

Fachartikel aus:

Zeitschrift für digitale Geisteswissenschaften, Heft 10 (2025)

Titel:

Temporal Aspects of Structural Differences in Dramatic Genre

---

Autor\*in:

Botond Szemes

Kontakt: [szemes.botond@abtk.hu](mailto:szemes.botond@abtk.hu)

Institution: HUN-REN Research Centre for the Humanities, Institute for Literary Studies

GND: [1322519579](#) ORCID: [0000-0002-0637-6776](#)

Contribution (CRediT): [Conceptualization](#) | [Formal analysis](#) | [Writing - original draft](#) | [Methodology](#)

Autor\*in:

Mihaly Nagy

Kontakt: [nagy.lukin.mihaly@gmail.com](mailto:nagy.lukin.mihaly@gmail.com)

Institution: Eötvös Lorand University, Doctoral School of History

GND: [1337490776](#) ORCID: [0000-0002-0577-8706](#)

Contribution (CRediT): [Data curation](#) | [Writing - review and editing](#)

---

DOI des Beitrags:

[10.17175/2025\\_007](#)

Nachweis im OPAC der Herzog August Bibliothek:

[1927466989](#)

Erstveröffentlichung:

05.06.2025

Lizenz:

Sofern nicht anders angegeben 

Letzte Überprüfung aller Verweise:

04.06.2025

Format:

PDF ohne Paginierung, Lesefassung

GND-Verschlagwortung:

[Drama](#) | [Gattungstheorie](#) | [Handlung \(Literatur\)](#) | [Informationsfluss](#) | [Netzwerkanalyse](#) | [Literaturwissenschaft](#)

Empfohlene Zitierweise:

Botond Szemes / Mihaly Nagy: Temporal Aspects of Structural Differences in Dramatic Genre. In: Zeitschrift für digitale Geisteswissenschaften 10 (2025). 05.06.2025. HTML / XML / PDF. DOI: [10.17175/2025\\_007](#)

Botond Szemes / Mihaly Nagy

# Temporal Aspects of Structural Differences in Dramatic Genre

## Abstract

Computational drama analysis often treats plays as static entities, overlooking both their processual nature and the influence of literary history in their formation. This paper aims to address these limitations. Building on our previous research, which quantitatively examined the structural differences between dramatic genres (comedies and tragedies), we introduce the dimension of temporality in two key ways. First, we explore the impact of a play's creation time on its classification. While our findings suggest that the time of creation does not significantly influence structural characteristics, they reveal a gradual decline in the distinctiveness of generic structures over time. These results contribute to ongoing debates in genre theory, particularly regarding structuralism versus historicism, as well as to discussions on the cultural evolution of dramatic genres. Second, we investigate how structural features evolve within individual works as their plots unfold. Our analysis emphasises the processual nature of genre-specific relationships between characters, offering a dynamic perspective on dramatic form.

Computergestützte Dramenanalyse behandelt Theaterstücke häufig als statische Gebilde und übersieht dabei sowohl ihren prozesshaften Charakter als auch den Einfluss der Literaturgeschichte auf ihre Entstehung. Dieser Beitrag befasst sich mit jenen Einschränkungen. Aufbauend auf unserer früheren Forschung, in der wir die strukturellen Unterschiede zwischen dramatischen Gattungen (Komödien und Tragödien) quantitativ untersucht haben, führen wir die Dimension der Zeitlichkeit in zweierlei Hinsicht ein. Erstens untersuchen wir den Einfluss der Entstehungszeit eines Stücks auf seine Klassifizierung. Unsere Ergebnisse deuten zwar darauf hin, dass die Entstehungszeit keinen signifikanten Einfluss auf die strukturellen Merkmale hat, doch zeigen sie eine allmähliche Abnahme der Unterscheidbarkeit der Gattungsstrukturen im Laufe der Zeit. Diese Ergebnisse leisten einen Beitrag zu den laufenden Debatten in der Gattungstheorie, insbesondere in Bezug auf Strukturalismus kontra Historismus, sowie zu den Diskussionen über die kulturelle Evolution dramatischer Gattungen. Zweitens untersuchen wir, wie sich strukturelle Merkmale innerhalb einzelner Werke im Verlauf ihrer Handlung entwickeln. Unsere Analyse hebt den prozessualen Charakter der gattungsspezifischen Beziehungen zwischen den Figuren hervor und bietet dadurch eine dynamische Perspektive auf die dramatische Form.

## 1. Introduction

<sup>1</sup>The strength of studies on the structural differences between dramatic genres, we believe, lies in their ability to reveal the connection between fundamental storytelling modes and forms of community organization. By examining character interactions (mostly in forms of networks) and their distribution, computational approaches can highlight structural features that relate to thematic characteristics of different genres.<sup>2</sup> This motivation also drives the present investigation. Such studies, however, often rely on measurements insensitive to historicity and the processual nature of the plot, which may narrow the scope of the research. Therefore, this time we focus on the *temporality* of the features in order to uncover the dynamics that underlie broad structural distinctions. [1]

On the one hand, genres are frequently treated as historical constants in computational drama analysis, overlooking the fact that they have been historically developed and subject to change. Our paper addresses this issue by linking the analysis to genre theory, particularly discussions on historicism vs. structuralism.<sup>3</sup> The central question is whether the differences in dramatic genres evolve over time and to what extent the time of creation influences the classification based on structural features. The results offer a nuanced and complex picture. While the structural distinction between comedies and tragedies has persisted for centuries, [2]

<sup>1</sup> Botond Szemes was supported by the ÚNKP-22-4 and ÚNKP-23-4 New National Excellence Program of the Ministry for Culture and Innovation (Hungary) from the source of the National Research, Development and Innovation Fund. The authors are grateful to the reviewers for their insightful comments on the original draft.

<sup>2</sup> Cf. Algee-Hewitt 2017; Szemes / Vida 2024; Trilcke et al. 2015b.

<sup>3</sup> See also in the field of computational literary studies: Underwood / The NovelTM Research Group 2016.

the *degree* of distinction has gradually diminished. In our interpretation, this finding simultaneously reinforces the idea of consistency of genres and illustrates the mechanisms of cultural evolution, whereby artistic forms are transformed (and often blurred) through transmission and imitation.<sup>4</sup>

On the other hand, the commonly used metrics (mostly derived from network theory) treat plays as a whole rather than capturing how different features evolve throughout the plot. While this approach has its advantages, it also has limitations. Franco Moretti famously attributes the epistemic power of character networks to their ability to spatialise the chain of events that would otherwise develop over time:

[3]

»But before analyzing spaces in detail, why use networks to think about plot to begin with? What do we gain, by turning time into space? First of all, this: when we watch a play, we are always in the present: what is on stage, is; and then it disappears. Here, nothing ever disappears. What is done, cannot be undone. Once the Ghost shows up at Elsinore things change forever, whether he is on scene or not, because he is never not there in the network. The past becomes past, yes, but it never disappears from our perception of the plot.«<sup>5</sup>

Acknowledging the significance and spatialising power of diagrams, we argue that incorporation of temporal changes can further enhance such studies. Our second hypothesis is that the characteristics of dramatic texts and genres emerge through the evolving structure of character networks – that is, through the transformation of the represented communities. The paper demonstrates that some of the main structural differences between dramatic genres are not present from the outset of the plays, nor are they merely a consequence of contrasting endings. Rather, these differences emerge as *processes* unfolding in opposite directions. By analysing changes in character networks, we find that comedies tend to depict the formation of dense communities, whereas tragedies illustrate their gradual disintegration.

This approach not only refines our understanding of genre distinctions but also provides a framework for close reading of individual plays. To illustrate this, the paper includes brief interpretations of Shakespearean plays that exemplify or deviate from their respective genres. In addition to a detailed examination of these cases, we extend our analysis to a broader corpus of German dramas to ensure statistical robustness and validate our findings across different literary traditions.

[4]

## 2. Related Works and Corpora

In an earlier paper, we contributed to the line of research that seeks to capture structural differences between comedies and tragedies.<sup>6</sup> The methodology of computational drama analysis can help not only to answer such questions in relation to a large number of texts, and to identify their features that are not reflected in the reading process, but also to look at known phenomena from a new perspective. Gaining this new perspective was the most important result for us, leading to a new kind of description – and not to an apparently objective, final decision on the questions of genre.<sup>7</sup> Our results have made it possible to describe the difference between comedies and tragedies not in terms of linguistic<sup>8</sup> or thematic aspects<sup>9</sup> or simply in terms of positive / negative endings, but from the point of view of *the system of relationships between the*

[5]

<sup>4</sup> For an overview, cf. Sobchuk 2023.

<sup>5</sup> Moretti 2011.

<sup>6</sup> Szemes / Vida 2024.

<sup>7</sup> Cf. »Digitally based research does not offer us the impossible dream of objective humanities research, but it does offer us the possibility of applying subjective humanities-based insights in a consistent way to test their applicability and utility across a large number of instances« (Hope / Witmore 2010, p. 365).

<sup>8</sup> Cf. Hope / Witmore 2014.

<sup>9</sup> Cf. Asmuth 2016; for a computational approach, see Schöch 2017. The thematic aspect can be traced back, of course, to Aristotle, who describes the difference between genres in terms of the ordinariness of the characters and the historicity of the events (i.e. whether they really happened before). Cf. Aristotle 2012.

*characters*. We could capture these relationships both through the metrics describing character networks created along the appearances in common scenes<sup>10</sup> and through the distribution of words and time spent on stage among them. These are collectively referred to as structural features.

While character networks can inform us about dramatic form<sup>11</sup> and the structure of the plot,<sup>12</sup> our primary focus was their interpretability in terms of social relations. This does not mean revisiting the debate on how character networks correspond to real-world networks;<sup>13</sup> rather, it underscores that dramas inherently represent different types of communities. Examining these representations enriches both textual interpretation and our understanding of real social structures. Likewise, while characters should not be equated with real people, analysing them can reveal meaningful patterns of personality and behaviour or function in the networks. Ultimately, we aimed to uncover the underlying community structures that define stories as comedies or tragedies and to explore what insights they offer for understanding our shared world. The analysis of speech distribution and stage time is also relevant in this context, as these relational metrics reflect community structures from another perspective (e.g. how equal a chance do the characters have to shape the discourse of the play?). For this reason, we did not strictly separate language and network-based metrics in our research.<sup>14</sup>

[6]

The accuracy of a *Support Vector Machine (SVM)* classification<sup>15</sup> in *leave-one-out (LOO)* scenarios<sup>16</sup> suggests that comedies are characterized by denser networks and fewer prominent characters who control the exchange of information, talk notably more than their peers, or have considerably more connections than others. In such a network multiple pieces of information circulate simultaneously, not only leading to ineffective communication and misunderstandings (also as a source of comical situations) but preventing the will of a single character and a single worldview from prevailing in the play. This structure thus avoids the operation of tragedies, in which networks are broken down into subgroups, one or two key figures act as the link between them, and the majority of the characters play only a peripheral role in the development of the plot (both in terms of number of links and amount of speech). While these arrangements result in more effective communication by allowing the spread of a single or a small number of pieces of information (thus limiting misunderstandings but increasing vulnerability and the possibility of deliberate deception), they are also more fragile, since the failure of this ›one truth‹ can lead to the disintegration of the whole network.<sup>17</sup>

[7]

<sup>10</sup> In such a network the nodes represent the characters, and the edges are created by characters appearing in the same scene (in the case of weighted graphs, the more often the characters appear together, the stronger the connection – visually: the thicker the edge – between them).

<sup>11</sup> See e.g. Trilcke et al. 2024.

<sup>12</sup> See e.g. Moretti 2011.

<sup>13</sup> For a critique of such an approach, see: Labatut / Bost 2019; Trilcke 2013, p. 207.

<sup>14</sup> We filtered out metrics that were highly correlated with cast size and other metrics (in this case, we chose the one that better described structural organisation), which resulted in the following 13 features: first of all we used metrics describing the character networks: *average cluster coefficient* (the extent to which nodes connected to a node would also be connected to one another), *density* (a ratio of the number of actual edges to the number of possible edges), *diameter* (how far are the two furthest nodes from each other), *maximum betweenness score* (betweenness: how many shortest paths pass through a node, i.e. how much it acts as a link between subclusters in the network) and the *ratio of maximum degree and the network size* (the node with the highest degree interacts with what percentage of the cast). We also calculated the proportion of characters with *high, medium and low degree* in the networks after grouping them algorithmically. The same was done based on words spoken by characters and the portion of groups with *high, medium and low speech* was also used in the classification, as we considered these proportional aspects of the characters to be measures of structure too. This was supplemented by the calculation of *the average length of utterances* (in words) and *the average number of characters in a scene*.

<sup>15</sup> This procedure can be understood by imagining that each play is represented as a point in a 13-dimensional space, where each dimension corresponds to one of the measured features. A Support Vector Machine (SVM) then tries to find the best dividing line – or, in higher dimensions, a flat surface called a *hyperplane* – that separates the plays into different categories. The *kernel* is the mathematical function that determines how this dividing line is drawn. We used a *linear kernel*, which means the SVM draws a straight, flat boundary without bending or curving to fit the data more closely. This lack of *curvature* helps keep the model simple and prevents it from overfitting – that is, from paying too much attention to minor quirks in the data that might not apply to other, unseen examples.

<sup>16</sup> In a leave-one-out scenario, the model is trained on all but one data point, which is then used for testing, and this process is repeated for each data point in the dataset.

<sup>17</sup> Cf. Szemes / Vida 2024.

However, similar approaches to structure ignore one important element: time. Specifically, they do so in two senses. On the one hand, the temporality of the drama history is ignored when the characteristics of each genre are considered as historical constants; on the other hand, so are the development of the relations between the characters during the plot when the metrics refer to a drama as a whole. In what follows, we aim to fill this gap and implement (1) the historicity of genre differences and (2) the evolution of character networks within a play.

[8]

At the same time, it is important to note that examples of both approaches can be found, albeit sporadically, in the field of computational drama analysis. For the former, we can mention Mark Algee-Hewitt's influential research, in which he also discussed the chronological development of certain metrics in order to draw conclusions for English drama history (e.g. the distinction between central and peripheral characters in dramas becomes smaller, while the number of non-main characters with important connecting functions increases between the 16th and 20th centuries)<sup>18</sup> and Trilcke et al. who did the same for German language plays from 1730 to 1920 with the metrics of *density*, *average path length*, *number of characters*, *maximum degree* (the highest degree of an actor of a drama network) and *average degree*.<sup>19</sup> It is also worth mentioning Hope and Witmore's study on the language of Shakespeare's plays, in which they show that the texts exhibit a shift from concrete references to the external world toward a greater emphasis on internal discursive relations.<sup>20</sup> Finally, Moscato et al. have created statistical models also based on the variation in word frequencies to determine the first performance of Shakespearean-era plays.<sup>21</sup>

[9]

For the latter the research by Fischer et al. offers a good example, who also start from the premise that »the sequential dimension of literary texts, as a consequence of their temporality, usually remains in the dark: what is extracted, visualised and analysed are static networks. Plot, however, is essentially a concept supposed to theoretically encompass the temporality of narrative (as well as dramatic) texts.«<sup>22</sup> In order to capture the temporality of the plot, Fischer et al. examine the process of progressive structuration using event-based and progression-based measures: an event-based approach, for example, is to identify the point in the plot at which all the characters in the play have already appeared on the scene; a progression-based measure is to determine the dynamic of the change of characters between scenes (the ratio of the number of new or leaving characters between two scenes and the total number of characters in the scenes). Similarly, Nalisnick and Baird focus on the changes in a single drama, but they capture the emotional arc of the characters in Shakespeare's plays – and draw conclusions about their roles and the turning points in the plot.<sup>23</sup> Finally, we can also draw on computational studies examining how different genres conclude. Alison Findlay has analysed the final utterances of Shakespeare's plays to identify distinctive features of discourse at endings.<sup>24</sup> Similarly, Reiter and Willand have explored the characteristics of tragic and comedic endings, as well as shifts in characters' speech patterns toward the conclusion of a play.<sup>25</sup>

[10]

In the following analysis we use the same corpora as in our previous study to achieve comparability: the August 2022 version of the *German*<sup>26</sup> and the *Shakespeare* sub-corpus of the DraCor database.<sup>27</sup> Beyond the fact that the Shakespeare collection attracts special attention in (quantitative) drama analysis, it also provides a well-defined sub-corpus which is small enough to allow a closer look on the results. Accordingly, in this paper we also provide micro-interpretations of some plays by Shakespeare. GerDraCor, by contrast, supplies a large number of well-encoded texts ideal for statistical analysis – from which we selected those

[11]

<sup>18</sup> Algee-Hewitt 2017.

<sup>19</sup> Trilcke et al. 2015a.

<sup>20</sup> Hope / Witmore 2014.

<sup>21</sup> Moscato et al. 2022.

<sup>22</sup> Fischer et al. 2017.

<sup>23</sup> Nalisnick / Baird 2013.

<sup>24</sup> Findlay 2020.

<sup>25</sup> Reiter / Willand 2022.

<sup>26</sup> The data about the examined version can be found [in our GitHub repository](#).

<sup>27</sup> Fischer et al. 2019.

that have genre labels (comedy or tragedy),<sup>28</sup> more than five characters and more than two scenes: in all other cases, character networks cannot be said to be indicative of dramatic structure of distinct genres. Applying these reductions, the GerDraCor collection finally contained 253 dramas. The use of two different collections strengthens the representativeness of the results, but more extensive research is needed in the future to reach general findings – at present, our findings are valid only for the traditions and periods covered by the two corpora.

### 3. Time as Drama History

Table 1 presents the accuracy<sup>29</sup> of genre classification for GerDraCor, the larger and thus statistically more reliable dataset from our previous study. The relatively high F1 scores indicate fundamental differences in the composition of genres. However, when visualizing the relationship between dramas in two-dimensional space using *principal component analysis (PCA)*, the genre distinction becomes less pronounced (see Appendix, Figure 11). This suggests that the observed metric differences cannot be attributed solely to genre. Other factors, such as the time of a drama's creation, may also influence the positioning of data points.

[12]

	Precision	Recall	F1
Comedy (n=136)	0.73	0.82	0.77
Tragedy (n=117)	0.76	0.65	0.71

Table 1: Accuracy score for the LOO SVM classification by genre.

This is supported by the large chronological extent of the examined version of GerDraCor corpus: the earliest dramas are by Andreas Gryphius from 1650 (*Horribilicribrifax Teutsch*, *Carolus Stuardus*, *Leo Armenius*), while the latest is Henry von Heiseler's tragedy *Die Kinder Godunófs* from 1929. Although the classification has shown that even at such a long period of time, the structural differences between comedies and tragedies can be grasped, it is reasonable to assume that these specificities are not realised in the same way in different historical periods. In other words, we aim to supplement the structuralist point of view with aspects of literary history. Our first hypothesis therefore is as follows: metrics carry information not only about the genre of a play, but also about its position in the history of drama. However, by marking the date of creation in the figure (more precisely, whether the year assigned to the drama is earlier or later than the median year of the dataset, 1829), groups of works by period do not really emerge (see Appendix, Figure 12). At the same time, it is also worth bearing in mind that these visualisations may not be able to convey all aspects of the variance in the data due to dimensionality reduction. We therefore decided to conduct another experiment to test the hypothesis.

[13]

First, we clustered the GerDraCor dramas according to their metrics using *k-means clustering*,<sup>30</sup> and then calculated (1) the absolute difference between the number of comedies and tragedies in each group – a measure of the extent to which genre determines clustering: the larger the difference, the more over-

[14]

<sup>28</sup> The metadata in GerDraCor's corpus also specify the so-called »normalized genre tags« which standardise the different designations that refer to the same genre tradition (e.g., »Lustspiel« and »Komödie«).

<sup>29</sup> *Precision* is the proportion of true positive predictions (correctly identified positives) out of all the positive predictions made by the model. It answers the question: »Of all the positive predictions, how many were actually correct?« *Recall* is the proportion of true positive predictions out of all the actual positive instances in the data. It answers the question: »Of all the actual positives, how many did the model correctly identify?« *F1 score* is the harmonic mean of precision and recall. It combines both metrics into a single number, balancing the trade-off between precision and recall.

<sup>30</sup> K-means clustering is an algorithm that groups data points into K number of clusters based on their similarity. It starts by selecting K random points as cluster centers (centroids), then assigns each data point to the nearest centroid. Afterward, the centroids are updated by calculating the geometric centre of the points in each cluster. This process repeats until the centroids stabilize, meaning the algorithm has found the best grouping of the data.

represented a genre is in the cluster;<sup>31</sup> and (2) the *standard deviation (SD)* of years<sup>32</sup> – which measures the extent to which clustering is determined by the time period: the smaller the SD, the more works from one period are included in the cluster. Finally, we took the average of the values for both calculations, so that we could characterise the whole grouping with only two numbers.<sup>33</sup>

These were compared with the results of random clustering as a second step. For this the same number and size of clusters as in the k-means procedure were created in random sampling, and the absolute difference in the number of genres and the variance of years in these clusters were calculated in the same way. The sampling was repeated 1000 times and the mean of the values was calculated taking all 1000 cases into account. We were then able to examine the extent to which the metrics of the groups clustered by k-means differed from those of the random sampling. If the difference in the number of genres is considerably larger compared to the random groups, then clusterisation is also based on genre; if the variance in years is considerably smaller compared to the random groups, then clusterisation is also based on the time of writing. [15]

The experiment was also performed for 4, 6, 8 and 10 clusters. The results are shown in Figure 1. While there is a major difference between the two ways of grouping in terms of genre (the smallest difference with 10 clusters is also present – see Figure 2), there is just a faint difference in terms of temporality (the largest with 10 clusters is almost non-existent – see Figure 2). Thus, we must reject our first hypothesis: the dramas clustered by structural features do not typically come from the same period. This also means that the genre difference is a long-standing phenomenon, even over centuries (at least in relation to GerDraCor). This reinforces an ahistorical view of genre. The cluster that is most likely to refer to a single period (for k = 10) includes 20 dramas, mainly from the first third of the 19th century, which have many characters who appear in only one episode of the story held together by the journey of the hero(es); such as the variations of the Faust story (see Table 3 in the Appendix). [16]

<sup>31</sup> One might reasonably ask why we did not use an established evaluation metric like *purity* instead of defining our own measure. Purity is a common way to assess how well clustering aligns with known labels – in this case, genres – by measuring the proportion of items in each cluster that share the same label. However, because we have only two genres (comedy and tragedy) but multiple clusters, even perfectly genre-homogeneous clusters would not yield a maximum purity score of 1, since each genre would inevitably be distributed across several clusters. This makes purity less suitable as a direct measure of genre over-representation in our case. Still, it can provide a general sense of genre consistency: our k-means clusters are ‘purer’ than random clusters, even if purity does not capture the full extent of the genre effect in this specific setup. See the calculations in the [GitHub repository](#).

<sup>32</sup> DraCor’s interface assigns a «normalized year» to each work: this is normally the date of first publication or first performance (whichever is earlier). However, if more than 10 years separate the writing and the performance or printing of the work, the normalised year is the date of writing. Cf. <https://dracor.org/doc/faq>.

<sup>33</sup> Although the k-means procedure gives slightly different results with each new run, these differences were found to be negligible in the experiment.

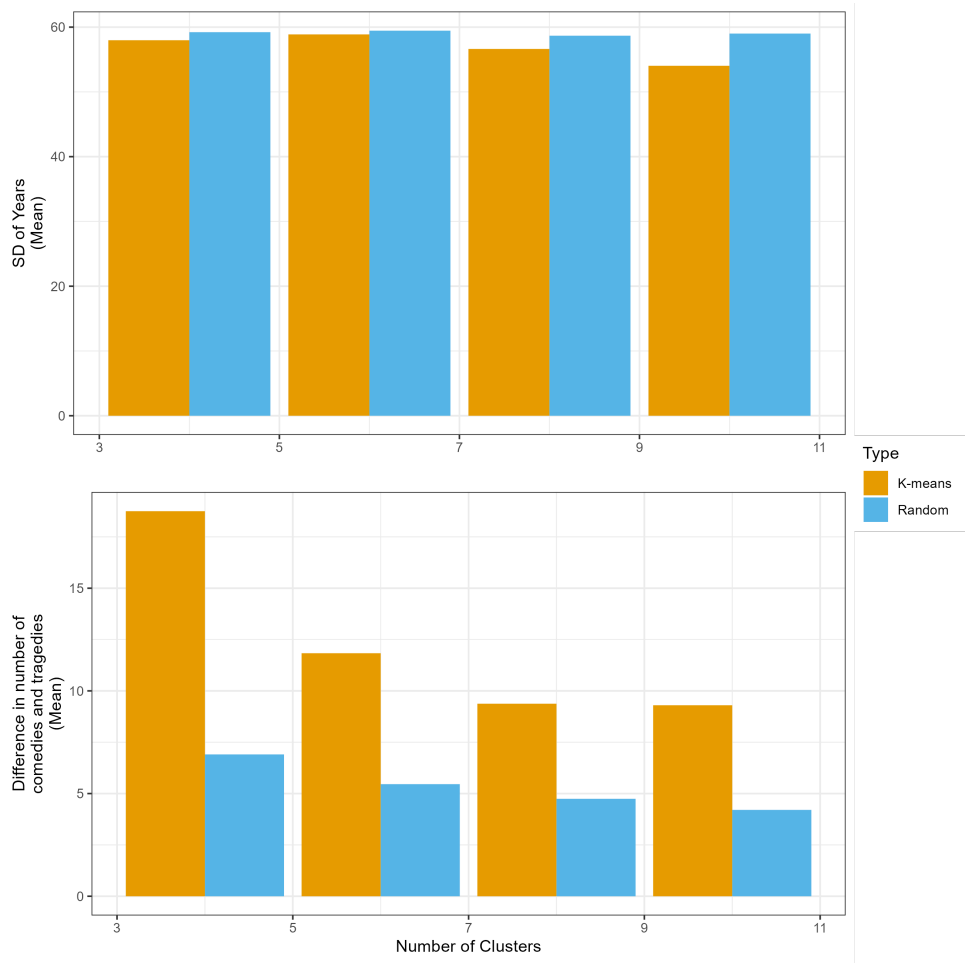


Figure 1: Differences between k-means and random clustering with respect to time of creation (top) and genre (bottom) – GerDraCor. [Charts: Szemes / Nagy 2025]



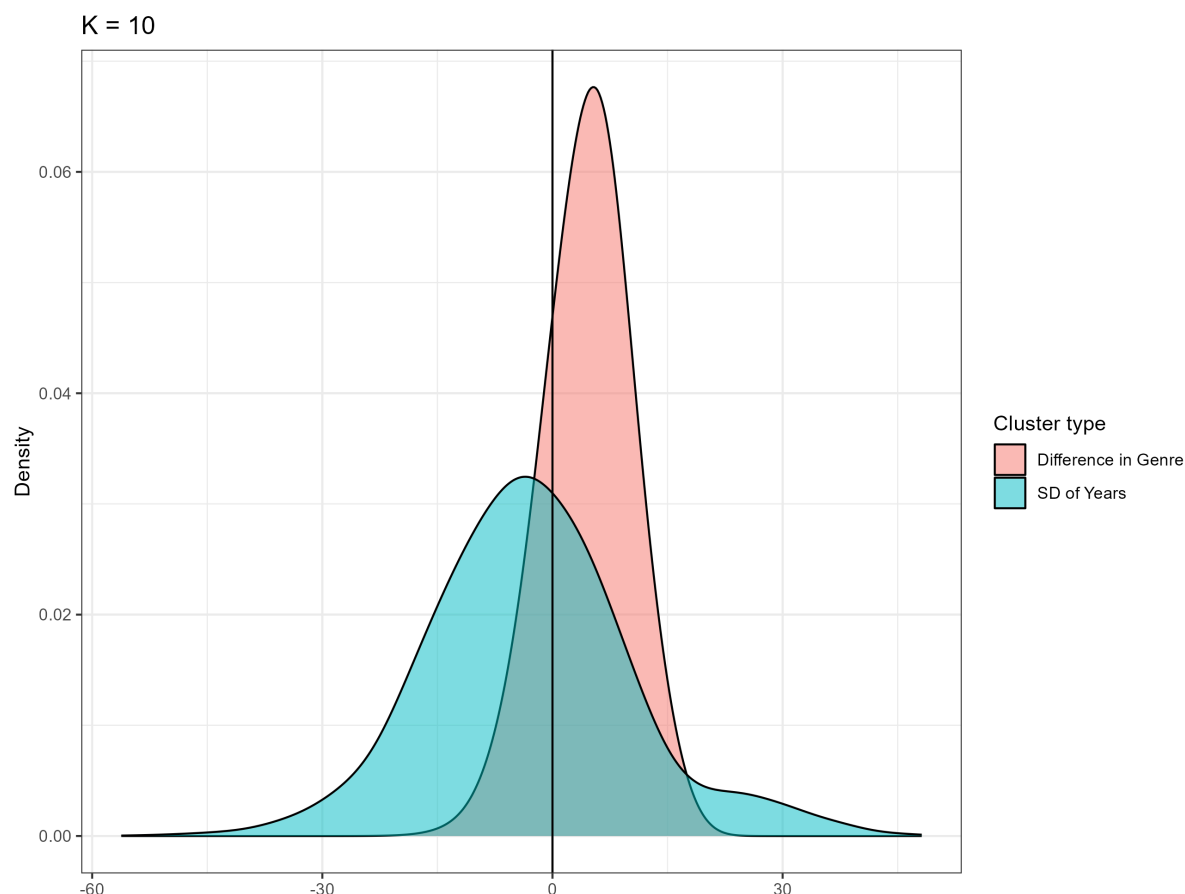


Figure 2: We first calculated the difference for each metric (standard deviation by year and genre) between the types of clusters (k-means or random); we then plotted the distribution of these differences. The more the peak of the curves approaches zero, the smaller the difference between the two types of clusterisation. For genre, the more it is in the positive range, the more the metric is characteristic of the k-average clustering. For time, the more it is in the negative range, the more the date of writing is an important factor. Finally, the ›sharper‹ the curves, the more the data are distributed around a single value. [Chart: Szemes / Nagy 2025]

Yet this is not the only aspect in which historicity can be relevant when grouping dramas. Perhaps, although the different periods do not play an important role in the separation of texts, the *genre distinction* itself is not necessarily constant across the history of drama. If it is true, the boundaries between the structural features of tragedies and comedies are not similarly sharp in each period. This leads us to our second question: is the link between genre and the structure of drama weakening over time, or is it being strengthened by cultural transmission? Ted Underwood expected the latter when he looked at the sub-genres of novels – but he found no evidence for it; on the contrary, he registered a slight decrease in genre differences.<sup>34</sup> Šeja et al. arrived at a similar result when they studied the relationship between the meter and theme of poems. According to their research, the relationship is stronger in earlier periods than over time: »Historical differences in classification accuracy also suggested semantic accumulation in metrical forms and a diffusion of meter's ›meaning‹ over time without swamping it beyond recognition.«<sup>35</sup> The authors explain this phenomenon within a framework of cultural evolution: initially, a strong link between form and meaning becomes established and widely followed; over time, as this association becomes part of the tradition, more variation and deviation are tolerated – without completely eroding the original connection. The gradual blurring of differences may suggest that the boundaries between categories have been firmly established through the process of cultural transmission, making it unnecessary for authors to reiterate these distinctions in every case. In the context

[17]

<sup>34</sup> Underwood 2019.

<sup>35</sup> Šeja et al. 2020.

of drama history, Mark Algee-Hewitt can be mentioned again. Although somewhat speculatively, he suggests that the trend toward a more equal distribution of character prominence in English drama networks may be linked to a gradual narrowing of genre distinctions: »Over time, as drama becomes more distributed in its protagonism, as plays lose their peripheries, and as they are subdivided into multiple competing networks of interaction, they become more structurally comedic. Again, as they leave the throne room (the space of history and tragedy) and enter the drawing room (the native space of comedy), even tragedies look morphologically more like comedies.«<sup>36</sup> Later, however, he makes other arguments to explain the historical processes and does not verify the claim.

The question was tested with a simple experiment. We divided the data set into two parts, for works written before and after the median of the year assigned to the dramas (1829). This date also correlates with the end of German Classicism, which is traditionally considered to have ended with Goethe's death (1832).<sup>37</sup> There are 69 comedies and 57 tragedies in the period before 1829, and 66 comedies and 58 tragedies after that (there are 3 plays from 1829). Then, the SVM classification used in our previous research (with a linear kernel<sup>38</sup>) was performed again also in a leave-one-out scenario, but this time on the two data sets separately. This allowed us to compare the performance of the classification in the two periods: we compared separately the extent to which the genres can be separated according to the feature set. The results are shown in Table 2 (and in Figures 13 and 14 in the Appendix). It shows that classification is more effective for dramas before 1829 than after. Figure 13 shows a greater degree of separation compared to Figure 14 or even Figure 11, as well as the earlier period showing a higher F1 score. In other words, in this case too, we can identify a process in literary history in which first a close link between thematic and formal-structural levels is established, and then deviations from this rule become more frequent, without the link being broken completely – see the F1 scores for the classification of dramas after 1829. Given that researchers in both poetry and prose have reached similar conclusions, it may be possible to view these findings as part of a broader mechanism in literary history.

With regard to the two experiments on the 253 dramas of GerDraCor, we can therefore say that, although the features describing the relationships between the characters do not provide information on the time of writing, the characteristics of the genre that can be determined on the basis of these features become less and less clear in the course of the drama's history, and to this extent historicity also affects the classification.

	Precision	Recall	F1
<b>1650–1829</b>			
<b>Comedy (n=69)</b>	0.8	0.81	0.81
<b>Tragedy (n=57)</b>	0.97	0.75	0.85
<b>1829–1929</b>			
<b>Comedy (n= 66)</b>	0.73	0.7	0.71
<b>Tragedy (n=58)</b>	0.77	0.71	0.73

Table 2: The performance of SVM classification relating to the two data sets (before and after 1829).

<sup>36</sup> Algee-Hewitt 2017, p. 771.  
<sup>37</sup> For more about this dramatic era, see Williams / Hamburger 2008.  
<sup>38</sup> Using *radial basis function (RFB)* leads to similar results in terms of the difference between the genres and periods – see Appendix, Table 4. For the basic statistical concepts see James et al. 2013, pp. 337–373.

## 4. Time as Plot

The measures used so far always describe dramas as a whole and therefore fail to capture the temporality of the plays themselves. This kind of temporality refers to the development of the relationships between the characters as the plot unfolds. It is therefore this development that we wish to trace in the following. This is also a continuation of our previous research, in which we investigated how the measures of character networks are influenced by the outcome of the plot of the dramas.<sup>39</sup> Here we start from the commonly held belief that, especially in the case of Shakespeare, the difference between the two genres can be traced back to whether a play ends with a wedding (comedy) or with the gradual death of the characters (tragedy).<sup>40</sup> In the former case, the plot ends with a scene of a crowd in which most of the characters meet, resulting in a denser network than in tragedies without such an ending. The research therefore investigates the effect of the last act in five-act plays on the metric of density.<sup>41</sup> [20]

We have calculated the value both for the whole drama and without the last act: if the difference between the two results is positive, the character network of the drama as a whole is denser than that of the first four acts – that is, the last act establishes new relationships between the characters already presented on stage. A negative result indicates the opposite: the emergence of new characters in the last act, in contact with only a small proportion of the already presented characters. The results are shown in Figures 3 and 4. [21]

In the Shakespeare corpus, the last act of the comedies indeed increases the density of the networks, while the tragedies decrease it (Figure 3). This is not so clear in the GerDraCor dramas, but for the comedies of both corpora, on average only the last act has a positive effect on density (Figure 4) – which can be explained by the fact that new characters rarely appear in the final act of the comedies. However, the picture is nuanced by the fact that the *Wilcoxon test*<sup>42</sup> shows that the difference in density between genres remains significant even without the last act. In other words, it can be said that the structural features of the plays are influenced by plot development, but not exclusively. From this point of view, the plot is only one component of the characteristic of the genre in terms of the relations between the characters. [22]

---

<sup>39</sup> Szemes / Vida 2024.

<sup>40</sup> In the field of computational drama analysis this is raised for example by Trilcke et al. 2015b. This is supported by the fact that a significantly higher proportion of the cast appears in the last scene of an average comedy than in that of a tragedy – see Fischer et al. 2017.

<sup>41</sup> We focused on five-act plays to ensure comparability – in terms of length – between works and their internal sections. While it would also be possible to rely solely on changes between scenes, comparing plays with vastly different structures – for example, those with 3 or 20 scenes – would have been difficult, even if their development could be expressed proportionally (e.g. by comparing x% of the plays). To avoid these complications, we chose a simpler and more transparent approach: analysing plays with the same number of acts. Additionally, since Shakespeare's plays are traditionally structured in five acts, we did not include other formats such as four- or six-act plays from the GerDraCor corpus as well. This resulted in a selection of 128 five-act plays with genre labels in that respect. Furthermore, in doing so, we align ourselves with an important forerunner of quantitative drama analysis, see Yarkho 2019.

<sup>42</sup> The Wilcoxon test is a statistical method used to compare two sets of data, especially when the data does not follow a normal distribution. It looks at the differences between paired observations (like measurements taken before and after a treatment) or compares two independent groups to see if they are significantly different from each other. For details see Corder / Foreman 2024, pp. 39–69.

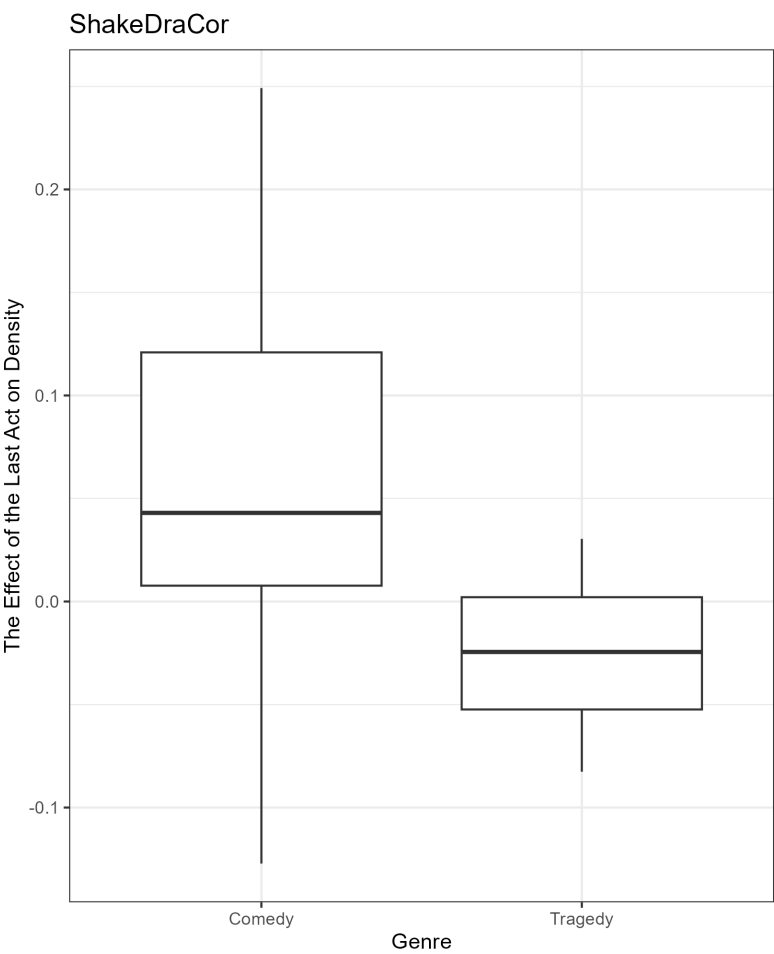


Figure 3: Effect of the last act on density – ShakeDraCor. [Chart: Szemes / Nagy 2025]

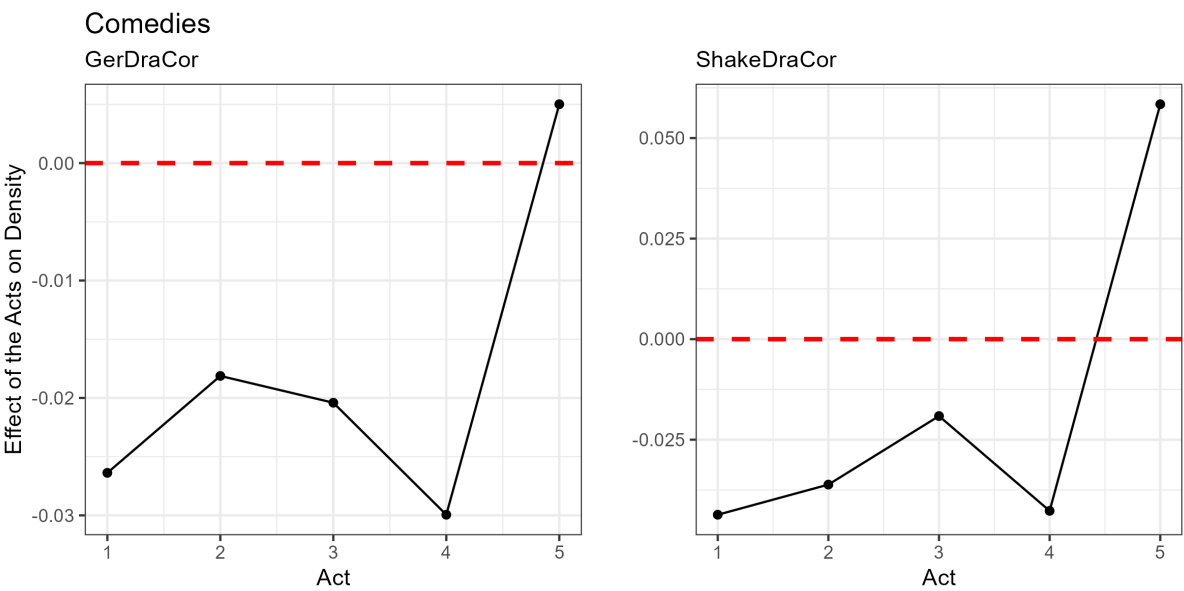


Figure 4: Effect of the acts on density in five-act comedies – GerDraCor (left), ShakeDraCor (right). [Charts: Szemes / Nagy 2025]

Another aspect, however, does not only capture the development of the plot from the point of view of the ending, but also makes it possible to examine it as a process unfolding over time. For this purpose, metrics of the character networks were calculated cumulatively. In other words, we first calculated the values for the first act, then for the first two acts together, and so on until the fifth act, describing the drama as a whole. In this way, the development of the final character networks can be plotted, enabling us to see the dramaturgical structure of the dramas in a new way. To do this, we have created interactive visualisations for density and average clustering coefficients that allow us to select which plays in the corpus we wish to compare with each other. In this case, the comparison is based primarily not on specific values, but rather on trends across the acts. See [Figures 5 and 6](#) or [the GitHub repository of the research](#).

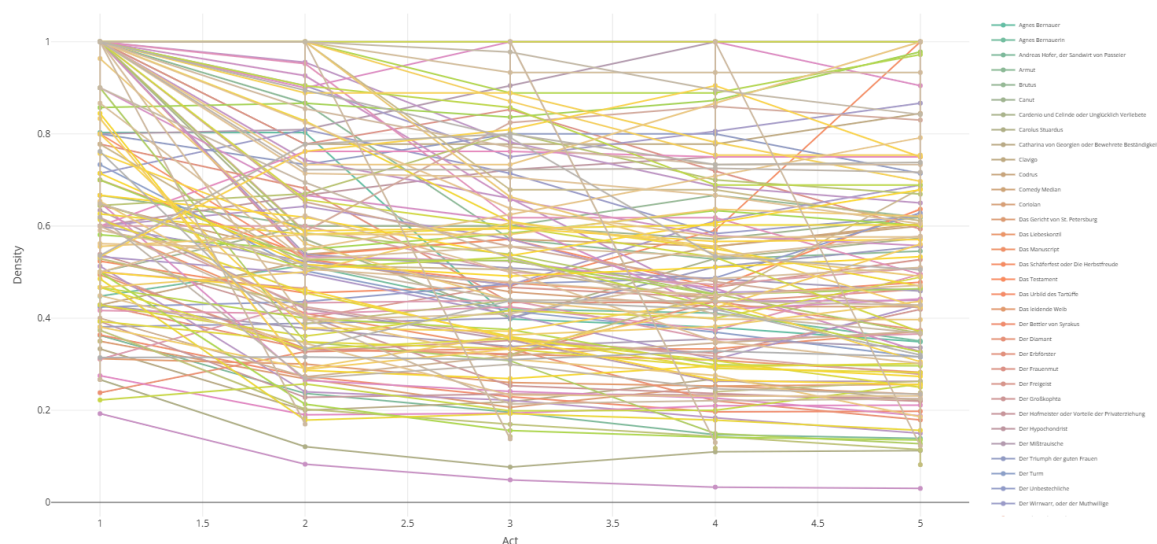


Figure 5: Plot of cumulative density in five-act plays – GerDraCor. Click [here](#) for the interactive visualisation. [Chart: Szemes / Nagy 2025]

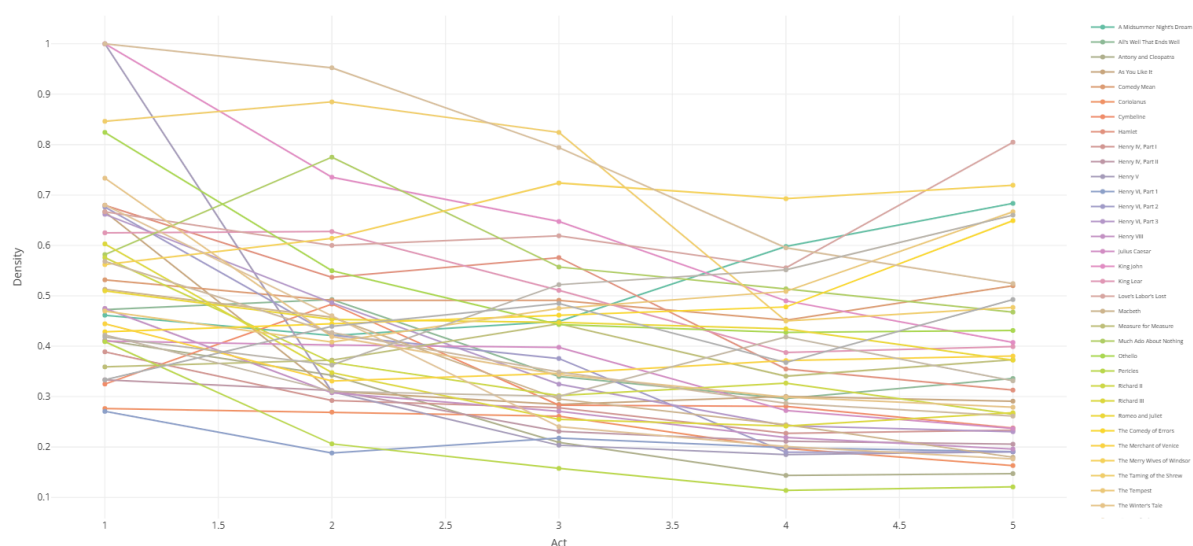


Figure 6: Plot of cumulative density in five-act plays – ShakeDraCor. Click [here](#) for the interactive visualisation. [Chart: Szemes / Nagy 2025]

At the same time, static graphs showing the median values of each genre can also be informative. Figure 7 shows the evolution of density in the Shakespeare and German collections, which indicates that it is not only the ending of the story that distinguishes the genres. Although the character network of comedies loses density after the first act in the same way as that of tragedies, it becomes increasingly dense thereafter, or at least the measure does not show much fluctuation (for prototypical comedies in the Shakespeare corpus, the graph clearly shows an upward trend – see e.g. *A Midsummer Night's Dream* in Figure 8.) The networks of tragedies are initially even denser than those of comedies but soon lose this coherence. In both cases, the changes begin, and a significant part of them take place before the last act. The plot of comedies can be characterised as a process from loosening relationships that create misunderstandings around the first act to the clarification of those at the end; since in order for a community to be able to bring contradictory information into conflict with each other, increasingly dense networks are required. In comedies, temporality thus refers to the slow and laborious process of building up a network of relationships in which (by the end of the drama) it is possible to verify and harmonise information. In contrast, the character network in tragedies becomes increasingly ›looser‹. It is noteworthy that the initial density is mostly loosened in the second and third acts; that is, these networks ›fall apart‹ before the ending of the plot – one could say that *the tragic ending is the consequence of this loss of density*.

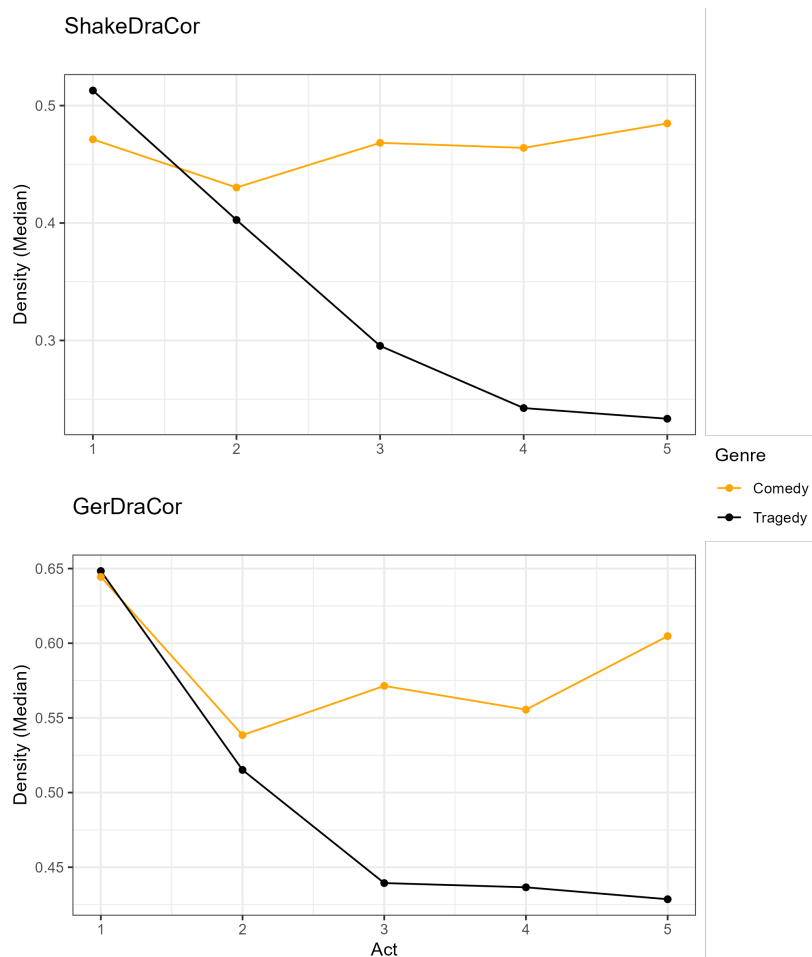


Figure 7: Changes in the median of density in five-act plays – ShakeDraCor (top), GerDraCor (bottom). [Charts: Szemes / Nagy 2025]

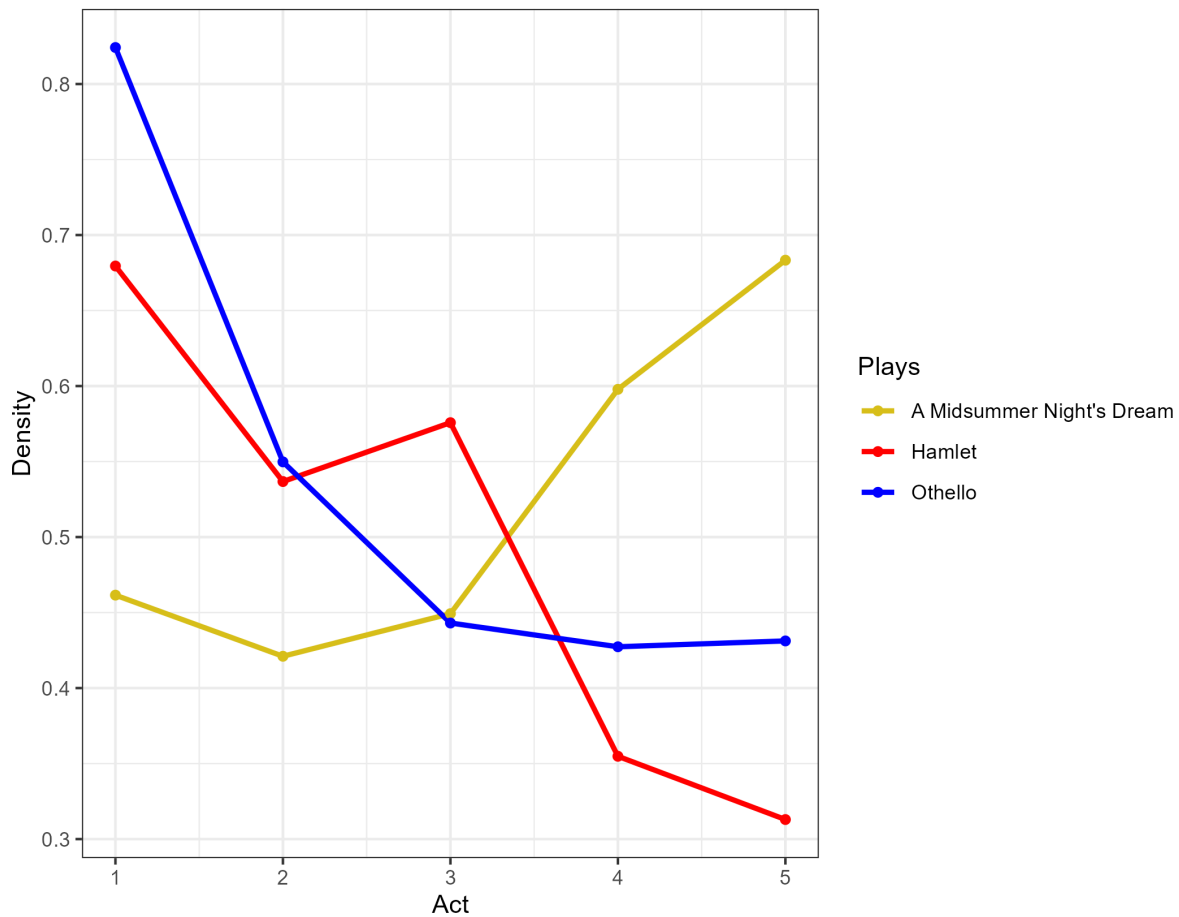


Figure 8: Changes in density in three Shakespeare plays. [Chart: Szemes / Nagy 2025]

This is supported by the development of the plot of *Othello* in Figure 8. Here a very cohesive community is disintegrating at the beginning of the drama, and although the network of relationships does not lose much of its density later on, this disintegration is sufficient to cause the tragic events to occur.<sup>43</sup> The turning point is clearly the move of Othello and his companions to Cyprus. In Venice, the society run by the Duke is still characterised by dense relationships (similar to that of the comedies at the end of the plot), in which, even though Iago tries to manipulate and misinform the others, in the public space each piece of information is tested against each other. This is what happens to Brabantino, who takes his grievances to the Duke; but the Duke also listens to the accused instead of serving justice, thus creating the possibility that the truth reveals itself among the participants (see when he points out that a statement is not in itself evidence, and will only be a vague suggestion until it is supported by external facts: »To vouch this is no proof / Without more wider and more overt test / Than these thin habits and poor likelihoods / Of modern seeming do prefer against him.« I / 3). The way the Duke maintains the flow of information is best illustrated in the opening of Act I, Scene 3, in which he evaluates and compares the news of the approaching Turkish fleet. This section amounts to a summary of the essence of such organisation of community and state:

[25]

<sup>43</sup> This remains important even though the reduced values in later acts still indicate relatively high density compared to other dramas, particularly tragedies. One of *Othello's* defining characteristics is that it situates a tragic narrative within a comedic network (cf. Szemes / Vida 2024). From this perspective, it becomes even more apparent that the tragedy unfolds as a consequence of the network's disintegration after the first act. However, even when excluding the first act (and its characters), the drama continues to exhibit a decreasing trend in density scores.

»DUKE (*reading a paper*)

There's no composition in these news  
That gives them credit.

FIRST SENATOR (*reading a paper*)

Indeed, they are disproportioned.  
My letters say a hundred and seven galleys.

DUKE

And mine, a hundred forty.

SECOND SENATOR (*reading a paper*)

And mine, two hundred.  
But though they jump not on a just account  
(As in these cases, where the aim reports  
'Tis oft with difference), yet do they all confirm  
A Turkish fleet, and bearing up to Cyprus.

DUKE

Nay, it is possible enough to judgment.«

[26]

Othello's flaw is precisely that he cannot create a similar network when he arrives in Cyprus – he acts as a soldier and not as a statesman: as soon as Iago informs him of an event, he makes instant decisions that only later turn out to be wrong. Following a Hungarian Shakespeare-scholar, Géza Kállay's analysis, we can describe this flaw in terms of private life and marriage, but also using the metaphor of the network: while in comedies the private lives of two people are always part of a larger network, Othello fails in the impossible enterprise of making his love for Desdemona exclusive, an enterprise to which Iago contrasts external perspectives.<sup>44</sup>

[27]

More generally, the following can be said about genres and the worldview they convey according to density. In the networks of comedies, which at first become sparse and then increasingly dense, individual opinions circulate in parallel, independent of each other – but by gradually linking them, it is finally possible to evaluate their truthfulness. In tragedies, the flow of information is being controlled more and more by one or two characters in connecting position. These two models could be related to Jürgen Habermas' theory of the public sphere of civil society and communicative action.<sup>45</sup> Accordingly, comedies would represent a similar community as the public sphere of civil society, organised not by a central authority but by community-controlled and community-accepted positions based on the differences of opinion of individuals. Tragedies demonstrate the possibility of such manipulation of the public – which, according to Habermas, is characteristic of the age of mass media. What is different, however, is that in Shakespeare's comedies, the publicity that allows the clash of opinions is always linked to the state power: as in *Othello*, the Duke, so in the comedies a state leader gathers the characters together to resolve complications. An even more important difference is that comedies do not heroise this process – it is precisely by emphasising the *processual nature* of the construction of networks and the pettiness resulting from misunderstandings and mistakes; that is, the steps through which we reach the final network.<sup>46</sup> From this point of view, the ending of *A Midsummer Night's*

[28]

<sup>44</sup> Cf. Kállay 1996.

<sup>45</sup> Habermas 1989; Habermas 1984. This is not to say that Shakespearean comedies are the historical precursors of civil society, but that they show similar ways of organising communities. In other words, this is not a historical argument, it only points to the structural similarity of the described communities.

<sup>46</sup> This allows to realise the banal nature of the public sphere, rather than emphasising the individual efforts of the participants. It is in connection with the criticism of Habermas' theory that it is based solely on human intention and does not take into account the self-referential functioning of the systems underlying the public sphere, especially the media (cf. Luhmann 2000). From this



*Dream* is particularly frustrating, where the misunderstandings are only seemingly clarified: at the end of the drama, Demetrius is still under the influence of the magic potion given to him by Puck. This suggests that a world built on shared, controlled knowledge simultaneously normalises and homogenises individual desires and thought processes – highlighting a potential downside of the public sphere.

In addition to median values and curves for dramas considered prototypical, dramaturgies differing from the generic schema can also be identified by plotting metrics cumulatively. In Figure 8 we can observe the uniqueness of *Hamlet's* curve: the continuous disintegration of the network is interrupted in Act III, when the relationships between the characters become surprisingly denser. The reason for this is the Mousetrap scene in the act, the performance directed by Hamlet, in which the main characters of the court participate (Hamlet and Horatio are on stage, then comes the King, the Queen, Polonius, Ophelia, Rosencrantz, Guildenstern, other lords and later the actors). In other words, the characters are gathered in the middle of the drama – and, unlike in other dramas in which this usually occurs at the end of the plot, the truth is revealed, Claudius is confronted with his sin, and Hamlet is convinced of its validity. The tragic ending occurs in this knowledge; the turning point and realisation expected at the end of the drama takes place earlier. This is supported by the fact that the play-within-a-play alludes not only to Claudius's past fratricide but also to Hamlet's future revenge, insofar as he refers to the murderer in the play not as »brother« but as »nephew«, which mirrors Claudius's relationship to Hamlet.<sup>47</sup> However, at the end of the play, when Claudius is murdered, Hamlet does not mention his father's death, but only the fatal wound he received and the poisoning of his mother as the reason for his revenge. The motives for the actions seem to be limited to the present, ignoring the investigation of the whole play. This emphasises the dramaturgical shift: in the third act, the »investigator« becomes certain of the identity of the criminal, yet the story continues and ends in a rather accidental conclusion. The drama's unique dramaturgy is due to the fact that in the middle of the plot there is a major scene in which the most important questions are answered – but only to the main characters, since the others cannot interpret the performance in the same way. This means that the knowledge of the protagonists is not made public, allowing the tragic plot to develop further, which also means a further reduction in the density of the character network.

[29]

Figure 9 illustrates the change in the average clustering coefficient in the same cumulative way for the Shakespeare and GerDraCor sub-corpora. The clustering coefficient is a measure of the proportion of the neighbours of a node (the nodes with which it is connected) that are also connected with each other; the average of the measures for each node describes the clustering of the network as a whole. Several conclusions can be drawn from the figure. First, the high values for both genres support the interpretation of dramas as small worlds.<sup>48</sup> On the other hand, although comedies are characterised by an increasingly dense network, as the plot progresses, these networks become somewhat less clustered than in tragedies, which means that two characters are more often only connected through the mediation of a third. In the case of the tragedies, we again see the disintegration of a dense network: while initially they show higher values than the comedies, more and more characters come into contact with one another without meeting their neighbours. However, this interpretation is weakened by the fact that the difference between the two genres is not statistically significant for the acts (Wilcoxon test for the first act  $p = 0.25$ ; for the second act  $p = 0.31$ ). Thirdly, and perhaps most importantly, the two curves, especially in the case of GerDraCor, are converging; that is, the otherwise already insignificant difference between tragedies and comedies is completely eliminated by the end of the plot in terms of the average clustering coefficient. This is also supported by the gradual decrease in the variance of the measure in the acts for the corpus as a whole. In the case of GerDraCor, there is a significant difference between all the acts according to the Wilcoxon test, whereas in the Shakespeare collection, the difference is significant only between the beginning (first two acts) and the end (fourth to fifth) – see Figure 10. It is as if, regardless of genre, there is an arrangement of the characters towards which each

[30]

point of view, each genre is similar to such a system, whose aim is to enable the public acceptance of a communally accepted truth, while at the same time delaying this acceptance (in the case of comedies, by the simultaneous spread of parallel opinions, in the case of tragedies, by the manipulation of the flow of information).

<sup>47</sup> Cf. Calderwood 1983.

<sup>48</sup> For small worlds in dramatic networks, see Trilcke et al. 2024; Stiller / Hudson 2005; Stiller et al. 2003.

work ›strives‹. In such an arrangement, several routes can be taken simultaneously between nodes, yet the direction of travel cannot be chosen arbitrarily. This allows the characters to occupy similar, yet different positions in the information flow; and, by the end of the plot, to have a *shared* recognition through the conflict of information, which relates to precisely such a *difference* of positions. In other words, this arrangement allows the characters to affect each other, but not in the same way. In more clustered networks, there is less opportunity for the creation of individual positions, as well as less chance of misunderstanding; in less clustered networks, the impact of the main characters on the network is reduced, and it is also more difficult to make public the differences in opinion. Tragedies progressively reach this point (i.e. the difference between the positions in the network becomes greater), while comedies introduce new characters on average in Act II who meet only part of a group and who only then come into contact with the other characters.

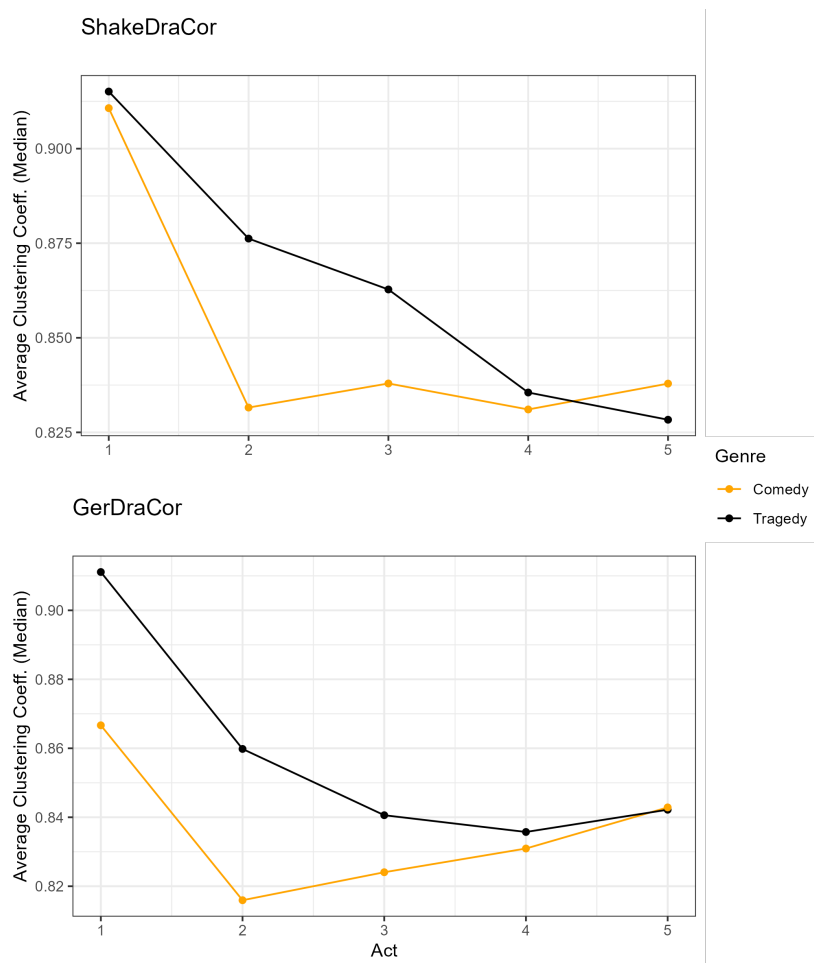


Figure 9: Changes in the median of average clustering coefficient in five-act plays – ShakeDraCor (top), GerDraCor (bottom).  
[Charts: Szemes / Nagy 2025]

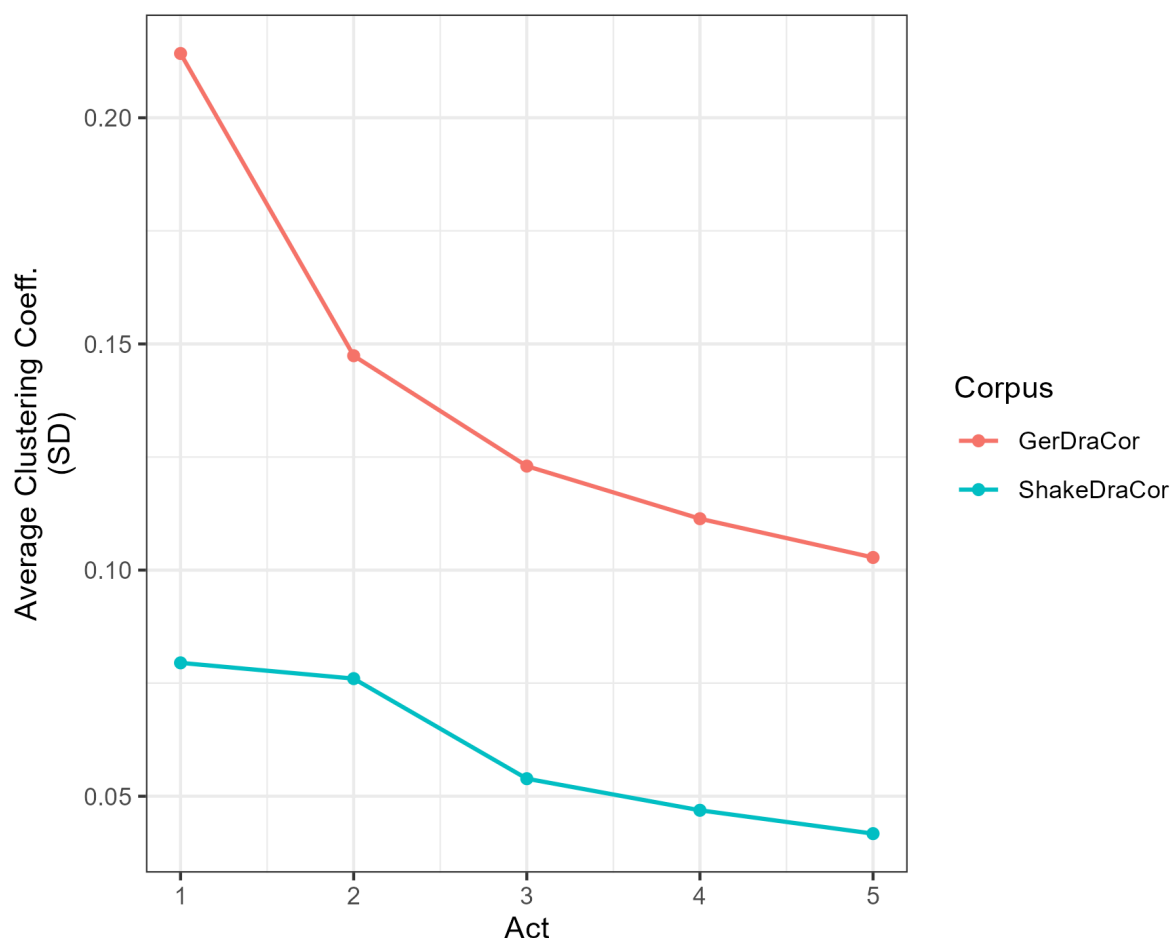


Figure 10: Changes in the standard deviation of average clustering coefficient in five-act plays. [Chart: Szemes / Nagy 2025]

## 5. Conclusion

This paper aimed to contribute to the study of structural differences between comedies and tragedies. Such investigations are valuable as they reveal how fundamental storytelling modes correspond to patterns of character relationships and models of community organization. However, these approaches often risk becoming ahistorical, focusing solely on structural features while overlooking literary history. Additionally, the metrics used to describe these structures are themselves the outcome of a dynamic process – the development of a plot. Computational studies often ignore this component in their comparison as well. To address these limitations, this research explored two key aspects: the influence of a play's time of creation on genre clustering and the processual nature of genre-specific character network formation.

[31]

First, it has become apparent that the time of creation of dramas does not have a great influence on their structural features, but it does influence the extent of structural differences between genres. These findings support both sides of the structuralism vs. historicism debate. In doing so, they highlight the potential of computational methods to integrate formal, textual, and historical evidence, facilitating a dialogue between the different perspectives.<sup>49</sup> On one hand, the enduring stability of genre differences across centuries has become evident, as well as the fact that the time of creation has little effect on classification. Comedies and tragedies are long-lasting categories at least in German and English drama history. On the other hand, it seems that the link between structure and genre is stronger in the earlier stages of drama history, while later

[32]

<sup>49</sup> Cf. Underwood / The NovelTM Research Group 2016.

on more and more plays deviate from this pattern, which does not lose its relevance. This can be explained by the fact that the relationship between form and content must be reinforced by most of the participants at early stages of imitation, in order to establish a stable rule that no longer needs to be re-validated every time. Our results are consistent with the findings of research on other genres,<sup>50</sup> suggesting a more general phenomenon in literary history.

Secondly, it can be also illuminating to trace the evolution of network metrics within a drama. Here we can also identify genre-specific trends that may help to nuance our previous insights: the increasingly dense character networks of comedies and the increasingly ›loose‹ networks of tragedies inform us that the development of the relationships between the characters begins before the plot reaches its conclusion, while the tragic / comic ending tends to amplify these processes. From this point of view, comedies and tragedies are not only defined by a specific type of community, where the density and distribution of interactions vary, but also by the process through which these communities are formed. In fact, it is often the transformation of character networks that drives the development of comic or tragic narratives. Conversely, in the case of the average clustering coefficient, it is rather the similarity between the two genres that is in the foreground: the two kinds of work take different paths but arrive at a similar arrangement of characters, an arrangement that can be described as a ›dramaturgical zero point‹.

[33]

Finally, in addition to the trends in genres, we can also compare and explore the individual specificities of each play through its representation. More detailed interpretations based on these aspects, though, are still a task for future research. We hope that we have offered useful perspectives for such interpretations, and that we can thus contribute to the development of the results that build on each other – which is also the basis for the temporality of scholarship.

[34]

## Data Availability

Code and data are available in the GitHub repository of the research: [https://github.com/SzemesBotond/drama\\_cluster\\_genre](https://github.com/SzemesBotond/drama_cluster_genre)

[35]

---

<sup>50</sup> Šeja et al. 2020; Underwood 2019.

# Appendix

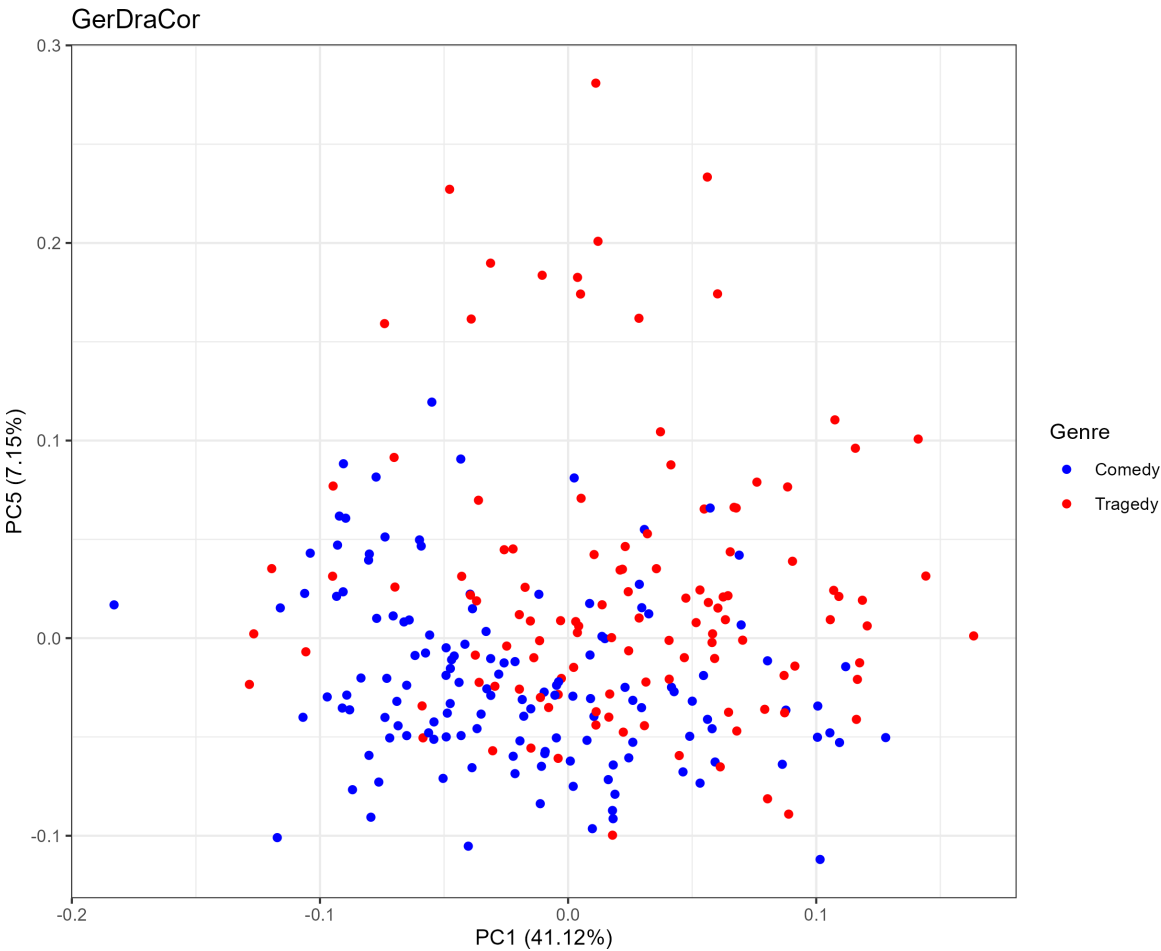


Figure 11: The location of GerDraCor's dramas on the PCA graph based on structural features. [Chart: Szemes / Nagy 2025]

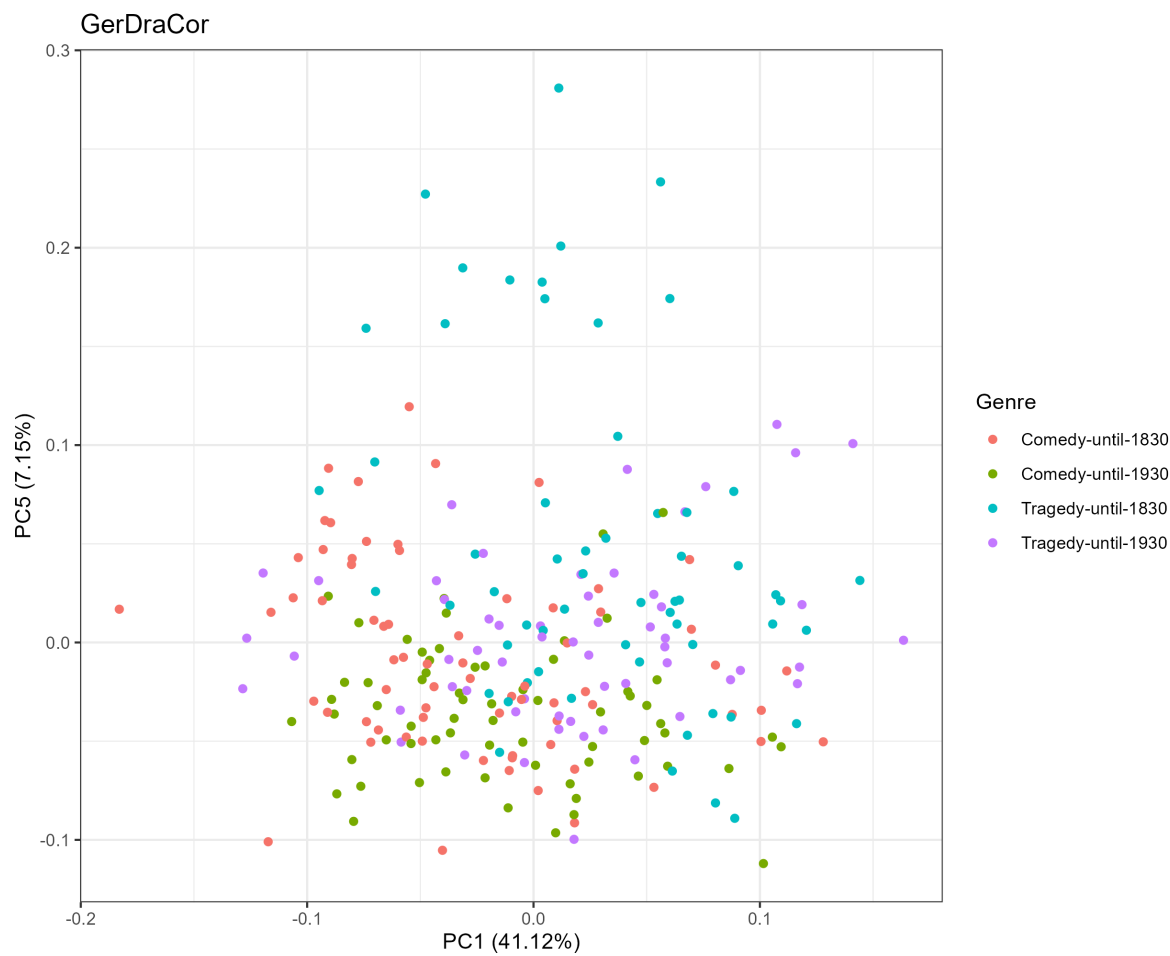


Figure 12: The location of GerDraCor's dramas on the PCA graph based on structural features with a separate colour for comedies and tragedies written before and after 1830. [Chart: Szemes / Nagy 2025]

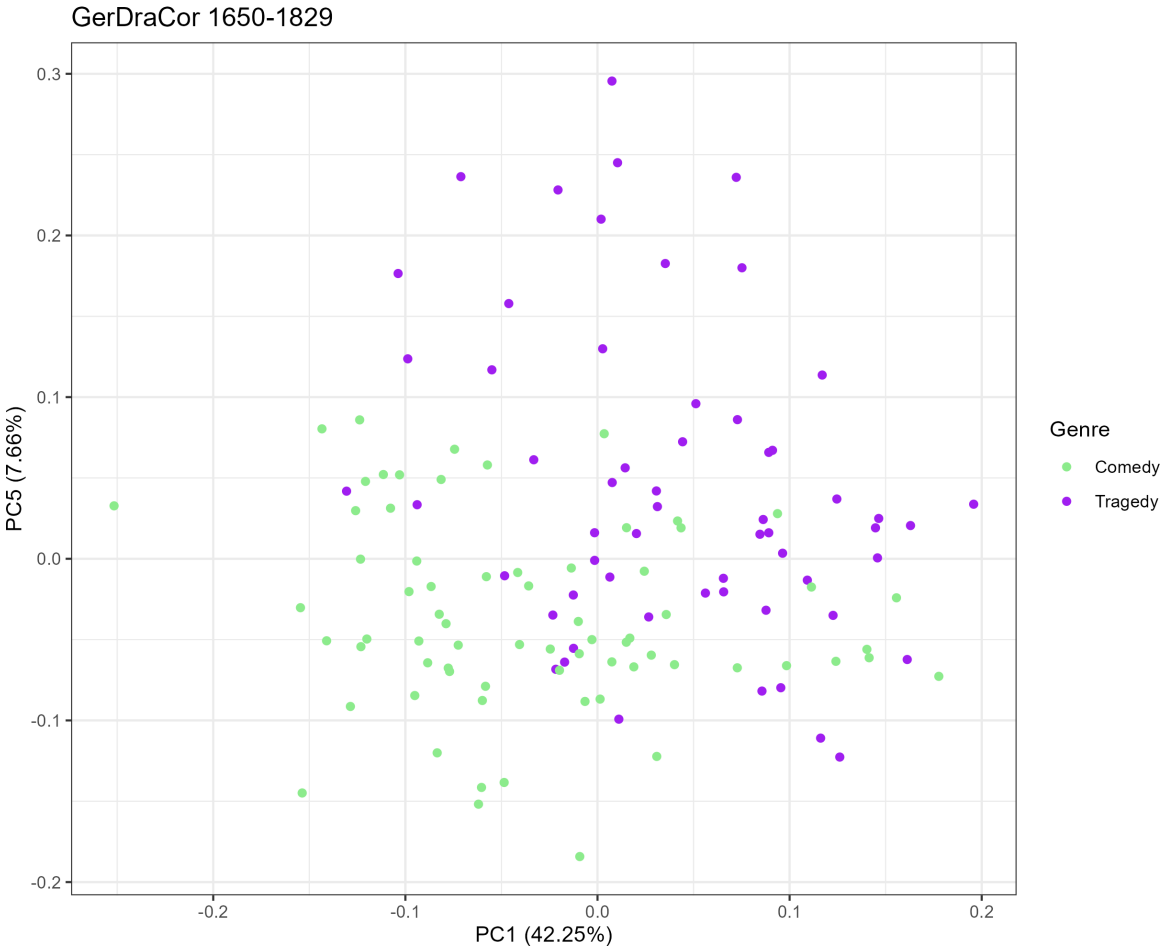


Figure 13: The location of GerDraCor's dramas on the PCA graph based on structural features (1650–1829). [Chart: Szemes / Nagy 2025]

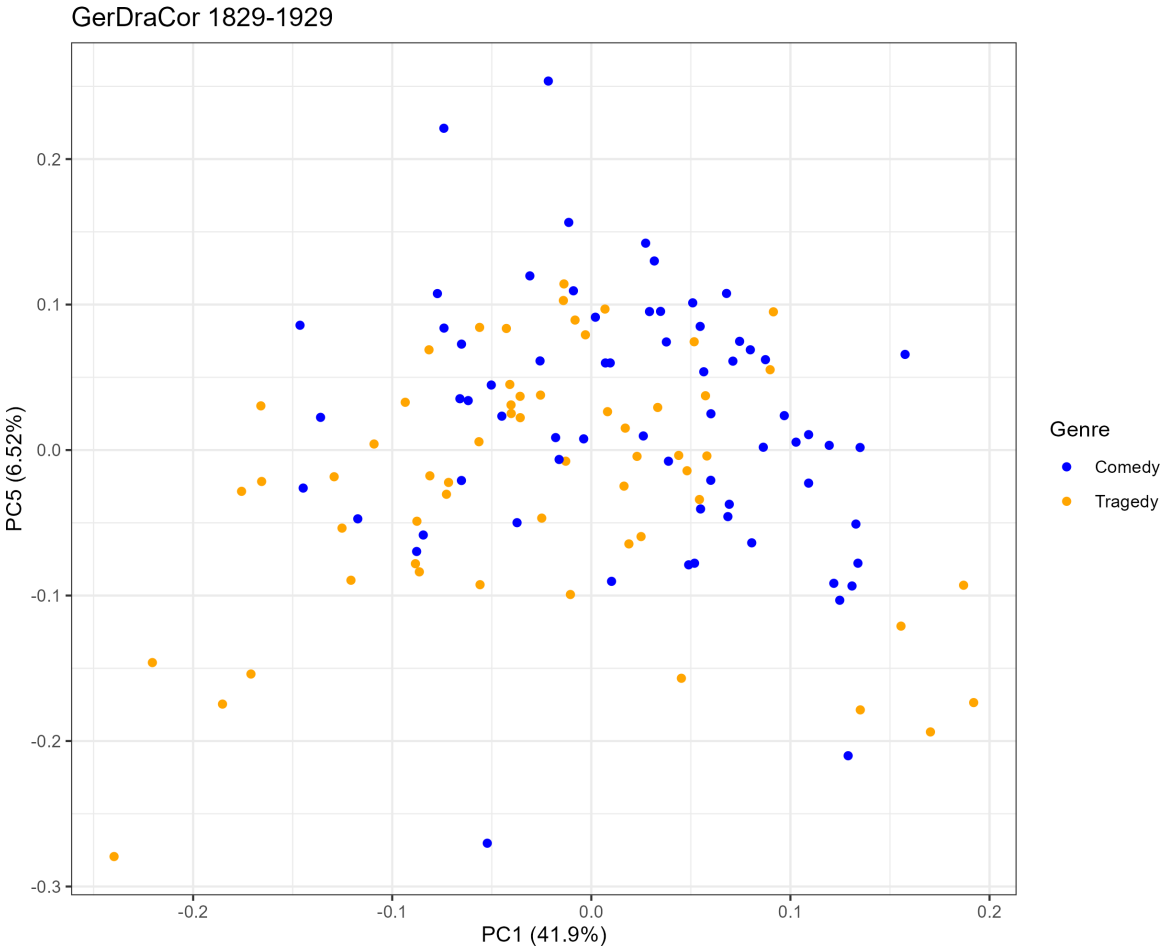


Figure 14: The location of GerDraCor's dramas on the PCA graph based on structural features (1829–1929). [Chart: Szemes / Nagy 2025]



Author	Title	Genre	Year
Johann Nestroy	Zu ebener Erde und erster Stock oder Die Launen des Glücks	Comedy	1835
Johann Nestroy	Der böse Geist Lumpazivagabundus oder Das liederliche Kleeblatt	Comedy	1833
Johann Nestroy	Der Talisman	Comedy	1843
August Klingemann	Faust	Tragedy	1812
Ferdinand Raimund	Der Barometermacher auf der Zauberinsel	Comedy	1823
Heinrich von Kleist	Penthesilea	Comedy	1808
Eduard von Bauernfeld	Industrie und Herz	Comedy	1842
Johann Wolfgang Goethe	Der Großkophta	Comedy	1791
Karl Johann Braun von Braunschweig	Faust	Tragedy	1835
Friedrich Schiller	Die Jungfrau von Orleans	Tragedy	1801
Julius von Voss	Faust	Tragedy	1823
Karl Immermann	Andreas Hofer, der Sandwirt von Passeier	Tragedy	1835
Jakob Michael Reinhold Lenz	Der neue Menoza oder Geschichte des cumbanischen Prinzen Tandil	Comedy	1774
Clemens Brentano	Ponce de Leon	Comedy	1803
Friedrich Schiller	Die Verschwörung des Fiesco zu Genua	Tragedy	1782
Hermann Sudermann	Der Bettler von Syrakus	Tragedy	1911
Karl Immermann	Das Gericht von St. Petersburg	Tragedy	1832
Ludwig Tieck	Leben und Tod der heiligen Genoveva	Tragedy	1800
Jens Baggesen	Der vollendete Faust oder Romanen in Jauer	Comedy	1808

Table 3: The group created by the k-average clustering procedure that contains the most dramas from one period (k = 10).

	Precision	Recall	F1
<b>1650–1829</b>			
<b>Comedy</b>	0.8	0.84	0.76
<b>Tragedy</b>	0.72	0.68	0.78
<b>1829–1929</b>			
<b>Comedy</b>	0.75	0.85	0.67
<b>Tragedy</b>	0.63	0.53	0.67

Table 4: The performance of SVM classification with RFB kernel (before and after 1829).

## Bibliography

- Mark Algee-Hewitt: Distributed Character: Quantitative Models of the English Stage, 1550–1900. In: *New Literary History* 48 (2017), no. 4, pp. 751–782. DOI: 10.1353/nlh.2017.0038
- Aristotle: *Poetics*. Ed. by Leonardo Tarán / Dimitri Gutas. Leiden etc. 2012. [[Nachweis im GVK](#)]
- Bernhard Asmuth: *Einführung in die Dramenanalyse* (= Sammlung Metzler, 188). 8th edition. Berlin etc. 2016. DOI: 10.1007/978-3-476-05472-2
- James L. Calderwood: *To Be and Not To Be: Negation and Metadrama in Hamlet*. New York 1983. DOI: 10.7312/cald94400
- Gregory W. Corder / Dale I. Foreman: *Nonparametric Statistics. A Step-by-Step Approach*. 2nd edition. New York 2014. [[Nachweis im GVK](#)]
- Alison Findlay: Epilogues and Last Words in Shakespeare: Exploring Patterns in a Small Corpus. In: *Language and Literature: International Journal of Stylistics* 29 (2020), no. 3, pp. 327–346. 22.08.2022. DOI: 10.1177/0963947020949442
- Frank Fischer / Mathias Göbel / Dario Kampkaspar / Christopher Kittel / Peer Trilcke: Network Dynamics, Plot Analysis. Approaching the Progressive Structuration of Literary Texts. In: Rhian Lewis / Cecily Raynor / Dominic Forest / Michael Sinatra / Stéfan Sinclair (eds.): *Digital Humanities 2017. Conference Abstracts* (DH2017, Montréal, 08.–11.08.2017). PDF. [[online](#)]
- Frank Fischer / Ingo Börner / Mathias Göbel / Angelika Hecht / Christopher Kittel / Carsten Milling / Peer Trilcke: Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama. *Digital Humanities 2019: Complexities* (DH2019, Utrecht, 09.–12.07.2019). DOI: [10.5281/zenodo.4284001](#)
- Jürgen Habermas: *The Theory of Communicative Action*. Vol. 1: Reason and the Rationalization of Society. Boston 1984. [[Nachweis im GVK](#)]
- Jürgen Habermas: *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Cambridge, US-MA 1989. [[Nachweis im GVK](#)]
- Jonathan Hope / Michael Witmore: The Hundredth Psalm to the Tune of »Green Sleeves«: Digital Approaches to the Language of Genre. In: *Shakespeare Quarterly* 61 (2010), no. 3, pp. 357–390. DOI: 10.1353/shq.2010.0002
- Jonathan Hope / Michael Witmore: Quantification and the Language of Later Shakespeare. In: Christophe Hausermann / Dominique Goy-Blanquet (eds.): *La langue de Shakespeare. Actes des congrès de la Société française Shakespeare* 31 (Paris, 21.–23.03.2013). Paris 2014, pp. 123–149. DOI: [10.4000/shakespeare.2830](#)
- Gareth James / Daniela Witten / Trevor Hastie / Robert Tibshirani: *An Introduction to Statistical Learning with Applications in R*. New York etc. 2013. [[Nachweis im GVK](#)]
- Géza Kállay: *Nem pusztá szó. Shakespeare Othelloja nyelvfilozófiai megközelítésben*. Budapest 1996. [[Nachweis im GVK](#)]
- Vincent Labatut / Xavier Bost: Extraction and Analysis of Fictional Character Networks: A Survey. In: *ACM Computing Surveys* 52 (2019), no. 5. DOI: 10.1145/3344548
- Niklas Luhmann: *The Reality of Mass Media*. Stanford 2000. [[Nachweis im GVK](#)]
- Franco Moretti: *Network Theory, Plot Analysis* (= Literary Lab Pamphlets, 2). Stanford 2011. PDF. [[online](#)]
- Pablo Moscato / Hugh Craig / Gabriel Egan / Mohammad Nazmul Haque / Kevin Huang / Julia Sloan / Jonathon Corrales de Oliveira: Multiple Regression Techniques for Modelling Dates of First Performances of Shakespeare-Era Plays. In: *Expert Systems with Applications* 200 (2022). DOI: 10.1016/j.eswa.2022.116903
- Eric T. Nalisnick / Henry S. Baird: Character-to-Character Sentiment Analysis in Shakespeare's Plays. In: Hinrich Schuetze / Pascale Fung / Massimo Poesio (eds.): *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Volume 2: Short Papers* (ACL 2013, Sofia, 04.–09.08.2013). Stroudsburg, US-PA 2013, pp. 479–483. [[online](#)]
- Nils Reiter / Marcus Willand: What Are They Talking About? A Systematic Exploration of Theme Identification Methods for Character Speech in Dramatic Texts. In: Fotis Jannidis (ed.): *Digitale Literaturwissenschaft. DFG-Symposium* 2017. Stuttgart 2022, pp. 473–509. DOI: [10.1007/978-3-476-05886-7\\_20](#)
- Christoph Schöch: Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama. In: *Digital Humanities Quarterly* 11 (2017), no. 2. [[online](#)]
- Artjoms Šeļa / Boris Orekhov / Roman Leibov: Weak Genres: Modeling Association Between Poetic Meter and Meaning in Russian Poetry. In: *CHR 2020: Workshop on Computational Humanities Research* (= CEUR Workshop Proceedings, 2723; Amsterdam, 18.–20.11.2020). Aachen 2020. PDF. [[online](#)]
- Oleg Sobchuk: Evolution of Modern Literature and Film. In: Jamshid Johari Tehrani / Jeremy Kendal / Rachel Kendal (eds.): *The Oxford Handbook of Cultural Evolution*. Oxford 2023. DOI: 10.1093/oxfordhb/9780198869252.013.45
- James Stiller / Matthew Hudson: Weak Links and Scene Cliques Within the Small World of Shakespeare. In: *Journal of Cultural and Evolutionary Psychology* 3 (2005), no. 1, pp. 57–73. DOI: 10.1556/jcep.3.2005.1.4
- James Stiller / Daniel Nettle / Robin Ian MacDonald Dunbar: The Small World of Shakespeare's Plays. In: *Human Nature* 14 (2003), no. 4, pp. 397–408. DOI: 10.1007/s12110-003-1013-1
- Botond Szemes / Bence Vida: Tragic and Comical Networks: Clustering Dramatic Genres According to Structural Properties. In: Melanie Andresen / Nils Reiter (eds.): *Computational Drama Analysis*. Berlin etc. 2024, pp. 167–188. DOI: [10.1515/9783111071824-009](#)
- Peer Trilcke: Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft. In: Philip Ajouri / Katja Mellmann / Christoph Rauen (eds.): *Empirie in der Literaturwissenschaft* (= Poetogenesis, 8). Leiden 2013, pp. 201–247. DOI: 10.30965/9783957439710\_012
- Peer Trilcke / Frank Fischer / Mathias Göbel / Dario Kampkaspar (2015a): 200 Years of Literary Network Data. In: *Network Analysis of Dramatic Texts*. 25.06.2015. HTML. [[online](#)]
- Peer Trilcke / Frank Fischer / Mathias Göbel / Dario Kampkaspar (2015b): Comedy vs. Tragedy: Network Values by Genre. In: *Network Analysis of Dramatic Texts*. 31.07.2015. HTML. [[online](#)]
- Peer Trilcke / Evgeniya Ustinova / Ingo Börner / Frank Fischer / Carsten Milling: Detecting Small Worlds in a Corpus of Thousands of Theater Plays. A DraCor Study in Comparative Literary Network Analysis. In: Melanie Andresen / Nils Reiter (eds.): *Computational Drama Analysis*. Berlin etc. 2024, pp. 7–33. DOI: [10.1515/9783111071824-002](#)
- Ted Underwood / The NovelTM Research Group: Genre Theory and Historicism. In: *Journal of Cultural Analytics* 2 (2017), no. 2. 25.10.2016. DOI: [10.22148/16.008](#)

Ted Underwood: The Life Spans of Genres. In: Ted Underwood: Distant Horizons. Digital Evidence and Literary Change. Chicago 2019, pp. 34–68. [\[Nachweis im GVK\]](#)

Simon Williams / Maik Hamburger (eds.): A History of German Theatre. Cambridge, UK etc. 2008. [\[Nachweis im GVK\]](#)

Boris Isaakovich Yarkho: Speech Distribution in Five-Act Tragedies (A Question of Classicism and Romanticism). In: Journal of Literary Theory 19 (2019), no. 3. DOI: 10.1515/jlt-2019-0002

## List of Figures and Tables

- Table 1: Accuracy score for the LOO SVM classification by genre.
- Figure 1: Differences between k-means and random clustering with respect to time of creation (top) and genre (bottom) – GerDraCor. [Charts: Szemes / Nagy 2025]
- Figure 2: We first calculated the difference for each metric (standard deviation by year and genre) between the types of clusters (k-means or random); we then plotted the distribution of these differences. The more the peak of the curves approaches zero, the smaller the difference between the two types of clusterisation. For genre, the more it is in the positive range, the more the metric is characteristic of the k-average clustering. For time, the more it is in the negative range, the more the date of writing is an important factor. Finally, the »sharper« the curves, the more the data are distributed around a single value. [Chart: Szemes / Nagy 2025]
- Table 2: The performance of SVM classification relating to the two data sets (before and after 1829).
- Figure 3: Effect of the last act on density – ShakeDraCor. [Chart: Szemes / Nagy 2025]
- Figure 4: Effect of the acts on density in five-act comedies – GerDraCor (left), ShakeDraCor (right). [Charts: Szemes / Nagy 2025]
- Figure 5: Plot of cumulative density in five-act plays – GerDraCor. Click [here](#) for the interactive visualisation. [Chart: Szemes / Nagy 2025]
- Figure 6: Plot of cumulative density in five-act plays – ShakeDraCor. Click [here](#) for the interactive visualisation. [Chart: Szemes / Nagy 2025]
- Figure 7: Changes in the median of density in five-act plays – ShakeDraCor (top), GerDraCor (bottom). [Charts: Szemes / Nagy 2025]
- Figure 8: Changes in density in three Shakespeare plays. [Chart: Szemes / Nagy 2025]
- Figure 9: Changes in the median of average clustering coefficient in five-act plays – ShakeDraCor (top), GerDraCor (bottom). [Charts: Szemes / Nagy 2025]
- Figure 10: Changes in the standard deviation of average clustering coefficient in five-act plays. [Chart: Szemes / Nagy 2025]
- Figure 11: The location of GerDraCor's dramas on the PCA graph based on structural features. [Chart: Szemes / Nagy 2025]
- Figure 12: The location of GerDraCor's dramas on the PCA graph based on structural features with a separate colour for comedies and tragedies written before and after 1830. [Chart: Szemes / Nagy 2025]
- Figure 13: The location of GerDraCor's dramas on the PCA graph based on structural features (1650–1829). [Chart: Szemes / Nagy 2025]
- Figure 14: The location of GerDraCor's dramas on the PCA graph based on structural features (1829–1929). [Chart: Szemes / Nagy 2025]
- Table 3: The group created by the k-average clustering procedure that contains the most dramas from one period (k = 10).
- Table 4: The performance of SVM classification with RFB kernel (before and after 1829).