 'doing deals' extends to the tossing of a coin for Larisa by Knurov and Vozhevatov, bringing home to Larisa the fact that all four men treat her only as a 'thing'. This is supported by the fact that even though Larisa and Vozhevatov share the leadership in their speaking activity and are old childhood friends, they do not have either a strong network edge between them, nor many 'dialogues'.

In the end, *Without a Dowry* turns out to be a men's play, where Larisa, despite appearing to be the main character of the play while reading it, becomes only the object of the men's deals. Even though Larisa appears in the top section of the speaking activity rankings, she does not dominate the characters of the play in terms of her speaking activity, unlike Katerina. While in both plays – *The Storm* and *Without a Dowry* – men have more complex speech, they also have more substantial and stronger connections between each other in the latter play.

Conclusion

In this chapter, the use of digital methods has allowed me to quantitatively and qualitatively survey the representation of women in Russian plays of the 19th century. We have seen that the majority of the 19th-century authors included more male than female characters in their plays. On average, women only 'occupy' one third of the 19th-century Russian plays.

Drawing upon the quantitative analysis of the RusDraCor, I was able to identify the most frequently represented author of the corpus – Ostrovsky. Relying on the great number of his plays in the corpus and the general statistics of how well women are represented in his plays, it became possible to name him as one of the most 'equal' playwrights of his time.

The analysis of *The Storm* and *Without a Dowry* provided additional understanding of Ostrovsky's tendency to dedicate his plays to the women's question, among other things. While neither *The Storm*, nor *Without a Dowry* appeared to have a greater number of female characters and female speech-acts during the preliminary analysis of Ostrovsky's plays, it was necessary to take a closer look at these two plays to identify that *The Storm* is quite a female-oriented play. Looking at both the general 'female' statistics of his plays and the study of *The Storm*, it is reasonable to say that Ostrovsky indeed provided a fairer platform for women in his drama. I suggest that further thorough analysis of a greater number of his plays would help to establish that.

New Drama

Until now, the focus of this work has been only on plays written before the 20th century. While they required extra work in terms of the TEI-encoding and their quantitative analysis, such as additional segmentation in *A Month in the Country* and *Uncle Vanya* or narrowing down the 'core' characters in *The Storm*, the 19th-century plays could still be reliably analysed using the proposed methodology, both on a large and small scale.

In this final chapter we will consider the possibilities of studying modern Russian drama – or plays written in Russian – with the use of digital methods, mostly social network analysis (SNA). I will also explore more recent plays from the 21st century, and specifically some of the interesting features of 'New Drama', which at times differs greatly from the plays that I have been analysing so far.

As we have seen previously, the SNA method in general works well with plays, since they offer a structured representation of the communication between characters. However, one of the major developments in Russian-language drama of the late 20th-21st centuries has been their less conventional structure. Plays written since the beginning of the 21st century have mostly lost their familiar configuration, together with clearly defined characters and actions. The research question shaping this chapter is what potential there is in using current methodology for studying these plays and what the adjustments and assumptions are that need to be made in order to adapt it to contemporary plays.

The later plays in the RusDraCor

The Russian Drama Corpus currently (February 2022) features 212 plays from 1747 to 1947. The 20th-century authors whose plays are included in the RusDraCor are Lev Tolstoy, Sergey Naydyonov, Maksim Gorky, Anton Chekhov, Aleksandr Blok, Leonid Andreev, Vladimir Belsky, Dmitry Merezhkovsky, Nikolay Gumilev, Mikhail Artsybashev, Velimir

Khlebnikov, Vladimir Mayakovsky, Mikhail Bulgakov, Isaac Babel, Dmitry Kedrin, Aleksandr Afinogenov, and Evgeniy Petrov.

Although it is a great open-source resource, from which the present dissertation and other research projects have greatly benefited, it has its limitations. As discussed in the previous chapter, it lacks representation of female writers. It also does not cover any modern plays, because of copyright issues. For this chapter I therefore had to independently compile a subcorpus of more recent plays. These 23 plays were written between 1966 and 2020. The choice of the authors and plays was based on subjective impressions of how well-known they are, as well as – for the most recent plays – their influence in the 'New Drama' movement.

This movement, that emerged in the 2000s, took its name from the festival organised by Elena Gremina and Mikhail Ugarov (founders of Teatr.Doc) and Eduard Boyakov (founder of the Golden Mask festival) in 2002. It is a complex phenomenon. And even though there have been many attempts to define the criteria for a 'New-Drama' play, Freedman notes that the term "applies more to a time period than to any specific manner of writing" (Freedman 2014, p.7). Nevertheless, many of the most typical features that characterise 'New Drama' will be discussed in this chapter.

In order to link the 20th and the 21st centuries, I have included some earlier authors who either stood on the border line between conventional plays of the late Soviet era and experimental drama, or have shifted towards 'New Drama' in their later works. Thus, the subcorpus that I have compiled contains 20th-century plays by Vampilov, Petrushevskaya and Razumovskaya.

Furthermore, in this chapter I introduce the term 'Russian-language drama'. During the 2010s, tensions in international relations in the post-Soviet space escalated, with the most important event being the annexation of Crimea by Russia in 2014. This intensified national

movements in post-Soviet countries such as Ukraine, where civil society increasingly aimed at distancing itself from Russian culture and the Russian language. Russian, Ukrainian and Belarusian theatres and theatremakers, that once formed a close and collaborative community, began to separate. Some playwrights started using their national languages, and focused on individual issues within their own countries. However, many of the topics remained similar and drama, as one of the most reactive forms of art, concentrated on current politics and social problems. Thus I believe that it is appropriate to include some plays by Ukrainian and Belarusian authors when discussing 'New Drama' written in the Russian language. Even though this chapter covers only one play by a non-Russian author, that is merely a coincidence and not an intentional disregard.

Within and beyond Socialist Realism

The 19th century was extremely fruitful for the Russian theatre and brought to us the plays of Gogol', Pushkin, Griboyedov, Turgenev, Chekhov and Ostrovsky. The latter three are the most associated with the development of Russian realism. Realism on stage was also necessary for the productions of historical drama. The playwrights turned to the genre to avoid the censorship of the court and to continue to discuss the pressing political and social issues of the country.

As a reaction against the realist movement, an experimental, modernist theatre then emerged. The Revolution of 1917 provided further conditions for experimental explorations in drama during the reshaping and re-evaluating of the theatre and its purposes for the country. Unfortunately, this experimental freedom did not last for a long time, and in 1923 the Glavrepertkom (the repertoire committee) was formed to oversee the repertoire of Soviet theatres.

By the 1930s, Stalin's grip became tighter, and the artistic community faced multiple arrests, the closure of theatres, and the banning of plays and productions. Writers and directors found themselves forced to choose between complying to the party line, escaping the country, or being arrested and, possibly, executed. By suppressing what was termed 'formalism' and the avant-garde, the party propagated Socialist Realism in art. The theatre was supposed to 'ideologically transform' the people, educate them 'in the spirit of socialism' and motivate them to work for the greater good of the Soviet Union – all in all, to manipulate people. There was nothing real about Socialist Realism.

In the 1950s, with Khrushchev becoming the leader of the party, the Soviet Union went through a period of Thaw. Literary critics appealed to writers to express the truth and expose the imperfections of the Soviet reality. In 1953 *Новый Мир* published Vladimir Pomerantsev's essay on sincerity in literature (Pomerantsev 1953). Pomerantsev criticised the obviously 'staged' ("состроенные") characters in carefully constructed novels and plays. He believed that authors should reinstate the emotional lives of their characters as the focus of literature. Viktor Rozov and Aleksandr Volodin (the author of such famous plays as *Пять вечеров*, *С любимыми не расставайтесь* and *Дочки-матери*) were among the important playwrights of that period. They were driven by the idea of breaking away from the ideological nature of Soviet literature, and their plays contained elements of romanticism by focusing on the characters' evolution and the tensions between internal and external conflicts.

However, the Thaw did not last for a long time either and was followed by another wave of repressions, that led to artistic stagnation in the 1970s. The theatre lacked the decent drama that would correspond to the party's criteria and turned to adapting prose. At that time, the plays of Aleksandr Vampilov (1937-1972) were discovered. While some of Vampilov's plays were staged in Latvian theatres, the most important were the productions by Oleg Yefremov in the Moscow Art Theatre in the late 1970s. Vampilov is an important post-Stalinist author, because

he, among few others, bent the rules of Socialist Realism by shifting the thematic focus of his plays on to deep psychological issues and, as we will see, experimented with different ways of narrating a story.

In the subcorpus that I have collected for this chapter, I have included the four longest plays out of approximately twenty plays written by Vampilov.

Most of the source texts for this subcorpus have been gathered from the online Theatre Library assembled by Sergey Efimov.[23] Some of the plays that were not available in the Efimov library were taken from other online libraries, or from a collection of plays published by Teatr.Doc.

The encoding process was partly automated by using simple Python scripts that went through the text of the play with basic assumptions about the structure of the texts (for example, the character name being the first word(s) on the line followed by a full stop). All source files were either in Word or PDF format. Since, in the case of some plays, the files contained additional formatting, such as bold font or all capitals for characters' names and italic font for stage directions, it simplified the encoding. Despite that, the XML files still required manual validation in all cases.

Apart from small errors that the formatting or the scripts did not account for, the manual editing addressed conceptual questions, some of which we have already touched upon before.

For instance, during the encoding process it became clear that newer plays contain more stage directions and authors' remarks than the earlier plays. Moreover, encoding certain types of stage directions required making a subjective decision.

[23] https://theatre-library.ru/

To illustrate this, let us look at several speech-acts from Aleksandr Vampilov's *Farewell in June* (*Proshchaniye v iyune*) (1964). This is a two-act comedy about the love-story of a young student Kolesov and Tanya. Tanya's father – Repnikov – is the rector of the university where Kolesov studies. The two men have a conflict that compromises the relationship between Tanya and Kolesov. But the play finishes with Tanya and Kolesov trying to 'restart' their relationship.

```
<sp who="#kolesov">
    <speaker>Колесов </speaker>
    <stage>(перебивает).</stage>
    <p> Ладно.</p>
</sp>
```

Figure 25. Extract from the first act of *Farewell in June*

```
<sp who="#repnikov">
    <speaker>Репников.</speaker>
    <p>Ага... <stage>(Веселому.)</stage> Вы, видимо, жених?
            <stage>(Садится.)</stage></p>
</sp>
```

Figure 26. Extract from the first act of *Farewell in June*

```
<sp who="#gomyra">
    <speaker>Гомыра.</speaker>
    <p> Ладно... пью последнюю. За цивилизацию. <stage>(Пьет и
        выходит.)</stage></p>
</sp>
<stage>За Гомырой - Серьезный.</stage>
```

Figure 27. Extract from the first act of *Farewell in June*

For reference, in Figure 25 we see a classic case of a stage direction, where the remark relates to the character's speech and is situated inside the speech-act. Figures 26 and 27 demonstrate one kind of decision that I have made during the encoding. If a stage direction is at the end of a speech-act (i.e., not followed by any further words by the character), but it relates to the character's speech or actions relevant to the speech-act (which is normally indicated by parenthesis), the stage direction is included in the <sp> element (Figure 27). If a stage direction

is at the end of a speech-act and does not relate to the character, it is put outside of the <sp> element and is considered relevant to the text of the scene or act in general. One might take a different perspective on that and exclude the stage directions from the speech-act in both cases. Thus, such details need to be taken into account when talking about the proportion of stage directions per speech-act.

Talking about the network measures of *Farewell in June*, the play has a conventional network structure and benefits from the methodology, since its network graph illustrates well the division of the characters into two clusters. The green nodes correspond to the community of Kolesov, the rector and the Repnikov family. The pink nodes represent the student community, who do not have many interactions with the Repnikovs. Despite that division into two clusters, most of the characters in the play 'meet' each other at least once and the density of the play's network is 0.99.

While most of the nodes in the graph have the same degree, the graph highlights how, unlike Repnikov, his wife Repnikova, Tanya's mother, is barely involved in the communication in the play. Repnikov, on the other side, together with Tanya and Kolesov, is in the top-three characters in the weighted degree ranking. This makes the three of them the most communicative characters in the play. Moreover, the graph and its network metrics show that Tanya and Kolesov have the strongest connection because the edge between their nodes has the highest weight.

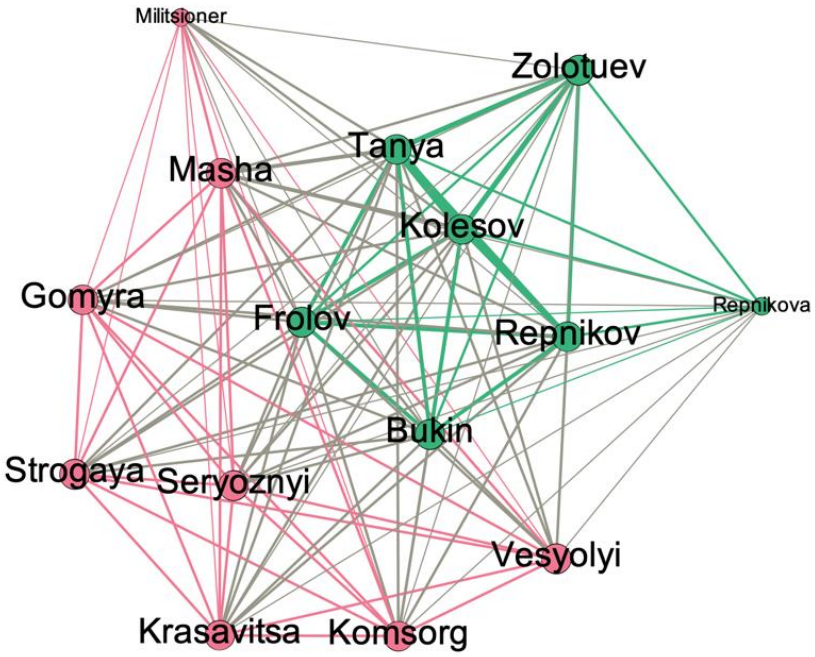


Figure 28. Network graph of *Farewell in June* by Vampilov

The second of Vampilov's plays in the subcorpus is *The Elder Son* (*Starshii syn*) (1967) – another two-act comedy revolving around the story of the fraud perpetrated by a young student, Busygin, who pretends to be the illegitimate son of Sarafanov. Sarafanov believes Busygin and lets him stay with him and his family (the younger son Vasen'ka and daughter Nina). Busygin falls in love with his 'sister' and eventually confesses that he is not Sarafanov's son, but they grow fond of each other, so Busygin stays with the Sarafanovs after all.

The network of *The Elder Son* is slightly less dense (its density is 0.87) than the network of *Farewell in June*. Still, this play also has a conventional network structure and does not differ much from the plays of the 19th century. The clustering in the graph separates the two young women that Busygin and his friend Sil'va meet in the beginning of the play, together with *Sosed* – the Sarafanov's neighbour – who has more interactions with Busygin, Sil'va and the two women than with the rest of the characters.

Even though the love-story of Busyin and Nina plays an important part in the plot development and is one of the key reasons Busygin decides to confess to Sarafanov, it does not appear to be quantitatively more important than the communication between Busygin and Sarafanov. Nina and Kudimov – her fiancé – have the lowest degree in the play, and the characters with the highest weight of edges between their nodes are Busygin, Sarafanov, Sil'va and Vasenk'a.

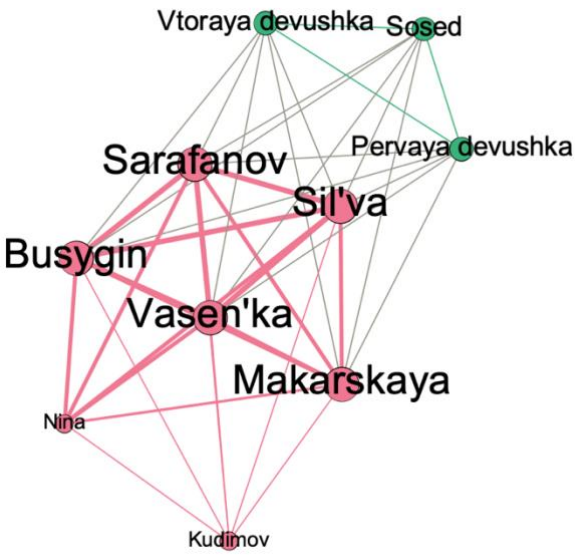


Figure 29. Network graph of *The Elder Son* by Vampilov

Vampilov's plays are examples of drama which responsed to Pomerantsev's call to leave behind Socialist Realist conventions in drama. In his plays Vampilov explores the personalities of his characters, focusing on their motivations and feelings. Kolesov in *Farewell in June* grows from an unscrupulous young man to a man of principle, seeking independence from the university's authority. In *The Elder Son* Busygin turns from being an impostor to a legitimate member of the Sarafanov family.

Besides the content, we can see the resemblance of Vampilov's plays to 19th-century drama from their network graphs. These plays have not offered any new challenges for the previous encoding scheme. They have more than two characters, with act and scene

segmentation. The segmentation might allow us to see stronger connections based on the number of segments where the characters appear simultaneously, which provides material for further interpretation of the plays.

However, the next play by Vampilov play we are going to discuss – *Duck Hunting* (*Utinaya okhota*) (1967) – is different. It is one of Vampilov's most famous works. The play has three acts and is about a 30-year-old man, Zilov. It is a mixture of Zilov's reminiscences and imagination, together with the actual events of the play. Through these we learn about Zilov's life and his character. Zilov is an adventurous and confident man who fails in his relationship with his wife Galina over a young student Irina. He does not treat his work seriously any more, and fights with and embarrasses his friends – Sayapin, Kuzakov and Dima. Zilov is deeply disappointed in his life, which he considers senseless. His reflections lead him to the only option he sees for himself – to take his own life. He notifies his friends about his funeral, and they are the ones who save him by interrupting his suicide attempt. Eventually, Zilov decides to take part in the annual duck hunting he organises together with his friends. The ending leaves us with hope for the future of Zilov, who represents a Soviet man of the time, and his feelings of frustration and uncertainty.

Duck Hunting has an experimental structure which is not apparent on the graph provided below. Even though it appears to have a familiar network structure with the density of 0.89, it can be encoded in a different way.

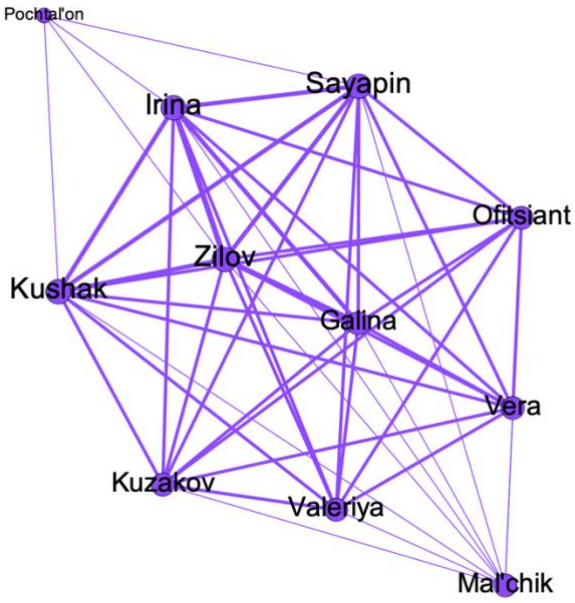


Figure 30. Network graph of *Duck Hunting* by Vampilov

Considering the unusual narration in the play, the dense network graph above essentially demonstrates the connections between the characters from Zilov's perspective, as if he was the narrator. From a broader perspective, the connections on the graph should look different, as some of the characters do not always meet each other on stage. Some of the communication between the characters in the first scene of the first act only occurs in Zilov's imagination. On top of that, Irina and Galina never actually meet each other, they only talk and hug in Zilov's fantasy about his funeral. This play can be classified as meta-drama, as the action happens both on stage in the real world, and within Zilov's mind. Here Vampilov acted as one of the pioneers of such complex and confusing structures, which turned out to be a very inspiring experiment in Russian drama.

The fourth play by Vampilov play featured in this chapter is *Last Summer in Chulimsk* (*Proshlym letom v Chulimske*) (1971). Its network graph is similar to the graph of *Duck Hunting*; however, we will see that this fact does not reflect the similarity of their network structures. *Last Summer in Chulimsk* is a two-act drama about the life of a small community in a village in the Siberian Taiga. A young woman, Valentina, works in the town's tavern. She is hopelessly in

love with Shamanov, the older police officer. At the same time, the tavern owner's son Pavel tries to get Valentina's attention, which makes Shamanov realise that he might like her too. The play finishes with Valentina and Pavel coming home from a night at the town discotheque. Their dialogue suggests that they have engaged in intercourse which was non-consensual from Valentina's side. Valentina tells her father that she was with another man – the accountant Mechetkin, who gets the approval of Valentina's father to marry her – and demands that Pavel and Shamanov never bother her again. Here, again, Vampilov takes us through the transformation of Valentina's personality by showing her grow from a naïve young woman to a disillusioned and undeceived person.

Similarly to *Farewell in June*, the play contains unusual – from what we have seen before – stage directions. In Figures 31 and 32, both of the types of stage directions discussed earlier are situated within the speech-act, and so there is only one way of encoding such speech-acts, even though the second stage direction relates to other characters.

```
<sp who="#valentina">
    <speaker>ВАЛЕНТИНА.</speaker>
    <p> Уже не боюсь... <stage>(Снова возится с калиткой.)</stage> Помогите,
        пожалуйста. <stage>Еремеев спускается с крыльца.</stage> Подержите мне
        ее... Вот так... Вдвоем они наладили калитку. Ну вот, большое вам
        спасибо... Я вас разбудила?</p>
</sp>
```

Figure 31. Extract from *Last Summer in Chulimsk*

```
<sp who="#dergachev">
    <speaker>ДЕРГАЧЕВ.</speaker>
    <p> Да, прозвали так. Седьмой секретарь, иначе его тут не величают. <stage>
            Хороших появляется в буфете одна. Валентина выходит на веранду,
            вытирает со стола, за которым завтракал Мечеткин.</stage> Илья, у
        тебя деньги есть?</p>
</sp>
```

Figure 32. Extract from *Last Summer in Chulimsk*

Cases like this are quite frequent in the subcorpus of newer drama that I have collected. This contrasts with the 19th-century plays where, in the text, the actions of a character do not

usually overlap with speech-acts by another character. The early post-Stalinist literary world saw many works still being written to the party's order, in the spirit of Socialist Realism. Many plays of that time were lifeless and uninspiring, so several theatres chose to start adapting prose for their productions. Some authors and critics at the same time strained after the literary psychological realism that had been endorsed by Stanislavsky, in addition to his acting methods, but which was left behind during the strictest period of censorship. The shift we have observed in the approach to stage directions among the authors of the 20th – 21st centuries contributes to the more dynamic character of their plays. Moreover, this shift might derive from a more cinematic sense of performance, due to the frequent screen adaption of Soviet plays.

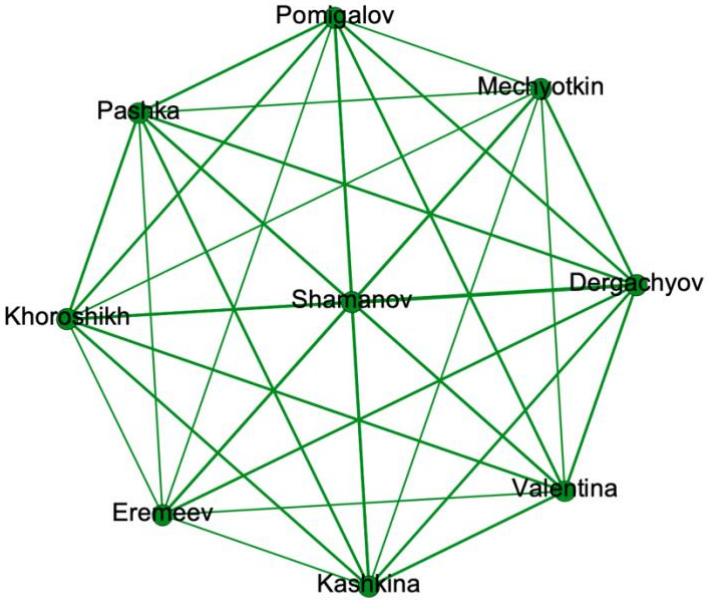


Figure 33. Network graph of *Last Summer in Chulimsk* by Vampilov

As mentioned, the graph above looks a lot like the graph of *Duck Hunting*. The density of this graph equals 1 and all the characters have the same degree. Unlike in *Duck Hunting*, this network graph reflects the unity of place in *Last Summer in Chulimsk* and the equal importance of each storyline of the play. One could argue that the graph does not represent well enough the central story of Valentina, and I believe it can be readjusted by introducing the segmentation by the change of scenes as discussed on page 79, instead of the segmentation by acts. However, that

would not drastically change the network structure of the play, which is familiar to us from the 19th century.

Thus, judging from the four plays discussed above, Vampilov often followed quite a conservative composition of characters in his plays. But *Duck Hunting* shows us that he was one of the playwrights who contributed to the transitional period in Russian drama and inspired other authors to continue the experiments.

In the second half of the 1980s, Mikhail Gorbachev, who became the General Secretary of the Communist Party of the USSR in 1985, advocated for the political and cultural movement of *glasnost'*. He was committed to making the government's structure transparent and to lifting the censorship, so that the Soviet people could learn about Stalin's repressions. Together with that, more information about the Western world and its culture was filtering into the Soviet Union. During that time the theatre started losing its special function. The country was on its way towards *perestroika*, which brought a liberalisation of the media and the attractions of popular culture to the Soviet Union. This provoked the decline of traditional theatre. Playwrights confronted yet another period of artistic stagnation, and were forced to focus on a much-needed transformation of theatre.

One of Vampilov's contemporaries, Lyudmila Petrushevskaya, lived through all these changes in Soviet Russia, and continues to write until this day. Smelyansky believes that Vampilov and Petrushevskya were 'the most gifted post-Stalinist playwrights' (Smeliansky 1999, p. 21). They were able to shrug off the burden of decades of censorship, and to develop original theatrical ideas despite the stagnation.

Petrushevskaya experiments with form in her plays, poetry and prose. She started out with short stories and fairytales in her thirties, but she has become best-known for her plays. Her work, which some might call *chernukha*, is harsh, raw and unapologetic. Her plays often do not

abide by structural conventions. Together with their flamboyant content this makes Petrushevskaya one of the most exciting playwrights of the second half of the 20th century. For my current subcorpus, I have chosen four plays by Petrushevskaya: two longer texts, *Three Girls in Blue* (*Tri devushki v golubom*) and *The Moscow Choir* (*Moskovskii khor*); and two short plays with few characters: *Cinzano* (*Chinzano*) and *The Isolation Ward* (*Izolirovannyi boks*).

The structure of Petrushevskaya's larger plays appears to be quite conventional, and we are able to map the network graphs on to the storyline of the plays. For instance, the graph in Figure 34 corresponds to the plot of *Three Girls in Blue* (1980). It is a two-act comedy – a family drama about three female middle-aged cousins, Ira, Svetlana and Tatyana. They all live together with their three sons at a country house. Ira starts a love affair with their married neighbour Nikolay and travels to Koktebel', where he is on vacation with his family. After Nikolay abandons her, she has to return home to her son Pavlik. In the meantime, the country house where the cousins live has been flooded and they and their children continue living there all together, in the only remaining dry room of the house.

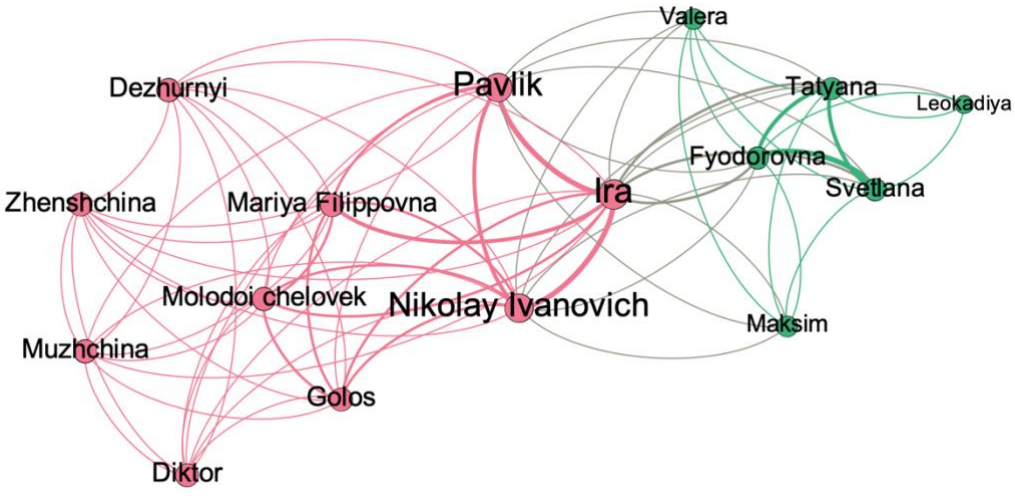


Figure 34. Network graph of *Three Girls in Blue* by Petrushevskaya

The density of the play's network is 0.62, which emphasises a certain level of disconnectedness within the family and reflects the non-unity of time and place in the play. The

network graph above highlights the central love story of Ira and Nikolay and distances the rest of the family from their affairs. Thus, the nodes of Ira, Pavlik and Nikolay have the highest degree centrality. They are also part of a cluster which is separate from the rest of the family, because the two other cousins, Tatyana and Svetlana and their families, as well as the owner of the country house Fyodorovna, do not meet the characters present during the Koktebel' scenes. Besides that, we can see that the weight of edges between Ira, Pavlik and Nikolay and between Tatyana, Svetlana and Fyodorovna is higher than between the rest of the pairs of characters, which aligns with the play's plot. Thus, the network graph of *Three Girls in Blue* serves its purpose by providing us with a reliable representation of the play and allowing us to draw certain conclusions about the interactions of the characters, even without reading the play.

The graph representation works well with another large play by Petrushevskaya – *The Moscow Choir*, a two-act play written in 1984. Here Petrushevskaya again recreates the life of a dysfunctional family and explores their complicated sense of community. Set in the period after Stalin's death, the play tells us the story of Lika and Neta (the older sisters), their children and grandchildren, while ultimately criticising Stalin's policies. Neta and her daughter Lyuba are returning to Moscow from 17 years of exile in Ufa. Lika, by contrast, was lucky to stay in Moscow in her apartment with her immediate family (her son Sasha, his wife Era and their daughter Olga). Lika also arranged an apartment in Moscow for Katya, Neta's younger daughter, who was not exiled. Even though Neta's and Lyuba's hardships were not the fault of their family, who had been trying to help them for the whole time, they are unable to reunite and forgive each other. Katya is rejected by her mother and sister. Sasha is leaving Era for another woman pregnant with his child. Their daughter Olga gives birth to an illegitimate child. Neta and Lyuba end up writing an official complaint (*донос*) against the family of Lika, who herself decides to leave Moscow for good, together with Sasha. The whole action of the play is accompanied by the Moscow choir (and later in the play the Dresden choir joins in), which, as

Kablukova claims, represents the unity of people and their time, that in reality fails to be achieved (Kablukova 2011).

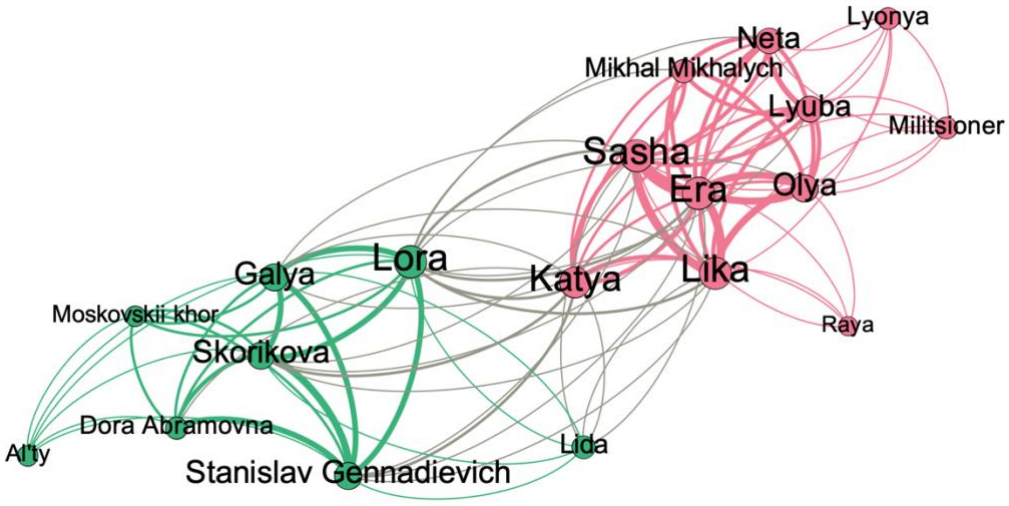


Figure 35. Network graph of *The Moscow Choir* by Petrushevskaya

As we can see from the graph above, there is also no network unity of characters in the play.

The network graph of *The Moscow Choir* clearly shows that the two communities appear on stage in different scenes: Lika and her family, and the choir together with Lora – Katya's daughter, who sings in the Moscow choir. It also reflects appropriately the activity and importance of the characters (by their node size), as well as the strength of their connections (by the thickness of the edges). The Moscow part of the family (Lika, Sasha and Era, and Katya and Lora) has the highest degrees in the graph, which makes sense since they are more established in the society of the play and have more connections. Lora also has the highest betweenness centrality because the majority of the 'choir nodes' are connected with the rest of the characters via Lora's node. The nodes with the lower degree on the periphery of the graph represent the characters who are less involved in the play's action, such as Raya, Sasha's mistress, and

Militsioner (the Policeman). Thus, the network representation of *The Moscow Choir*, in a similar way to *Three Girls in Blue*, is accurate, and provides an essential overview of the play's structure which matches its plot.

Even though Petrushevskaya explores various forms, she frequently turns to short and small plays, where the SNA approach turns out to be not as effective. Among such plays are *Cinzano* and *The Isolation Ward*.[24] These are dialogue-based plays with two or three characters (rarely more), where the main focus is on the content of the characters' conversations, through which we may learn the plot development, rather than more diverse interactions. Moreover, the dialogues themselves in Petrushevskaya's plays are often "limited and confused" (Dalton-Brown 2000, p.139), which contributes to the 'noncommunicativity' of her characters.

Cinzano (1973) is a one-act play, which consists of a dialogue between three friends (middle-aged men) – Kostya, Valya and Pasha. We learn about Pasha's divorce from his wife, Kostya's love affairs and, in the end, the fact that Valya's mother has died. The play could be summed up as the tragicomic testimony of drinking companions.

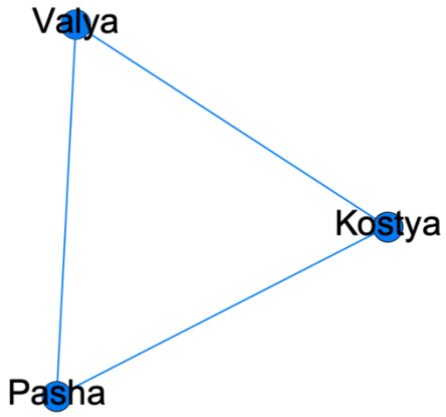


Figure 36. Network graph of *Cinzano* by Petrushevskaya

[24] See, for example, *Квартира Коломбины* (1981) with three characters, *Стакан воды* (1978) with two characters, *Скамейка-премия* (1983) with two characters, *Я болею за Швецию* (1977) with three characters

Even though all three characters in *Cinzano* are connected to each other and the play retells their conversation, Pasha's confession about his mother's death gets lost in the nonsensical mundane chatter of his drinking companions. The constant references to memories, and confusions of time and space result in the fact that the characters seem to be talking to themselves rather than with each other.

There is no such issue in *The Isolation Ward* (1988), however, because it is simply a short dialogue between two dying women in a cancer hospital, A and B. The network graph below obviously illustrates the structure of the play correctly. But since the play consists only of one dialogue, the graph cannot offer anything in terms of the interpretation.

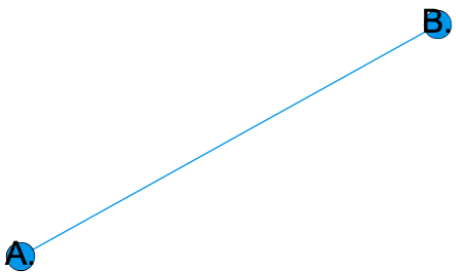


Figure 37. Network graph of *The Isolation Ward* by Petrushevskaya

Thus, not only is the graph representation of such plays simply uninformative from a structural perspective, but it is also not entirely applicable in cases like *Cinzano*. The basic assumption behind an edge between the nodes, as we have defined it in the beginning of this dissertation, does not stand its ground. We will see that this is one of the major issues that the current methodology needs to address when talking about contemporary plays.

Perestroika

During and after the perestroika period (1985-1991), the mood of frustration in the Russian theatre world continued. Many directors turned back to productions of the classics: plays

by Chekhov, Tolstoy and especially Ostrovsky, whose drama corresponded well to the critical times, filled the stage. Moreover, after the collapse of the Soviet Union, the Russian culture was swamped with Western artistic and media products of all kinds of quality. The audience became surfeited with variety and freedom, as ironic as that may sound, and became harder to impress. Still, there were playwrights who continued to experiment and tried to adapt to the rapidly changing world.

Another female author to note in this context is Lyudmila Razumovskaya. Like Petrushevskaya, she explored different forms, and has written both conventionally structured larger plays, as well as short dialogue-based plays. She also concentrates on the topics of dealing with change, grief and community. The four plays by Razumovskaya that I have chosen to survey in this chapter are *Vladimir Square* (*Vladimirskaya ploshchad'*) (1997), *The Life of Yura Kurochkin* (*Zhitiye Yury Kurochkina i ego blizkikh*) (2000), *Without a Dowry* (*Bespridannik*) (2000) and *Halloween for Russians* (*Khellouin dlya rossiyantsev*) (2004).

Vladimir Square is a two-act play with two characters – Vera Ivanovna and Pavel Sergeyevich. They meet one night in Vladimir Square in St. Petersburg, where Vera trades cheap vodka. Vera recognises her school friend in the homeless man Pavel. They are both over sixty, with a tough background, which was not uncommon for many in Russia during the Perestroika period. Over several months, Vera and Pavel get carried away by the memories and hopes of their youth, and eventually get married in a church. However, they have nowhere to live together and their ways part again, because of their inability and unwillingness to compromise.

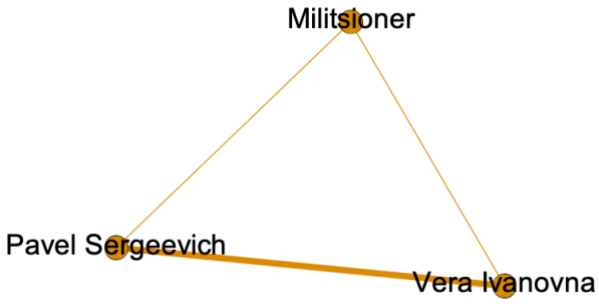


Figure 38. Network graph of *Vladimir Square* by Razumovskaya

Vladimir Square presents the same challenges for the methodology as some of Petrushevskaya's plays discussed earlier. There is nothing else to derive from its graph apart from using it in a large-scale analysis for demonstrating structural changes that occur in Russian-language drama over this time. The SNA methodology is more applicable to the larger plays by Razumovskaya, although this is not always the case, and I will discuss this further.

The Life of Yura Kurochkin is a two-act comedy (*комедия из народной жизни*). Its first act is set in the eighties and describes the life of several families in a communal flat. Razumovskaya immerses us in the twists and turns of the *коммуналка* life: the love affairs, the housing problems, the scandals with an alcoholic neighbour. In the second act, set in the nineties, Yura, who once lived in the flat, becomes a real-estate agent and receives an instruction from his boss (*Хозяин*) to buy the flat's residents out of this prestigiously located apartment in the centre of Saint Petersburg. He is amazed to see how his neighbours, who represented his family at one time, have changed. Ninka, who he had had an affair with, is terminally ill with cancer. Her daughter Tatyana is married to a German man but is unhappy and still loves Yura. Nikolay is paralysed and his wife Zinka is pregnant with yet another child. Mariya, who they had a connection with, ended up marrying Sergey and giving birth to a son. Sergey and Auntie Shura have died. All of the residents agree to receive a private flat in a new-build in exchange for their *коммуналка* shares. On the day of the deal, Yura's boss arrives to see the flat, and Roza, who

survived the Siege of Leningrad, is drawn into a conflict with the boss and his bodyguards. In the end, Roza dies from a heart attack and Yura calls the police to report a murder.

Although the network of *The Life of Yura Kurochkin* is dense (its density is 0.93), its graph works well in representing the structure of the play, and separates the characters who only appear in the first act (in the past) from the characters who meet Khozyain (the Boss). There is a cluster of six characters (Yura, Nina, Roza, Mariya, Zinka, Tatyana) who are equally the most influential in the network, judging from their centralities values, because they connect Khozyain with the rest of the neighbours. Khozyain, in accordance with the plot, is the least influential character, which is an interesting detail, since it is his actions that influence the *коммуналка* community greatly and are, partly, the reason of Roza's death. Taking into account the fact that Khozyain is making his deals behind the scenes via Yura, without involving the rest of the characters directly, we could at least expect Yura to be the most influential in the network. But all the six characters who communicate with Khozyain share the same centrality values. Nevertheless, Mariya and Nina have the highest weighted degree, which means that they are the most communicative in the network and speak to other characters more frequently. They have the highest weight of the edges between themselves and with the majority of the characters, including Yura. But Yura's influence over the community's life is not reflected in the graph, since it is recognisable from the content of the play, but is not reflected in its text from a quantitative perspective.

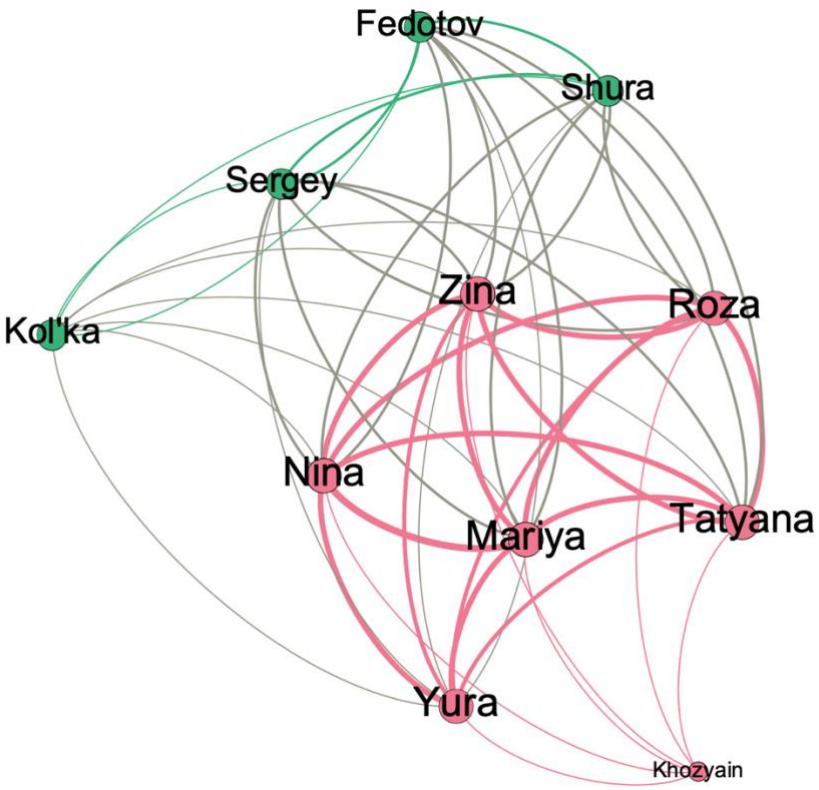


Figure 39. Network graph of *The Life of Yura Kurochkin* by Razumovskaya

As with the larger Petrushevskaya plays, *The Life of Yura Kurochkin* by Razumovskaya structurally resembles the 19th-century plays, and in a similar way to them, benefits from the network representation which illustrates the division of the characters into clusters based on the level of connections between them, as well as bringing to light the mismatch of the actual (i.e. based on the plot) influence of certain characters as opposed to their influence in the quantitative network of the play.

Without a Dowry, at first glance, is another quite conventional play with a dense network of characters (with the density of 1), that reflects the unity of time and place. However, this is not entirely true. The time, place, and the characters themselves, as well as their relationships, change drastically throughout the play, but the graph below does not reflect that.

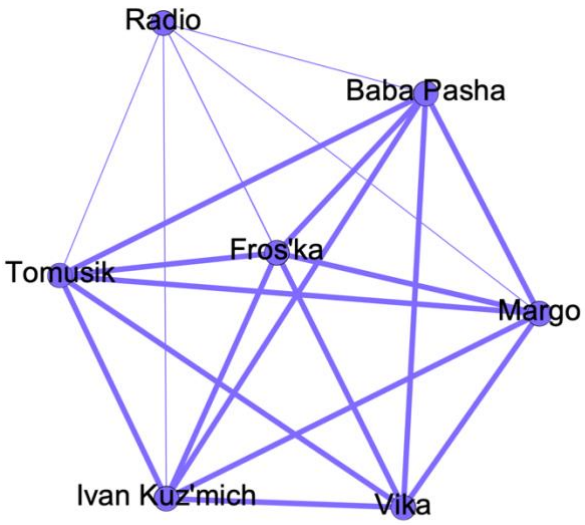


Figure 40. Network graph of *Without a Dowry* by Razumovskaya

This is also a two-act comedy – a satire on post-Soviet Russia in the nineties. The play is set within one day at the marriage bureau, where four women, Baba Pasha – a seventy-five-year-old ex-milkmaid, Margo – a sixty-year-old 'famous' writer, Fros'ka – a fifty-year-old businesswoman and Tomusik – a forty-five-year-old librarian, meet the prize catch – Ivan Kuz'mich – a fifty-nine-year-old bachelor. The women immediately start competing over this bachelor who has no fortune, but the plot takes many surrealistic turns. The characters' lives are changing on the fly: two explosions at Fros'ka's kiosks, Tomusik's attempt to emigrate to the Emirates, and Russia being banned "from everything" by some European committee of rights.[25] In the end, Vika, the twenty-five-year-old agent at the bureau, decides to leave for Europe while it is still possible, and the rest stay as the defenders of the fatherland, with Fros'ka leading the group.

Another unconventional feature of *Without a Dowry* is the radio announcements that are formatted as speech-acts and are accompanied by stage directions (*включается*, *выключается*). It is an open question whether these prompts should be included in the network of characters.

[25] "Европейская комиссия по правам исключила Россию из всего за многочисленные нарушения всего и объявила нашу любимую родину матушку Русь, в скобках бывшую Святой, страной-изгоем со всеми вытекающими последствиями."

We will come across a similar issue in *One Hour and 18 Minutes* by Elena Gremina, where the announcements are made not by an electrical device, but by a theatre worker.

The reasons I believe that the network graph of *Without a Dowry* does not align with the play's structure are the following. First of all, at some point during the day Ivan receives a letter from his ex-wife. Because of a faulty translation, all the characters believe it is from his American wife who has died and left him an inheritance of two million dollars. Ivan's promised fortune makes Vika herself consider him as a partner, and she tries to lure him. But it finally turns out that the letter is from another of his ex-wives, who now lives in the Emirates, and she is telling him he has a daughter, and that the daughter is Vika.

In addition, Tomusik's attempt to leave for the Emirates, following the example of Vika's mother, violates the unity of place. Even though at the end she returns to the bureau, we learn that she has been on a plane on the way to the Emirates. The plane Tomusik was on diverted its course during the flight and returned to St Petersburg.

Finally, during the play the whole world around the characters is collapsing: a coup is on its way, and the characters are forced to adapt to the unfolding changes in haste. The whole disposition of the play changes during this one day, which hardly identifies it with the same category of canonical plays with a similar network structure. Thus, even though, all the characters meet and speak to each other, I believe that the graph representation is not sufficient to capture important features of the play.

The last play by Razumovskaya I am going to discuss, however, has a similar dense network structure which accurately reflects the nature of the play. *Halloween for Russians* is a two-act play – and describes *первоапрельские шутки* (April Fool tricks) as the subtitle says. While Razumovskaya's *Without a Dowry* shares its title (although with a gender switch) with the famous play by Ostrovsky already discussed in this work, this play is built on the fabula of

Gogol's *The Government Inspector*. It follows a Gogolian scenario set in the criminal world of post-Soviet Russia. The mayor of a small town, Filip Filipovich, and his wife Aleftina fall victim to his colleague's joke about an inspection coming from Moscow to assess Filip Filipovich's work. The mayor throws a banquet to which he invites the inspectors, Naden'ka – the owner of the town hotel and the mayor's mistress – and the 'honoured citizens' (schoolteacher Elena Kazimirovna, the bandit Zatravkina, the state officials Pal Palych and Sidor Sidorych, and the long-lived Uncle Petya). Major Kovalyov (another Gogolian character), who is having an affair with the mayor's wife, and the journalist Zemfira also join the banquet. At the banquet Filip Filipovich believes Vasya and Innokentiy and makes corrupt deals with them, but eventually he learns about his friend's prank. Vasya and Innokentiy, however, are able to disguise themselves again and pretend to be the other, real Moscow inspectors. The twists and turns of this play remind us of Razumovskaya's *Without a Dowry*, as the play ends with Zatravkina bombarding the hotel and killing the mayor. She turns out to be playing a double game as a bandit and a police officer.

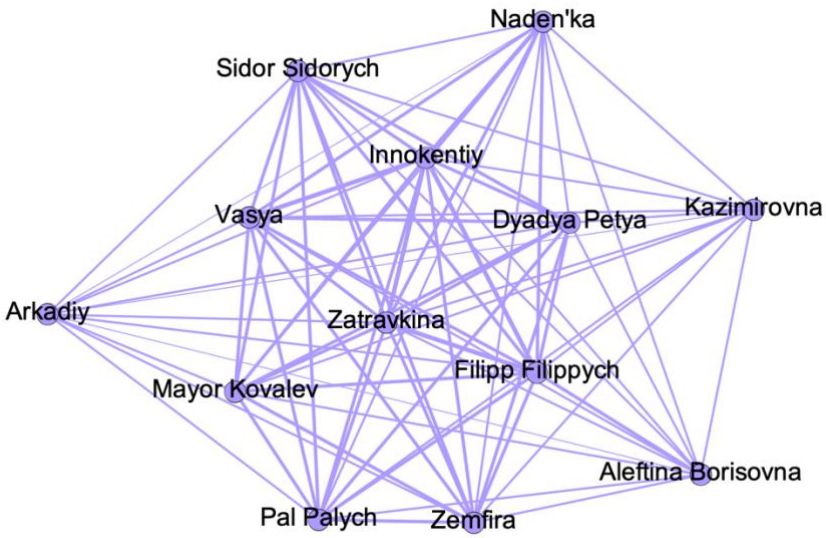


Figure 41. Network graph of *Halloween for Russians* by Razumovskaya

The graph of *Halloween for Russians* is also quite informative and shows us that, unlike in *The Government Inspector*, all the characters in the play communicate with each other as the density of the graph is 1. Moreover, the play loops back to the beginning, at the point when Vasya and Innokentiy are believed to be the ("real") Moscow inspectors for the second time. This also contributes to the overall connectedness of the characters.

Thus, while we have seen that the discussed methodology may face some issues with the plays by the 'transitional' authors (in the cases of dialogue-based plays and the plays with less straightforward plots), it mostly works well for the larger pieces and provides us with the same kind of insights that I discussed in the previous chapters, analysing the earlier plays. However, the authors of Russian drama have continued to experiment, and the plays written in the 21st century differ drastically in their structure and composition, which possesses even more challenges for the SNA approach.

The 21st Century

The disruption of the unity of time and place is a common feature in contemporary Russian drama. Vampilov's experiment in *Duck Hunting* continues to resonate in more recent plays. And while this disruption is not obvious from the network graph of Razumovskaya's *Without a Dowry*, the graph of *Terrorism* (*Terrorizm*) by the Presnyakov brothers reflects it well.

Oleg and Vladimir Presnyakov are playwrights and theatre producers. They have written their plays as a duo and have become quite successful on the contemporary Russian stage, as well as outside of Russia. We are going to discuss two of their best-known plays – *Terrorism* (*Terrorizm*) (2002) and *Playing the Victim* (*Izobrazhaya zhertvu*) (2002).

Terrorism is short six-act play that premiered at the Moscow Art Theatre, directed by Kirill Serebrennikov, and was also staged by the Royal Court Theatre in London, as well as across Europe. It starts with the conversation of three airport passengers who are waiting outside

the airport because it is believed to be booby-trapped. The second act is a scene of a love affair between a married woman and her lover. The third act tells us the story of an office where a female employee hanged herself. The fourth act is a gossipy conversation between two older women at the playground – the grandmother with her grandson and her friend, who suggests the grandmother should kill her son-in-law. The fifth act is a scene from some military department, where men torture their colleague for fun and discuss the events of the day. The sixth act is another scene with the airport passengers, now in the sky, flying away in an unknown direction. Through the conversations of these characters in the seemingly independent scenes, we learn that one of the passengers is the husband of the woman from the second act. He comes back home from the airport to find his wife tied to the bed, with her lover asleep. The husband turns the gas on and leaves. Later the little boy from the playground will be running away from his grandmother to the entrance hall of an apartment building and pressing the doorbells of the flats just for fun. In one of the flats the doorbell ring will trigger an explosion. The man and woman will die, the destiny of the boy is unknown. The woman who hanged herself at her workspace, was probably the wife of the man who cheated on her with the passenger's wife. Thus, the alleged bomb threat at the airport is not the main terrorist act of the play. We are told multiple stories that are connected to each other by a casual, inter-personal, day-to-day terrorism.[26]

If we read the play as a sequence of individual stories, each story and each set of characters would form a separate network, and thus we would see a disconnected graph as below. There are five subgraphs in this graph, because the first and the last scene feature the same characters – the passengers at the airport.

[26] https://mxat.ru/authors/playwright/presniakovi/24624/.

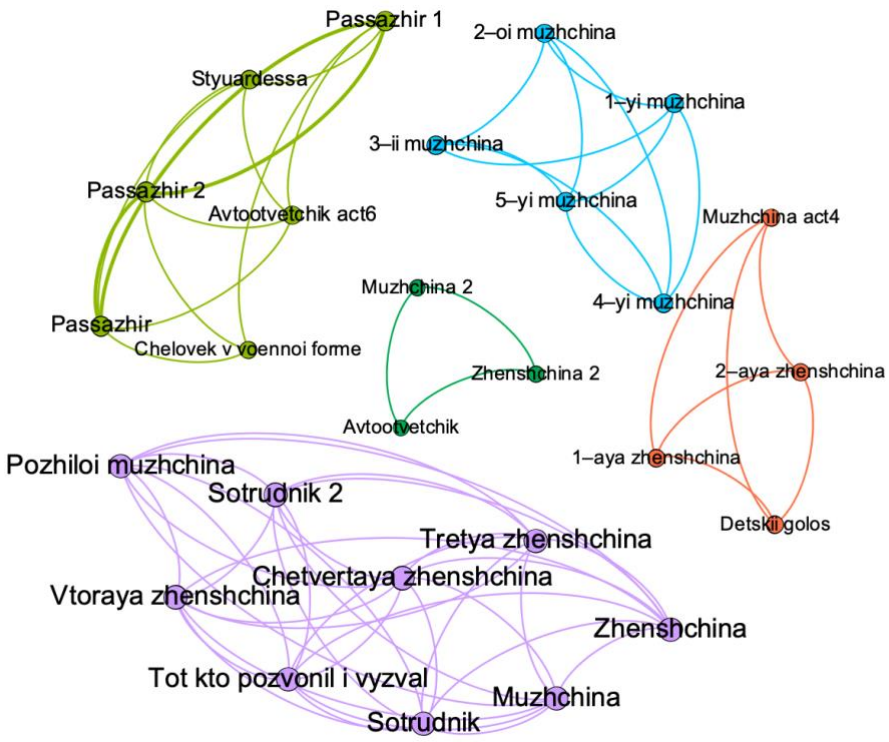


Figure 42. Network graph of *Terrorism* by the Presnyakov brothers, version 1

However, it is safe to assume that the man (*muzhchina*) from the fourth act is the passenger (*passazhir*) from the first and the last acts. The same *avtootvetchik* (voicemail machine) connects him, his wife and her lover (it is a separate question whether these devices – such as a voicemail machine or radio – should in fact be considered characters in the TEI). On that basis, the resulting graph would look quite different:

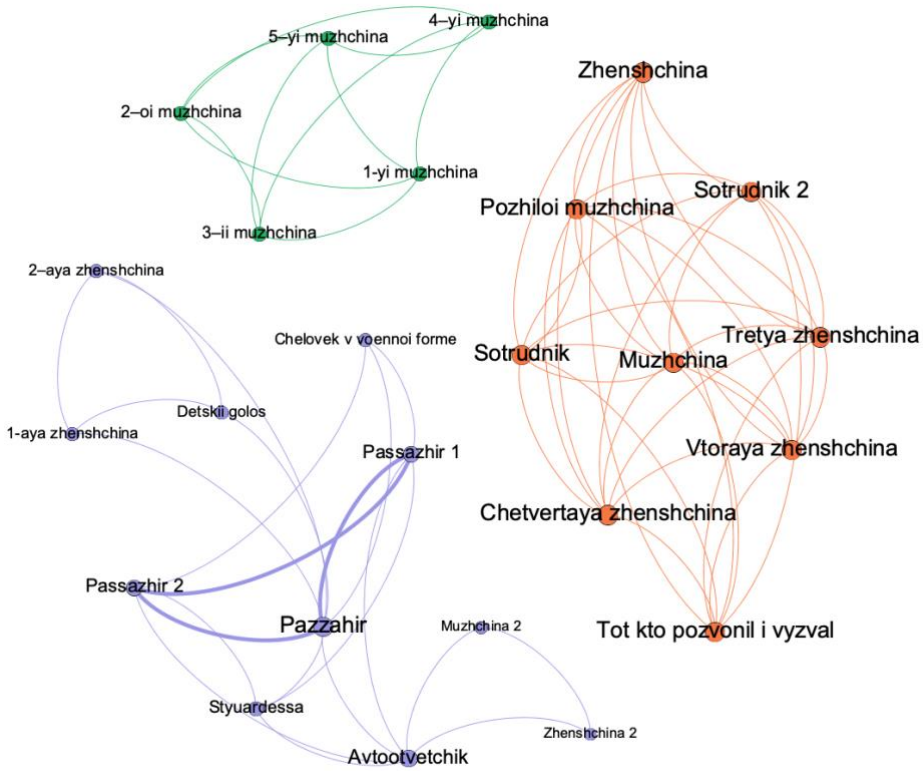


Figure 43. Network graph of *Terrorism* by the Presnyakov brothers, version 2

Moreover, another reader might notice or suggest other connections and overlaps between the characters that would provide yet another version of the network graph. Static graphs, unfortunately, do not allow for this freedom of interpretation. Perhaps the solution to this would be to use a dynamic graph that would demonstrate how a reader's understanding of the characters' network changes throughout the play. A dynamic graph is a set of static graphs (states) where each state can correspond to a segment (scene or act) of a play. In this way, the graph would represent the process of reading the play and each state of the graph would correspond to the network of characters that we can reconstruct at a certain point of the play (act by act).[27]

Even though Fischer *et al.* argue that the usage of dynamic graphs is "unsatisfactory when it comes to actual corpus-based analysis" (Fischer *et al.* 2017), I believe that in cases such

[27] See, for example: https://maladesimaginaires.github.io/intnetviz/.

as with *Terrorism*, the usage of a static graph is insufficient, especially for a corpus-based analysis. While we have come across similar issues of the possibility of several nodes of the network graph corresponding to a single character (mentioned on page 141), the case of *Terrorism* is different. When minor characters are not properly identified in a graph, this could still change the network structure. But the passazhir-muzhchina (male passenger) in *Terrorism* is one of the protagonists of the play from the beginning, and becomes the most influential node, as per the network graph in Figure 43. Thus, the correct representation of this device used by the authors becomes essential for both close and distant analyses of Russian drama.

Playing the Victim also possesses a challenge for automatic encoding, though it is a more common issue that we have already discussed in the second chapter. It is a one-act play where the segmentation is introduced by using stage directions, setting the place and time, instead of by division into scenes. Thus, the TEI XML version of it required additional manual editing. The play tells us the story of a young man Valya, who works with the police as an actor at crime re-enactments. He always plays the victims, and explains his choice of profession as a means of overcoming his fear of death. The play shifts between the scenes from Valya's work, his conversations with his mother and girlfriend Olga, and his Hamletian dialogues with his dead father. In 2006, *Playing the Victim* was turned into a successful film by Kirill Serebrennikov.

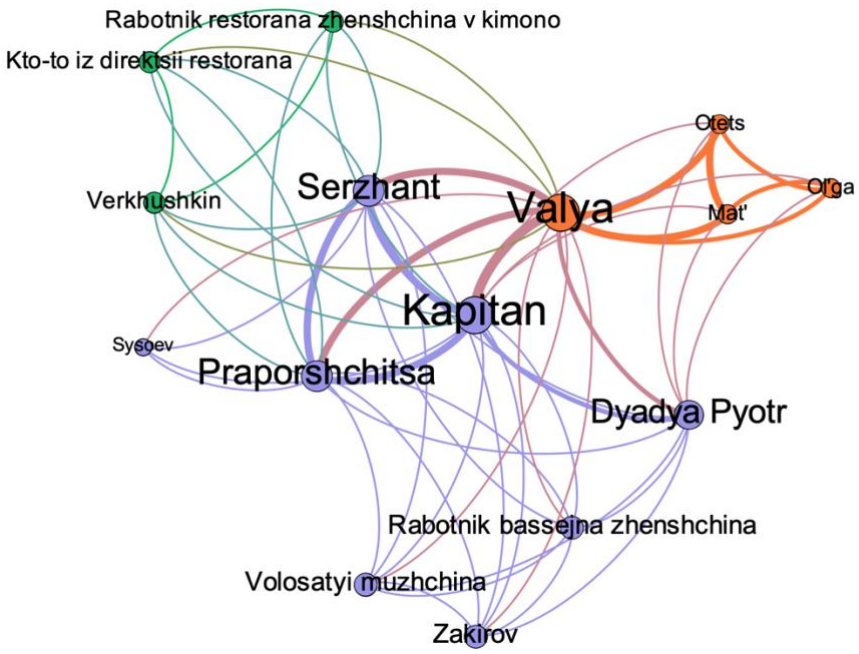


Figure 44. Network graph of *Playing the Victim* by the Presnyakov brothers, version 1

While the graph above highlights correctly the network structure of the play and the dominant role of Valya in it, it is not entirely accurate because of the Otets node and its connections with the other characters. As mentioned in the summary of the play above, Valya talks to the spirit of his father. Previously we operated within the assumption that the characters who appear on stage together, even if they do not interact directly with each other, should be connected in the graph. Such an assumption is based on the idea that even if these characters do not share any dialogues between themselves, they might hear each other talking to the other characters and, in this way, they share the same communication space with each other. But such an approach is not suitable in this case, because Valya is the only one who hears and sees his

father. A more accurate version of the network graph of the play would be as below.

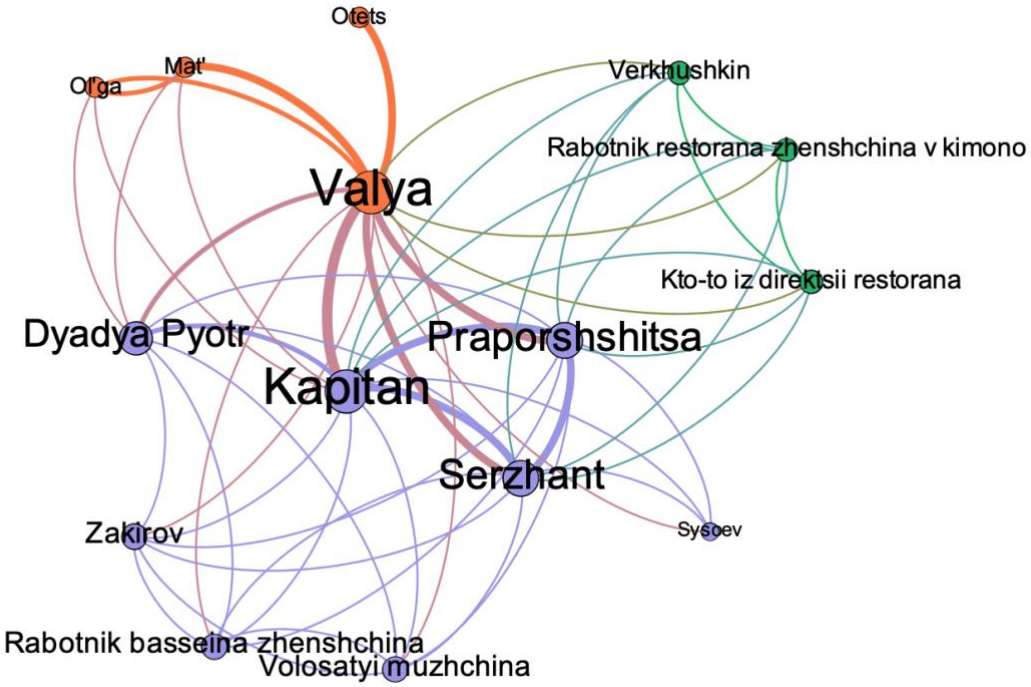


Figure 45. Network graph of *Playing the Victim* by the Presnyakov brothers, version 2

We can clearly see how the Presnyakov brothers' plays can benefit from the graph representation. For instance, in *Terrorism*, the character with the highest centralities values is *passazhir*, because he is, in my interpretation of the play's text, the only character, who appears in three scenes out of six. And, as previously mentioned, he connects the two separated subgraphs of the network. In *Playing the Victim*, the network metrics highlight for us the core of the plot, emphasising the characters of Valya and his colleagues. Despite that, the methodology clearly needs to account for the difficulties and unusual techniques of the playwrights that were mentioned above and continue to emerge in other 21st-century plays.

For instance, the difficulty of justifying the approach to formalising the connection between the characters, that we have seen in *Playing the Victim*, is also present in the plays written by a contemporary of the Presnyakov brothers – Ivan Vyrypaev. Vyrypaev is one the most exciting writers in the New Drama movement, as well as a theatre and cinema director. His

plays have been translated and staged worldwide, and he is known for experimenting with genre and style by neglecting many of the traditional strategies for creating a drama text. In his drama, Vyrypaev focuses on the verbal component of the plays (Vyrypaev 2019, pp.6-7). Thus, for this chapter I have chosen to analyse two of his plays that are deeply rooted in their texts (Kurant 2020) – *Oxygen* (*Kislorod*) (2002) and *July* (*Iyul'*) (2006).

Oxygen is Vyrypaev's most famous play. It premiered on the stage of Teatr.Doc in Moscow, and later in 2009 Vyrypaev also directed a film based on it. *Oxygen* is an unusual play, a proclamation of a love story structured as a variation on the Ten Commandments. There are two narrators – He and She, who tell us the story of two Sashas, a young man from Serpukhov and a young woman from Moscow. The male Sasha falls in love with the female Sasha and beats his wife to death because he realises 'there is not enough oxygen' in her. The play describes the two Sashas' love story, together with a social and political critique of Russia and the whole world in general.

As for the play's structure, *Oxygen* has a unique segmentation. As mentioned previously, the play is structured by the Ten Commandments. In addition to that, Vyrypaev approaches the play's structure from a musical perspective: instead of acts or scenes, there are ten compositions in the play, each split into verse and chorus. I have decided to encode the commandments as acts, and the verse and chorus within them as scenes. Due to the unusual structure of the play, it had to be encoded into TEI format manually.

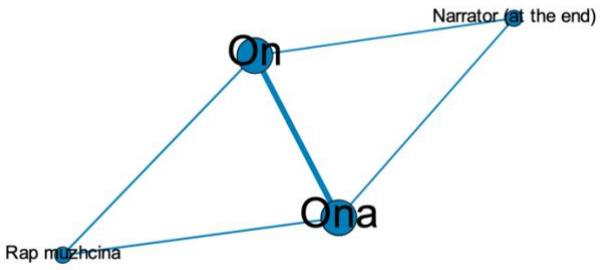


Figure 46. Network graph of *Oxygen* by Vyrypaev, version 1

It can be argued that *Oxygen* is a dialogue between Him (*Он*) and Her (*Она*). Their responses, of course, resonate, as they are parts of the same narrative. However, they do not appear in the same act (composition) until the fourth composition. Being the narrators, they are also not connected by any real-life interaction, as characters traditionally are. Moreover, twice in the play additional characters (narrators) appear. Firstly, there is a rap inset in the play that is attributed neither to Him nor to Her. Secondly, after the last response in the play, a response by Her, there is a final narrative section that loops us back to the beginning of the play, by offering an alternative story about two Sashas. This bit is also not attributed, either to Him or to Her. Thus, even though we see a connected graph of four characters, this should perhaps instead be four independent nodes or two connected and two independent nodes, because two of the characters – On and Ona – are engaged in a dialogue, as Julie Curtis argues (Dudrah *et al.* 2020). Moreover, one can identify the narrators with the events that they are describing and believe that the narrators actually represent the characters of the story (the two Sashas). As we can see, this is a very controversial play that compromises our understanding of the nature of a character in general. When creating a network graph of such a play, one is obliged to choose which dimension to focus on: is this a network of the characters in the story, or a network of the narrators, who nevertheless appear as characters played by actors on stage for the audience? Depending on this choice, the network graph of *Oxygen* might also look like one of the following:

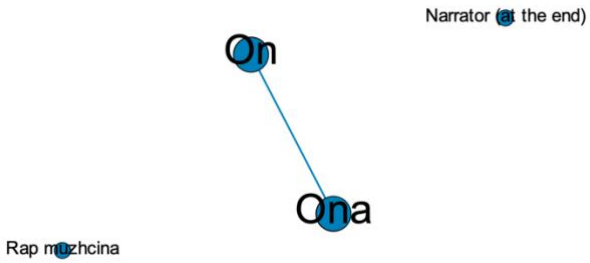


Figure 47. Network graph of *Oxygen* by Vyrypaev, version 2

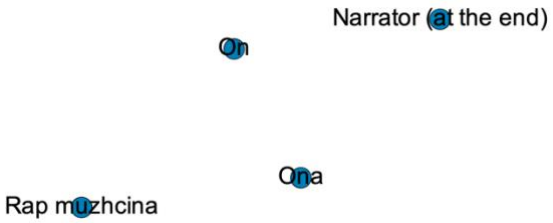


Figure 48. Network graph of *Oxygen* by Vyrypaev, version 3

Thus, Figure 48 creates a completely different network vision of the play. If we were to perform a large-scale analysis of Russian drama looking at the networks' connectivity and density, it would be important to establish what interpretation of the play should be reflected by its graph.

July (2006) – another play by Vyrypaev – also raises the issue of the disconnectedness of characters in monologue-based plays. It is the monologue of a sixty-something man, a maniac who murders and eats people, intertwined with short monologues and the responses of the nurse in the psychiatric ward and the man's sons. Vyrypaev presents us with reflections on the essence of God through the violent, schizophrenic perspective of the character who ends up eating God, or what he sees as God, in order to place him inside himself.

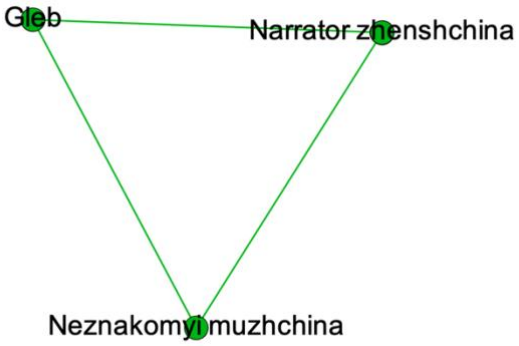


Figure 49.Network graph of *July* by Vyrypaev, version 1

Even though is it about the life of a man, Vyrypaev has requested that the play should be narrated by an actress. The gender difference between the narrator and the character himself (Petr, the murderer and cannibal in his sixties) does not allow us to draw a parallel between them as we could in *Oxygen*. And, as Mutc reminds us in her essay, Vyrypaev requests the actress to "perform a text, not a role" (Curtis 2020, p.109). Besides that, we again face the same difficulty as in *Duck Hunting*. First of all, the responses of Nelya, the woman whom Petr kills, are inserted into the main narrative. They are not, however, formatted as speech-acts within the text, so there is no corresponding node for Nelya.

Furthermore, the narrator plays the part of Gleb, Petr's son. In addition, it is necessary to note that the play contains author's remarks that are also read by the same female narrator. In this way, she announces that we are about to hear Gleb's voice, and then she plays Gleb herself.

When the narrative brings us to the part where the main character dies and his sons arrive at the hospital to collect his body, we are retold the conversation between Gleb and the hospital worker (*незнакомый мужчина*). Since Petr is dead, it is unclear whether this conversation is part of his narrative or is independent. This might mean that the narrator shifts from the main character's perspective to the observer's point of view (perhaps this is what happens in *Oxygen* too?). Therefore, the narrator's node consists of Petr, Nelya, and sometimes Gleb, as well as the

narrator herself. There is also a question, again, whether the network graph should be connected, because Gleb and the hospital worker do not meet the narrator, but they do meet Petr (alive or dead), whom she sometimes represents. Furthermore, since the responses of Gleb and the hospital worker are embedded within the main narration, there is a question of whether their characters should be represented as nodes in the network graph at all, or if they, similarly to Nelya, are 'meta'-characters in the play.

As we can see, this a very complex play in terms of the characters' network, and the SNA method here turns out to be more of a disadvantage when it comes to any alternative representation of the structure and plot of the play. Similarly to *Oxygen*, the network graph of the play can look differently, depending on one's understanding of it:

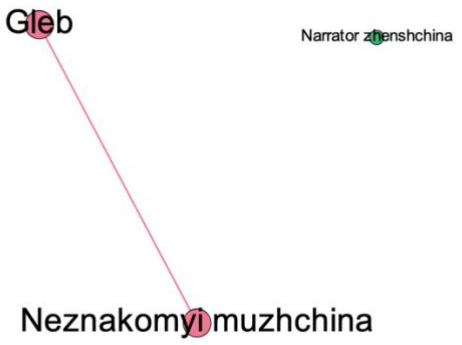


Figure 50. Network graph of *July* by Vyrypaev, version 2



Figure 51. Network graph of *July* by Vyrypaev, version 3

The plays by the Presnyakov brothers and Vyrypaev are experimental and innovative and may frustrate an audience which is not familiar with contemporary Russian theatre. The reason behind such an avant-garde approach to drama in the Russian theatre of the beginning of the 21st century is perhaps the lack of it during the Soviet period. For example, the Theatre of Absurd, that was well-established in the European drama of the second half of the 20th century, did not gain significant popularity in Russian theatre until the 1980s.[28] In addition to the lack of absurdist and avant-garde drama in Russian theatre, the unexpected inflow of Western art and media on the Russian market during and after Perestroika forced literary authors to come forward with extremely unexpected ideas and to explore new possibilities for theatre. This also explains the great diversity of genres and topics in New Russian Drama and brings us to the next set of plays written within the domain of documentary theatre.

Documentary Theatre

While Soviet playwrights turned to historical drama and allegorical allusions to talk about political and social issues, contemporary Russian writers became free (up to a certain point) to discuss pressing issues directly, and so this popularised the verbatim mode of theatre. As a result, New Drama authors have frequently turned to the monologue form in their plays. Above all this tendency brings 21st-century Russian drama closer to prose, and the subject matter comes to dominate the theatrical elements in the plays. It becomes especially popular to use dialogues and monologues in documentary drama, to imitate a person narrating their own story, which is what we will see in the next two plays by Elena Isaeva.

Elena Isaeva is a poet, screenwriter and playwright. Isaeva took part in some of the most important festivals of contemporary drama in Russia – Lyubimovka, Novaya Drama and

[28] With the exception of a few small-scale productions by the OBERIU collective and the Sovremennik Theatre.

Dokumentalnaya Drama. Many of her plays were staged in Teatr.Doc in Moscow. In her plays, Isaeva often turns to the genre of verbatim, hence the dominance of monologues in her works.

My First Man (*Pervyyi muzhchina*) is a play that consists of simultaneous narratives by three women. It was written in 2003 and is one of Isaeva's plays that were staged in Teatr.Doc. The women (pervaya, vtoraya and tret'ya) in the play talk about their childhood and youth, and their experiences with men, often socially unacceptable and violent.

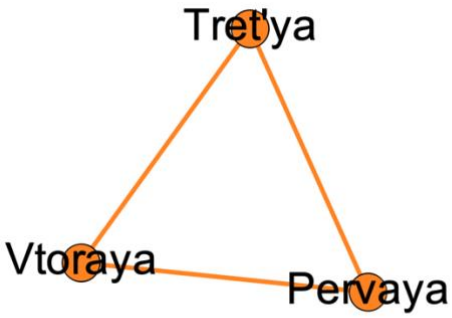


Figure 52. Network graph of *My First Man* by Isaeva

Despite the network graph above, it is unclear whether the characters – or, better, the narrators – should be connected to each other. The graph corresponds to the logic described in the introduction of this thesis, but in reality there may have been no connection intended between the three women, even if they all were present on stage during each other's monologues. In this way, the network graph of the play, encoded by the standards established in the DraCor and this work, is not representative or, to put it more simply, is false. It is clear, as mentioned earlier, that such small-network plays do not benefit from the SNA approach (i.e. no new information about a play can be gathered via the network graph that could not be obtained from reading the play). However, this approach is closely connected to the big data and distant reading ideas that use digital methods for the analysis of a large number of texts to provide a general insight into the evolution of Russian drama, as discussed in the Introduction chapter. Thus, it is important still to

include the graphs of the small-scale plays in a general discussion about the development of Russian drama, for comparison, and to observe the structural evolution of the plays. And without some correct formalised representation of the plays, such analysis turns out to be unsustainable.

The methodology is again challenged in another monologue-based play by Isaeva – *Doc.Tor* (2006). Isaeva received the Grand Prix of the Novaya Drama festival for this play in 2006. It features a monologue by a surgeon (*khirurg*), sharing with the audience a doctor's anecdotes from the hospitals he worked in. As in *Playing the Victim*, the play is structured as a collection of stories, rather than sequential scenes, each labelled with a subtitle.

Even though it is a monologue by one character, similarly to *Duck Hunting* the text of the play is composed as a mixture of the doctor's responses in real life with the responses of his interlocutors in his memories. Hence, the people from Khirurg's stories are presented in the play's text as independent characters and, due to the logic of the TEI encoding, are represented on the graph below as individual nodes. In that sense, the nature of Khirurg's node is different from the other nodes in the network graph. Khirurg himself appears to be the character, narrator and director at the same time. He often addresses the audience directly ("*в зал*"), and the other characters simply play along.

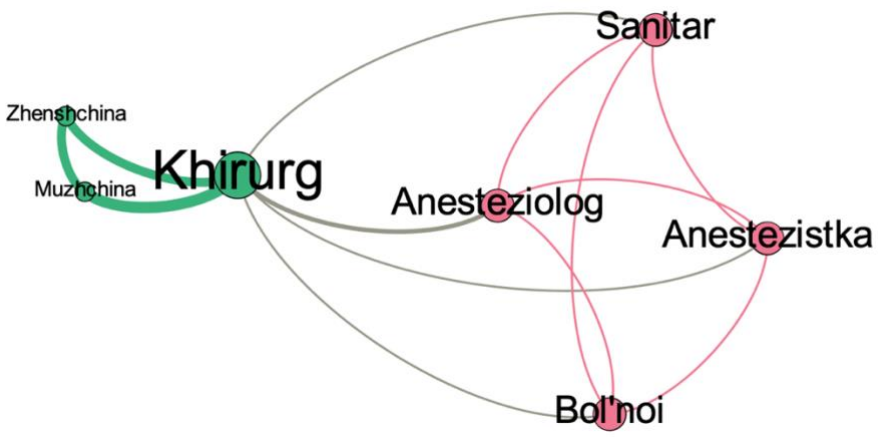


Figure 53. Network graph of *Doc.Tor* by Isaeva

Perhaps one solution in order to highlight the secondary nature of some characters would be to introduce a different way of encoding such characters, and to visualise their nodes and edges on the graph respectively, for instance, by using transparent, less visible nodes and lines. Another solution would be only to include the node of Khirurg in the network graph, since the other characters exist only within the space of his memories and are introduced to the audience via his story and not their own. At this point, it is necessary to address and reevaluate the core principles behind the encoding of the plays. If before, working with earlier plays, it was safe to distance the approach from performance, purely focussing on the textual component of drama for the sake of a unified approach, this is no longer possible with contemporary drama. The issue that first emerged in Vampilov's *Duck Hunting* becomes more crucial in the plays by the Presnyakov brothers, Vyrypaev and now Isaeva.

Apart from the question of how we should define a node in the network graph of a play, another issue in *Doc.Tor* is the ill-defined number of characters, which we have also seen in the Presnyakov brothers' *Terrorism*. Isaeva makes an observation at the beginning of the play that the director is free to add more characters (men and women) on stage. This suggests that the responses attributed to *muzhchina* and *zhenschina* in the play may be made by more than two characters. Such freedom in the number of characters cannot be logically transcribed in the TEI and visualised on the network graph and, therefore has to be ignored by the current approach.

Another interesting feature of contemporary drama, that adds to the compromising of the SNA approach when applied to plays, is the breaking of the fourth wall. As mentioned, Khirurg in *Doc.Tor* addresses the audience directly. Because of that, instead of prioritising his node in the network graph, his character might be split into two characters – the one that narrates the story and the one that takes part in the stories and interacts with the other characters in them. This, again, brings us to the importance of performance over text when analysing contemporary plays.

The breaking of the fourth wall is, indeed, a popular device in documentary drama, and becomes even more obvious and challenging in Elena Gremina's play *One Hour and 18 Minutes* (*Chas 18*). Elena Gremina, together with her husband Mikhail Ugarov, was one of the founders of Teatr.Doc – one of the most significant Moscow theatres for the New Drama movement. Moreover, she was one of the founders of the Lyubimovka and Novaya Drama festival, already mentioned several times in this work, as important hubs for contemporary Russian-language playwrights.

In recent times, the importance of the documentary drama in Russian theatre has been indisputable. The rising numbers of political arrests in Russia, the elimination of the Memorial centre and, most recently, Russia's invasion of Ukraine, could not leave the theatre world aside. Drama has become politicised, and the authorities of the Russian government have started paying close attention to Teatr.Doc productions.[29] These circumstances have created an absolute necessity for authors to document these events and to keep these narratives alive or even construct them, when there are no other means for their expression. *One Hour and 18 Minutes* is, as its subtitle puts it, "суд, которого не было, но который должен быть". It was written in 2010. This semi-documentary play is based on selections from the diaries and letters written during his jail-time by Sergey Magnitsky, as well as witness testimonies and the report by the Public Monitoring Committee (*OHK*) about detention facilities in the Russian Federation. Sergey Magnitsky was arrested after he tried to expose a huge corruption scheme in Russian governmental institutions. According to the evidence which Gremina used for the play, Magnitsky was tortured to death in a detention centre, where he was kept for almost a year, without leaving him a chance to even appear before court. The case of Magnitsky gained international resonance and led to the introduction of The Magnitsky Act by the U.S., which is still now used by the U.S. Government globally to issue sanctions against anyone who "is

[29] The police and the nationalist movement SERB have on multiple occasions interrupted screenings and performances in Teatr.Doc in 2017, 2018, 2019 and 2021.

responsible for extrajudicial killings, torture, or other gross violations of internationally recognized human rights committed against individuals […]".[30] In the play Magnitsky's mother shares her grief over the injustice, or rather the justice that did not happen, to her son. The judges, the doctors, the investigator and other people involved in the case try to justify themselves in the form of court testimonies. This play, as with Vyrypaev's *Oxygen*, had to be encoded manually. It is structured as a court session, includes fourteen judgement paragraphs and one interlude. I have again treated the paragraphs as acts.

As mentioned earlier, *One Hour and 18 Minutes*, among other things, is interesting for the absence of the fourth wall in the play – indeed, this is a common technique for the documentary theatre which is one of the typical features of 'New Drama'. The audience is provided with instructions (short reference information on the judges and witnesses)[31] and the characters occasionally address the audience and request them to read the instructions. Apart from that, a theatre worker (*человек от театра*) appears on stage to read the instructions himself and, again, to ask the audience to read the instruction.

[30] https://www.govtrack.us/congress/bills/114/s284/text.
[31] For instance: *Алексей Криворучко, судья Тверского суда г. Москвы. Дважды санкционировал продление ареста Магнитского до суда. Отклонил просьбы относительно лечения и в смягчении условий содержания. Отклонил просьбу Магнитского выдать ему кипяток в суде. Сказал, что это «не его забота».*

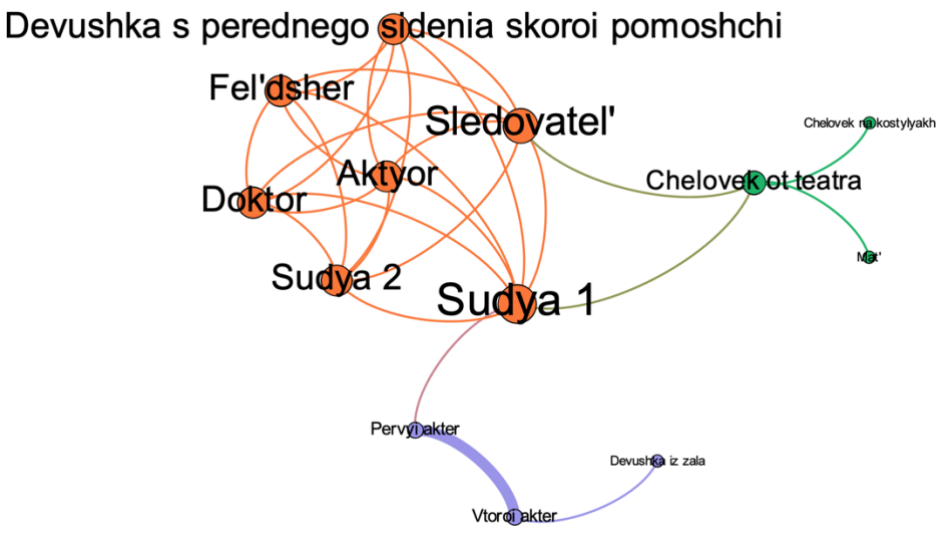


Figure 54. Network graph of *One Hour and 18 Minutes* by Gremina

While the disappearance of the fourth wall does not directly affect the graph representation of the play, the techniques that are used to achieve this effect do. The actors who do not represent any characters from the plot (people connected to the Magnitsky's case) alter the network graph structure. For instance, without the node of *chelovek ot teatra* Magnitsky's mother and the person on crutches would be disconnected from the main network. This would, however, highlight the monologue-based nature of the play, that mimics court testimonies. Therefore, there are two questions here, whether such nodes as *chelovek ot teatra* (*человек от театра*) should be included in the graph at all, since they do not correspond to proper characters in a play without the 'fourth wall', and, if so, whether they should be connected with other characters. Despite these issues, it is not true to say that all documentary plays possess such a challenging and unusual structure. In some cases, the authors abide by a conventional structure and present to the audience plays with a familiar and simple composition.

For instance, in *The Two in Your House* (*Dvoe v tvoem dome*) (2012) written by Gremina, we see a simple network of characters, similar to 19th-century plays. What differentiates this play from 19th-century drama, is the fact Gremina uses subtitles instead of a

division into acts and scenes, as we have seen in other plays discussed in this chapter. This once again brings into focus the prose-like nature of contemporary Russian plays.

The Two in Your House is based on the story of the writer and poet Vladimir Neklyaev, who ran for president in Belarus in 2010. He was arrested when leading a protest group after the presidential elections and put under home arrest. The arrest required Neklyaev and his wife, Olga, to live in their flat together with KGB officers. The play's dialogues between Vladimir, Olga and the KGB officers convey a range of emotions from anger and fear to compassion. While we learn about Olga's attitude to Vladimir's arrest and the government in general, we also see that even among the KGB officers their positions on the (still present) Lukashenko and his policies are quite different. Thankfully, it is a play with a relatively happy ending: Neklyaev was released without serving his full two-year prison sentence.

The network graph below truthfully represents the unity of place in *The Two in Your House*. The network density is 1, and all the characters of the play are connected since they all interact with each other. The weight of the edges highlights the distribution of the communication in the play, with Olga and Vladimir having the strongest connection. We can also note that Vladimir communicates more frequently with the first policeman (Pervyi), while Olga talks more often with the second one (Vtoroi). The third policeman (Tretii) has the weakest connection to the other characters, because he does not live at the apartment of Vladimir and Olga, and talks to the other characters only on a few occasions during the play.

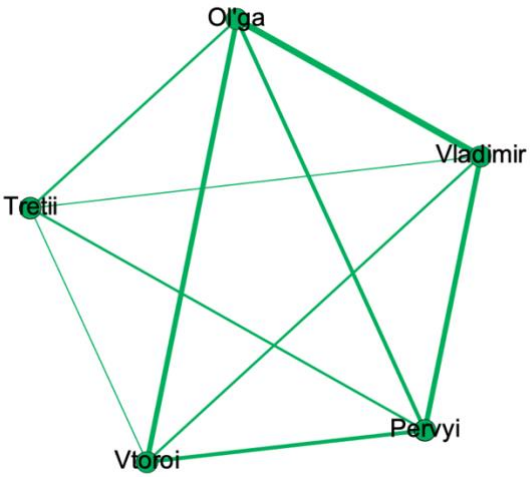


Figure 55. Network graph of *The Two in Your House* by Gremina

Fiction and mockumentary plays

Despite the popularity of documentary drama, the 21st-century authors did not cease to write fictional plays. And, as mentioned in the beginning of this chapter, there were also Ukrainian authors in the New Drama movement. Thus, I am going to discuss one play by Anna Yablonskaya. She represents the last years of unproblematic collaboration between Russian-language playwrights in different nations. The only Ukrainian author in the subcorpus,[32] she was part of the New Russian Drama community, with some of her plays being read for the first time at the Lyubimovka festival in Moscow. Her plays have been staged in Ukraine, Russia and Belarus. Unfortunately, she died young in a terrible terrorist act at Domodedovo Airport, Moscow in 2011.

Yablonskaya was famous for writing rough, emotional drama. I will discuss one of her best-known plays – *Pagans* (*Yazychniki*). It is a two-act family drama about the crisis of faith, written in 2010. Marina and Oleg, husband and wife, have a student daughter Kristina. They are unhappy but their life is settled until Oleg brings his pious mother Natalya Stepanovna to live with them. She is shocked to see how her irreligious son and his family live. Natalya is

[32] Not on purpose, though, as explained in the beginning of this chapter.

determined to impose her own order in the family. Gradually, she succeeds in turning the neighbour Sergey (*Botsman*), Marina and Oleg towards the church. During that time, their daughter Kristina is going through a teenage heartbreak after her relationship with her university tutor (*uchitel'*) ends. She gets suspended from the university, behaves obnoxiously and refuses to respect the beliefs of her family. Natalya insists on Kristina being baptised, in order to exorcise the demon out of her. Unable to stand the pressure and the communication breakdown in her family, together with her broken heart, Kristina jumps out of the balcony window. In the hospital, her father learns the true reason behind Kristina's suicide attempt and realises that his wife and he completely missed out on what was happening to their daughter. At the end, we learn that Natalya dies, Kristina gets better, Oleg leaves Marina, and Sergey returns to his sailing job. They are each happier in their own way, with a transformed attitude to religion.

This play does not have any specific structural and compositional features that would challenge the proposed methodology. On the contrary, as with some of the plays discussed earlier, it benefits from the network representation, which complements our understanding of the play. The network density of *Pagans* is 0.64 and, with the clustering in the network graph below, we can see a clear division between Kristina's life and the rest of the family. It is also interesting that Kristina and Svyaschennik (the priest) are the most influential nodes in the graph, which highlights the play's topic of the transformation of faith in the family and Kristina being the central point of their focus.

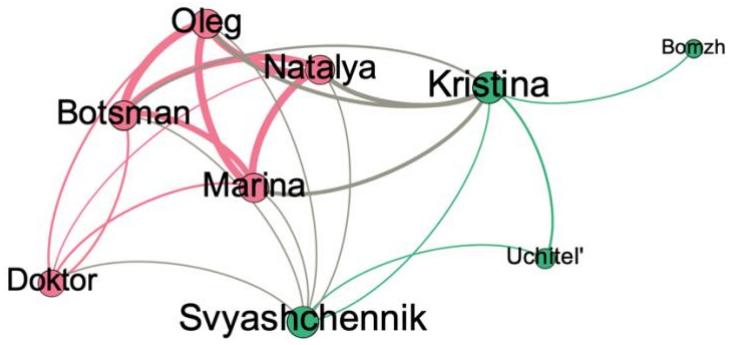


Figure 56. Network graph of *Pagans* by Yablonskaya

This play by Yablonskaya uncovers the process of transformation through religious faith – a topic that many can relate to, regardless of their nationality and history. But since the events of 2014, there has been a great shift in Ukrainian drama, with a transformation and affirmation of Ukrainian national identity in opposition to Russia, and a revival in the use of the Ukrainian language in drama. And we should expect even greater changes in the aftermath of 2022.

Fictional and documentary drama produced an interesting blend that can be called mockumentary drama. There are two such plays in the subcorpus, compiled for this chapter, that were written by Dmitry Danilov. Danilov has written prose, poetry and drama. In 2018 he received the Golden Mask award as the best playwright, and in 2020 his play *A Man from Podol'sk* (*Chelovek iz Podol'ska*) (2016) was turned into a film. It is a comic play about a young man Nikolay who lives in Podol'sk, a town outside of Moscow, works in Moscow and dreams about going to Amsterdam. He gets arrested by the police who threaten and torture him in order to make him see the beauty of Podol'sk.

The pattern of network simplicity in some of the modern plays we have discussed earlier continues to be seen in Dmitry Danilov's works. The star-shaped network graph of *A Man from Podol'sk* below is identical to the ones we can find frequently in the 19th-century drama: it has five characters, and they are all connected to each other without any of them being more

influential in the play. Thus, this graph demonstrates the unity of time, place and action in Danilov's play.

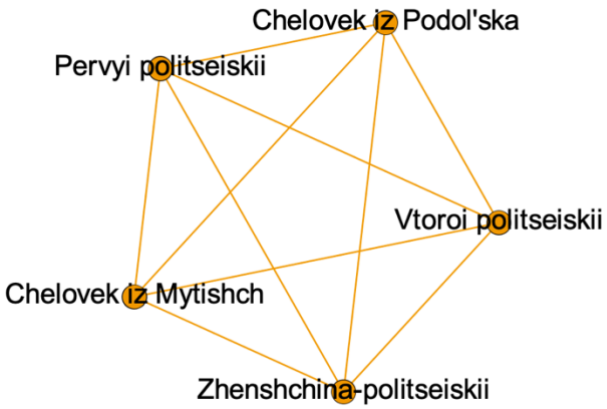


Figure 57. Network graph of *A Man from Podol'sk* by Danilov

The same applies to the last and the most recent play in the subcorpus, also written by Danilov – *Choose Three* (*Vybrat' troikh*). It was written in 2020 and features an online video conversation between five family members, the mother Nelli Maksimovna, the father Nikolay Stepanovich, their two children Pavel and Sveta, and Nikolay's mother Aglaya Pavlovna. The play is set in our time, during the COVID-19 pandemic. The family gathers over their screens, in a fashion which is already so familiar to us, to discuss the new rule that allows the citizens to be in close contact with only three people. The discussion is filled with funny awkwardness when the relatives try to explain and justify to one another why they are not going to include the others in their COVID-19 social bubble.

The play reminds us that within the scope of documentary theatre there is also space for documenting not only political events and their effect on the society, but also important global developments, such as pandemics, environmental crises, etc.

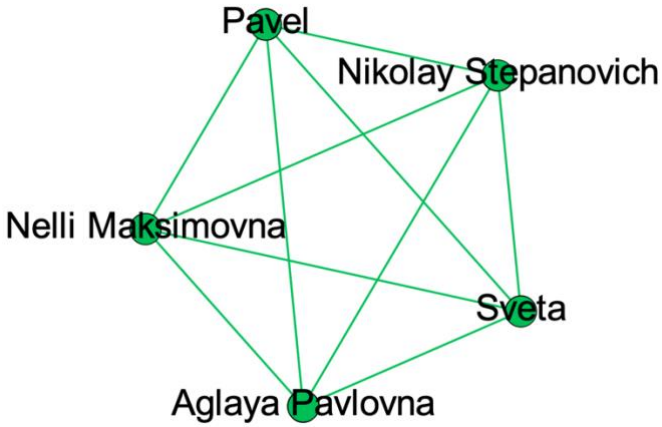


Figure 58. Network graph of *Choose Three* by Danilov

The network graph above demonstrates, again, the unity of time, place and action in the play and the fact that all the characters are equally important in the play from a network-perspective.[33]

Thus, even though some plays, including Danilov's works, fall within the domain of documentary theatre, their unusual content (compared to the earlier plays) does not interfere with the structure of the plays, keeping it simple and straightforward. This emphasises one meaningful insight into the development of Russian-language drama, which turns out to be its drift back (in some cases) towards simplicity.

Conclusion

Apart from highlighting the technical issues and problems in the analysis of contemporary Russian drama, where many plays have required manual editing and encoding, the goal of this chapter was also to demonstrate the conceptual issues a researcher will face when working with New Drama using the means of digital humanities.

[33] This might be proven wrong with the other kind of quantitative and qualitative analysis applied to the plays in the previous chapters.

We have seen that even structurally conventional plays by Vampilov need to be carefully examined, due to his experimental approach to creating a character's narrative. Petrushevskaya and Razumovskaya have greatly contributed to the promotion of plays consisting mainly of dialogues where, as already mentioned several times, the SNA approach becomes redundant. The Presnyakov brothers and, especially, Vyrypaev, are the main innovators in the 21st-century Russian theatre. It is not possible to analyse Vyrypaev's ground-breaking plays 'distantly'.[34] It is, in fact, hard for his readers and his audience to agree on the structure (and understanding) of his plays after reading them ('closely'). Documentary theatre, by its very nature, also brings us plays with an inconvenient composition from a quantitative and digital perspective. However, as already noted, we can currently see a tendency in Russian-language drama to return to uncomplicated structures, prioritising its thematic component.

Thus, my main argument is that while it is possible to adjust the current approach to the New Drama, it requires a more thorough investigation of each play. When working with the 19th-century plays, quantitative methods complement or challenge the traditional literary criticism. Contemporary Russian plays, on the other hand, can only become more confusing when one is trying to formalise them. However, it is still important to include New Drama in the general discussion of Russian drama, at least for the sake of the previously mentioned large-scale analysis. Otherwise, we would not be able to see the tendency in Russian drama to return to simple network structures without comparing the modern plays to the earlier plays in the RusDraCor. I will demonstrate this in greater detail in the Conclusion chapter.

[34] Referring to the idea of distant reading.

Conclusion

By exploring several case-studies based on different kinds of material, a fuller understanding of how and why digital and quantitative methods can be used to analyse drama has emerged from this study. At the same time, a significant focus of this dissertation was to discover the limitations of the proposed methodology. It thus concentrated on the applicability of certain methods and their benefits to traditional literary studies. A number of Russian-language plays of very different genres written in the 19th, 20th and 21st centuries were closely examined, and the insights gathered by using quantitative methods for analysing them were discussed. The RusDraCor project provided the principal data resource for my own work, but I have also encoded a significant number of new works in order to extend the scope of this data, to reflect writing by women and by more contemporary authors. This dissertation therefore hopes to contribute new insights to English-language scholarship in this field.

It has become apparent that the proposed methodology is particularly suitable for the accurate representation of 19th-century plays. It offered valuable additional discoveries, enriching the traditional interpretation of these plays. For instance, when analysing *Boris Godunov* by Pushkin, one of the unexpected conclusions is the fact that Tsar Boris and Lzhedmitriy not only never actually meet, but also that their connection in the play depends solely on Gavrila Pushkin, who acts as an intermediary between the two of them. This conclusion was drawn simply from the network graph of the play, and did not require reading its text.

In *Masquerade* by Lermontov the network metrics highlighted the crucial competition between Knyaz' Zvezdich and Arbenin, leaving Nina to one side. However, the SNA approach does not always provide exceptional insights like that. Therefore, other methods,

such as NLP, can complement the analysis of the plays. For instance, in Gogol's *The Government Inspector* the network graph reflected the unity and equality of the major characters. However, looking at the basic quantitative metrics (number of scenes and speech-acts), we can see that Khlestakov is the most active character in the play, and the basic NLP analysis shows us that he is more self-involved in comparison with the Gorodnichii, while his speech is slightly more engaging. Thus, there is a higher probability of Khlestakov grasping the audience's attention. And this reflects the reader's intuition when reading the play, since the story of Khlestakov and the events around him are the core elements of the plot.

The discussed methodology is not only useful for analysing and comparing the characters within plays, but also for comparing the structures of several plays. For instance, my analysis has provided a quantitative basis for the differences between the two plays with essentially the same *fabula – A Month in the Country* by Turgenev and *Uncle Vanya* by Chekhov. It emphasised the core storyline in the earlier play, while highlighting the decentralisation of the characters' network that Chekhov pursued in *Uncle Vanya*. The comparison of these two plays can be taken further by including another play with a similar plot – *The Stepmother* by Balzac, since this is believed to be the play that inspired Turgenev when writing *A Month in the Country*. Furthermore, the availability of more than a dozen drama corpora in different languages in the DraCor enables scholars to recreate this sort of exercise of comparing certain plays that were inspired by one another or adopted from one another, to see how close they are from quantitative and structural perspectives.

Even though, the SNA approach turned out to be useful for understanding the difference between the two plays, the analysis was not fully automatic, because the TEI files of the plays required additional manual work. In this part of the dissertation, we first encountered the necessity of still undertaking 'close' readings of the plays, despite the

availability of the digitised versions of the texts that allow us to perform 'distant' reading. The segmentation of *A Month in the Country* and *Uncle Vanya* had to be revised in order to carry out this analysis. And it would not be possible to come to certain conclusions about the two plays without modifying the standard segmentation in them, which required semi-manual re-encoding of the texts.

The idea that the 'distant' reading is not possible without the 'close' reading is what lies at the heart of the discussion of the limitations of the methodology in this work. For instance, we come across this issue again in the survey of the representation of women in 19th-century Russian drama, and in Ostrovsky's plays more specifically. By performing a large-scale analysis of the plays, I have established that the majority of 19th-century authors devoted less than half of the space in their plays to female characters, with female characters and speech-acts only taking up a third of the plays on average. Drawing conclusions from the texts available in the RusDraCor, it is safe to say that Ostrovsky is one of the authors whose plays were focused both on male and female characters more equally than in the plays of other authors. However, based on the 'female statistics' of his plays, this is not always the case even in Ostrovsky's works. Thus, we can see that his most famous play *The Storm* can scarcely be included in the list of his 'equal' plays. But a detailed analysis of *The Storm* and its revised network of characters has shown us that the play actually focuses heavily on its female characters and the relationships between them. His play *Without a Dowry*, by contrast, appears to be more of a 'male play', where women merely act as yet another item for the men to negotiate over. However, both plays revolve around the merchant society, and we have seen that the speech of male characters in both plays is more complex than the speech of their female characters. Moreover, both Katerina from *The Storm* and Larisa from *Without a Dowry* have the least diverse speech. This can, however, indicate the persistent, obsessive

focus of their narratives, instead of the limits of their intellectual capacities that might be linked to a more limited vocabulary.

Thus, while the large-scale analysis provided essential insights concerning the issue of women in 19th-century Russian plays and in Ostrovsky's drama, it still required a further, closer examination of the plays. It would be interesting to study other plays by Ostrovsky to see whether their overall statistics regarding the representation of female characters would match with the results of a more detailed analysis. Besides that, a similar large-scale analysis, which, in this case, would be dedicated to the representation of women, can be performed with the focus either on different genres, or on different groups of characters. For instance, it would be interesting to examine whether the representation of women differs in comedies and tragedies (or dramas). In addition to that, one could similarly analyse the representation of working people in comparison to the representation of noble citizens in the plays of the 19th-century. Such an analysis could also include the NLP approach to studying the topics of the plays and the speech of certain groups of characters. Thus, there is a great potential for researchers to use digitised drama texts together with the proposed methodology, in order to survey structural and textual trends in Russian drama.

The methods discussed in this work appear to be quite suitable for the 19th-century conventionally structured plays, and these could benefit even more from further quantitative linguistic analysis. But one will ultimately doubt whether this approach can be applied to the more experimental, avant-garde plays of the following centuries. It was a challenging task to review contemporary Russian-language drama, due to the greater variety of its genres and themes. Nevertheless, I believe it was an essential step for this research, since Russian drama has not ceased to develop and change after the 19th and early 20th centuries. While I was focussing on the modern plays, it was also necessary to include the plays of transitional

playwrights of the second half of the 20th century in the survey of contemporary Russian-language drama.

The modern plays that were surveyed in the final section of this work can be divided into three groups: the plays with conventional network structure, the dialogue-based plays, and the plays with experimental structure, which go so far as to compromise the basic definitions of a dramatic character. There is no doubt that in order to work with modern plays the current methodology needs to be adjusted and revised on multiple levels. Despite that, the distant reading approach still proved to be useful to some extent, even if it failed to offer an accurate representation of the plays in all cases. My chronological review of contemporary Russian-language drama has shown the increasing tendency of the playwrights towards simplicity, that has highlighted the prevalence of thematic component of the plays over their theatrical elements. This becomes more apparent if we look at the poster below, that includes the network graphs of all the plays in the RusDraCor (212 plays available in April 2022) together with the 23 contemporary plays that I have encoded into TEI myself.[35][36]

[35] This is an extended version of the 'Co-occurrence Networks of 144 Russian Plays in Chronological Order (1740s–1940s)' by Fischer, F. and Skorinkin, D., https://doi.org/10.6084/m9.figshare.12058179.v1.
[36] The graph includes the initial versions of the network graphs of contemporary plays before their modification discussed in the last chapter.

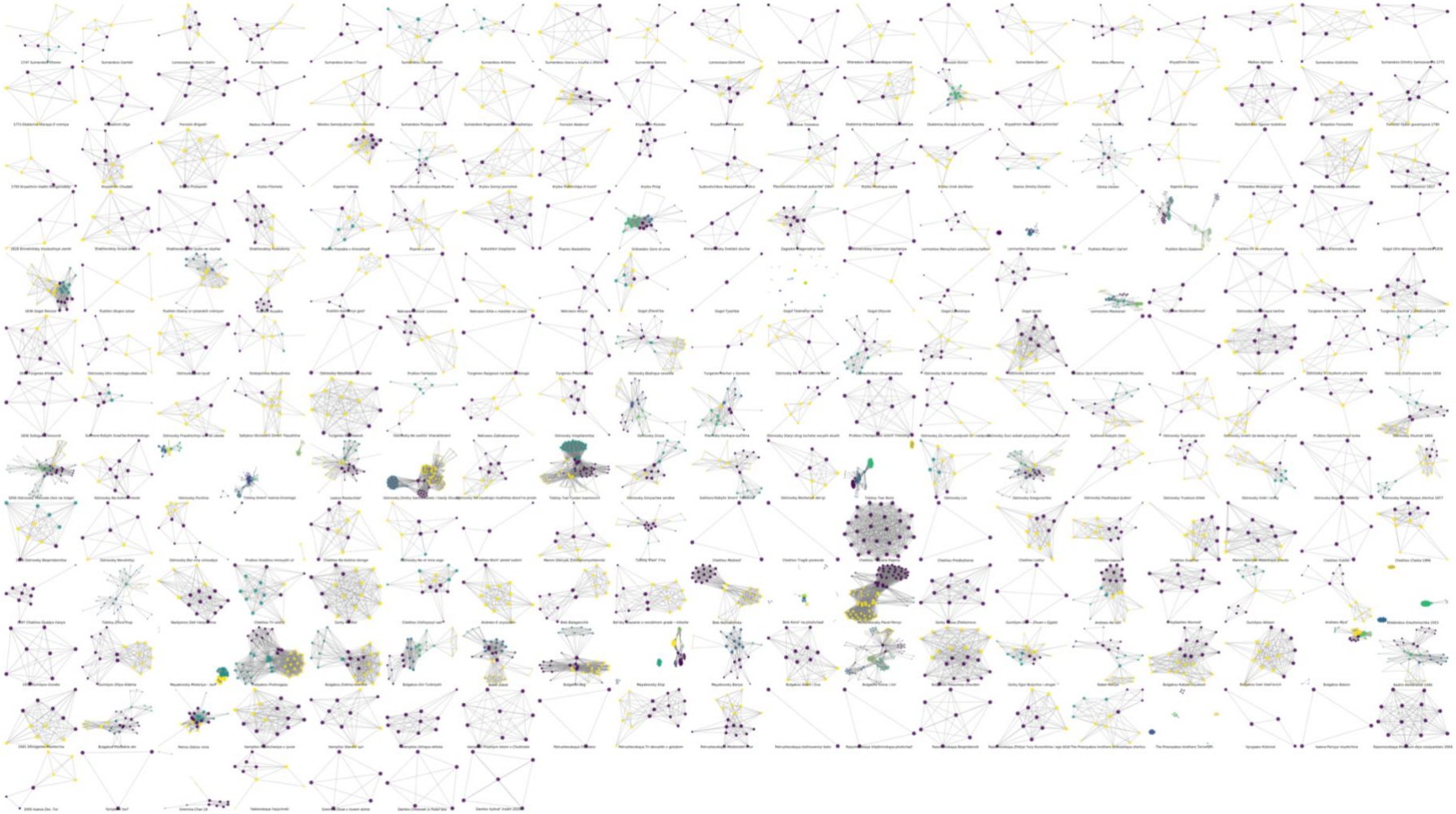


Figure 59. ‘Superposter’ of Russian-language plays

We can see that Russian-language drama reached the peak of its structural complexity between the second half of the 19th century and the beginning of the 20th century.[37] The latest plays in the graph, by contrast, look more similar to the graphs of the plays written in the 18th century and the beginning of the 19th century. It is true, however, that modern plays are underrepresented in this selection of Russian-language drama, and that many plays of the Soviet period have been left out of the picture. In this way, it is important to continue growing the RusDraCor, in order to create the possibility of studying the development of Russian-language drama in greater depth.

A large-scale visualisation of this kind is a canonical example of 'distant' reading that Digital Humanities offer us for analysing the trends in literature. As already mentioned, there are multiple drama corpora available in TEI at the moment, and thus the next step in such an analysis could be to compare and contrast the development of drama in different languages, to establish both common tendencies and deviations across them.

To conclude, the quantitative and digital approach to studying literature opens new research opportunities and offers new perspectives for working with literary texts, and with plays in particular. The choice of methods and the thoroughness of the analysis depend on the research questions and the scale of the research itself. Thus, working with only a few texts requires great attention to details and much revision of the digitised versions of the texts. A large-scale analysis, on the other hand, can tolerate a certain level of inaccuracy without disturbing the general representation of literary trends. Nevertheless, the methods discussed in this work have proven to be a valuable addition to traditional literary studies, by offering a new perspective for looking at Russian drama and by raising challenging questions about contemporary Russian-language plays.

[37] See, for instance, the plays written in that period by Ostrovsky, Tolstoy, Chekhov, Gumilyov, Mayakovsky, Tretyakov and Bulgakov, that include large numbers of characters divided into several communities.

Bibliography

Algee-Hewitt, Mark. 2017. 'Distributed Character: Quantitative Models of the English Stage, 1550–1900', *New Literary History*, vol. 48 no. 4, pp. 751-782.

Andreev M. L. *Metasyuzhet v teatre Ostrovskogo* (Moscow: Rossiyskiy gos. gumanitarny universitet, In-t vysshikh gumanitarnykh issledovaniy, 1995).

Belyaeva, I.A. *Sistema zhanrov v tvorchestve Turgeneva* (Moscow: MPGU, 2005).

Berdnikov, G.P. *Turgenev i teatr* (Moscow: Iskusstvo, 1953).

Biber, D., Gray, B. 'Challenging stereotypes about academic writing: Complexity, elaboration, explicitness', *Journal of English for Academic Purposes,* Vol.9 (1), 2010, pp.2-20.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. 'Fast unfolding of communities in large networks', *Journal of Statistical Mechanics: Theory and Experiment*, 2008: P10008.

Bondi, S., 'Dramaticheskiye proizvedeniya Pushkina' in *A. S. Pushkin, Sobraniye sochinenii v desyati tomakh*, vol. 4, (Moscow: 1975), p. 492.

Briggs, A.D.P. 'Two Months in the Country: Chekhov's Unacknowledged Debt to Turgenev', *New Zealand Slavonic Journal*, 1994, pp. 17-32.

Busa, R. 'The annals of humanities computing: The index Thomisticus', *Computers and the Humanities*, 1980, 14: 83-90.

Chekhov, M. *Literaturnoe nasledie v dvukh tomakh. T. 2: Ob iskusstve aktera* (Moscow: Iskusstvo, 1986).

Chudakov, A.P. *Poetika Chekhova* (Moscow: Nauka, 1971).

Curtis, J.A.E. (ed.) *New Drama in Russian: Performance, Politics and Protest in Russia, Ukraine and Belarus* (London: Bloomsbury Academic, 2020).

Dalton-Brown, S. *Voices from the Void. The Genres of Liudmila Petrushevskaya* (Oxford, New York: Berghahn Books, 2000).

Dereza, O., Kayutenko, D., Fenogenova, A. 'Automatic Morphological Analysis for Russian: A Comparative Study', in *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2021"* (Moscow, 2021).

Dudrah, R., Curtis, J.A.E., Bullock, P., Birksted-Breen, N. (2020) 'A Breath of Fresh Air… Ivan Vyrypaev's *Oxygen* (2002) from Moscow to Birmingham via Oxford' in Katrin Kohl (et al.) *Creative Multilingualism: A Manifesto* (Cambridge: Open Book Publishers, 2020), pp. 87-108.

Dugast, D. 'Sur quoi se fonde la notion d'étendue théoretique du vocabulaire?'. *Le français moderne* Vol.46 (1), 1978, pp. 25-32.

Eikhenbaum, B. 'Kak sdelana «Shinel'» Gogolya', in *Поэтика* (OPOYAZ: Petrograd, 1919), pp. 151-165.

Ermilov, V.V. *Chekhov* (Molodaya Gvardiya, 1946).

Fischer, F., Göbel, M., Kampkaspar, D., Kittel, C., Trilcke, P. 'Network Dynamics, Plot Analysis: Approaching the Progressive Structuration of Literary Texts', in *Digital Humanities 2017: Book of Abstracts* (Montréal: McGill University, 2017).

Fischer, F., Skorinkin, D. (2021) ‘Social Network Analysis in Russian Literary Studies’ in: Gritsenko, D., Wijermars, M., Kopotev, M. *The Palgrave Handbook of Digital Russia Studies* (London: Palgrave Macmillan, Cham, 2020), pp. 517-536.

Fischer, F., Trilcke, P., Kittel, C., Milling, C., Skorinkin, D., ‘To Catch a Protagonist: Quantitative Dominance Relations in German Language Drama (1730–1930)’, in *Digital Humanities 2018: Book of Abstracts / Libro de resúmenes* (Mexico: Red de Humanidades Digitales A. C., 2018), pp. 193-201.

Freeborn, R. ‘Turgenev, the Dramatist’, *Transactions of the Association of Russian-American Scholars in the U.S.A.*, New York, 1985, pp. 57-74.

Freedman, J. *Real and Phantom Pains: An Anthology of New Russian Drama* (Washington, D.C.: New Academia Publishing, 2014).

Gerasimov, Iu., L. M. Lotman, and F. Ia. Priima. 1982. *Istoriia russkoĭ dramaturgii : XVII--pervaia polovina XIX veka* (Nauka: Leningrad).

Gogol, N.V. ‘Revizor’ in *N. V. Gogol. Sobraniye sochinenii v 9 t. T. 4.* (Russkaya Kniga: Moscow, 1994).

Gogol, N.V., ‘Avtorskaya Ispoved’’ in *Sochineniya N.V. Gogolya, naidennyye posle yego smerti* (Moscow: 1855).

Grossman, L. *Teatr Turgeneva* (St Petersburg: Izdatel′stvo Brokgauz-Efron, 1924).

Haug, D. ‘Syntactic discontinuities in Latin — A treebank-based study’ in *Bergen Language and Linguistics Studies*, Vol. 8, Issue 1 (2017).

Heldt, B. *Terrible Perfection : Women and Russian Literature* (Bloomington, Ind: Indiana UP, 1987).

Henley, N. ‘Ostrovskij’s Play-Actors, Puppets, and Rebels’, *The Slavic and East European Journal*, Vol.14 (3), 1970, pp. 317-325.

Henley, N. *Without a dowry and other plays* (New York, NY; London: Ardis, 2014).

Herdan, G. *Type-Token Mathematics: A Textbook of Mathematical Linguistics* (The Hague: Mouton and Co., 1960).

Hoover, David L., Jonathan Culpeper, and Kieran O'Halloran. 2014. *Digital literary studies: corpus approaches to poetry, prose, and drama* (Routledge: New York).

Jockers, Matthew Lee. 2013. *Macroanalysis [electronic resource] : digital methods and literary history* (University of Illinois Press: Urbana).

Johnson, W. ‘Studies in Language Behavior: I. A Program of Research’, *Psychological Monographs*, Vol. 56 (2), 1944, pp. 1-15.

Kablukova, N. ‘Sotsialnaya istoriya i istoriya chastnoi zhizni v piese Petrushevskoi «Moskovskii khor»’, Filologicheskii klass, no.1 (25), 2011, pp. 85-90.

Knuth, Donald Ervin. *The Stanford GraphBase: a platform for combinatorial computing* (Association for Computing Machinery: New York, 1994).

Kurant, E. *Teatr chudes. Avtorskiye strategii v dramaturgii Ivana Vyrypaeva* (Krakow: LIBRON – Filip Lohner, 2020).

Lermontov, M. ‘Maskarad’ in *M. Yu. Lermontov. Sochineniya v 2-kh tomakh. T. 2.* (Pravda: Moscow, 1990).

Litinenko, N. *Pushkin i teatr: formirovaniye teatralnykh vozzrenii* (Iskusstvo: Moscow, 1974).

Lomunov K., '«Maskarad» Lermontova kak sotsial'naya tragediya' in *Lermontov M. Yu. Maskarad* (BTO: Leningrad, 1941), pp. 43 - 92.

Lotman L., 'Istoriko-literaturnyi kommentarii' in *Пушкин А. С. Борис Годунов.* (Gumanitarnoye Agenstvo «Akademicheskii Proyekt»: Saint Petersburg, 1996), pp. 129-359.

Magarshack, David, *Chekhov the dramatist* (London: J. Lehmann, 1952).

Mann, IU. *Tvorchestvo Gogolia : smysl i forma* (Izd-vo Sankt-Peterburgskogo universiteta: Sankt-Peterburg, 2007).

Meister, Charles W. *Chekhov criticism: 1880 through 1986* (London : McFarland, 1988).

Mendenhall, T. C. 'The Characteristic Curves of Composition' in *Science*, 1887, ns-9: 237.

Mendenhall, T.C. 'A Mechanical Solution of a Literary Problem' in *Popular Science, Monthly*, 60, 1901, pp. 97-105.

Miles, J. *Major Adjectives in English Poetry: From Wyatt to Auden* (London: Cambridge University Press, 1946).

Montgomery, Guy. *Concordance to the poetical works of John Dryden* (University of California Press: Berkeley, 1957).

Moretti, Franco. 'Network Theory, Plot Analysis', *New Left Review*, 2011, NLR68: 80.

Moretti, Franco. *Distant Reading* (Verso: London, 2013).

Nasseri, M., Thompson, P. 'Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences', *Assessing Writing*, Vol. 47, 2021.

Ogata, Takashi, and Taisuke, Akimoto. *Computational and cognitive approaches to narratology* (IGI Global, 2016).

Ostrovsky, A. N. 'Puteshestviye po Volge ot istokov do Nizhnego Novgoroda', *Morskoy sbornik*, 1859.

Peace, R. 'A.N. Ostrovsky's "The Thunderstorm": The Dramatization of Conceptual Ambivalence', *The Modern Language Review*, 1989, Vol.84 (1), p. 99.

Pitcher, H. *The Chekhov Play* (Berkeley: University of California Press, 1985).

Pomerantsev, V. 'Ob iskrennosti v literature', *Novy Mir*, no. 12, 1953.

Propp, V. *Morphology of the Folktale* (Berkley: University of California Press, 1968).

Propp, V., 'Morfologiia skazki' in *Voprosy Poetiki*, issue 7 (Academia: Leningrad, 1928).

Pushkin A., 'Boris Godunov' in *Pushkin A. S. Polnoye Sobraniye Sochinenii v 20 tt. T. 7: Dramaticheskiye Proizvedeniya*, (Nauka: Saint Petersburg, 2009).

Rayfield, D. *Chekhov. The Evolution of his Art* (London: Paul Elek, 1975).

Seeley, F.F. *Turgenev: a reading of his fiction*. (Cambridge: University Press, 1991).

Semanticheskaya razmetka khudozhestvennykh tekstov dlya kolichestvennykh issledovanii v filologii (na primere romana "Voina i mir" L. N. Tolstogo)

Semyonov, S.T. 'O vstrechakh s A.P. Chekhovym', *Put'*, issue №2, 1913.

Sherman, L. A. *Analytics of literature* (Boston: Ginn & Co., 1901).

Skorinkin D., *Semanticheskaya razmetka khudozhestvennykh tekstov dlya kolichestvennykh issledovanii v filologii (na primere romana "Voina i mir" L. N.*

Tolstogo) (unpublished doctoral thesis, Higher School of Economics, Moscow, 2018).

Smeliansky, A. The Russian Theatre After Stalin (Cambridge University Press, 1999).

Sverdlov, M. I. '«Krasota»: Konflikt znacheniy (analiz «Grozy»)'. *A. N. Ostrovsky, A. P. Chekhov i literaturnyi protsess XIX-XX vv*., Moscow: Intrada, 2003, pp. 88-108.

Temple, Emily, 'How Many Books Will You Read Before You Die?' <https://lithub.com/how-many-books-will-you-read-before-you-die/>, 2017 [accessed 1 May 2019].

Trilcke, P., Kittel, C., Reiter, N., Maximova, D., Fischer, F. 'Opening the Stage: A Quantitative Look at Stage Direction in German Drama', in *Digital Humanities 2020: Book of Abstracts* (Canada: University of Ottawa and Carleton University, 2020).

Turgenev I. *A Month in the Country*, trans. Berlin, I. (London: The Hogarth Press, 1981).

Turgenev I., Savina M. *Turgenev i Savina. Pis'ma. Vospominaniya* (St Petersburg: Izdatel'stvo Gosudarstvennykh Teatrov, 1918).

Vishnevskaya, I. L. *Teatr Turgeneva* (Moscow: Nauka, 1989).

Vyrypaev, I. *P'esy* (Moscow: Tri kvadrata, 2019).

Yarkho, B. *Metodologiya tochnogo literaturovedeniya: Izbrannyye trudy po teorii literatury* (Yazyki Slavyanskikh Kul'tur: Moscow, 2006).