

Graphs and Comparative Literary Studies

Irina Golovacheva and Mikhail Zhuravlev

St Petersburg State University, 7/9 Universitetskaya Emb.,
St Petersburg 199034, Russia

1 Introduction

Mathematical methods in literary studies have been drawing the philologists' attention for more than a century now, especially after the success of statistical methods in stylometry was widely recognized. A few scholars demonstrated their interest in mathematical methods, which is mainly traced in the works employing mathematical and physical metaphors. Such words as 'entropy', 'continuum', 'spatial and temporal dimensions' are not uncommon in literary criticism. 'The statistical turn' was followed by the 'digital turn', i.e. by Digital Humanities in literary studies exemplified primarily in the works of Franco Moretti (Moretti, 2007) and his followers.

While the success of statistical approach in stylometry or the growing application of Moretti's 'distant reading' (Moretti, 2013) is apparent, other mathematical methods have been shadowed. Even so, over the past decades scholars of literature have been trying to broaden the spectrum. One of the examples is the study of love dynamics by means of ordinary differential equations (Rinaldi et al., 2015). Since love/seduction is a common element of plots, one can easily find a sufficient number of texts to research. Such studies offer a rather complicated description of love dynamics demonstrating this type of the direct application of mathematics in literary criticism (Golovacheva et al., 2017; Zhuravlev et al., 2015).

One of the most valuable contributions of mathematics into literary theory as well as to the studies of myth and folklore is the application of graph theory. An early example was Pierre Maranda's structuralist chapter "Cendrillon: Théorie des graphes et des ensembles" (Maranda, 1973). It was soon followed by L. O'Tool's paper offering an analysis of the structure and 'semiotic space' in *The Book of Genesis* and "Eveline," a section of James Joyce's *Dubliners* (O'Toole, 1980). O'Tool employed R. Atkin's Q-analysis (Atkin, 1977), the method based on graph theory which, as the latter thought, could be useful in social studies, political science, organizational analysis, and other "soft sciences". Besides, R. Atkin applied Q-analysis for the study of some literary texts. In our own research (Golovacheva et al., 2018), we used the method to compare two masochistic novellas, "Venus in Furs" by Leopold von Sacher-Masoch and "Torrents of Spring" by Ivan Turgenev.

2 Method and Texts

In the present work, we use the tools of graph theory in order to compare canonic doppelgänger short fictions and critically reconsider each of them. In its main features, our approach is similar to R. Atkin's Q-analysis which analyzes simplices combined into simplicial complexes. However, there are some important differences in our method as compared with Atkin's. We will present our approach here without discussing specific aspects of literary theory. Still, the most important issues of the latter will be commented on.

First of all, there is a serious methodological question: what does mathematics have to offer in terms of the analysis of literary texts? We suggest the following general scheme that addresses the theoretical issue and leads to a novel comparative interpretation of a set of works of fiction. The point of departure is to quantify some uniform meaningful characteristics of a text or a group of the related texts (of a set). The next step is to apply a most adequate mathematical method to investigate them. Finally, the obtained results are to be assessed in the framework of literary criticism. It is apparent that mathematics in this approach does not provide a "complete and ultimate" proof of some preliminary hypothesis or theses. Still, it can highlight the regularity or the uniqueness of phenomena and thus direct our searchlight to the interesting points in the field of literary theory and comparative research.

The first issue to be tackled is the following one: what exactly are we going to quantify? Only few characteristics of the text can be quantified without raising additional questions. If one quantifies the number of words in a sentence or the number of stressed syllables in a poem's line, such quantification does not raise any serious question. However, we intend to quantify certain *qualitative* characteristics of the text. In such cases, the consensus among the researchers becomes an important argument in favor of the chosen quantification.

In the present research, we will compare the doppelgänger fictions - the ones in which the protagonist encounters his double - quantitatively. So, we make sure to select not only "important" but also the undisputed characteristics (attributes) of the doppelgänger texts. To reduce the arbitrariness of the selection, we use the generally recognized characteristics of doppelgänger plots explicated in Otto Rank's *The Double* (Rank, 2019) (first edition "Der Doppelgänger. Eine psychoanalytische Studie", 1925, Internationaler Psychoanalytischer Verlag). Also, we add some characteristics which are not found in Rank's psychoanalytical investigation of the archetypal phenomenon. The reason for such inclusion is that these characteristics are either important features of a particular text or they can be found in several doppelgänger texts. We chose the following texts for our comparative study: "Der Doppelgänger" (1821) by E.T.A. Hoffmann, "William Wilson" (1839) by E.A. Poe, "The Double" (1846) by Fyodor Dostoevsky, "The Jolly Corner" (1908) by Henry James, "The Secret Sharer" (1910) by Joseph Conrad, "The Encounter" ("L'entrevue," 1919) by Henri de Régnier, and *Despair* (1932-34) by V. Nabokov.

3 Quantitative Similarity Between the Texts

We divide the identified characteristics¹ of the doppelgänger texts into four groups (sub-sets) of uniform attributes (invariant motifs) grouped as 'psychological attributes

¹ Rigorously, what we called 'characteristics' or 'attributes' can be classified as 'topoi' or 'motifs' in literary theory, the difference between these terms being insufficient for the present study.

Table 1: Incidence matrix for the group ‘psychological attributes of doubling’.

	Horror	Madness	Aggression	Narcissism	Anxiety
William Wilson	1	1	0	0	1
Jolly Corner	1	0	1	0	1
Secret Share	1	1	1	0	1
L’entrevue	0	0	0	0	0
Double	1	1	1	1	1
Doppeltgänger	1	1	0	0	1
<i>Despair</i>	1	1	1	1	1

	Malevolence	Assistance	Double’s mocking imitation of the original	Disgust when encountering a double
William Wilson	0	1	1	1
Jolly Corner	0	0	0	1
Secret Sharer	0	1	0	0
L’entrevue	0	0	0	0
Double	1	1	1	1
Doppeltgänger	0	0	0	0
<i>Despair</i>	0	0	0	0

of doubling’, ‘physical manifestations of doubling’, ‘elements of biography in doubling’, and ‘mysterious attributes of doubling’. Our research starts with the construction of incidence matrix for each group of the attributes. Table 1 represents incidence matrix for ‘psychological attributes’ (U_{psych}). The name of this group signifies and generalizes the qualities of the original character (protagonist) or/and those of his double as well as the specific features of their interaction. Each psychological characteristic in this matrix is found in Rank’s work. Each element of the matrix is positioned at the cross of a row corresponding to a specific text and a column corresponding to a specific attribute. The element equals one if this attribute is present in this text, otherwise it equals zero. Each attribute determines a simplex. For our purpose, we can consider a simplex as a graph whose vertices are the texts and whose every two vertices are connected by an edge. Some vertices (texts) may fall within more than one simplex. Combined together, these simplexes form a simplicial complex which represents a structure of the set of texts in relation to the attributes. For a small number of vertices, a simplicial complex can be drawn. The analysis of the structure can be carried out basing on incidence matrix (without picturing it). E.g., Figure 1 presents a simplicial complex formed by the simplices determined by such attributes as Madness, Aggression and Assistance/Opposition. Hereafter, we use the following notations “SS” stands for “The Secret Sharer,” “WW” - for “William Wilson,” “D(D)” - for Dostoevsky’s “Double,” “D(H)” - for Hoffmann’s “Doppeltgänger,” “JC” - for “The Jolly Corner,” “De” - for *Despair*, and “L’entr” – for “L’entrevue.”

How can one estimate the similarity between each pair of the texts in the set? We propose the following method: firstly, we define “index of similarity” of the texts as the sum of coincidences between the corresponding lines of the incidence matrix. Therefore, coincidence is the simultaneous presence or absence of an attribute in texts. For instance, the index of similarity between “William Wilson” and “The Jolly Corner” equals 5 because of five coincidences between the corresponding lines. The symmetrical matrix of the indexes of similarity (‘matrix of similarity’) I_{psych} can be

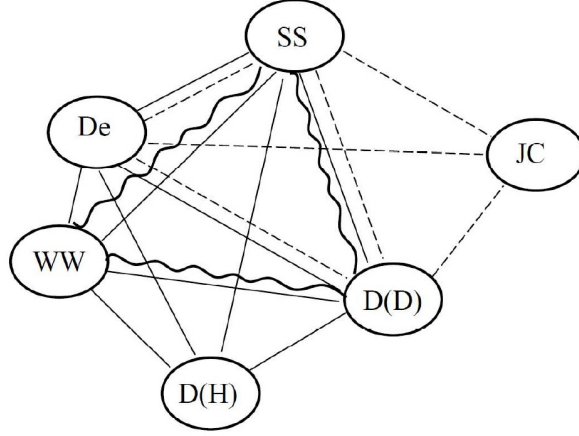


Figure 1: Simplicial complex formed by three simplices determined by such attributes as Madness (—), Aggression (---), and Assistance/Counterwork (~~~~). "L'entrevue" does not appear in this simplicial complex because the above-listed attributes are lacking in de Régnier's story.

Table 2: The matrix of similarity of the texts in relation to 'psychological attributes of doubling'.

	WW	JC	SS	L'entr	D(D)	D(H)	De
WW	9	5	6	3	6	6	4
JC	5	9	6	5	4	6	6
SS	6	6	9	4	5	7	7
L'entr	3	5	4	9	0	6	4
D(D)	6	4	5	0	9	3	5
D(H)	6	6	7	6	3	9	7
De	4	6	7	4	5	7	9

obtained as the product of incidence matrix by its transpose $I_{Psych} = U_{Psych}U_{Psych}^T$ if we multiply the elements 0, 1 not as the numbers but according to the following rule: $1 \cdot 1=1$, $0 \cdot 0=1$, $1 \cdot 0=0$, $0 \cdot 1=0$. The diagonal elements of I_{Psych} are equal to the number of the attributes.

The matrix of the indexes of similarity (Table 2) shows the degree of similarity between a pair of texts in relation to the chosen psychological attributes. Also, we can calculate the total number of coincidences for the n -th text like $\sum_{n \neq m} (I_{Psych})_{nm}$.

Considering I_{Psych} (Table 2), one concludes that ("The Secret Sharer," "Der Doppelgänger"), ("The Secret Sharer," *Despair*), and ("Der Doppelgänger," *Despair*) are the pairs of most similar texts in relation to 'psychological attributes'. Also, we conclude that ("L'entrevue," "The Double") is the pair of less similar fictions.

We perform these procedures for each group of attributes. Table 3 presents the matrix of incidence for 'physical manifestations of doubling'.

Fig. 6 shows the incidence matrix for the 'elements of biography', and Fig. 7 displays the corresponding matrix of similarity.

The corresponding matrix of similarity I_{Phys} (Table 4) shows that the most similar pairs of texts are ("William Wilson," "The Secret Sharer") and ("William Wilson," *Despair*), whereas ("William Wilson," "L'entrevue"), ("The Jolly Corner," "The Double"), ("The Jolly Corner," "The Secret Sharer"), ("The Jolly Corner," "Der Doppelgänger") demonstrate less similarity inside each pair. The presence of so many pairs of dissimi-

Table 3: Incidence matrix for the group 'physical manifestations of doubling'.

	Reality of the double	Unreality of the double	Real similarity between the character and his double	Imaginary similarity between the character and his double	Unusual voice of the double	Objects/signs related to doubling
WW	1	0	0	1	1	1
JC	1	1	0	1	0	0
SS	1	0	0	0	1	1
L'entr	0	0	0	0	0	0
D(D)	1	0	1	0	0	1
D(H)	1	0	1	0	0	1
De	1	0	0	1	0	1

Table 4: Matrix of similarity of the doppelganger texts in relation to 'physical manifestations of doubling'.

	WW	JC	SS	L'entr	D(D)	D(H)	De
WW	6	3	5	2	3	3	5
JC	3	6	2	3	2	2	4
SS	5	2	6	3	4	4	4
L'entr	2	3	3	6	3	3	3
D(D)	3	2	4	3	6	6	4
D(H)	3	2	4	3	6	6	4
De	5	4	4	3	3	3	6

Table 5: Incidence matrix for the 'elements of biography in doubling' group.

	<i>Childhood</i>	<i>Alternative life</i>	<i>End/Begining</i>	<i>Ruining of the plans</i>	<i>Rivalry for the affection</i>	<i>Tanatos</i>
WW	1	0	1	1	0	1
JC	1	1	1	0	0	0
SS	0	0	1	1	0	0
L'entr	0	0	0	0	0	0
D(D)	0	0	1	1	1	1
D(H)	1	0	1	1	1	0
De	1	0	1	1	0	1

Table 6: Matrix of similarity of doppelganger texts in relation to ‘the elements of biography in doubling’.

	WW	JC	SS	L’entr	D(D)	D(H)	De
WW	6	3	4	2	4	4	6
JC	3	6	3	3	1	3	3
SS	4	3	6	4	4	4	4
L’entr	2	3	4	6	2	2	2
D(D)	4	1	4	2	6	4	4
D(H)	4	3	4	2	4	6	4
De	6	3	4	2	4	4	6

Table 7: Incidence matrix for the group ‘mysterious attributes of doubling’.

	<i>Dream</i>	<i>Secret</i>	<i>Mysterious Building (Home, Castle)</i>	Supernatural	Sign of Doom
WW	1	0	1	1	1
JC	1	0	1	1	0
SS	1	0	0	0	0
L’entr	0	1	1	1	0
D(D)	1	0	0	1	1
D(H)	1	1	1	1	0
De	1	0	0	0	1

lar texts in this group of attributes could be hypothetically explained by the intention of every writer to create a new protagonist who would not resemble any other character in somebody else’s doppelganger prose. However, we suggest what we think a more adequate explanation: ‘physical attributes’ are less uniform as compared to ‘psychological’. Moreover, the ‘doppelganger poetics’ does not entail the usage of all or too many physical features related to the doppelganger’s appearance or setting.

Table 5 shows the incidence matrix for the ‘elements of biography’, and Table 6 displays the corresponding matrix of similarity.

The italicized attributes in Table 5 are introduced by ourselves. E.g. *Childhood* signifies the description of the early years of the protagonist. *Alternative life* is the attribute found in “The Jolly Corner” alone. Spencer Brydon’s double in James’s story is the visible embodiment of an alternative life in America that was not chosen by the hero. *End/Beginning* attribute is meant to signify that the finale, or the final event in the development of the plot, is at the same time the beginning of a new phase of the protagonist’s existence: his new life or death. The corresponding matrix of similarity is presented in Table 6.

Table 6 shows that the most similar texts are (“William Wilson,” *Despair*), whereas (“The Jolly Corner,” “The Double”) is the pair of less similar writings in relation to ‘the elements of biography in doubling’. Table 7 and Table 8 present the incidence matrix and the matrix of similarity for the group of characteristics named ‘mysterious’ (gothic).

As in previous Figures, we mark in italics the attributes which we introduced ourselves. E.g., *Secret* signifies the presence of some specifically gothic mystery in a doppelganger text.

There are several pairs of similar texts (the index of similarity equals 4), namely, (“William Wilson,” “The Jolly Corner”), (“William Wilson,” “The Double”), (“The Jolly corner,” “Der Doppeltgänger”), (“The Secret Sharer,” *Despair*), (“L’entrevue,”

Table 8: Matrix of similarity of the doppelgänger texts in relation with ‘mysterious attributes of doubling’.

	WW	JC	SS	L’entr	D(D)	D(H)	De
WW	5	4	2	2	3	3	3
JC	4	5	3	3	2	4	2
SS	2	3	5	1	2	2	4
L’entr	2	3	1	5	2	4	0
D(D)	3	2	2	2	5	1	3
D(H)	3	4	2	4	1	5	1
De	3	2	4	0	3	1	5

“Der Doppeltgänger”). The less similar pairs are (“The Secret Sharer,” “L’entrevue”), (“L’entrevue,” “The Double”), (“Der Doppeltgänger,” *Despair*). The analysis of the similarity between the texts based on all four groups of attributes has revealed two ‘frontrunners’ – “William Wilson” and *Despair*. All other texts of our doppelgänger set are close to these two. Undoubtedly, Poe’s tale had a tremendous influence on the succeeding doppelgänger fictions of American, European, and Russian prose. Its influence has far exceeded that of the earliest doppelgänger text in our set, i.e. Hoffmann’s “Der Doppeltgänger.” The reason of the vital role played by Poe’s doubles is found in a cluster of ‘genetic traits’, in a great collection of doppelgänger devices created by the American romanticist, which was passed on as an international literary heritage (Golovacheva, 2021). The similarity of Nabokov’s novella with other doppelgänger texts has a completely different explanation. All attributes of the double discourse are accentuated and exaggerated there. *Despair* is a palimpsest, the result of Nabokov’s reflection on and playing with the conventions of doppelgänger imagery of his predecessors.

4 Complexity

There still are characteristics which can be calculated in line with standard Q-analysis. Thus, we calculate first structure vector which will allow us to estimate “the complexity” of simplexes. As we stated above, each attribute determines a simplex. The first structure vector \mathbf{Q} for a simplicial complex K has $\dim K + 1$ components, $Q_0, Q_1, \dots, Q_{\dim K}$. The dimension of complex K , $\dim K$, is the maximal dimension of the simplexes that constitute the complex (the dimension of a simplex equals the number of its vertices minus one). The j -th component of \mathbf{Q} is determined as follows (Atkin, 1977; Casti, 1979). The vertices can belong to different simplices. As a consequence, some simplices can have common faces. For instance, in Fig. 2 the simplex “Madness” and simplex “Aggression” have common “SS”, “D(D)”, and “De” vertices. These simplices share the triangle which is a two-dimensional figure, meaning that these simplices are two-connected. Therefore, some simplices can be connected through a chain of other simplices. In such a chain, two neighboring simplices share some q -dimensional face. We say that two simplices σ_1 and σ_2 are q -connected if there is a chain with σ_1 and σ_2 at the ends of the chain, and each pair of neighboring simplices in the chain shares a face whose dimension equals q or higher. A simplicial complex can contain several q -connected chains. The number Q_j of such components (chains) for the given j is precisely the j -th component of structure vector \mathbf{Q} .

In Table 9 we show the first structure vectors for each group of the attributes in the

Table 9: First structure vectors and complexity.

Psychological Attributes	8 7 6 5 4 3 2 1 0 (1 1 1 1 1 1 1 1 1)	1.0
Physical Manifestations	3 2 1 0 (1 3 1 1)	1.6
Elements of Biography	3 2 1 0 (3 2 1 1)	2.1
Mysterious Attributes	3 2 1 0 (2 1 1 1)	1.4

following form $\begin{matrix} dimK & \dots & 1 & 0 \\ (Q_{dimK} & \dots & Q_1 & Q_0) \end{matrix}$

First structure vector serves as a base for further Q-analysis calculations. In particular, we calculate the complexity of simplicial complex by the expression (Casti, 1979)

$$\Psi(K) = 2 \sum_{k=0}^{dimK} \frac{(k+1)Q_k}{(dimK+1)(dimK+2)}.$$

These characteristics represent the structural complexity of a simplicial complex. The calculated values of the complexity are displayed in the third column of Table 9. The ‘simplest’ of all four simplicial complexes turns out to be that of ‘psychological attributes’. This result highlights, we think, the regularity, if not triviality, of such psychological elements of doppelganger discourse as anxiety, rage, etc. The most complicated one appears to be the simplicial complex of ‘biographical attributes of doubling’. We suggest the following explanation: in any of such fictions the protagonist’s life-story is never restricted to the doppelganger episode alone. Any writer would broaden the range of biographical details instead of following the archetypical doppelganger patterns of myth and folklore. Besides, the chosen stories were written within five national and four successive literary traditions – Romanticism, Realism, Symbolism, and Modernism. For this reason, the heroes could not possibly have similarly arranged biographies, as their creators pursued various types of verisimilitude according to the changes in aesthetics.

5 Conclusion

The analysis of the obtained data through critical literary lenses led to a number of unpredicted conclusions related to various topical interconnections between the chosen texts. Our method allowed us to view the doppelganger poetics as a whole and each of the chosen seven texts in particular. For example, it demonstrated that de Régnier’s “L’entrevue” outstands in our set of texts due to its dissimilarity. Actually, this singularity makes one reconsider its nature and see it as a pseudo-doppelganger text: it is not de Régnier’s protagonist but the late owner of a Venetian palace who is doubled, split into a mirror ghostly presence and a marble bust. The analysis of the obtained graphs also revealed the similarity of “William Wilson” and *Despair*. We argue that Poe’s tale is one of the pre-texts of *Despair*, such intertextuality previously neglected in Nabokov studies. Besides, our calculations have revealed the previously undiscovered basic difference between the complexity of simplicial complexes constructed for the previously defined groups of attributes: the most intricate simplicial complex was

obtained not for the ‘psychological’ elements, or the ‘mystical’ (gothic) ones, as could be expected, but for the ‘biographical attributes of doubling’ simplicial complex. As seen from the results of our case-study, at least two methods of graph theory can bring new knowledge into the sphere of literary criticism.

The research was supported by Saint Petersburg University grant “Literary texts and their language: quantitative, corpus and digital methods and their cross-testing (Nabokov in comparative research)” ID 70286404.

References

- Atkin, R. (1977). *Combinatorial Connectivities in Social Systems: An Application of Simplicial Complex Structures to the Study of Large Organizations*. Birkhauser.
- Casti, J. L. (1979). *Connectivity, Complexity, and Catastrophe in Large-Scale Systems*. J. Wiley.
- Golovacheva, I. (2021). From Poe to James via Dostoevsky: Doppelgangers in American and Russian Short Story. In *Connection and Influence in the Russian and American Short Story*, Volume 9781793629883, pp. 35–46. London: Lexington Books.
- Golovacheva, I., A. Stroev, M. Zhuravlev, and P. de Mauny (2018). An Invitation to Mathematical Modelling of Artistic Space in Literary Criticism: Masochism Reconsidered. In *Mathematics as a Bridge Between the Disciplines: Proceedings of MACAS 2017 - Symposium*, Copenhagen, pp. 125–139. University of Southern Denmark.
- Golovacheva, I., M. Zhuravlev, and P. de Mauny (2017). Love Formulas: Mathematical Modelling of Literary Plots. *Novoe literaturnoe obozrenie (in Russian)* 5(147), 151–169.
- Maranda, P. (1973). Cendrillon: Théorie des ensembles et théorie des graphes. In C. Chabrol, *Sémiotique narrative et textuelle*, Volume 1, pp. 122–136. Paris: Larousse.
- Moretti, F. (2007). *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso.
- Moretti, F. (2013). *Distant Reading*. London: Verso.
- O’Toole, L. M. (1980). Dimensions of Semiotic Space in Narrative. *Poetics Today* 1(4), 135–149.
- Rank, O. (2019). *Double: A Psychoanalytic Study*. University of North Carolina Press.
- Rinaldi, S., F. Della Rossa, F. Dercole, A. Gragnani, and P. Landi (2015). *Modeling Love Dynamics*. World Scientific.
- Zhuravlev, M., I. Golovacheva, and P. de Mauny (2015). What Issues of Literary Analysis Can Differential Equations Clarify? *International Journal of Applied Evolutionary Computation* 6(3), 49–63.