



TECHNISCHE UNIVERSITÄT DARMSTADT  
**DIGITAL HUMANITIES COOPERATION**  
UNIVERSITÄT KONSTANZ

**Thomas Weitin**

**Average and distinction**

The Deutscher Novellenschatz between literary history and corpus analysis

**LitLab**

**Pamphlet #6**

**March 2018**

Ed. Thomas Weitin



# Thomas Weitin

## **Average and distinction.\***

The *Deutscher Novellenschatz* between literary history and corpus analysis

### **Abstract**

In the second half of the 19<sup>th</sup> century, prefiguring and pioneering realism, the novella became the epitome of mass production in the German-speaking literary world. The two editors of the *Deutscher Novellenschatz*, Paul Heyse and Hermann Kurz, were fully aware of the key problem of the epoch: how can a single writer make a difference? The edition of the *Novellenschatz* gives clear advice for a distinctive style in novella writing and sets forth ideas for a new method of writing literary history driven by principles of form. The article extrapolates two major principles from the *Novellenschatz* to let the analysis of the corpus be inspired by them: distinctivity and relationality. The stylometric methods in use here can be derived from the historical terminology of the novella collection. Network analysis may be seen as an appropriate method where the very understanding of individuality emerges as a relational concept to be measured by distances between individual texts and larger quantities. As a result, the understanding of being similar or distinct needs to be rethought in favor of a typology that can describe the difference between global and local similarity.

©2018 Thomas Weitin, [weitin@linglit.tu-darmstadt.de](mailto:weitin@linglit.tu-darmstadt.de)

ISSN: 2629-7027

Ed. by Thomas Weitin

---

\*An earlier German version of this text was published in: Archiv/Fiktionen. Ed. Daniela Gretz, Nicolas Pethes. Freiburg 2016. The network model has been thoroughly reworked.

Thomas Weitin

## Average and distinction. The *Deutscher Novellenschatz* between literary history and corpus analysis

### 1 Paul Heyse, a »virtuoso of the average«

In the second half of the nineteenth century, prefiguring and pioneering realism, the novella became the epitome of mass-produced literature. In a letter to Paul Heyse, Gottfried Keller voices his concern that novella writing had turned into a general leveling, which brought »no more enjoyment and soon no more honor either«.<sup>1</sup> Heyse, who was one of the most popular writers in the German language at the time, was not the worst addressee for this complaint. The prolific writer, who had been awarded the Nobel Prize late in life, mustered 177 novellas altogether, and at the same time, as editor of the *Deutscher Novellenschatz* (1871–1876), his ambition was to preserve »what promises endurance amongst the mass of lively ephemera, and is worth preserving« against the »growing number of daily and weekly newspapers«.<sup>2</sup>

Today, Heyse has long been a case for the irony of literary history; researchers of realism and the novella only know and remember him as an »archivist of the genre«,<sup>3</sup> whose collections<sup>4</sup> ensured the »breakthrough of novelistic narrative as *the* popular reading matter«.<sup>5</sup> The author's own oeuvre, which included a large number of plays and leastwise eight novels, remains forgotten,<sup>6</sup> although Heyse bequeathed a number of manuscripts and collections from his private library to the Bayerische Staatsbibliothek while he was still alive, and his bequest, which they still have archived on 35 linear meters of shelf space, was considered one of their most important collections around 1930.<sup>7</sup> Literary history at that time admittedly demoted the man as an author, though it had once deemed him a fit eponym for

---

<sup>1</sup>»Gottfried Keller an Paul Heyse am 7. September 1884«, in: ›Du hast alles, was mir fehlt...‹. Gottfried Keller im Briefwechsel mit Paul Heyse, ed. Fridolin Stähli. Zurich 1990, pp. 249–252, here: p. 250. This and all following quotes from German-language sources translated by Dr. Claudia Rapp.

<sup>2</sup>Paul Heyse and Hermann Kurz: »Einleitung«, in: *Deutscher Novellenschatz*, eds. eid. Munich 1871, p. 10, p. 22.

<sup>3</sup>Claudia Stockinger: *Das 19. Jahrhundert. Zeitalter des Realismus*. Berlin 2010, p. 115.

<sup>4</sup>After Hermann Kurz had died in 1871, the year the *Deutscher Novellenschatz* was published, further collections followed under the joint editorship with Ludwig Laistner: *Novellenschatz des Auslandes* (1877–1884) and *Neuer Deutscher Novellenschatz* (1884–1887).

<sup>5</sup>Albert Meier: *Novelle. Eine Einführung*. Berlin 2014, p. 90.

<sup>6</sup>With respect to the ›de-canonizing‹ of Heyse compare Christoph Grube: *Warum werden Autoren vergessen? Mechanismen literarischer Kanonisierung am Beispiel von Paul Heyse und Wilhelm Raabe*. Bielefeld 2014.

<sup>7</sup>Cf. Ingrid Rückert: »Paul Heyse (1830–1914). Genese und Präsentation eines Schriftsteller-nachlasses«, in: *Bibliotheksforum Bayern 8* (2014), pp. 270–273.

the epoch following the Goethe era, or ›Goethezeit‹. Writing from a distance of two generations, Oskar Walzel, who updated Wilhelm Scherer's *Geschichte der deutschen Literatur* beyond the Goethezeit, was able to see how »posterity« had canonized Keller at the expense of »both his companions« Storm and Heyse.<sup>8</sup> Especially the latter of the two seemed to have already »taken a backseat«, although Walzel at least recognized Heyse as theoretician of the novella genre: »While Storm accomplished the expression of his demands for the novella, it was Heyse who developed the new doctrine of the novelistic art.«<sup>9</sup>

In the preface of the *Novellenschatz*, which he edited jointly with Herman Kurz, a Swabian writer of village stories, Heyse drafts a genuinely realist genre theory for the novella, which however suffered an unfortunate fate under the moniker ›Falkentheorie‹, or ›falcon theory‹. To this day, students learn to ascribe this term to Boccaccio's novella »Federigo's Falcon« and to connect it to the idea that every novella must feature an object symbol (»Dingsymbol« in German), which consolidates its content and gets to the heart of its substance. Indeed, that is what Heyse postulates, but in his theory of the novella this requirement is aimed at a gain in distinction, which needs to be understood within the context of the literary communication of his time. The novella represents the epitome of mass or popular literature, and for the male and female writers working in this field that means facing the fact that there will always already exist a large quantity of very similar material to what they are producing. How do you write with the awareness of mass similarity; how can you still vary and differ at all? Posterity attached the devaluing attribute »virtuoso of the average«<sup>10</sup> to Heyse the writer, as if to highlight his failure in this respect. And yet the novella collector was fully aware of how unsuitable an elitist canon in this of all genres had to be. Heyse turns out to be a modern theorist when he abruptly complements the gesture of selecting from the available mass and building a canon with a liberal attitude, according to which everyone, authors as well as editors, would have to come to terms with this mass. To this effect Heyse states: »While ›might is right‹ does not apply to it, the prevailing taste of each epoch is a power that mustn't be impeded by doctrinaire obstinacy«.<sup>11</sup>

If you produce a mass product such as the novella, and at the same time steward its publication as an editor, you can't afford ignorance, you need the ability to set yourself apart amidst the bulk and

---

<sup>8</sup>Wilhelm Scherer and Oskar Walzel: *Geschichte der deutschen Literatur*. Berlin 1921, p. 603.

<sup>9</sup>Ibid., p. 604.

<sup>10</sup>Klaus Jeziorowski: »Der Virtuose des Durchschnitts. Der Salonautor in der deutschen Literatur des 19. Jahrhunderts, dargestellt am Beispiel Paul Heyse«, in: *Eine Iphigenie rauchend. Aufsätze und Feuilletons zur deutschen Tradition*, ed. id. Frankfurt am Main 1987, pp. 114–129.

<sup>11</sup>Heyse and Kurz: »Einleitung«, p. 14.



recognize differences or variants. As the »virtuoso of the average«, Heyse had those skills and was thus predisposed to put the realist theory of the novella genre in a nutshell. The extensive correspondence of the two documents his collaboration with Hermann Kurz in detail; hence we know that Heyse left the perusal of shoals of magazines to his less fortunate colleague.<sup>12</sup> Kurz processed what was eligible, Heyse decided what would be included in the *Novellenschatz* in the end. To the extent that the selection was his responsibility, he honed his theory of the novella, which prefaced the collection, to focus on the criterion of distinction. How exactly that is done in the preface merits a more elaborate account.

Heyse formulates the application of his theory of the novella fully aware of the epoch he was in, following the methodological development an increasing empiricism in both the natural sciences and the humanities had brought. He claimed that the »turn towards realism« had aided the breakthrough of the genre in the German-speaking countries, at a »time that first found footing again on the solid ground of the actual and factual in politics and philosophy, that based its method on the study of sources in historiography, and on the experiment in physics and chemistry«.<sup>13</sup> Traditionally, the novella possessed »a certain property right for the mere factual«, based on which writers now knew how to »poetically exploit everything that is unique and peculiar, even whimsical or fanciful, and veering on the ugly«.<sup>14</sup> In the theory however, neither the subject matter nor the esthetics of the genre matter that much, but it is primarily interested in its *modus operandi*, by which »even the smallest form is capable of big effects«. »Contrary to the novel«, the novella's achievement for Heyse is found in its ability to »condense the impression, put it in a nutshell, and thus intensify it to the highest degree«.<sup>15</sup>

The story is the main thing here, not the conditions, nor the event, nor the worldview mirrored in it; because even the deepest ideational content of the individual case will only retain a relative merit, due to its one-sidedness and separateness – the isolation of the experiment, as the natural scientists would call it –, while the breadth of the novel allows for the thorough illumination of a vital human or moral issue, exhaustively and from all angles.<sup>16</sup>

The poetics of condensation, which seeks to approximate the natural sciences like the positivist school in literary studies does, develops concrete ideas for the operationalization of what is postulated as the characteristic nature of the novella and as a writing strategy under the conditions of mass-produced

---

<sup>12</sup>Cf. Monika Walkhoff: *Der Briefwechsel zwischen Paul Heyse und Hermann Kurz in den Jahren 1869–1873 aus Anlass der Herausgabe des ›Deutschen Novellenschatzes‹*. Munich 1967.

<sup>13</sup>Heyse and Kurz: »Einleitung«, pp. 9–10.

<sup>14</sup>Ibid., p. 15.

<sup>15</sup>Ibid., p. 17.

<sup>16</sup>Ibid., p. 18.

or popular literature. Its hypothesis – novellas condense – is operationalized in concrete ideas of the method of this condensation. The qualifiers listed for the ›falcon‹ indicate a range of such operations on different levels: The »test of the excellence of a novelistic motif« consists of checking whether its summary may be formulated as a sequence of actions, the way Boccaccio prefaces his novellas.<sup>17</sup> The idea of a »basic motif« is prevalent also in the »painterly language«, where the »strong silhouette« as a mental representation originates with the reader and thus targets the cognitive act of comprehension.<sup>18</sup> Finally the »simple form« should be tested for »the specific«, »that which distinguishes this story from a thousand others«.<sup>19</sup> A timely concept of originality is developed in the awareness of the necessary distinction on the literary mass market, an originality that defines itself differentially as a measure of the individual text's distance to the aggregate corpus of the genre.

## 2 Problems of literary historiography

The theory of the novella, which is formulated as a recommendation, is systematically connected to the principles of text selection for the *Novellenschatz*. In the same way as writers are urged to pen their respective texts with the awareness of very many similar texts, the editors find themselves confronted with a mass of texts as they search suitable novellas. The original intent of assembling a chronological literary history of novella writing in the German language, with which they approach their collection, designed to consist of 24 volumes, turns out to be excessive from the beginning. Only the first three volumes, which quite conventionally place Goethe at the beginning of a development, and then lead to realism via Kleist, the Romantics, and Tieck, dignify each novella with a critical commentary; afterwards they rely on coherent, brief biographic author portraits, in a moderately positivist manner. The collection's introduction ascribes the failure of the original approach to the quantitative dimension of the project: »Considering the incalculability of the material, it would take decades before we could begin in good conscience, if we wanted to stringently group the texts according to historical and esthetic criteria.«<sup>20</sup> The third volume conveys an idea of how such a ›stringent grouping‹ might be measured, when it assembles Tieck's *Des Lebens Überfluss*, Eichendorff's *Die Glücksritter*, Adolf Widmann's *Die*

---

<sup>17</sup>The sequence for the falcon novella reads as follows: »Federigo degli Alberighi loves without finding his love reciprocated; he squanders all his belongings in his chivalric suit, retaining only a single falcon; when the lady in question visits his house by chance, he cooks the animal, since he has nothing else to serve her as a meal. She finds out what he has done, suddenly changes her mind, and rewards his love by making him her master, giving him her hand and her fortune«. (Heyse and Kurz: »Einleitung«, p. 19).

<sup>18</sup>Ibid.

<sup>19</sup>Ibid., p. 2.0

<sup>20</sup>Ibid., p. 22

*katholische Mühle*, and finally Keller's *Romeo und Julia auf dem Dorfe* successively. Halfway through this group, the comments claim, there runs the boundary between romanticism and realism,<sup>21</sup> the different epochs clearly defined by the different ›tone‹ of the texts. *Die katholische Mühle* and *Die Glücksritter* variegates the novella motif of misguided love as tragic culmination with a negative outcome (Widmann), and as rococo comedy of errors with a happy ending (Eichendorff), respectively. Both are suffused with »Waldeinsamkeit« – woodland solitude – and a mood that matches the »keynote of Eichendorff's poetry«.<sup>22</sup> They claim however that it is just as apparent that Widmann seizes romanticism's »territory of subject matter and mood« for »modern realism«, namely »with characteristic local undertones«, which represent the »basic appeal« of the novella.<sup>23</sup> Heyse lists both geographic details and linguistic aspects, in the form of the Swabian dialect the author employed even more clearly in the original version. In the following comments on Keller it is once again the ›tone‹ that tips the balance, a tone that creates »temperate objectivity«<sup>24</sup> and is thus »akin to Goethe's tone«. Part of the »stylistic appeal« of Keller is a method of »contrast«, which effects the »heightening of the everyday to the typical«.<sup>25</sup>

Because of the time required, the editors of the *Novellenschatz* don't think stylistic groupings are feasible within the framework of a sustained chronological order. In their selection practice for the collection they do note the early signs of a new form of literary historiography, for which neither »illustrious names« nor the principle of »including only the flawless«<sup>26</sup> are the sole pivotal criterion. If you do not want to act ›doctrinaire‹ vis-à-vis the prevailing taste, you must make your selections otherwise, namely in a way that ensures »nobody remains completely unrepresented, and everybody joins the *chain of development* with his or her most idiosyncratic work«.<sup>27</sup> Following this principle, Heyse and Kurz hope to have compiled a »model collection«, which provides the »groundwork for a future history of the novella«.<sup>28</sup>

The literary-historical ambitions of the *Deutscher Novellenschatz* are unmistakable. From today's

---

<sup>21</sup>The comments on Widmann state: »We intentionally followed the final romanticist with this novella writer« (Heyse and Kurz: *Deutscher Novellenschatz*, p. 163 (Vol. 3)).

<sup>22</sup>Ibid.

<sup>23</sup>Ibid., p. 164.

<sup>24</sup>Ibid., p. 235.

<sup>25</sup>Ibid., p. 235 (1st & 2nd quote), p. 236 (3rd quote).

<sup>26</sup>Heyse and Kurz: »Einleitung«, p. 23, p. 22.

<sup>27</sup>Ibid., p. 23, [my italics]

<sup>28</sup>Ibid., p. 24.

perspective, the appeal of the collection and its follow-up projects for research lies in the fact that they seamlessly connect the small section of canonical texts, which have been thoroughly penetrated hermeneutically and thus become representative within ›their‹ epoch, with all but forgotten works of authors that, as representatives of the actual supply of books and periodicals available in their time, are part of the ›great unread‹ of literary history.<sup>29</sup> The *Novellenschatz* project was developed with an awareness of the impact the then-current method transfer between natural sciences and humanities was having on literature and literary historiography. In this respect, the »chain of development« Heyse mentions towards the end of his theoretical sketch is noteworthy. It is not generated by the great minds of literature as sovereign creators, whose biographies fit into a general chronology along with those of the ›arbiters‹ of history, but by texts that are ascribed a formal similarity or commonality via the novella genre, which allows for the observation of changes in style and form. The ›chain of development‹ is a line of the literary form, not a line of personages. Admittedly, this remains almost entirely a concept and marks only one feature of this theory of the novella, one that cannot be reconciled with others. The comments on the Romantic authors never tire of using individual psychologizing to confirm the image of a ›sick romanticism‹.<sup>30</sup> Moreover, the modern (procedural) poetics of isolation, pursuant to the scientific experiment, is pitted against more traditional concepts of esthetic history without any attempt at mediation. These conspicuous dissonances show that while the methodological transformation in question may already be grasped in its early stages, it cannot be fully implemented just yet.

The correspondence provides us with decent insight into the editors' selection practice. It reinforces the idea that the *Novellenschatz* is supposed to be a model collection, not a »sourcebook«.<sup>31</sup> The first volume of the collection accordingly opens with Goethe's *Die neue Melusine*, the text Heyse at least considered »invaluable as a style pattern«.<sup>32</sup> Apparently Kurz had initially protested the inclusion of this novella due to its fairytale character, but was then persuaded to change his mind, which in turn served to diametrically change his opinion on another author afterwards. He thought the novellas of Wilhelm Hauff were each and all unsuitable for the realistically oriented *Novellenschatz*, until meeting

---

<sup>29</sup>Cf. the passage ›The Great Unread‹ in: Franco Moretti: *Distant Reading*. London 2013, pp. 87–89.

<sup>30</sup>The introduction to Achim von Arnim is paradigmatic, stating that he squandered his poetic talent »not because of his illness, but purely on a whim« (Heyse and Kurz: *Deutscher Novellenschatz*, p. 165 (Vol. 1). The »morbidness of the time« is made responsible for the imprint of »form- and characterlessness« (Ibid., p. 166) in his works.

<sup>31</sup>Paul Heyse to Hermann Kurz on February 6, 1870 (quoted from Walkhoff: *Der Briefwechsel zwischen Paul Heyse und Hermann Kurz in den Jahren 1869–1873 aus Anlass der Herausgabe des ›Deutschen Novellenschatzes‹*, p. 28).

<sup>32</sup>Paul Heyse to Hermann Kurz on June 6, 1870 (quoted from *ibid.*, p. 39).

with Heyse and taking them into account again in relation to the accepted style pattern of Goethe's text. His new verdict was: »Wilhelm Hauff. Phantasien im Bremer Rathskeller. Is a novella as well, and an exceptional one at that, viewed by the light of *Melusine*«. <sup>33</sup> The relationality of texts plays a vital part in the selection process, which not only involves the relation of individual texts to each other, but also the relation of the individual text to various quantities, and even to the sheer mass of other texts. How the criterion of relationality is linked with that of distinctivity, or the discriminability within the mass, becomes clear when Heyse explains to his co-editor Kurz why he picked the *Weinhüter von Meran* from his own, quite substantial novella oeuvre for the *Novellenschatz*: »For my own part, I resolved to add the *Weinhüter*, which, considered from all angles, best represents my novelistic writing.« <sup>34</sup> Universally compatible in his own, unique way, Heyse does seem to view himself in his double function as author and editor as ›virtuoso of the average‹. Another decision in the selection process shows how strongly he conceived of the *Novellenschatz* as an organ of reproducible mass literature. Both editors formed their opinion on the basis of their reading, which they repeated under the impression of the increasing bulk of text that grew with the collection, which in turn led to the revision of these opinions – see Kurz' assessment of the *Neue Melusine* and the *Phantasien im Bremer Rathskeller* – or their consolidation. Somehow this practice made Heyse realize that the ›style templates‹ of the collection themselves began to function like a mechanized formative principle. Referring to a novella by Karl August Varnhagen von Ense intended for volume 15, he writes to Kurz: »I reread ›Reiz und Liebe‹ (Varnhagen) and confirmed my impression that despite everything, this pin should not be missing on the barrel of our novella street organ«. <sup>35</sup>

Many pragmatic and economic considerations played a part in the selection of the novellas; after all, the publisher – Oldenbourg Verlag – had to sell the volumes, which were published in 5,000-strong editions, purposely higher-quality bound than the spare editions of the classics, and anything but cheap at a price of 15 silver *groschen* per volume. The division of the 24 volumes into four series of six volumes each was purely a publishing measure, while ideas of creating thematic cycles and reserving for example the last six volumes exclusively to village stories foundered. Variety was in high demand, especially during the »travel and bathing season«, <sup>36</sup> which is why Heyse issued the slogan: »the more varicolored the better«. <sup>37</sup> It seems that precisely because there was no content-based arrangement

---

<sup>33</sup>Paul Heyse to Hermann Kurz on August 12, 1870 (quoted from *ibid.*, p. 67).

<sup>34</sup>Paul Heyse to Hermann Kurz on May 4, 1873 (quoted from *ibid.*, p. 88).

<sup>35</sup>Paul Heyse to Hermann Kurz on January 2, 1872 (quoted from *ibid.*, p. 32).

<sup>36</sup>Paul Heyse to Hermann Kurz on January 8, 1872 (quoted from *ibid.*, p. 35).

<sup>37</sup>Paul Heyse to Hermann Kurz on June 4, 1871 (quoted from *ibid.*, p. 13).

plan, the editors were sensitive to the functional side of volume construction and to the text ratios in their totality, with the principles of relationality and distinctivity shining through again and again. One example would be Heyse reasoning: »We are doing the right thing, closing each volume with one of the most modern novellas«. <sup>38</sup>

### 3 The *Novellenschatz* as corpus – a network analysis

The *Deutscher Novellenschatz* is interesting as a testimony to 19<sup>th</sup> century literary historiography, both for what the collection fails at by its own account, and for developing rudimentary alternatives. On that score, we can distinguish three principles of text compilation. As obvious an excessive demand as the ›stringent grouping‹ of the texts in the framework of a chronological order turns out to be, considering the plethora of text, the fact that the editors stick to their grouping by ›tone‹, i.e. by the style of the respective texts, is just as obvious (1). In doing so, they notice significant differences within individual volumes, between the different epochs for example. What all texts have in common is their genre, which allows for the generation of formal lines, and on the basis of great similarity, these make the differences apparent (2). As a matter of principle, the understanding of each individual text is honed by the awareness of its relation to (very many) other texts (3).

Historically those three principles condense within the literary historiography of realism. Oskar Walzel's expanded and modernized edition of Scherer's *Geschichte der deutschen Literatur*, which sold 29,000 copies between 1917 and 1928, <sup>39</sup> applies the grouping by ›tone‹, which the editors of the *Novellenschatz* envisaged, on a grand scale. And parallel to the literary history he inherited from Scherer, Walzel develops theoretical ideas for a history of literary forms as well as a literary analysis that is supposed to operate under the guiding concepts of ›Gehalt‹ (content, substance) and ›Gestalt‹ (form, shape). Walzel wanted to combine hermeneutic analyses of content and quantitative analyses of linguistic form. He contemplated a synergetic method that would allow him to analyze both individual texts and large quantities of text. <sup>40</sup> The approach was widely received by the Russian formalists who, like Walzel, developed a strong interest in the novella. <sup>41</sup> From there it found its way into structuralism. The fact that the *Novellenschatz* is thus part of the history of literary historiography, which in turn

---

<sup>38</sup>Paul Heyse to Hermann Kurz on April 24, 1871 (quoted from *ibid.*, p. 29).

<sup>39</sup>Cf. Rainer Rosenberg: *Zehn Kapitel zur Geschichte der Germanistik*. Berlin 1981, p. 109.

<sup>40</sup>Cf. Oskar Walzel: *Gehalt und Gestalt im Kunstwerk des Dichters*. Potsdam 1929.

<sup>41</sup>For more details compare Thomas Weitin: »Literarische Heuristiken: Die Novelle des Realismus«, in: *Komplexität und Einfachheit. Veröffentlichung zum gleichnamigen DFG-Symposium 2015*, ed. Albrecht Koschorke (forthcoming).

is connected to the history of quantitative methodology, makes the collection interesting as a corpus. Due to its popularity, it was an important showpiece on the literary market for the contributors. Their correspondence with the editors also affords interesting glimpses of the irritations with respect to the ambition of the large-scale undertaking. Herman Grimm for example, whose novella *Das Kind* was included in the sixth volume of the *Novellenschatz*, specifically asked Heyse for clarification.

I am not completely clear about the endeavor. You are writing about magazines: is it supposed to be a corpus that accrues by installment, or a magazine that reveals itself as that only later?<sup>42</sup>

Heyse's answer has not survived, but it is clear that while periodicals were an important source for the collection, the aim was not that it would become one itself. Literary studies have produced a wide spectrum of studies and interpretations of those individual texts that ended up achieving the canonization intended by the editors – examples would be *Die Judenbuche* or *Romeo und Julia auf dem Dorfe* –; studies in which the context of the collection did not play any part. The fact that only one text per author was included, except for four cases, kept the interest of those philologists who focus on specific authors low from the first. It is more astonishing however that even the bustling research of the relationship between literature and periodicals in the 19<sup>th</sup> century has not concerned itself with the *Novellenschatz* as collection.

Today, the methods of digital humanities enable us to study the collection as precisely that, i.e. as a corpus, in order to follow up on the already formulated problems of literary history and the resulting questions of literary theory.<sup>43</sup> The functional description of ›virtuoso of the average‹ that was bestowed on Heyse is a good starting point for this, since it encapsulates the basic tension between mass similitude and individual distinction. We saw that writing under the conditions of mass similitude characterized the situation of the novella writers, their esthetics of production. Heyse provided them with a handout that formulated the realist(ic) poetics of the novella genre, how to vary and differ within the mass. On top of that, the editors were confronted with the problem of bulk when selecting and arranging their texts, to which they responded by orienting their selection on the criterion of style (›tone‹) and honing their awareness of the relationality of the texts within groups and large quanti-

---

<sup>42</sup>»...fern von Madrid«. Ein Brief Herman Grimms an Paul Heyse zum ›Deutschen Novellen-schatz‹, in: *Zwischen den Wissenschaften. Beiträge zur Literaturgeschichte*, ed. Gerhard Hahn and Ernst Weber. Regensburg 1994, pp. 389-394, here 391. Herman Grimm was the second son of Wilhelm Grimm.

<sup>43</sup>I deliberately ignore the debates around quantitative approaches and digital humanities at this point and refer to my own programmatic thoughts on the integrated approach instead: Thomas Weitin: »Digitale Literaturwissenschaft«, in: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* (DVJs) 89.4 (2015), pp. 651–656.

ties. Taken together, the style-oriented compilation and its reliance on the principles of distinctivity and relationality make the stylometric approach, which has been the preferred approach for a large majority of digital quantitative text analyses so far, the obvious choice. Before we discuss in greater detail why the approach is suitable in this case, we will look at a stylometric visualization of the entire *Novellenschatz* corpus with its 86 texts.


In figure 1 (see appendix), we see a *consensus tree*, which is named that because it retains the »consensus« of a large number of cluster analyses, which were generated on the basis of lists of the most frequently used words of each individual text, resulting in lists of varying length.<sup>44</sup> In the example shown, those analyses were conducted with gradually expanded word frequency lists, starting with 500 and up to 2,000 *most frequent words* (MFW). In this process, only those correlations between texts were retained that proved to be stable in the iterations. These are the altogether 18 pairs or triples in the right half of fig. 1 (see appendix), which are rated as *nearest neighbors* with a constant close range or proximity. The stylometrically determined similarity between them can be regarded as a fixed correlation, which distinguishes them from the texts in the left half. Their arrangement does not provide information on similarities between individual texts; we are dealing with bulk similarity here, within which such specific pairs cannot be stabilized. This is a rather unedifying result for a stylometric analysis interested in classifications; the difficulty of distinguishing fixed individual correlations in the face of mass similarity reminds us of the problem we took as our starting point. We can modify the experiment to get more dramatic results if we change the setting R-Studio's Stylo package, which we use to perform these analyses, designated for »*consensus strength*«. Its default value is set at 0.5, which means that the correlations between text pairs must occur in at least half of the borne-out clusters iterated with increasing MFW number. In other words, the actual »consensus« of all individual analyses is 50 percent. If we increase the consensus strength to its maximum value of 1 (which equals 100 percent), the extent of seemingly indistinguishable similarity mass increases enormously, while only 5 pairs of stable nearest neighbors remain (figure 2, see appendix).

Surprisingly, only two of the four authors represented in the *Novellenschatz* with two of their respective texts are among the remaining pairs, namely Ludwig Tieck and August Kopisch. Even at half the consensus strength, Jeremias Gotthelf's novellas do not behave as might be expected: They

---

<sup>44</sup>The »culling« value indicates that we took all those words into account that are present in at least 20 % of the texts.



$$Delta - Score = \frac{1}{n} \sum_{i=1}^n \left| \left( \frac{x_i - \bar{x}}{s} \right) - \left( \frac{y_i - \bar{y}}{s} \right) \right|$$


Average value of all z-score difference figures

Fig. 3: Formula for calculations in the delta method.

do not form a stylometric similarity pair, and the same goes for Leo Goldammer, if we ramp up the consensus requirements to the maximum value.

We don't want to provide theses from the observation of individual texts for this phenomenon,<sup>45</sup> but remain on the corpus level to sneak a peek at what is under the hood of R-Studio's Stylo package, which enjoys an incredible popularity with a quickly growing community within literary studies. What is actually being calculated here? That should always be the question when literary scholars visualize data generated from texts with the help of quantitative tools. The fact that they may not always understand the answer does not invalidate the question. In the present case however, we are in the fortunate position of being familiar with the underlying method, since it is part of the basic knowledge taught in any introduction to statistics.<sup>46</sup> It is the delta method, first named thusly by John Burrows, which may be used to calculate distance measures between texts.<sup>47</sup>

The delta score is a distance measure based on word frequency. The formula in figure 3 shows how the distance is calculated, namely on the basis of the z-scores for each individual word in the list. Since what we are looking for is a relation between two texts x and y, we calculate the difference, i.e. subtract, and for the frequency value of each word ( $x_i$  and  $y_i$ , respectively) we take the z-score as a

<sup>45</sup>With respect to Gotthelf, the editors suggest an explanation in their introduction to the novellas, pointing out that *Der Notar in der Falle* contains a disproportionately higher amount of dialect words than the medieval knight story *Kurt von Koppigen*, which was much less typical for Gotthelf. Cf. Heyse and Kurz: *Deutscher Novellenschatz*, vol. 7, pp. 5-43; vol. 12, pp. 5-194.

<sup>46</sup>The author completed this introduction in the winter semester 2015/16 at Konstanz university with the linguist Gerold Schneider, whom he wants to thank as sincerely as his brilliant R teacher Nathaniel Phillips from the psychology department. In this context, he highly recommends *YaRrr! The Pirate's Guide to R* (<https://bookdown.org/ndphillips/YaRrr/>).

<sup>47</sup>Cf. John F. Burrows: »Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship«, in: *Literary and Linguistic Computing* (LLC) 17 (2002), pp. 267-287.

basis, which we get by subtracting the arithmetic mean for all texts of the corpus from the absolute frequency and then divide this difference by the standard deviation. That way we calculate the z-score for each two compared texts for each word of the list (e.g. for 500 *most frequent words*), calculate the difference, and the values for all differences, to end up with a fixed zero point. We finally divide all those difference values by the sum of all monitored words (in our example: by 500), and get the delta score, an arithmetic mean that measures the distance between the two texts.

This simple method is suitable for our purpose in several respects. By calculating the relation as difference between/of the texts (therefore: delta method) via the z-scores, we simultaneously include the corpus relation of each individual text in the intertextual relation. As the difference between absolute word frequency and corpus mean divided by the standard deviation, the z-score indicates by how many standard deviations the respective value deviates from the average. In statistics, this normalizing abstraction is used to compare values from different data sets with a hypothetical normal distribution. In our case it is word frequencies that are compared in this way. The standard deviation itself indicates how strongly an individual value deviates on average from the mean of all values, and thus the deviation we can expect for each new value.<sup>48</sup>

If the calculation is carried out like that, the method corresponds strictly to the questions we are interested in. Texts are measured in their relationality with other texts, with the distance measure indicating the similarity between texts simultaneously incurring the similarity or distinctivity of each individual text within the entire corpus. We are working with both intertextual similarity and distinctivity within a large quantity of texts, hence with a double relationality. The only problem is that this double relationality finds no easily interpreted visual expression in the cluster analyses that are currently in vogue in literary stylometry, and that may be iterated into consensus trees in terms of validation as shown above. Therefore, my research group and I searched for a way to measure and capture the similarity between texts and the ›corpus property‹ of each individual text, in the sense of its distinction from the whole, in one and the same quantitative model. Inspired by the considerations of Ulrik Brandes and Maciej Eder,<sup>49</sup> we ended up with a network analysis of the *Novellenschatz* corpus with the

---

<sup>48</sup>In a data set with perfect normal distribution, 68 % of all values lie within the scope of single standard deviation, and 95 % within the scope of double standard deviation.

<sup>49</sup>Cf. Ulrik Brandes, Pádraig Cunningham, Conrad Lee, and Bobo Nick: »Simmelian Backbones: Amplifying Hidden Homophily in Facebook Networks«, in: *Proc. ASONAM* (2013), pp. 525–532; Maciej Eder: »Visualization in Stylometry: Cluster Analysis Using Networks«, in: *Digital Scholarship in the Humanities* (2015), pp. 1–15.

aid of the network tool »Visone«. This tool was developed for the analysis of social networks,<sup>50</sup> where *nodes* usually represent individuals and *edges* represent the connections between them. We imported a double data matrix of the delta scores of the *Novellenschatz* into Visone (figure 4, see appendix). One consists of the delta scores for the distance measure of all texts among one another; for each individual text there are 85 values for the relation to each different available individual text. The second matrix was supposed to measure the relation of each individual text to the whole corpus, to which end we conflated all 86 texts for the word count into one file and determined its distance to each individual text.

For the network modeling, we attributed the ›corpus delta‹ of each individual text to the nodes, and the delta values for the relation from text to text to the connections or edges, resulting in a network whose nodes are descriptive of the distinctivity of the texts within the corpus, while the edges represent their relationality among each other. Color, size, and thickness of the nodes and their labels, as well as the edges were scaled in grayscale corresponding to their delta values. Please note that figure 5 (see appendix) does not show the entire network, but the part of it that contains the largest paired similarities, which were filtered out with the help of a delta threshold value filter. In the mapped experiment all connections with a higher delta value than 0.775 were filtered out; the remaining pairs yield the pictured network.

Because of the grayscale in the visualization, we worked with »reversed delta«, which means that the scale was reversed, so the dark nodes and edges represent a low delta value and thus a strong similarity. As a trial, you can look at the connection between the two Tieck texts, which as expected are very similar (as they also were in the consensus tree). Generally, you must not be misled by the high number of light gray connections, since what is light in this partial corpus with the highest paired similarities would still be a strong (i.e. dark) connection. We already know that there is strong similarity within our *Novellenschatz* corpus. In this regard, the particularly dark nodes are interesting. The node embodies the delta value for the relation of individual text to whole corpus. Dark nodes possess a low delta value. Hence, those texts display a high corpus similarity; in other words, in terms of the whole they are particularly average. If we check for which authors that is the case, we find the editor Heyse, and with Heinrich Zschokke and Carl Immermann two further writers that were very popular in the 19<sup>th</sup> century and had high sales figures. The second editor Hermann Kurz on the other hand is markedly less average within the corpus.

---

<sup>50</sup>»Visone« stands for *v i s u a l s o c i a l n e t w o r k s*: [www.visone.info](http://www.visone.info).

The quantitative experiments for the *Deutscher Novellenschatz*, which should be merged with the qualitative research of literary history in a »scalable reading«,<sup>51</sup> are still in their early stages. Nevertheless, important decisions have already been made, particularly that of working with »Visone«, whose algorithms were abstracted and programmed in a theoretical order that accommodates the problems of literary studies which interest us. According to the current progress of our work, this applies especially to the algorithm that was deliberately christened »Simmelian backbone« by its creators,<sup>52</sup> named after the cultural scientist and sociologist Georg Simmel, whose analyses of social communities of the 19<sup>th</sup> and early 20<sup>th</sup> century focus on the need for distinction under the conditions of mass similarity. We are back at our underlying historical theme.

Contrary to a globally discriminating delta threshold value filter, which was used in the experiments shown this far, the »Simmelian backbone« enables a locally adaptive filtering. It identifies connections in locally dense environments, where similarity dominates. That means for texts: It searches for connections between texts that have many ›mutual neighbors‹, between which the delta distance is low in each case. Thereby local similarity groups become visible. Connections that do not belong to such a dense environment are filtered out – yet the layout retains those connections without which the network would fall apart, thus the term ›backbone‹. Developed for the analysis of social networks, the aim of this algorithm is to identify hidden traits that establish local communities within a global network.

Figure 6 (see appendix) shows the Simmelian backbone of the *Deutscher Novellenschatz*. In the example the network was transformed with the operational configuration that for each text, 6 neighbors are taken into account, 3 of which have to be mutual. You can see how groups are formed and how the backbone of the whole network emerges. But that is not the only interesting thing to observe. What also emerges is those texts that for very many other texts are the one with which they have the most mutual neighbors. Computationally, those asymmetrical similarity centers or hubs possess a high indegree centrality, which may be visually displayed by many incoming connections at their node. Therefore, Maciej Eder has found an apt metaphor when he calls them »*absorbing hubs*«. <sup>53</sup> It

---

<sup>51</sup> Compare the report from the inaugural conference: Cornelius Eggert: »Scalable Reading. Paul Heyeses ›Deutscher Novellenschatz‹ zwischen Einzeltext und Makroanalyse«, in: *Zeitschrift für Germanistik* 1 (2016), pp. 140–142.

<sup>52</sup> Cf. Brandes et. al.: »Simmelian Backbones: Amplifying Hidden Homophily in Facebook Networks«, p. 526.

<sup>53</sup> Cf. Maciej Eder: »Visualization in Stylometry«, p. 11.

is also possible to put this in a way more in tune with literary studies: An edge directed from A to B indicates that B is one of the works A has the most similar works with. In our *Novellenschatz* corpus, the editor Paul Heyse is such a candidate. He has the second highest number of incoming connections, right after Heinrich Zschokke. According to Eder, it is up to further research to find out what these »absorbing hubs« really mean. The interpretation he ventures picks up on the idea that informs the characterization of Heyse as the ›virtuoso of the average‹: »They might also reflect texts stylistically ›average‹, typical for their times rather than exceptional«. <sup>54</sup>

## 4 Discussion and outlook

With regard to our research, does this mean that, after considerable effort, we merely find quantitatively confirmed what from a literary-historical perspective was clear from the start? The resulting conclusion would be along the lines of this: Heyse's virtuoso averageness, which in the era of mass-produced literature made him the epitome of the pertinent genre, may be corroborated by stylometry. One way or another, perspectives for further research arise from a negative answer to this question. What the experiments show is not a sufficient result; rather they themselves require further validation. The pictured networks were produced with data that was based on a single word list length (500 MFW). An iterative analysis with many different word list lengths is necessary, a task for which Visone offers excellent preconditions, such as the possibility of creating network collections and visualizing them in dynamic layouts. The question of the historical plausibility of applying the stylometric delta method in the present case could be answered by pointing out the interconnected problems of mass similarity, necessary distinction, and self-conscious relationality identified by historical literary theory and sketched here. As an additional benefit of the analysis, it becomes apparent that the currently most frequently used method of quantitative literary studies may also be used for approaches that do not even primarily focus on the allocation of texts to authors, genres, or epochs. Questions of literary theory that are not oriented towards attribution may also be discussed with its help. At the same time, network analysis in particular must not be limited to dealing with word frequencies. The mass similarity of novellas is not only a matter of style, let alone one of stylometry. Recurring topics, constellations of characters, and points of view or narrative modes are no less responsible for it. Moreover, to fully exhaust the knowledge potential of the Simmelian backbone in particular, the minable attributes from the historical metadata of the texts need to be read into the already existing networks.

---

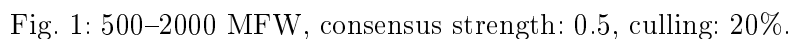
<sup>54</sup>Ibid., p. 13.

Thinking along those lines however, we might say that even what we already have contains more than the mere confirmation of what we knew. This becomes apparent if we compare the networks we received through working with the global delta threshold value filter on the one hand, and the locally adaptive Simmelian backbone on the other. While some of the texts that show a high corpus similarity also turn up as absorbing hubs in the backbone, that is not the case for others. Interestingly, this distinction runs between the two editors of the *Novellenschatz*. In the partial network with the highest paired similarities (fig. 5, see appendix), Paul Heyse has a very dark node, which reveals a low corpus delta of the text, and thus a great similarity with the average of all texts. The Simmelian backbone (fig. 6, see appendix) shows him as a text with which very many texts have the most mutually similar texts. In his indegree centrality (14) he is only surpassed by Heinrich Zschokke (15), who is likewise both highly similar to the corpus and constitutes a local similarity center. Another one who is very similar to the corpus and thus average in relation to the whole is Immermann, whose local similarity value is however much lower (indegree centrality: 6). The opposite is true for the second editor Hermann Kurz. Compared with Heyse, Zschokke, and Immermann, he gets a markedly lighter node if we inquire after global averageness (fig. 5, see appendix), but constitutes a local similarity center, which approximates Heyse and Zschokke, and has twice the value of Immermann (indegree centrality: 12).

author	indegree centrality
Paul Heyse: Der Weinhueter von Meran	14
Karl Immermann: Der Carneval und die Somnambule	6
Hermann Kurz: Die beiden Tubus	12
Heinrich Zschokke: Der tote Gast	15

**Table 1:** Indegree centrality for different types of similarity.

Thus, Heyse the virtuoso, who masters the average both globally and locally, is joined by a second editor, whose text is a lot less similar to the corpus but who still has a group-forming effect. Future questions of research in this field will include what distinguishes his group, and in what sense the emerging typology of averageness may add to the analysis of strategies of distinction in an era of mass similarity.



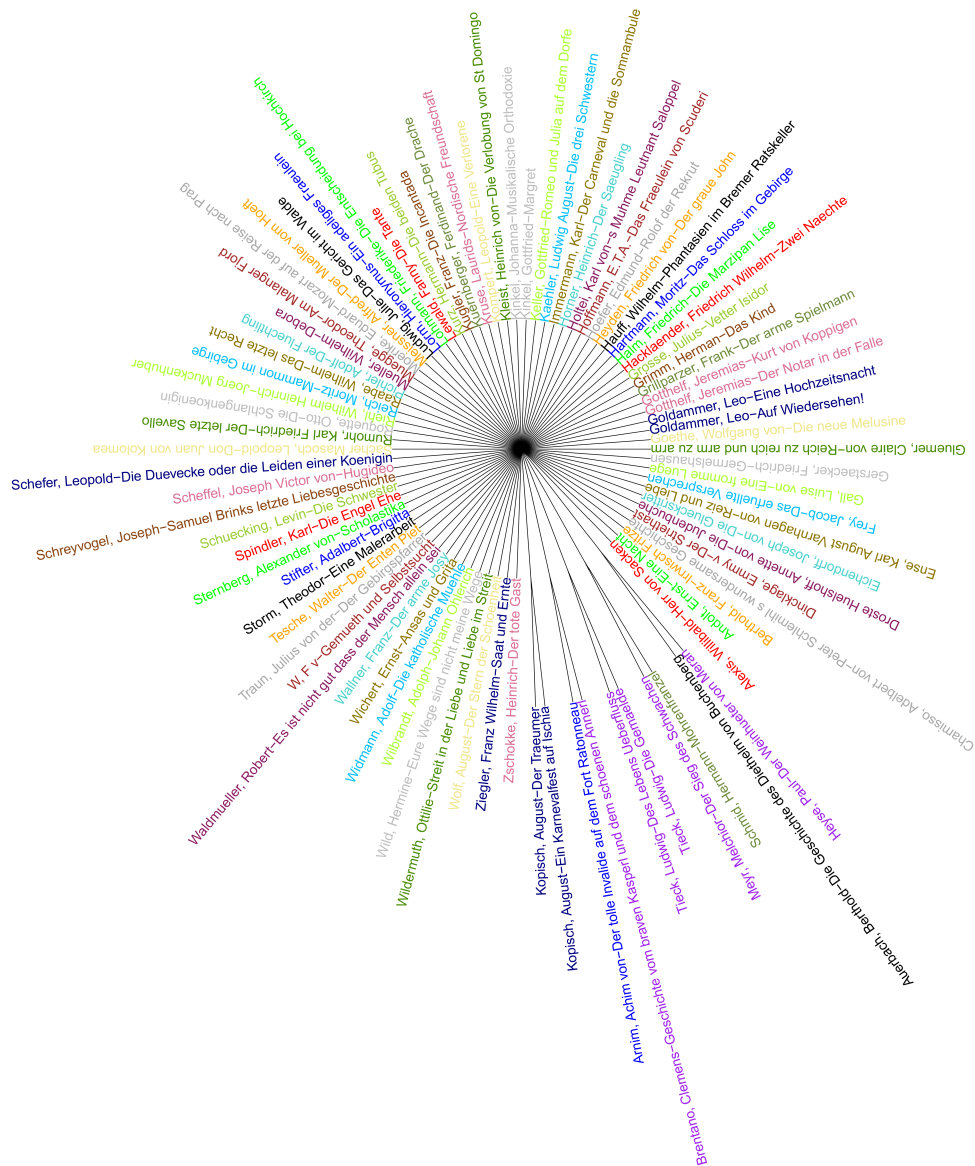


Fig. 2: Consensus tree of the *Novellenschatz*, 500–2000 MFW, consensus strength: 1.0, culling: 20%.



A	B	C	D
	Gesamtkorpus	"Alexis, _Willibald-Herr_von_Sacken"	"Andolt, _Ernst-Eine_Nacht"
1. "Alexis, _Willibald-Herr_von_Sacken"	1. 0.619370385476375	1. 0.0	1. 0.888390287467405
2. "Andolt, _Ernst-Eine_Nacht"	2. 0.776006707359245	2. 0.888390287467405	2. 0.0
3. "Arnim, _Achim_von-Der_tolle_Invalide_auf_der"	3. 0.824751070853411	3. 0.997620415137901	3. 1.05991920112622
4. "Auerbach, _Berthold-Die_Geschichte_des_Diethelm"	4. 0.614837468654176	4. 0.869943117211169	4. 1.06993394612992
5. "Berthold, _Fran-Irrwisch_Fritze"	5. 0.652947422923074	5. 0.875483695478742	5. 1.07278083497206
6. "Brentano, _Clemens-Geschichte_vom_braven_H"	6. 0.842077407664849	6. 1.07357855445429	6. 1.16385662386254
7. "Chamisso, _Adelbert_von-Peter_Schlemihl_s_w"	7. 0.774693230712353	7. 0.967342858825897	7. 0.979635596609223
8. "Dincklage, _Emmy_v-Der_Striethast"	8. 0.773823329892287	8. 0.913370171977517	8. 1.09423583429564
9. "Droste_Huelshoff, _Anette-von-Die_Judenbuche"	9. 0.623905482177761	9. 0.853372781823891	9. 0.993843325696704
10. "Eichendorff, _Joseph_von_Die_Gluecksritter"	10. 0.905334608054505	10. 1.11572894199668	10. 1.21374365775525
11. "Ense, _Karl_August_Varnhagen_von_Reiz_und"	11. 0.846595197756998	11. 1.101472751549670	11. 0.970167548527492
12. "Frey, _Jacob-Das_erfuellte_Versprechen"	12. 0.673941346089979	12. 0.851026362481082	12. 0.951724138420099
13. "Gall, _Luise_von-Eine_fromme_Luege"	13. 0.740594737213398	13. 0.922301796947545	13. 1.02965427977052
14. "Gerstaecker, _Friedrich-Germelshausen"	14. 0.96149842875762	14. 1.17456434705691	14. 1.26058324533583
15. "Gluemer, _Claire_von-Reich_zu_reich_und_arm"	15. 0.763035861079044	15. 0.962020385423918	15. 1.14070500655163
16. "Goethe, _Wolfgang_von-Die_neue_Melusine"	16. 0.917810434565349	16. 1.11814248761361	16. 1.03637751466392
17. "Goldammer, _Leo-Auf_Wiedersehen!"	17. 0.987730962687695	17. 1.07966480496991	17. 1.17462347677721
18. "Goldammer, _Leo-Eine_Hochzeitsnacht"	18. 1.11388210110537	18. 1.25202629809527	18. 1.36918838703706
19. "Gotthelf, _Jeremias-Der_Notar_in_der_Falle"	19. 0.933818447951573	19. 1.08226695553142	19. 1.28278618511723
20. "Gotthelf, _Jeremias-kurt_vonKoppingen"	20. 0.68826541917926	20. 0.856260701191899	20. 1.05683579065317
21. "Grillparzer, _Frank-Der_arme_Spielmann"	21. 0.771196268879711	21. 1.0139510608834	21. 1.10688891837978
22. "Grimm, _Herman-Das_Kind"	22. 0.736224632104404	22. 0.978337622842476	22. 1.8226881635912
23. "Grosse, _Julius-Vetter_Isidor"	23. 0.708368222843048	23. 0.927674301400118	23. 0.965602326455242
24. "Hacklaender, _Friedrich_Wilhelm-Zwei_Naechte"	24. 0.795826666291009	24. 0.944144345758205	24. 1.02545274408755
25. "Halm, _Friedrich-Die_Marzipan_Lise"	25. 0.775978108640869	25. 0.949170354801434	25. 0.996513050232035
26. "Hartmann, _Moritz-Das_Schloss_im_Gebirge"	26. 0.844107820306773	26. 1.06026141084127	26. 0.97212439299578
27. "Hauff, _Wilhelm-Phantasien_im_Bremer_Ratskeller"	27. 0.706734985943693	27. 0.908899597848194	27. 1.05522699959419
28. "Heyden, _Friedrich_von-der_graue_John"	28. 0.836226927237825	28. 0.884615372504986	28. 0.957400875267643
29. "Heyse, _Paul-Der_Weinhuetter_von_Meran"	29. 0.541994044020464	29. 0.770379212367372	29. 1.0120473179309
30. "Hoefel, _Edmund-Rolof_der_Rekrut"	30. 0.77497943292328	30. 1.00677221387146	30. 1.1342332522742
31. "Hoffmann, _E.T.A.-Das_Fraeulein_von_Scuderi"	31. 0.72061686548237	31. 0.92414876344386	31. 0.978351601500174
32. "Holtei, _Karl_von-s_Muhme_Leutnat_Saloppel"	32. 0.840092226917843	32. 0.960160864849423	32. 1.06449835479209

Fig. 4: Excerpt of the data matrix. Column B: distance ( $\Delta$ ) between text/corpus; column C: distance ( $\Delta$ ) between text X/Y.

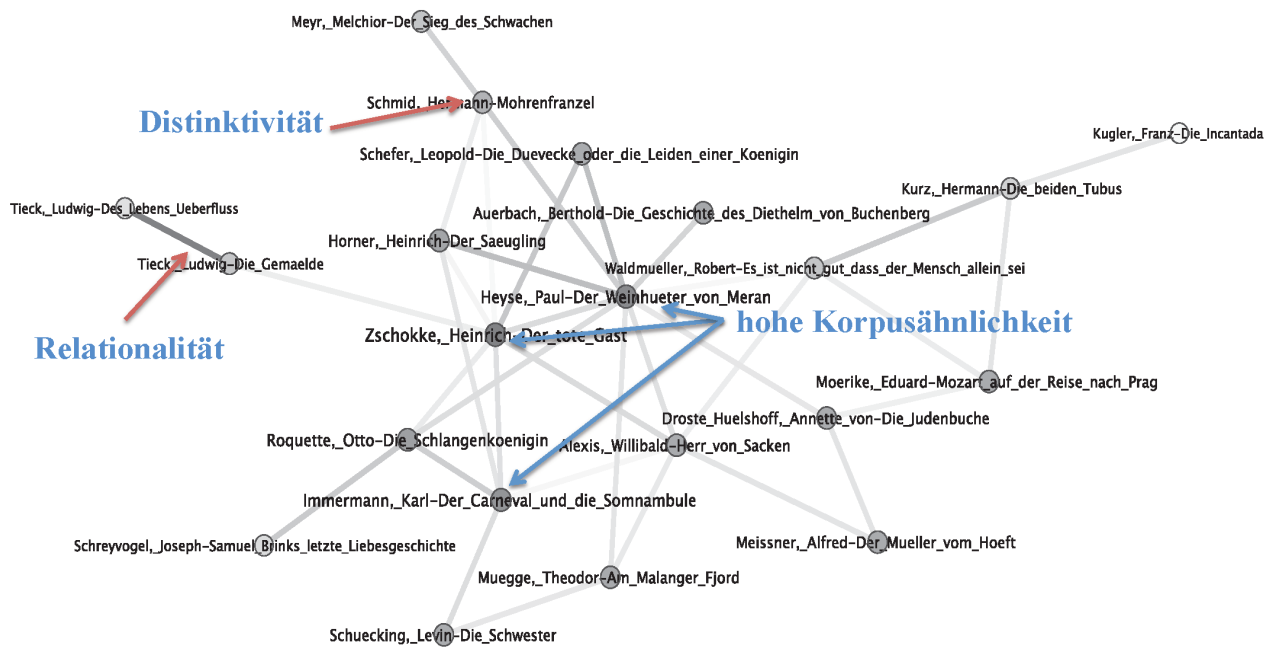


Fig. 5: *Deutscher Novellenschatz*: partial network showing the highest paired similarities. Color, size and thickness of the nodes, labels, and edges scaled according to the delta values of nodes and edges.

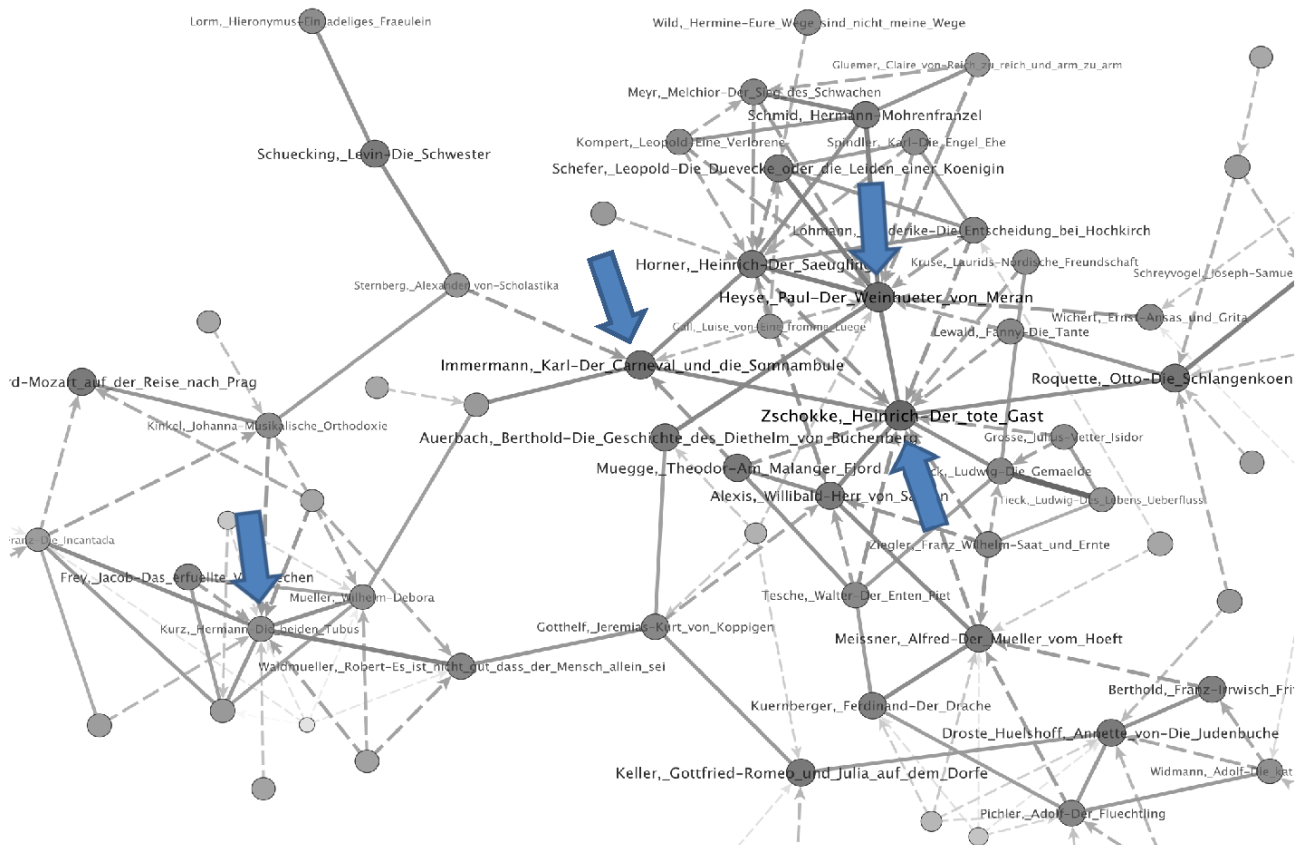


Fig. 6: Simmelian backbone network of the *Deutscher Novellenschatz*.