Detecting Small Worlds in a Corpus of Thousands of Theater Plays

A DraCor Study in Comparative Literary Network Analysis

Workshop Version – v0.8

2022-08-31

Peer Trilcke
University of Potsdam
trilcke@uni-potsdam.de

Evgeniya Ustinova
Charles University Prague
eugustinova@gmail.com

Frank Fischer
Free University of Berlin
fr.fischer@fu-berlin.de

Ingo Börner
University of Potsdam
ingo.boerner@uni-potsdam.de

Carsten Milling
University of Potsdam
carsten.milling@uni-potsdam.de

Abstract

Although homogenized TEI corpora of plays from different languages are becoming more and more available, research on plays with a comparative angle is still rare in the field of Computational Literary Studies (CLS). This is where approaches of formal network analysis come into play, which have been elaborated in recent years in particular with focus on theater plays. They bear huge potential for comparative research due to their modeling of texts as asemantic structures. An attempt to integrate the paradigm of such a formal analysis with general network research on the one hand and literary history on the other hand is the approach of a typification of networks with respect to the "Small World" concept. However, studies have so far remained limited to smaller and monolingual corpora. In this study, we conceptualize different operationalizations of the "Small World" concept and apply the measures to a larger, DraCor-based corpus of almost 3,000 plays. Looking at the results of these analyses, we examine how the different operationalizations of the "Small World" concept relate to each other and discuss how they could be used for a network-based typology of dramatic forms. We finally develop initial ideas for a network-grounded history of dramatic forms in a transnational perspective.

1 Introduction and Research Agenda

1.1 Literary Network Studies

Modelling and analyzing fictional artefacts – such as epics, novels, movies, or plays – as networks has become a widespread procedure in computational humanities (Labatut and Bost 2019). Accordingly, the extraction of network structures from those artefacts and their analysis has long since become more than 'just' a welcome exercise and optimization task for computer scientists (Elson, Dames, and McKeown 2010; Lee and Yeung 2012; Agarwal, Kotalwar, and Rambow 2013; Krug 2020), but is also being conducted within the scope of decidedly humanities-related research. Thus, for example, the importance of the Hamlet character for the structural stability of the network in Shakespeare's eponymous play was described (Moretti 2011, 4–5) or the role of the Mouse character in *Alice's Adventures in Wonderland* as a connector between many characters (Agarwal et al. 2012, 93–94). The networks of Homer's *Illiad* and of his *Odyssey* were (frequently) investigated (Kydros, Notopoulos, and Exarchos 2015; Miranda, Baptista, and Pinto 2018), and there are studies of the amalgamated network of the *Íslendinga sögur* and its subcomponents (Mac Carron and Kenna 2013) as well as of community structures in *Les Misérables* (Newman and Girvan 2004, 12–13). The network-based distinctions between rural and urban novels were discussed (Elson, Dames, and McKeown 2010; Jayannavar et al. 2015), and network measures were tested as a possible input for genre classification tasks (Coll Ardanuy and Sporleder 2015; Hettinger et al. 2015) and as measures of similarity (Reger 2016). Several papers have also proposed network-based concepts of "protagonism" as well as quantitative classifications of the characters of fictional texts (Park et al. 2013; Algee-Hewitt 2017; Fischer et al. 2018; Krautter et al. 2018).

So, while network analysis has already proven to be insightful for literary studies, some research has also shown that the "structural intuition" (Freeman 2004, 4) of network-based approaches and the related high level of formalization hold particularly high potential for comparative analyses. That is because the abstract modelling and the possibility of a (more or less) 'purely' topological comparison makes the comparative analysis highly independent of any domain- or language-specific origins of the networks. Three systematic approaches can be distinguished here: A comparative approach is possible on the one hand in the sense of *general network sciences*, in which a kind of universal morphology of all possible forms of networks is elaborated. In addition, there are what might be called *cultural* or (more narrowly) *literary network studies*, which focus on the analysis of fictional or even literary works, though against the backdrop of general network science, whose research especially on universal network morphology provides an important theoretical frame of reference. In this sense, networks of literary works and other fictional worlds are regularly discussed with respect to universal topological structures, such as in studies on mythological networks (Carron and Kenna 2012), on the Marvel universe (Alberich, Miro-Julia, and Rossello 2002) or on generative approaches to literary network structures (Moretti 2011).

1.2 Studying Small Worlds (in Literature)

One type of network structure that has driven and inspired comparative research is the so-called "small world network" (Watts and Strogatz 1998; Watts 1999a, 1999b), which differs from regular and random network types by certain topological characteristics and is suspected "to be widespread in biological, social and man-made systems" (Watts and Strogatz 1998, 442). Small world networks seem to be a perhaps universal form that empirical networks tend to take (in contrast to strictly mathematically constructed ones).

It does not come as a surprise that the small world concept is also used in the computational analysis of literary texts. However, these studies are usually limited to stating that the network constructed for the analysed work has topological characteristics similar to small world networks – and is therefore structured like a real world network (Kydros and Anastasiadis 2014; Miranda, Baptista, and Pinto 2018; Mac Carron and Kenna 2013). In contrast, Stiller, Nettle, and Dunbar (2003), guided by an elaborated theory, used the small world concept in their analysis of ten Shakespearean plays, in which they not only proved that the plays can be described as small worlds, but also offered evolutionary hypotheses as an explanation for this specific structure. However, even in this approach, which is further developed in Stiller and Hudson (2005), the focus of the analysis (following the paradigm of evolutionary history) lies on the similarity between real-world networks and small world structured plays.

As interesting as such analyses may be from the perspective of general network science or in light of evolutionary anthropology, from a literary studies point of view they lack at least the reflection of literary form, which has to be considered a major determinant of the networks we extract from literary texts. If, in contrast, one takes into account that the fictional worlds of literature only exist in their mediation through form, there are strong arguments for conceptualizing networks extracted from literary texts first as phenomena of form and not so much as representations of real-world social structures. In this sense, Trilcke (2013, 223–226) has suggested that methods of network science could be applied in literary studies in the context of analyzing form, referring in his own analyses to Klotz's morphological typology of drama (Klotz 1969).

With their corpus-based analysis of more than 450 German-language plays, Trilcke et al. (2016) have taken up this form-oriented approach of literary network analysis and related it to the small world concept. The small world test proposed by them, a procedure for the algorithmic classification of small world characteristics in the texts of a corpus, aims at analyzing the structural composition of plays and thus could be understood as a network-based contribution to the typology of dramatic forms.

In the following, we build on this typologically focused, form oriented research into small world phenomena in literature. Our overall objective is to explore different operationalizations of the small world concept as applied to dramatic texts to better understand the conceptual logics behind these different operationalizations. To this end, we will reimplement the above-mentioned small world test (Trilcke et al. 2016) and apply it to different research corpora, including a very big drama corpus (VeBiDraCor) with nearly 3,000 structurally fairly homogeneous plays, written in more than ten languages. Simultaneously, we will implement the measure of “Small-World-Ness” proposed by Humphries and Gurney (2008), which follows a different conceptual logic than the small world test. Our DraCor-based research corpora (Fischer et al. 2019) as well as both implementations and their corresponding implementations are described in chapter 2. In chapter 3, we will discuss the results of our analysis with regard to the potentials that the small world concept offers for a comparative typology of dramatic forms. Instead of an outlook, we will eventually sketch some ideas about the history of dramatic form from a network science point of view in chapter 4.

2 Operationalization and Corpora

2.1 Operationalizing Small Worlds

Small Worlds, theorized by Watts as a “deep feature of the social world” that constitutes “a family of graphs” (Watts 1999b, 493–494, 502), are usually characterized in research by two properties, each of which is defined relatively to a “regular” and a “random” network type. Phenomenologically described, small world networks (or more precisely: small world graphs) are, like regular networks, “highly clustered”, whereas random graphs are “poorly clustered”; at the same time, small worlds have “small characteristic path lengths, like random graphs” (Watts and Strogatz 1998, 440). The description already specifies the two network measures used for the formalization: the network average clustering coefficient C and the average shortest path length L. We will come back to these two measures in a moment.
Before we can apply network measures to dramatic texts, these texts must be *modeled as networks*, where G is a network with n nodes and m edges. This information is extracted from plays encoded in the TEI format (a subset of XML) as follows:

```
<div type="segment">
    <sp who="#speaker_A">
        <p>text</p>
    </sp>
    <sp who="#speaker_B">
        <p>text</p>
    </sp>
</div>
<div type="segment">
    <sp who="#speaker_B">
        <p>text</p>
    </sp>
    <sp who="#speaker_C">
        <p>text</p>
    </sp>
</div>
```

While each distinct speaker represents a node n, a relation m is established if the speeches `<sp>` of two or more speakers, assigned via the `@who` attribute (`"#speaker_A"`, `"#speaker_B"`) appear in the same segment `<div>`. The algorithms used for extracting our network data are open-source and can be viewed on our GitHub repositories.[1] it is important to note that this method of extracting networks from dramatic texts is only one possibility among many others and that the interpretation potential is limited. At the same time, this extraction mechanism has two merits: first, it can be implemented with a manageable effort, and thus can be applied quite quickly to large sets of texts; second, in line with our research interest, it strongly relies on the specific form of the dramatic text. Last not least, it is of particular importance for our approach that all dramatic networks are constructed in the same way to ensure comparability, which is possible thanks to the homogeneous structure of plays on the DraCor platform (Fischer et al. 2019).

Having extracted dramatic networks from plays in this way, we run a series of small world related analyses on the corresponding data.

First, our small world test (*swt*): Following the conceptualization by Watts and Strogatz (1998), this test relies on the two above-mentioned measures for network topology, the average clustering coefficient C and the average shortest path length L. We calculate C referring to Watts and Strogatz (1998), while L is implemented as a ratio of the sum of average path lengths for each node and the number of such paths.[2]

For our *swt*, we assume that each corpus Cor is a set of n plays P, to each of which corresponds a network G. We first calculate C and L for each G. For each G we then generate 1,000 random graphs (following the Erdös-Rényi model (E–R), (Bollobás 2001)) and calculate the mean for C and L, resulting in C_{rand} and L_{rand}. Next, for each P we calculate C^{dev} and L^{dev} with

$$C^{dev} = \frac{C_G}{C_{rand}} \tag{1}$$

$$L^{dev} = \frac{L_G}{L_{rand}} \tag{2}$$

As stated above, we understand small worlds in drama as a type of dramatic form that is shaped by a characteristic combination of *difference* from (i.e. higher clustering than random networks) and *similarity* to (i.e. average path length similar to random networks) another type. Although with C^{dev} and L^{dev} we have measures of similarity and difference, we do not yet have a procedure for classifying small worlds. Watts and Strogatz (1998) provide no rule as to when C^{dev} is high enough and when L^{dev}

1. The xQuery function of the *DraCor API* (https://github.com/dracor-org/dracor-api) *metrics:get-network-metrics* (https://bit.ly/3wDIJ5Y) extracts the segments of a given TEI file using the function *dutil:get-segments* (https://bit.ly/3AWraAK) and for each of these segments gets the distinct speakers with the function *dutil:distinct-speakers* (https://bit.ly/3CIvDYT). The network metrics are calculated based on these extracted features with the *DraCor metrics service* (https://github.com/dracor-org/dracor-metrics) using the Python package *networkx*.

2. For networks with more than one component, only the paths between connected nodes are counted, so the minimum path is 1.

is low enough to classify a network as a small world. Humphries and Gurney (2008, 2) therefore propose a “continuously graded notion of small-world-ness” and base the categorical concept of small world derived from this small-world-ness on a postulated threshold value (see remarks below).
Our approach to solve this problem is rooted in the domain focus of our study design. We assume that threshold values for the difference (or similarity) between C_G and C_{rand} (or L_G and L_{rand}), as indicated by C^{dev} (or L^{dev}), must be obtained by comparing them to the whole domain, i.e. in our case to the particular corpus studied. So for all P in Cor, we calculate C^{dev} and L^{dev} and average them. For our swt, we classify G as a dramatic small world if the following two criteria are met.

- **Criterion I**: C^{dev} of a single P has to be significantly higher than $\overline{C}^{dev}_{Cor}$ and
- **Criterion II**: L^{dev} of a single P must not differ significantly from $\overline{L}^{dev}_{Cor}$.

To decide if the values are significantly higher resp. do not significantly differ, we calculate a simple deviation test, where we consider anything above or below Mean ± 2 $\times$standard deviation to be a significant deviation. For **criterion I**, this means:

$$C^{dev} > \overline{C}^{dev}_{Cor} + 2\delta \tag{3}$$

And for **criterion II** this means:

$$\overline{L}^{dev}_{Cor} - 2\delta < L^{dev} \ < \overline{L}^{dev}_{Cor} + 2\delta \tag{4}$$

Second, in addition to swt, we carry out a scale-free test (sft), in which we check whether dramatic networks can be typified as scale-free networks following the conceptualization by Albert and Barabási (2002). In our interpretation, scale-free networks fulfill criterion I and criterion II *and* are characterised by having a node degree distribution that follows a power law (**criterion III**). Since strict power-law distributions are rare (Broido and Clauset 2019) especially in networks as small as our plays, we use the following operationalization: For the node degree distribution of each P that meets criterion I and criterion II, we calculate the coefficient of determination R^2 for a) a linear, b) a quadratic, c) an exponential, and d) a power-law fit. When R^2 of the power-law fit is highest, we consider criterion III to be fulfilled and the play has passed the sft.
Both tests, swt as well as sft, follow, as we have already stated, on the one hand a *categorical* logic: the network of a play is either a small world (or scale-free network) or it is not. On the other hand, our categorical attribution, which both tests make, is not an absolute one, but always only possible *relative* to a specific corpus. These two aspects of our operationalization distinguish them fundamentally from the approach of Humphries and Gurney (2008). The measure for “Small-World-Ness” S proposed by them, operates, in its conceptual basis, with a “continuously graded notion of small-world-ness” (Humphries and Gurney 2008, 2), so that a graph, in fact, can be more or less small-world-ish. Only in a second step the authors introduce a threshold value S equals 1 with which their operationalization can also be used for categorical attributions. More significant, however, is the difference between our operationalization and the one

proposed by Humphries and Gurney (2008) when it comes to the relational aspect of the term. In Humphries and Gurney's conceptualization, "Small-World-ness" S is an absolute measure that can be calculated without referring to a reference group (a "corpus" in our case). To be more specific, S – with G as any given graph – is calculated as follows (Humphries and Gurney 2008, 2):

$$\gamma_G = \frac{C_G}{C_{rand}} \tag{5}$$

and

$$\lambda_G = \frac{L_G}{L_{rand}} \tag{6}$$

so that

$$S = \frac{\gamma_G}{\lambda_G} \tag{7}$$

Thus, while our operationalization of *swt* proposes a *categorial-relative* term, Humphries and Gurney's measure of small-world-ness S proposes a *continuous-absolute* term (for which they at the same time indicate options to be applied in a *categorial-absolute* way, resulting in a small world type we will call *swn*). It must be noted, in addition, that Humphries and Gurney's low threshold S 1 for categorical attribution aims to show small worlds as a *widespread, general phenomenon*; our operationalization for use in literary studies, on the other hand, by relying on the significance test "Mean ± 2 $\times$standard deviation" (a test, which primarily targets outliers) conceptualizes small worlds more as a *rare, specific phenomenon.* Correspondingly, we expect a high number of dramatic small worlds of the type *swn*, but relatively few of the type *swt* (and, consequently, even fewer of the type sft).

2.2 Research Corpora

To understand how the different operationalizations behave and what typological and historical conclusions can be drawn from their application, we conduct a series of analyses on different corpora of plays (see Table 1). At the heart of our analyses is *VeBiDraCor* – our very big drama corpus, which we created by aggregating all individual corpora currently available through DraCor.[3] For a more detailed breakdown of the corpora aggregated for *VeBiDraCor* and their basic metadata, see Table 2.

From *VeBiDraCor*, we made two derivations, each of which we used to restrict the heterogeneity of dramatic forms somewhat. *VeBiDraCor_Struc* includes only plays whose number of segments is ≥ 5, so that we exclude plays which, due to their shape (i.e. few segments), tend to have rather exceptional networks. With *VeBiDraCor_Struc_Hist* we

3. *VeBiDraCor* was created using a dedicated, fully functional Docker image of DraCor (incl. metrics services and API functions), which we also use to version the state of the corpora at a given point in time to provide the identical data and API functionalities for use in replication studies. For more info, see "Data and Code Availability".

corpusAcronym	noOfPlays	differentAuthors	yearMin	yearMax	YearMean	YearSD	noOfSpeakersMean	noOfSegmentsMean
VeBiDraCor	2979	797	-472	2017	1719	332.2	14.3	19.9
VeBIDraCor_Stru	2328	702	-472	2017	1702	367.1	15.3	24.9
VeBiDraCor_Struc_Hist	2247	690	1508	2017	1766	91.8	15.4	25.3

Table 1: Research Corpora Overview.

corpusAcronym	corpusName	corpusTitle	noOfPlays	differentAuthors	yearMin	yearMax	YearMean	YearSD	noOfSpeakersMean	noOfSegmentsMean
AlsDraCor	als	Alsatian Drama Corpus	25	13	1816	1937	1914	24.8	12.0	16.2
BashDraCor	bash	Bashkir Drama Corpus	3	2	1917	1975	1952	30.8	18.7	3.7
CalDraCor	cal	Calderón Drama Corpus	205	1	1623	1760	1663	32.6	18.0	2.3
FreDraCor	fre	French Drama Corpus	1556	395	1170	2017	1747	85.6	10.0	19.3
GerDraCor	ger	German Drama Corpus	590	220	1650	1947	1838	57.3	22.3	24.8
GreekDraCor	greek	Greek Drama Corpus	39	4	-472	-388	-425	21.9	11.2	12.7
HunDraCor	hun	Hungarian Drama Corpus	41	25	1558	1970	1860	77.9	24.3	22.9
ItaDraCor	ita	Italian Drama Corpus	139	49	1449	1933	1657	123.2	10.9	26.1
RomDraCor	rom	Roman Drama Corpus	36	3	-215	96	-121	111.0	11.2	16.3
RusDraCor	rus	Russian Drama Corpus	212	56	1747	1947	1848	51.1	17.5	25.1
ShakeDraCor	shake	Shakespeare Drama Corpus	37	1	1591	1613	1600	6.5	38.7	20.8
SpanDraCor	span	Spanish Drama Corpus	25	9	1878	1945	1913	18.8	22.1	33.8
SweDraCor	swe	Swedish Drama Corpus	68	20	1880	1900	1889	5.7	11.6	14.4
TatDraCor	tat	Tatar Drama Corpus	3	1	1908	1912	1910	2.1	10.0	12.0

Table 2: Composition of VeBiDraCor.

add a second step of homogenization by restricting ourselves to plays published or first performed after 1500 (“yearNormalized”[4]).

2.3 Analyses

For all research corpora listed above we

- calculate S;
- perform a categorical application of the small-world-ness measure using the threshold 1 $\rightarrow swn1$ (Humphries and Gurney 2008, 2);
- perform a small-world test $\rightarrow swt$;
- perform a scale-free test $\rightarrow sft$.

Analyses were carried out based on the output of the DraCor API using a specific R script.[5]

3 Discussing Results from a Typological Point of View

3.1 Overview of the Results

In the following, we will present and discuss the results of our analyses and further inspect the received data. In our discussions and inspections, we will focus on the following

4. On calculating the “Normalized Year”, cf. https://dracor.org/doc/faq.
5. The code for the analyses is online, see ”Data and Code Availability”.

guiding questions: First, how do, in particular, the two small-world operationalizations *swn* and *swt* – the first from general network science, the second from literary network studies – relate to each other in quantitative and phenomenological respects? Second, if we assume that a larger or smaller group of the dramatic networks in our corpora are typified as small worlds: Can certain properties be specified for the dramatic networks that do not fall into this group, in other words, can other network-based types of dramatic form be described relative to the group of small worlds?

corpusName	N	S_{mean}	$swn1$	swt	sft
VeBiDraCor	2979	1.68	2270 (76.2%)	79 (2.7%)	32 (1.1%)
VeBiDraCor_Struc	2328	1.80	2071 (89.0%)	70 (3.0%)	31 (1.3%)
VeBiDraCor_Struc_Hist	2247	1.80	1990 (88.6%)	71 (3.2%)	31 (1.4%)

Table 3: Small World Attributions by Measure and Corpus.

Table 3 reports both the mean S as well as the absolute number and the percentage of dramatic networks typified as small world (or scale-free network) in the corresponding corpora according to the different measures. Figure 1 visualizes the relations of the sets of detected small world networks with respect to the different measures.

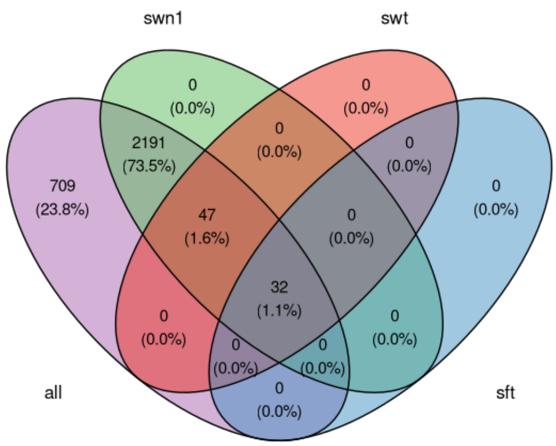

Figure 1: Relations of the Sets of Small Worlds, with "all" = *VeBidraCor* and "swn1", "swt" and "sft" as the corresponding subsets.

Let's first keep three observations:

- As expected, $swn1$ typifies a large part of the corpora as small worlds (between 76.2% and 89%), whereas swt seems to describe only a very specific type of dramatic network (between 2.7% and 3.2%). sft is even more rare.[6] We will discuss the striking dominance of type $swn1$ small worlds in the subsequent chapter 3.2.

- As can be seen in Figure 1 (and as can partly also be mathematically stated a priori), we do have a subset-inclusion order (i.e. a nested set collection), so that *VeBiDraCor* $\supseteq$ *VeBiDraCor_swn1* $\supseteq$ *VeBiDraCor_swt* $\supseteq$ *VeBiDraCor_sft*. In other words, the operationalization swt proposed by us in this context turn out to typify not so much a different type, but more something like an extreme type of a dramatic small world. We will discuss in chapter 3.3 why this extreme type is nevertheless not necessarily identical with plays that have the highest S value.

- Homogenizing the corpora in structural terms (see *VeBiDraCor_Struc*) leads to an increase in S and in the proportion of plays typed as small worlds, but mainly for swn, only slightly for swt. Combining structural and historical homogenization (see *VeBiDraCor_Struc_Hist*), on the other hand, has hardly any effect. We will come back to this observation in chapter 4.

3.2 *swn1*-Small Worlds as a Standard Type of Dramatic Form

Let us begin our discussion with the first observation, of which two aspects are noteworthy. First, according to $swn1$ the small-world-ness of dramatic networks is nothing special, but the rule: Described from a network analytical point of view, small worlds are a (or even: the) standard type of dramatic form. Secondly, however, it is important to note that not all drama networks can be classified as small worlds in the sense of Humphries and Gurney (2008), which raises the question of what other characteristics these networks have.

To recap: Small worlds are a type of network that occupies "a middle ground between regular and random networks", with "high local clustering of elements [...], but also short path lengths between elements" (Humphries and Gurney 2008, 1). Let's calculate the average of S for all 2,270 dramatic $swn1$-small worlds in *VeBiDraCor* and pick the play whose S is closest to this average (which is 1.85), a play called *Die Koralle*, written by Georg Kaiser,[7] which for now shall be taken as our average dramatic $swn1$-small world. If you look at the respective graph (cf. Figure 2), it actually proves: nothing spectacular. *Die Koralle* is a rather average dramatic network, as you get to see it again and again when browsing through DraCor. In this case, we can spot three somewhat 'more' highly clustered areas (top left, top right, and bottom), and also some nodes in the center of the network that connect these areas, ensuring a rather short path length between them. High clustering, short path length: Taken with Humphries and Gurney (2008) in a weaker sense (i.e. with a low threshold of S 1), this is all what small worlds are about, not more – but not less, either.

6. It is not only against this background that it is hard to understand how Moretti (2011, 4) came up with the idea that the "power-law distribution" is "characteristic of all networks".

7. https://dracor.org/id/ger000545.

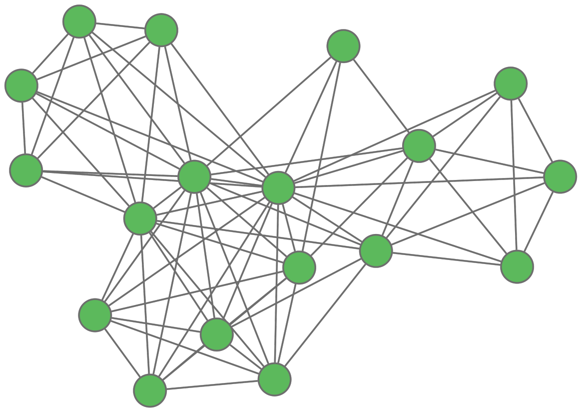

Figure 2: Georg Kaiser, *Die Koralle*, 1917.

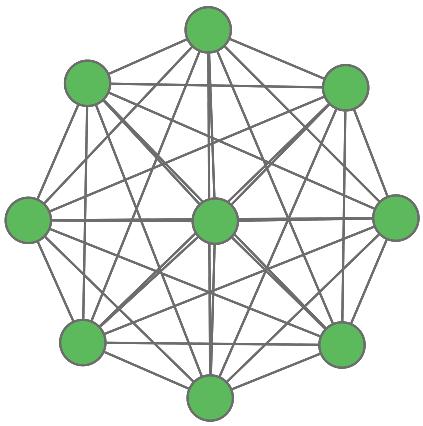

Figure 3: Calderón, *No hay burlas con el amor*, 1635.

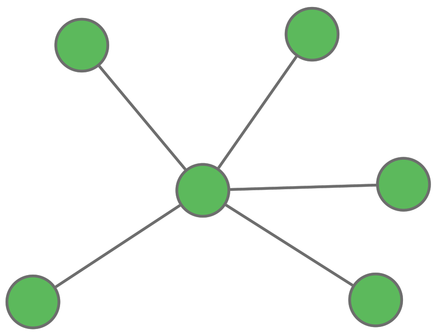

Figure 4: Louis de Boissy, *L'Apologie du Siècle*, 1734.

Now it might be quite obvious to conclude that the ubiquity of the small world phenomenon, as postulated by general network science, is thus also evident in the dramatic networks that can be typified by *swn*1. And indeed, there is nothing to argue against this interpretation. From the point of view of literary network studies, however, there is something essential to add. What manifests itself in an average small world network like *Die Koralle* is first of all an effect of *form*. That is: Highly clustering areas usually emerge as an effect of the form elements *scene* or *act*. And the connection of these areas by central nodes has something to do with the concept of "protagonism", understood as the idea of someone or something carrying the plot that (according to Aristotle) in a play regularly follows the "law of the single strand". In addition, the connections between the areas can also be understood as an effect of the traditional form principle of *liaison des scènes* (even if this principle is not applied in a completely strict interpretation).

Understanding the structure of dramatic networks (also) as effects of form, does not mean to principally reject the views of general network science. It does mean: to speak about dramatic small worlds not as a direct replica of real-world networks, but to always reflect on the literary form as an indispensable mediator of structure, world, society.

And sometimes form dominates (social) structure. After having discussed small worlds as a form-grounded standard type of dramatic networks, we will briefly look at the 709 plays from VeBiDraCor that were not typified as *swn*1-small worlds (see Table 3). Do the non small world plays in *VeBiDraCor* share any characteristics, or is every non small world un-small-world-ish in its own way?

A total of 149 plays from *VeBiDraCor*, for which our analyses did not come to any result at all ("NA") for various reasons,[8] cannot be adequately treated in the context of this question, however. Hence, we set them aside, so 560 plays remain.

Of these, 537 have (at least) two shared characteristics: For their networks, it is both $C = 1$ and $L = 1$, which, third, implies that the density D of the network = 1, which in turn indicates the structural property underlying all this, namely that the graph is fully connected. So, these fully connected (so to say full density) graphs are a second type of dramatic form; Figure 3 shows a randomly selected example, *No hay burlas con el amor*, written by Pedro Calderón de la Barca in 1635.[9] Rarely can such dramatic *full density networks* be understood as a kind of reflection of real-world networks; rather, their maximum clustering is primarily an effect of form, typically caused by specific conventions for structuring the scenes in a play (and often determined by the practical conditions of historical types of theatre stage).

So, 23 plays still remain. Most of them are variants of full density networks, in which only one or very few edges are missing for being fully connected (let's call them *high density networks*). Yet there is another small group that can be typified based on a common characteristic. This group of dramatic non small worlds is the antithesis of the full density networks in that their clustering coefficient is $C = 0$. These are the *dramatic*

8. Almost all reasons have to do with the fact that we have to divide by 0 at some step of the calculation, mostly because the dramatic network is just too small, and as a result L or $C = 0$. Sensibly, Watts also notes that applying the small world concept usually presupposes that a network is "numerically large" (Watts 1999a, 495).

9. https://dracor.org/id/cal000142.

star networks, as shown in Figure 4 using the example of *L'Apologie du Siècle* by Louis de Boissy, written in 1734.[10] All in all, we have identified eleven dramatic star networks (or slight variants of them) in *VeBiDraCor*.[11]

3.3 *swt*-Small Worlds as an Extreme Type of Dramatic Form

While the *swn*1-small world type was found to be the quantitatively dominant standard type of dramatic network, the results of the analysis in Table 3 show that the *swt*-small world type has a very low prevalence. The construction of the *swt* measure using an outlier test may suggest that the *swt* measure simply typifies those dramatic networks for which the highest small-world-ness value S was calculated in the *swn*1 analysis. This is not the case. In fact, the plays typified by *swt* are extreme types of dramatic small worlds, but their identification follows a somewhat different logic than the assignment of very high S values to dramatic networks. In other words, the extreme types typed by *swt* are not necessarily the small-world-iest plays in the sense of S. As is shown in Table 4, among the top 30 dramatic networks ranked by S, there are only eight networks (column *swt* = TRUE) that are also categorized by *swt* as dramatic small worlds.
We, in a first step, choose a phenomenological approach to the different conceptual logics of S and *swt* revealed in the ranking and take a look at two network visualizations. Figure 5 shows the play *Teatral'nyj raz"ezd posle predstavlenija novoj komedii* (*Leaving the Theatre after the Presentation of a New Comedy*) by Nikolai Gogol, published in 1842;[12] Figure 6 shows the first play also typified according to *swt* as a dramatic small world, Christian Dietrich Grabbe's *Hannibal* from 1835.[13] Both dramatic networks are complex; and both dramatic networks are notable at first glance for their numerous semi-autonomous clusters. However, while in Gogol's play the majority of these clusters are unconnected components, in Grabbe's play there are only four separate components in total. Gogol's play thus almost prototypically exemplifies what Watts has called the "caveman graph" (Watts 1999a, 500). Yet a caveman graph is not a small world.
In contrast, the dramatic network for Grabbe's play (Figure 6) also breaks down into several components. Overall, however, its combination of high clustering and short path length (generated by the bridging edges between the components) makes it much more small-world-ish.
So already from phenomenological inspection we can note that our implementation of Humphries and Gurney (2008)'s small-world-ness measure S seems to be quite insensitive to the disintegration of small world candidates into e.g. caveman graphs. But what is the computational reason that Gogol's play has extremely high small-world-ness according to *swn*1 resp. based on S, but is not typified as dramatic small world according to our *swt* measuring?

10. https://dracor.org/id/fre000152.
11. These plays deserve to be studied in more detail. Here is a list of their DraCor IDs: fre000029, fre000036, fre000098, fre000152, fre000223, fre000248, fre000317, fre000574, fre001117, fre001511, tat000001.
12. https://dracor.org/id/rus000053.
13. https://dracor.org/id/ger000393.

rank	id	author	title	yearNormalized	S	$swn1$	swt	C^{dev}	L^{dev}
1	rus000053	Gogol	Teatral'nyj raz"ezd posle predstavlenija novoj komedii	1842	26.9547931	TRUE	FALSE	10.4841483	0.38895302
2	ger000348	Gleich	Der Eheteufel auf Reisen	1821	14.6265535	TRUE	FALSE	8.20040285	0.56065175
3	fre000709	La Tessonerie	L'Art de Régner	1645	12.5818141	TRUE	FALSE	6.67135	0.53023753
4	ger000258	Voß	Faust	1823	11.3673313	TRUE	FALSE	9.56748261	0.8416648
5	fre001169	Quinault	La Comédie sans Comédie	1655	11.3403728	TRUE	FALSE	5.79964646	0.51141586
6	fre000518	Dorimond	La Comédie de la Comédie	1662	10.9781624	TRUE	FALSE	4.92200109	0.44834471
7	fre001014	Molière	Le Malade Imaginaire	1673	10.7359353	TRUE	FALSE	6.35111213	0.59157511
8	ger000393	Grabbe	Hannibal	1835	10.6441972	TRUE	TRUE	10.2893024	0.96665838
9	fre000612	Fagan	Momus à Paris	1770	10.2906717	TRUE	FALSE	4.66630579	0.45345007
10	fre001036	Monselet	L'Enfer des Gens de Lettres	1859	10.2804475	TRUE	FALSE	6.40007548	0.62254833
11	ger000123	Goethe	Götz von Berlichingen mit der eisernen Hand	1773	9.79729685	TRUE	TRUE	9.85566382	1.00595746
12	fre001206	Regnard	Le Carnaval de Venise	1699	9.08480838	TRUE	FALSE	6.53251932	0.71905967
13	fre000427	Dancourt	Nouveau Prologue, et Nouveaux Divertissements	1704	8.81015948	TRUE	FALSE	4.64985302	0.52778307
14	ger000302	Schaefer	Faustine, der weibliche Faust	1898	8.5222114	TRUE	FALSE	6.86318563	0.80532919
15	fre000868	Legrand	La Nouveauté	1727	8.19674483	TRUE	FALSE	5.85271167	0.71402878
16	fre000997	Molière	Le Bourgeois Gentilhomme	1671	8.13815303	TRUE	FALSE	5.89219207	0.7240208
17	ger000085	Büchner	Dantons Tod	1835	7.928771	TRUE	TRUE	8.8197507	1.11237299
18	rus000191	Bulgakov	Vojna i mir	1932	7.70106731	TRUE	TRUE	8.20582204	1.06554348
19	ger000378	Grabbe	Napoleon oder Die hundert Tage	1831	7.67036378	TRUE	FALSE	10.7148581	1.39691655
20	fre000993	Molière	Les Amants Magnifiques	1670	7.45739949	TRUE	FALSE	5.3957967	0.72354937
21	fre000068	d'Aure	Geneviève ou L'Innocence Reconnue Tragédie Chrétienne	1669	7.45371042	TRUE	FALSE	3.90681622	0.52414382
22	shake000037	Shakespeare	Pericles	1609	7.17801132	TRUE	TRUE	6.9280602	0.96517822
23	swe000047	Strindberg	Lycko-Pers resa	1882	7.17420499	TRUE	FALSE	5.81122167	0.81001612
24	ger000201	Goethe	Faust. Der Tragödie zweiter Teil	1832	6.96669002	TRUE	FALSE	8.27049454	1.18714835
25	ger000278	Avenarius	Faust	1919	6.91523967	TRUE	TRUE	6.41246687	0.92729496
26	ger000564	Büchner	Woyzeck	1837	6.86441757	TRUE	FALSE	5.44232487	0.79283127
27	ger000555	Baggesen	Der vollendete Faust oder Romanien in Jauer	1808	6.86145346	TRUE	FALSE	8.70179718	1.26821485
28	ger000532	Wolfram	Faust	1839	6.82541789	TRUE	TRUE	6.23355236	0.91328509
29	ger000149	Sorge	Der Sieg des Christos	1924	6.7601299	TRUE	FALSE	5.34690037	0.79094639
30	fre001180	Quinault	Roland	1685	6.67905609	TRUE	TRUE	6.2438816	0.93484491

Table 4: *VeBiDraCor* Plays Ranked by S, Top 30.

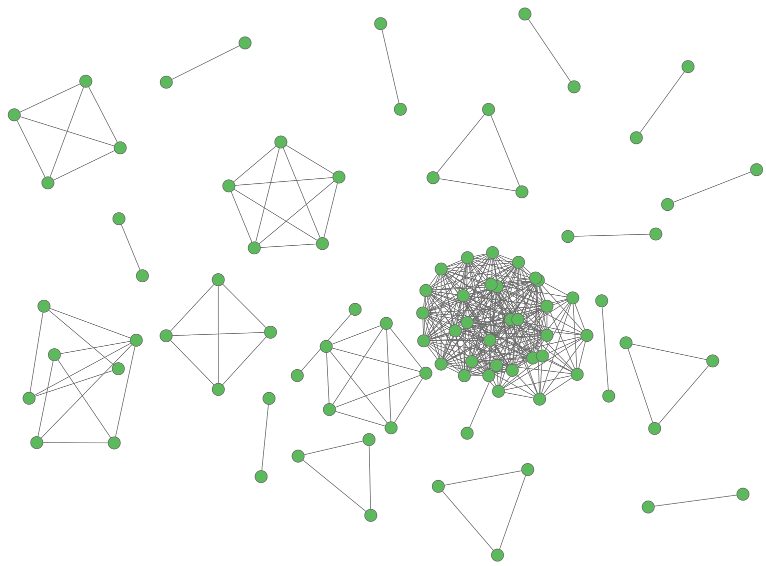

Figure 5: Nikolaj Vasilevich Gogol, *Teatral'nyj raz"ezd posle predstavlenija novoj komedii*, 1842.

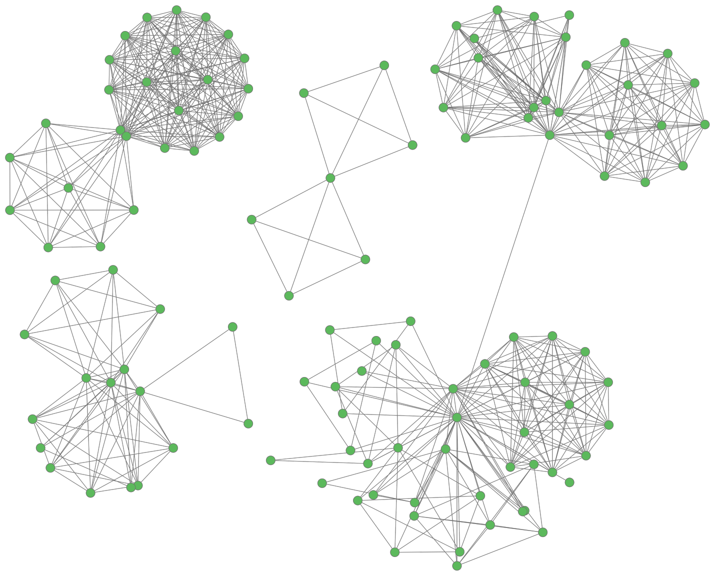

Figure 6: Christian Dietrich Grabbe, *Hannibal*, 1835.

The answer lies in the different way the two small world characteristics (high clustering; short average path length) are implemented in the measures. In our measure, which we constructed as a test where both characteristics are checked independently, both characteristics must be met (see equation 3 and equation 4). Humphries and Gurney (2008), on the other hand, relate the two measures to each other by division to create the integrative measure S (see equation 7). However, this makes the small-world-ness measure S susceptible to peaks of one characteristic, which may allow the other characteristic behaving inconspicuously at the same time. To give an example: On the basis of S, a dramatic network may be classified as small world, if it has an extremely high clustering (criterion I with extreme value fulfilled), but at the same time only a medium path length (characteristic 2 not fulfilled). This exactly is the case with Gogol's play (as well as with many other top-ranked plays according to S, which *swt* does not type as small world): the L^{dev} deviates here too much to pass the criterion II test (see equation 4) – and thus cannot be typified as a dramatic small world according to *swt*.
What follows from this high sensitivity of S to the usefulness of the measure from the point of view of a general network science is not and cannot be discussed here. From a literary network studies perspective, we suggest taking the two measures as a starting point for quite different directions of research. On the one hand, the idea of a measure of small-world-ness may be an interesting starting point for developing a *general morphology of dramatic networks* in which – as outlined in our first draft above – network structures are described and interpreted as effects of form. In this direction, small-world-ness could turn out to be something like *a general, at least transhistorical and transnational form property of dramatic networks*. On the other hand, dramatic networks of the *swt*-small world type offer an approach for *a network-based account to dramatic genres*, that is,

rather *historical forms* that, accordingly, emerge under certain historical conditions – and may disappear again.

4 Instead of an Outlook: Some Thoughts on a Network-Grounded History of Dramatic Form

In a last step, we want to discuss the potentials for research questions on literary history outlined in the considerations just presented at least briefly, namely with reference to a set of diagrams showing historical distributions. We first look at how the different network types that we typified in our analysis (i.e., $swn1$, swt, sft) are distributed historically. Then we examine the historical distribution taking into account the source of the dramatic small worlds from the different DraCor corpora (which are merged in *VeBiDraCor*).

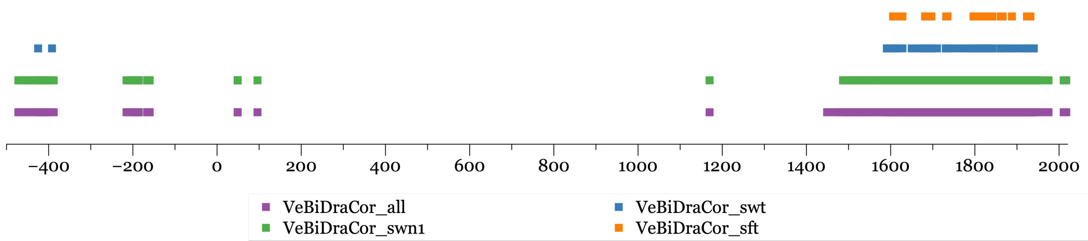

Figure 7: Historical Distribution of Dramatic Small Worlds by Different Measures.

Figure 7 allows a comparison of the historical occurrence of the dramatic networks typed according to different measures. While the purple "baseline" shows all plays (each represented as small square) from VeBiDraCor, the green line represents dramatic small worlds of type $swn1$, the blue line represents dramatic small worlds of type swt, and the orange line represents dramatic networks of type sft. As we have already noted, dramatic small worlds of the $swn1$ type can be understood as a historically more or less indifferent standard type; they describe a transhistorical form option of dramatic networks rather than a specific historical phenomenon.[14] In contrast, dramatic small worlds of the swt type turn out to be a primarily modern phenomenon. With the exception of two plays by Aristophanes (which would be worth a separate study),[15] the first dramatic small worlds of the swt type do not appear until the end of the 16$^{\text{th}}$ century. Not surprisingly, these first dramatic swt small worlds of the modern era are plays by Shakespeare, in specific, the two parts of *Henry VI*.[16] Furthermore, the dramatic scale-free networks of

14. At this point, at the latest, we however must emphasize that due to the composition of our corpus, we remain limited to a Western perspective in our analyses and thus in our findings. We would be very interested in breaking this Western perspective and including other, non-Western traditions of theatrical texts and so begin to argue from a genuinely global point of view.

15. See https://dracor.org/id/greek000027 and https://dracor.org/id/greek000032

16. See https://dracor.org/id/shake000020 and https://dracor.org/id/shake000021.

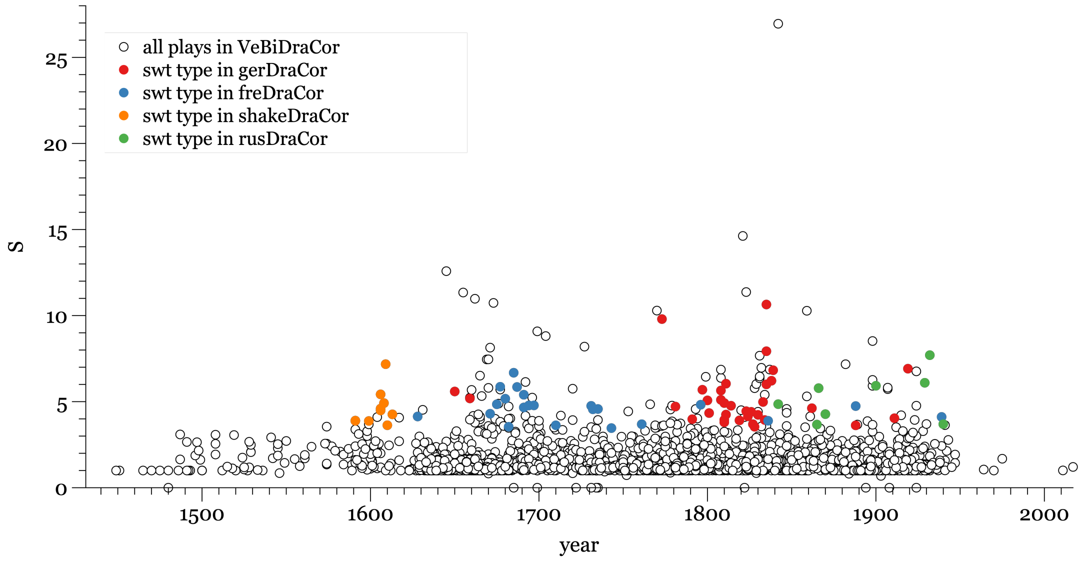

Figure 8: Historical Distribution of all Modern *swt* Type Dramatic Small Worlds Differentiated by Sub-Corpora.

the sft type, additionally characterized by a power-law distribution of the node degree distribution (see criterion III, defined in chapter 2.1), prove to be a genuinely modern phenomenon, at least on the basis of *VeBiDraCor*. Here, too, Shakespeare is the first: the networks of plays like *Timon of Athens*[17] and *Antony and Cleopatra*[18] turn out to be the first examples of dramatic scale-free networks in history. The extraordinary impact of Shakespeare on the history of dramatic form, which is repeatedly attested to him (just to mention studies on the open form in Klotz (1969)), seems to be confirmed by network-grounded analyses.

This, however, is where further research would now have to start, looking much more closely at the structures of these networks, at their realization through dramatic forms and at their relation to dramatic worlds. At this point, we would like to add just one more observation.

Figure 8 shows all dramatic networks from *VeBiDraCor* since 1450, with the small-worldness value S plotted on the y-axis and the year on the x-axis. The diagram suggests one historical hypothesis: Could it be that the form of dramatic networks founded by Shakespeare spread in a wave-like fashion through European literatures? Our analyses, of course, does not provide an adequate answer to this question. In fact, *VeBiDraCor* is far from balanced enough to do so. Moreover, many of the corpora we have brought together in *VeBiDraCor* assemble canonic plays. For a truly comprehensive transnational research on network-based history of dramatic forms, it turns out, even the already quite

17. See https://dracor.org/id/shake000029.
18. See https://dracor.org/id/shake000035.

big *VeBiDraCor* is still a far too small and, besides, not very representative. Future work in the DraCor context will be more devoted to the representativeness of old and new corpora.

Data and Code Availability

Data and code of this study are published on GitHub: https://github.com/dracor-org/small-world-paper. The *Very Big Drama Corpus (VeBiDraCor)* was compiled from data published on DraCor. The conference version relies on data returned from the DraCor API. Calculations were made using a script written in R, which can be found on the GitHub repository in the `conference-version` branch: https://github.com/dracor-org/small-world-paper/tree/conference-version.
The script consists of three parts: reading all necessary data and combining it into one dataframe, intermediate calculations of values for the criteria, and checking of conditions of different tests.
In the final version of the study we will use Docker images of *VeBiDraCor* (see https://github.com/dracor-org/vebidracor) and a research environment (*RStudio*), that will allow for a full reproduction of the study. The current working state can be found in the `development` branch https://github.com/dracor-org/small-world-paper/tree/develop.

Acknowledgements

We warmly thank all the members of and contributors to the DraCor community whose work on the corpora made *VeBiDraCor* – and thus this study – possible in the first place. An overview of all participants can be found at https://dracor.org/doc/credits and, with regard to the TEI files, in the corresponding repositories on GitHub, see https://github.com/dracor-org.
DraCor has, in the context of the project "Computational Literary Studies Infrastructure. CLS INFRA", recently received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004984.

References

Agarwal, Apoorv, Augusto Corvalan, Jacob Jensen, and Owen Rambow. 2012. "Social Network Analysis of Alice in Wonderland." In *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature,* 88–96. Montréal, Canada: Association for Computational Linguistics, June. Accessed August 8, 2022. https://aclanthology.org/W12-2513.

Agarwal, Apoorv, Anup Kotalwar, and Owen Rambow. 2013. "Automatic Extraction of Social Networks from Literary Text: A Case Study on Alice in Wonderland." In *Proceedings of the Sixth International Joint Conference on Natural Language Processing,* 1202–1208. Nagoya, Japan: Asian Federation of Natural Language Processing, October. Accessed August 8, 2022. https://aclanthology.org/I13-1171.

Alberich, R., J. Miro-Julia, and F. Rossello. 2002. *Marvel Universe looks almost like a real social network.* ArXiv:cond-mat/0202174, February. Accessed August 8, 2022. https://doi.org/10.48550/arXiv.cond-mat/0202174. http://arxiv.org/abs/cond-mat/0202174.

Albert, Réka, and Albert-László Barabási. 2002. "Statistical mechanics of complex networks." Publisher: American Physical Society, *Reviews of Modern Physics* 74, no. 1 (January): 47–97. Accessed August 2, 2022. https://doi.org/10.1103/RevModPhys.74.47. https://link.aps.org/doi/10.1103/RevModPhys.74.47.

Algee-Hewitt, Mark. 2017. "Distributed Character: Quantitative Models of the English Stage, 1500–1920." In *Digital Humanities 2017. Conference Abstracts,* 119–121. Montréal: McGill University & Université de Montréal. https://dh2017.adho.org/abstracts/103/103.pdf%E2%80%8B.

Bollobás, Béla. 2001. *Random graphs.* 2nd ed. Cambridge studies in advanced mathematics 73. Series Title: Cambridge studies in advanced mathematics. Cambridge: Cambridge University Press. ISBN: 978-0-521-79722-1 978-0-521-80920-7, accessed August 5, 2022.

Broido, Anna D., and Aaron Clauset. 2019. "Scale-free networks are rare" [in eng]. *Nature Communications* 10, no. 1 (March): 1017. ISSN: 2041-1723. https://doi.org/10.1038/s41467-019-08746-5.

Carron, Pádraig Mac, and Ralph Kenna. 2012. "Universal properties of mythological networks" [in en]. Publisher: IOP Publishing, *EPL (Europhysics Letters)* 99, no. 2 (July): 28002. ISSN: 0295-5075, accessed August 8, 2022. https://doi.org/10.1209/0295-5075/99/28002. https://doi.org/10.1209/0295-5075/99/28002.

Coll Ardanuy, Mariona, and Caroline Sporleder. 2015. "Clustering of Novels Represented as Social Networks" [in en]. *Linguistic Issues in Language Technology* 12 (October). ISSN: 1945-3604, accessed August 8, 2022. https://doi.org/10.33011/lilt.v12i.1379. https://journals.colorado.edu/index.php/lilt/article/view/1379.

Elson, David, Nicholas Dames, and Kathleen McKeown. 2010. "Extracting Social Networks from Literary Fiction." In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics,* 138–147. Uppsala, Sweden: Association for Computational Linguistics, July. Accessed August 8, 2022. https://aclanthology.org/P10-1015.

Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke. 2019. "Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama" [in en]. In *DH2019: »Complexities«. 9–12 July 2019. Book of Abstracts.* Publisher: Zenodo. Utrecht: Utrecht University, July. Accessed July 6, 2021. https://doi.org/10.5281/ZENODO.4284002. https://zenodo.org/record/4284002.

Fischer, Frank, Peer Trilcke, Christopher Kittel, Carsten Milling, and Daniil Skorinkin. 2018. "To Catch a Protagonist: Quantitative Dominance Relations in German-Language Drama (1730–1930)." In *DH2018: »Puentes/Bridges«. 26–29 June 2018. Book of Abstracts / Libro de resúmenes.* Mexico: Red de Humanidades Digitales A. C. https://dh2018.adho.org/to-catch-a-protagonist-quantitative-dominance-relations-in-german-language-drama-1730-1930/.

Freeman, Linton C. 2004. *The development of social network analysis: a study in the sociology of science.* Vancouver, BC: Empirical Press. ISBN: 978-1-59457-714-7, accessed August 7, 2022.

Hettinger, Lena, Martin Becker, Isabella Reger, Fotis Jannidis, and Andreas Hotho. 2015. "Genre Classification on German Novels." In *2015 26th International Workshop on Database and Expert Systems Applications (DEXA),* 249–253. ISSN: 2378-3915. September. https://doi.org/10.1109/DEXA.2015.62.

Humphries, Mark D., and Kevin Gurney. 2008. "Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence" [in en], edited by Olaf Sporns. *PLoS ONE* 3, no. 4 (April): e0002051. ISSN: 1932-6203, accessed August 2, 2022. https://doi.org/10.1371/journal.pone.0002051. https://dx.plos.org/10.1371/journal.pone.0002051.

Jayannavar, Prashant, Apoorv Agarwal, Melody Ju, and Owen Rambow. 2015. "Validating Literary Theories Using Automatic Social Network Extraction." In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature,* 32–41. Denver, Colorado, USA: Association for Computational Linguistics, June. Accessed August 8, 2022. https://doi.org/10.3115/v1/W15-0704. https://aclanthology.org/W15-0704.

Klotz, Volker. 1969. *Geschlossene und offene Form im Drama.* 4. durchges. und überarb. Aufl. Literatur als Kunst. Series Title: Literatur als Kunst. München: Hanser.

Krautter, Benjamin, Janis Pagel, Nils Reiter, and Marcus Willand. 2018. *Eponymous Heroes and Protagonists - Character Classification in German-Language Dramas.* Edited by Thomas Weitin. ISSN: 2629-7027, November. https://www.digitalhumanitiescooperation.de/wp-content/uploads/2019/06/p07_krautter_et_al_eng-1.pdf.

Krug, Markus. 2020. "Techniques for the Automatic Extraction of Character Networks in German Historic Novels" [in eng]. PhD diss., Universität Würzburg. Accessed August 8, 2022. https://doi.org/10.25972/OPUS-20918. https://opus.bibliothek.uni-wuerzburg.de/frontdoor/index/index/docId/20918.

Kydros, Dimitrios, and Anastasios Anastasiadis. 2014. “Social network analysis in literature. The case of The ”Great Eastern”,” vol. Proceedings of the 5th European Congress of Modern Greek Studies of the European Society of Modern Greek Studies, 681–702. Proceedings of the 5th European Congress of Modern Greek Studies of the European Society of Modern Greek Studies. October.

Kydros, Dimitrios, Panagiotis Notopoulos, and Georgios Exarchos. 2015. “Homer’s Iliad – A Social Network Analytic Approach.” Publisher: Edinburgh University Press, *International Journal of Humanities and Arts Computing* 9, no. 1 (March): 115–132. ISSN: 1753-8548, accessed August 8, 2022. https://doi.org/10.3366/ijhac.2015.0141. https://www.euppublishing.com/doi/abs/10.3366/ijhac.2015.0141.

Labatut, Vincent, and Xavier Bost. 2019. “Extraction and Analysis of Fictional Character Networks: A Survey.” *ACM Computing Surveys* 52, no. 5 (September): 89:1–89:40. ISSN: 0360-0300, accessed August 8, 2022. https://doi.org/10.1145/3344548. https://doi.org/10.1145/3344548.

Lee, John, and Chak Yan Yeung. 2012. “Extracting Networks of People and Places from Literary Texts.” In *Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation,* 209–218. Bali, Indonesia: Faculty of Computer Science, Universitas Indonesia, November. Accessed August 8, 2022. https://aclanthology.org/Y12-1022.

Mac Carron, P., and R. Kenna. 2013. “Network analysis of the Íslendinga sögur – the Sagas of Icelanders” [in en]. *The European Physical Journal B* 86, no. 10 (October): 407. ISSN: 1434-6036, accessed August 8, 2022. https://doi.org/10.1140/epjb/e2013-40583-3. https://doi.org/10.1140/epjb/e2013-40583-3.

Miranda, Pedro Jeferson, Murilo Silva Baptista, and Sandro Ely de Souza Pinto. 2018. “The Odyssey’s mythological network” [in en]. Publisher: Public Library of Science, *PLOS ONE* 13, no. 7 (July): e0200703. ISSN: 1932-6203, accessed August 8, 2022. https://doi.org/10.1371/journal.pone.0200703. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0200703.

Moretti, Franco. 2011. *Network theory, plot analysis.* Stanford Literary Lab: Pamphlets 2. https://litlab.stanford.edu/LiteraryLabPamphlet2.pdf.

Newman, M. E. J., and M. Girvan. 2004. “Finding and evaluating community structure in networks.” Publisher: American Physical Society, *Physical Review E* 69, no. 2 (February): 026113. Accessed August 8, 2022. https://doi.org/10.1103/PhysRevE.69.026113. https://link.aps.org/doi/10.1103/PhysRevE.69.026113.

Park, Gyeong-Mi, Sung-Hwan Kim, Hye-Ryeon Hwang, and Hwan-Gue Cho. 2013. “Complex System Analysis of Social Networks Extracted from Literary Fictions.” *International Journal of Machine Learning and Computing,* 107–111. ISSN: 20103700, accessed August 8, 2022. https://doi.org/10.7763/IJMLC.2013.V3.282. http://www.ijmlc.org/index.php?m=content&c=index&a=show&catid=35&id=280.

Reger, Isabella. 2016. “Figurennetzwerke als Ähnlichkeitsmaß” [in deu]. PhD diss., Universität Würzburg. Accessed August 8, 2022. https://doi.org/10/Reger_Isabella_Figurennetzwerke_MA.pdf. https://opus.bibliothek.uni-wuerzburg.de/frontdoor/index/index/docId/14910.

Stiller, James, and Mathew Hudson. 2005. “Weak Links and Scene Cliques Within the Small World of Shakespeare” [in en_US]. Publisher: Akadémiai Kiadó Section: Journal of Cultural and Evolutionary Psychology, *Journal of Cultural and Evolutionary Psychology* 3, no. 1 (July): 57–73. ISSN: 1589-7397, 1589-5254, accessed August 8, 2022. https://doi.org/10.1556/jcep.3.2005.1.4. https://akjournals.com/view/journals/256/3/1/article-p57.xml.

Stiller, James, Daniel Nettle, and Robin I. M. Dunbar. 2003. “The small world of shakespeare’s plays” [in en]. *Human Nature* 14, no. 4 (December): 397–408. ISSN: 1936-4776, accessed August 2, 2022. https://doi.org/10.1007/s12110-003-1013-1. https://doi.org/10.1007/s12110-003-1013-1.

Trilcke, Peer. 2013. “Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft.” In *Empirie in der Literaturwissenschaft,* edited by Philip Ajouri, Katja Mellmann, and Christoph Rauen, 201–247. Münster: Mentis.

Trilcke, Peer, Frank Fischer, Mathias Göbel, and Dario Kampkaspar. 2016. “Theatre Plays as ’Small Worlds’? Network Data on the History and Typology.” In *DH2016. Digital Identities: the Past and the Future,* 385–387. Kraków, Poland: Jagiellonian University, Pedagogical University of Krakow, July. Accessed August 8, 2022. https://doi.org/10.5281/zenodo.6974706. https://zenodo.org/record/6974706.

Watts, Duncan J. 1999a. “Networks, Dynamics, and the Small-World Phenomenon.” Publisher: The University of Chicago Press, *American Journal of Sociology* 105 (2): 493–527. ISSN: 0002-9602, accessed August 2, 2022. https://doi.org/10.1086/210318. https://www.jstor.org/stable/10.1086/210318.

———. 1999b. *Small worlds: the dynamics of networks between order and randomness.* Princeton studies in complexity. Series Title: Princeton studies in complexity. Princeton, NJ: Princeton Univ. Press. ISBN: 978-0-691-00541-6, accessed August 8, 2022. http://www.gbv.de/dms/hbz/toc/ht011180205.pdf.

Watts, Duncan J., and Steven H. Strogatz. 1998. “Collective dynamics of ‘small-world’ networks” [in en]. Number: 6684 Publisher: Nature Publishing Group, *Nature* 393, no. 6684 (June): 440–442. ISSN: 1476-4687, accessed August 2, 2022. https://doi.org/10.1038/30918. https://www.nature.com/articles/30918.