

# Netzwerkanalysen narrativer Texte - ein Vorgehensmodell

**Ketschik, Nora**

nora.ketschik@ims.uni-stuttgart.de  
Universität Stuttgart, Deutschland

## Einleitung

Die soziale Netzwerkanalyse ist in den Computational Literary Studies (CLS) seit mehreren Jahrzehnten als eine Methode etabliert, mit der Figurenbeziehungen in verschiedenen Textgattungen exploriert und analysiert werden. Der Fokus liegt dabei auf dramatischen Untersuchungsgegenständen (vgl. z.B. Szemes und Vida, 2024; Trilcke, 2013; Trilcke et al., 2024; Krautter und Vauth, 2023; Viehhauser, 2023); die Verwendung der Netzwerkanalyse für narrative Texte ist hingegen vergleichsweise selten. Dies liegt m.E. insbesondere daran, dass es um ein Vielfaches voraussetzungsreicher ist, die für Netzwerkanalysen benötigten Daten aus narrativen Texten zu extrahieren, als dies für Dramen der Fall ist. Während Netzwerkanalysen dramatischer Texte i.d.R. auf den Nebentextangaben basieren und daraus szenenbasierte Kookkurrenzen ableiten, bedarf es mehrerer komplexer Schritte, um die gleiche Art der Information aus Erzähltexten zu extrahieren. Die Mehrarbeit resultiert nicht nur aus komplexen Tasks wie Entitätenreferenzerkennung und Koreferenzauflösung, sondern auch aus dem Umstand, dass die relevanten Informationen mit anderen Aspekten auf *Discours*- und *Histoire*-Ebene zusammenhängen können. Beispielsweise ergibt es nur dann Sinn, die Beziehungen zwischen Figuren zu analysieren, wenn man differenzieren kann, welche Figuren auf welcher Erzählebene agieren, oder wenn man weiß, welche Figuren handeln und welche ggf. nur erwähnt werden (etwa, wenn andere Figuren über sie sprechen). Aspekte wie Erzählebenen (Gius et al., 2019b/2021), Erzählerkommentare (Weimer et al., 2022) und Figurenrede (Brunner et al., 2020a/b; Krug et al., 2016) sind als eigene Untersuchungsgegenstände in den CLS durchaus formalisiert, systematisiert und analysiert worden, bisher aber nicht in Workflows für netzwerkanalytische Datenerhebungen integriert worden. Im vorliegenden Beitrag stelle ich daher ein modulares Vorgehensmodell vor, das geeignet ist, kookkurrenzbasierte Netzwerke zwischen Figuren aus Erzählungen zu extrahieren, und dabei andere relevante Aspekte narrativer Texte, die die Netzwerkdaten entscheidend beeinflussen, mitberücksichtigt.

## Die Methode der Netzwerkanalyse in den Computational Literary Studies

Die ursprünglich aus den Sozialwissenschaften stammende Methode der (sozialen) Netzwerkanalyse wird bereits seit vielen Jahren in den CLS für die Analyse von Figurenrelationen, darunter Informations- und Machtstrukturen (Vauth, 2019; Krautter und Vauth, 2023), zur Klassifikation von Figurentypen (Krautter et al., 2020; Vauth, 2023) oder zur Unterscheidung dramatischer Formen (Trilcke, 2013; Szemes und Vida, 2024; Viehhauser, 2023) eingesetzt. Dem Gros der Beiträge ist gemein, dass sie sich dabei erstens auf dramatische Texte und zweitens auf die Szenen- oder Gesprächsstrukturen konzentrieren. Dies liegt u.a. daran, dass die notwendigen Informationen über die Textstruktur leicht greifbar sind: Die szenische Einteilung im Drama gibt Einheiten für Figurenkonfigurationen vor. Eine solch direkte und intuitive Datenerhebung ist für narrative Texte nicht möglich – weswegen es nach wie vor keinen gleichermaßen etablierten Zugang zur Erfassung von Kookkurrenzen zwischen Figuren gibt. Rochat und Kaplan (2014) verwenden beispielsweise einen Index bestehend aus Eigennamen und Seitenzahlen, um kookkurrenzbasierte Figurennetzwerke zu Rousseaus Autobiographie zu erstellen; sie nehmen immer dann eine Figurenrelation an, wenn zwei Figuren in einem Kontext von drei Seiten gemeinsam genannt werden. Argawal et al. (2012) stützen sich hingegen auf ein an Zeitungstexten entwickeltes Konzept von *Social Events*, um Interaktionen in *Alice in Wonderland* zu analysieren; Elson et al. (2010) extrahieren wiederum Konversationsnetzwerke basierend auf Dialog- und Sprecherannotationen in einem englischen Novellenkorpus. Ebenso stützt Vauth (2023) seine Analysen zu Kleist auf manuelle Sprecherannotationen (auch, um die Vergleichbarkeit mit den dramatischen Texten zu gewährleisten); Smeets (2021) erstellt wiederum Kookkurrenznetzwerke, verwendet dafür aber ausschließlich Eigennamen, die er sodann automatisch (aus einem entsprechend großen Korpus) extrahieren kann.

Die Beispiele zeigen, dass eine große Varianz dahingehend besteht, (erstens) welche Arten von Referenzausdrücken, (zweitens) welche Arten von Relationen, und (drittens) auf welche Weise die Informationen erfasst werden. Darüber hinaus werden Kontexte der Figurenvorkommen (z.B. innerhalb vs. außerhalb von Figurenrede) und Einflüsse von *Discours*-Elementen (etwa der Organisation der Erzählung auf verschiedenen Erzählebenen) mehrheitlich ignoriert. Dies führt zum einen zu undifferenzierten Netzwerken und verhindert zum anderen die Vergleichbarkeit zwischen Netzwerken über verschiedene Ansätze hinweg. Reflektiert werden die Auswirkungen verschiedener Datengrundlagen ansatzweise bei Jannidis et al. (2016), die an einem Novellenkorpus Möglichkeiten zur Evaluation automatisch erstellter Figurennetzwerke diskutieren,<sup>1</sup> sowie bei Andresen und Vauth (2018), die für einen Text ermitteln, wie die Art des Referenzausdrucks (Eigename/Appella-

tiv/Pronomen) und sein Kontext (in/außerhalb von direkter Rede) dessen Frequenz beeinflusst.<sup>2</sup>

Es ist anzunehmen, dass die fehlende Integration verschiedener, für die Erfassung von Figurenvorkommen und -relationen relevanter Aspekte in der Komplexität dieser Teilaufgaben begründet liegt. So wäre – analog zum Drama – die Einteilung eines Textes in „Szenen“, wie sie Gius et al. (2019a) vorschlagen, ein naheliegender Schritt für die Extraktion von Kookkurrenzen, der aber seinerseits extrem voraussetzungsreich ist, da er auf komplexen Kategorien wie Raum und Zeit/Chronologie aufbaut.<sup>3</sup> Auch Veränderungen in der Figurenkonfiguration (die nach Gius et al., 2019a ebenfalls einen Szenenwechsel indizieren) können nur identifiziert werden, wenn die Kontexte der Figurennennungen bereits bekannt sind. Wird beispielsweise lediglich über eine andere Figur gesprochen oder an eine andere Figur gedacht, führt dies nicht zu einer Veränderung in der Figurenkonfiguration im Sinne der Szenendefinition. Netzwerkanalysen zu narrativen Texten fußen damit – derzeit – entweder auf intensiver manueller Annotationsarbeit oder auf einer unvollständigen Datengrundlage. Umso wichtiger erscheint es, kritisch mit den möglicherweise lückenhaften Daten umzugehen und deren Auswirkungen auf die resultierenden Netzwerke abschätzen zu können.

## Textstatistiken im Kontext von Figurennetzwerken

Das hier vorgestellte Vorgehensmodell zur Extraktion von Kookkurrenznetzwerken aus narrativen Texten baut auf Analysen zu mittelhochdeutschen Romanen (Ketschik, 2024) auf und wurde im Rahmen des vorliegenden Beitrags an anderen narrativen Texten in neuhochdeutscher Sprache weiterentwickelt und erprobt. Für die Analysen und Statistiken wurden Texte aus dem Deutschen Romankorpus (DROC, Krug et al., 2018) verwendet, in denen Annotationen zu Figuren (inkl. Koreferenzresolution) und Figurenrede enthalten sind. Die Textstatistiken vermitteln einen Eindruck davon, inwiefern die Wahl des Figurenreferenzausdrucks, die Segmentierungsgröße und die Kontexte von Figurenreferenzen die resultierenden Netzwerke beeinflussen.

Text (Textauszug entspricht DROC)	Anzahl Tokens	Anzahl Figurenreferenzen	Eigenname (%)	Appellativ (%)	Pronomen (%)
Kafka_Amerika	2722	341	9,97	19,06	70,97
Dickens_OTwist	4094	545	21,83	20,92	57,25
Fontane_GMinde	3872	514	18,68	13,62	67,70
Fontane_MMöhring	3386	426	15,73	11,27	73,00
Raabe_Stopfkuchen	3252	293	10,92	19,45	69,62
Ahlefeld_Erna	6230	803	11,21	11,83	76,96
Aston_AusdemLeben	3546	436	11,93	20,87	67,20
Balzac_GlanzEiend	5689	795	19,12	24,40	56,48
Ebner-Eschenbach_Gemeindekind	6248	896	10,94	20,87	68,19
Verne_20000Meilen	2070	264	9,85	15,15	75,00
Durchschnitt	4110	531	14	18	68

Tabelle 1: Texte, Anzahl der Figurenreferenzen (absolut) und Art der Referenz (in Prozent).

Tabelle 1 zeigt, dass Eigennamen mit durchschnittlich 14% und Werten zwischen knapp 10% und gut 20% den geringsten Anteil der Figurenreferenzen im Textkorpus ausmachen, gefolgt von Appellativen mit durchschnittlich 18%. Die häufigsten Referenzausdrücke sind Pronomina, wobei ihr Anteil zwischen 57% und 77% liegt. Homodiegetische Erzählungen (hier *Stopfkuchen* und *Zwanzigtausend Meilen unter 'm Meer*) führen nicht zwangsläufig zu einem höheren Anteil pronominaler Referenzen gegenüber Texten mit heterodiegetischer Erzählerinstanz.

Die Verteilung ist aber innerhalb eines Textes figurenspezifisch. In *Oliver Twist* wird die Hauptfigur beispielsweise am häufigsten mit Pronomen, am zweithäufigsten mit Eigennamen und nur selten mit Appellativen benannt, für Jack hingegen gibt es zwar ebenfalls überwiegend pronominale Referenzausdrücke, aber häufiger Appellative als Eigennamen; bei Mr. Bumble wiederum sind nur knapp 50% der Referenzen pronominal, die Mehrheit hingegen Eigen- und Gattungsnamen (s. Anhang). Tendenziell haben Nebenfiguren mit niedriger Frequenz weniger pronominale Referenzen als Hauptfiguren (s. Anhang). So eine Ungleichverteilung wird immer dann relevant, wenn **nicht** alle Referenzausdrücke für die Extraktion der Entitäten berücksichtigt werden. Wenn also – oftmals aus Gründen der schnelleren Operationalisierbarkeit – auf pronominale Referenzen verzichtet wird, führt dies bei der Erfassung von Figurenvorkommen und -relationen zu Verzerrungen (tendenziell dahingehend, dass Nebenfiguren begünstigt und Hauptfiguren vernachlässigt werden). Darüber hinaus ist zu betonen, dass reine Named-Entity-Ansätze manche Figuren gar nicht erfassen können (insbesondere Figuren, die über ihre Rollen benannt sind, z.B. Vater/Arzt).

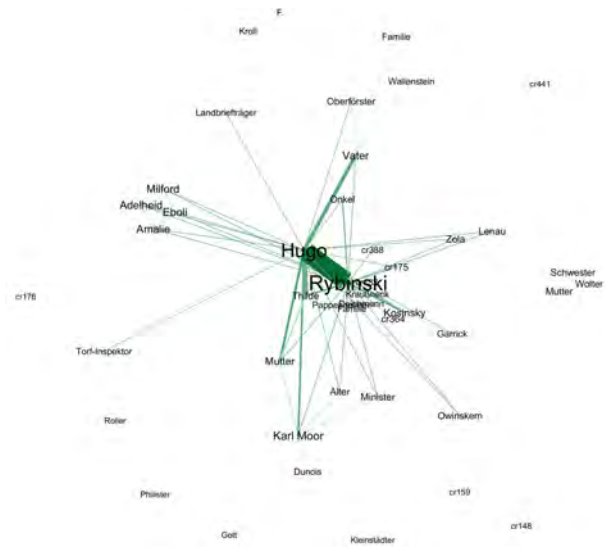


Abbildung 1: Netzwerk zu Fontanes *Mathilde Möhring* basierend auf allen Referenzausdrücken.

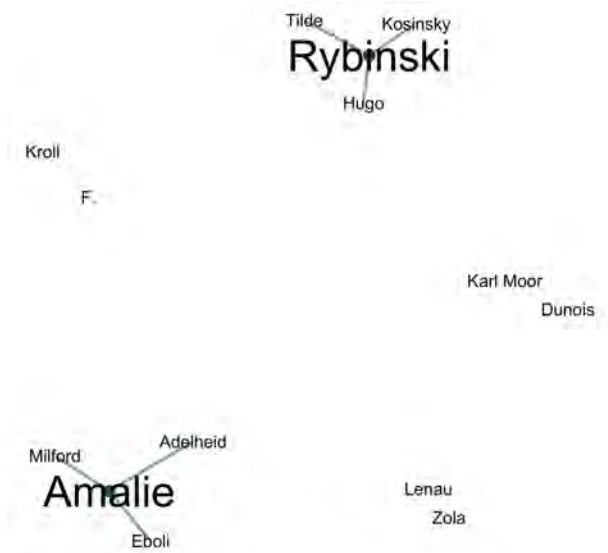


Abbildung 2: Netzwerk zu Fontanes *Mathilde Möhring* basierend auf Eigennamen.

Die Wahl des Referenzausdrucks hat konsequenterweise gravierende Auswirkungen auf das resultierende Netzwerk. Exemplarisch seien zwei Netzwerke zu einem Auszug aus Fontanes *Mathilde Möhring* gezeigt (Kookkurrenzen basieren auf Satzgrenzen): Abbildung 1 berücksichtigt alle Referenzausdrücke, Abbildung 2 nur Eigennamen.<sup>4</sup> Letztere umfasst, wie erwartbar, nur Bruchteile des Figureninventars. Besonders kritisch ist, dass eigentlich zentrale Figuren wie Hugo irrelevant und umgekehrt Randfiguren wie die nur erwähnte Amalie zentral erscheinen. Die verzerrenden Effekte, die mit einer solchen Restriktion auf Eigennamen einhergehen, fließen in Berechnungen wie Zentralitätsmaße ein (Tabelle 2):

Figur	Gradzentralität Netzwerk 1 (alle Referenzausdrücke)	Gradzentralität Netzwerk 2 (nur Eigennamen)
Rybinski	24	3
Hugo	21	1
Karl Moor	6	1
Amalie	5	3
Vater	5	-
Thilde	3	1

Tabelle 2: Gradzentralität für ausgewählte Figuren aus *Mathilde Möhring*.

Neben der Wahl der Referenzausdrücke hat die Segmentgröße als Grundlage für Kookkurrenzen einen entscheidenden Einfluss auf die Netzwerke. Die Größe der Segmentierung hängt (auch) mit der Wahl der Referenzausdrücke zusammen. Werden z.B. keine Pronomina berücksichtigt, ist es sinnvoll, die Segmente zu vergrößern. Welche Segmentgröße angemessen ist, lässt sich nicht pauschal beantworten. Hinweise können die Abstände zwischen Pronomen und Antezedent geben, da etwa ein Netzwerk ohne pronominalen Referenzen den „Weg“ bis zur nächsten nicht-pronominalen Referenz (Appellativ oder Eigennamen) überbrücken muss. Die Abstände sind wiederum stark textab-

hängig (und figurespezifisch), beispielsweise stehen in Fontanes *Mathilde Möhring* die Pronomen im Durchschnitt 130 Tokens von ihrem nicht-pronominalen Antezedenten entfernt, in Dickens *Oliver Twist* hingegen 37 Tokens (Tabelle 3).

Text (Textauszug entspricht DROC)	Abstand (in Anzahl Tokens), Median	Abstand (in Anzahl Tokens), Mittelwert
Kafka_Amerika	43	71
Dickens_OTwist	24	37
Fontane_GMinde	31	55
Fontane_MMöhring	80	130
Raabe_Stopfkuchen	49	157
Ahlefeld_Erna	74	135
Aston_AusdemLeben	38	48
Balzac_GlanzElend	21	37
Ebner-Eschenbach_Gemeindekind	30	55
Verne_20000Meilen	61	192

Tabelle 3: Abstände vom pronominalen Referenz Ausdruck zum vorangehenden Appellativ oder Eigennamen.

Das Beispielnetzwerk zu *Mathilde Möhring* lässt weitere Fallstricke für Netzwerkanalysen narrativer Texte erkennen. Erstens wird evident, dass Figuren im Netzwerk enthalten sind, die nicht Teil der fiktiven Welt sind, Karl Moor etwa ist Schillers *Räubern* entsprungen. Zweitens werden Figuren im Netzwerk abgebildet, die in der Textstelle (ein Dialog zwischen Hugo und Rybinski) nicht agieren; dies betrifft z.B. Hugos Vater, nach dem Rybinski sich erkundigt, der aber nicht anwesend und zum Zeitpunkt der Handlung bereits verstorben ist. Als weitere Kontexte von Figurenreferenzen sind neben Figurenrede Erzählerkommentare und narrative Ebenen zu berücksichtigen.

Die fehlende Differenzierung zwischen Kontexten der Figurennennungen führt nicht nur dazu, dass im Netzwerk ggf. Relationen zwischen Figuren visualisiert werden, die in der Handlung nie kookkurrieren, sondern auch, dass (etwa innerhalb von Figurenrede) erwähnte Figuren die Netzwerkmetriken (z.B. die Netzwerkgröße, Zentralitätswerte, Dichtemaße) in hohem Maß beeinflussen. So umfasst das Netzwerk zum *Mathilde-Möhring*-Ausschnitt lediglich drei Figuren, wenn erwähnte Figuren herausgefiltert werden.

## Vorgehensmodell

Die verschiedenen Aspekte, die für eine reflektierte Netzwerkanalyse von narrativen Texten relevant sind, werden nun in einem Vorgehensmodell zusammengefasst (Abb. 3). Das Modell kann unabhängig davon eingesetzt werden, ob die Schritte durch manuelle Annotation oder (teil-)automatisch umgesetzt werden. Bestimmte Arbeitsschritte (v.a. III–V) können je nach Textigenschaften oder Untersuchungsfrage ggf. wegfallen. Die einzelnen Schritte werden im Folgenden kurz erläutert.



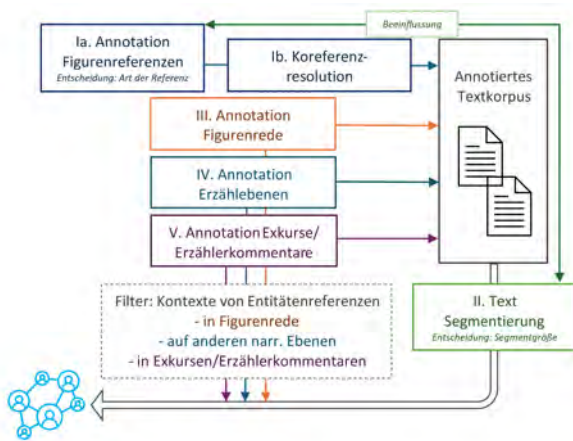


Abbildung 3: Vorgehensmodell zur Erstellung von Figuren-Kookkurrenznetzwerken aus narrativen Texten.

### I. Figurenreferenzannotation und Koreferenzresolution:

Um Figuren (oder andere Entitäten)<sup>5</sup> in Erzählungen zu erfassen, müssen diejenigen Ausdrücke, die auf eine Figur referieren, manuell oder (teil-)automatisch im Text annotiert werden. Referenzausdrücke lassen sich in „Eigennamen“, „Appellativ“ und „Pronomen“ einteilen. Koreferenzen müssen aufgelöst werden. Wenn es darum geht, Figurenpräsenzen zu finden, ist darauf zu achten, dass Referenzausdrücke verschachtelt sein können und eingebettete Ausdrücke herauszufiltern sind (z.B. [[Hugos] Frau] = Mathilde).

**II. Textsegmentierung:** Um Kookkurrenzen zwischen Figuren zu erfassen, muss ein Textfenster vorgegeben werden. Die Textsegmentierung kann auf Grundlage von Layout-Informationen (z.B. Absätzen, linguistischen Einheiten (z.B. Sätzen) oder narratologischen Einheiten („Szenen“ nach Gius et al., 2019a) erfolgen. Für Kookkurrenznetzwerke gibt die Textsegmentierung den Rahmen vor, in dem Relationen zwischen zwei oder mehr Figuren erfasst werden. Hierbei gilt grundsätzlich: Je größer das Segmentierungsfenster ist, desto mehr Kookkurrenzen werden erfasst, desto unschärfer werden aber auch die Treffer (mehr *false positives*). Darüber hinaus ist zu bedenken, dass ein Zusammenhang zwischen der Segmentierungsgröße und der Vollständigkeit der annotierten Referenzausdrücke besteht. Während es geeignet erscheint, satzbasierte Kookkurrenzen zu extrahieren, wenn sämtliche (Ko-)Referenzen erfasst sind, müssen größere Segmente gewählt werden, wenn keine pronominalen Referenzen annotiert sind. Korpusstatistiken zu den Abständen zwischen den verschiedenen Referenzausdrücken können helfen, ein geeignetes Kookkurrenzfenster festzusetzen.

**III. Figurenrede (Filter, optional):** Grundsätzlich ist es sinnvoll, präsente bzw. handelnde Figuren von solchen, über die „nur“ gesprochen wird, zu unterscheiden. Zu diesem Zwecke sind Passagen, die Redewiedergabe (dazu zählen direkte und indirekte Formen der Rede- und Gedankenwiedergabe) enthalten, entsprechend auszuzeichnen. Dies ermöglicht es, Referenzausdrücke, die innerhalb dieser Passagen liegen, herauszufiltern, wenn Kookkurrenzen zwi-

schen präsenten Figuren erfasst werden sollen. Im obigen Netzwerk zu *Mathilde Möhring* sind lediglich drei Figuren präsent, wohingegen die meisten anderen (darunter z.B. Hugos verstorbener Vater) ausschließlich in Figurenrede erwähnt werden. Gleiches gilt für Milada im *Gemeindekind*, s. Anhang.

**IV. Erzählebenen (Filter, optional):** Narrative Ebenen müssen grundsätzlich auseinandergehalten werden, da es sonst zu einer „Vermischung“ des Figureninventars verschiedener Ebenen kommt. Binnenerzählungen können z.T. über Figurenrede (Schritt III) erfasst werden. Eine Annotation der Erzählebenen ermöglicht es, die Figurenrelationen ebenenspezifisch zu analysieren, indem Figurenreferenzen aus anderen Erzählebenen herausgefiltert werden. Im *Stopfkuchen* besteht beispielsweise die Rahmenhandlung aus Eduard, der sich auf der Schifffahrt nach Kappstadt befindet, wohingegen (mehrere) eingebettete Erzählebenen Figuren aus seiner Vergangenheit und Heimatstadt umfassen. Die Ebenen sind räumlich und zeitlich voneinander getrennt.

**V. Exkurse/Erzählerkommentare (Filter, optional):** Unter Exkursen sind „Abschweifungen“ (Kommentare, Publikumsanreden) zu verstehen, die i.d.R. mit einer veränderten Sprechhaltung einhergehen und den Verlauf der Erzählung unterbrechen. Für Netzwerkanalysen ist die Annotation von Exkursen insbesondere dann relevant, wenn sie Referenzen auf Figuren/Personen umfassen, die **nicht** Teil der erzählten Welt sind. Oftmals wird im Rahmen von Exkursen das Erzählte kommentiert (also Vergangenes aufgegriffen), es werden Vergleiche zu anderen Figuren gezogen, intertextuelle Verweise oder realhistorische Bezüge hergestellt. Exkurse können damit Referenzen auf Entitäten enthalten, die i) zwar Teil der erzählten Welt, aber nicht zwangsläufig Teil der „Szene“, ii) Teil anderer fiktiver Texte, iii) Teil der realen Welt sind. Eine Annotation der Exkurse ermöglicht eine Differenzierung zwischen den verschiedenen „Status“ der Entitäteninstanzen.

**Sonstiges:** Je nach Organisation der Erzählung und Untersuchungsfrage können andere Aspekte, die das abgebildete Modell nicht berücksichtigt, relevant sein. Bei anachronischen Erzählungen kann es etwa sinnvoll sein, die Erzählung zunächst in eine chronologische Reihenfolge zu bringen, insbesondere wenn dynamische Netzwerke erstellt werden.

## Schluss

Das hier vorgestellte Vorgehensmodell umfasst relevante Schritte für die Extraktion kopräsender Figuren aus narrativen Texten. Ein Hauptanliegen des Beitrags ist es, aufzuzeigen, welche Aspekte bei der Netzwerkanalyse narrativer Texte (potenziell) eine Rolle spielen, und diese in die Datenerhebung zu integrieren – oder, sollte dies nicht möglich sein, zumindest das Bewusstsein für deren Einfluss auf die netzwerkanalytischen Daten zu schärfen. Zweifels- ohne können nicht alle Sonderfälle in einem möglichst generischen Modell berücksichtigt werden; vielmehr dient

das Modell als methodische Grundlage, die dazu befähigt, Kookkurrenznetzwerke aus Erzähltexten zu erstellen, die für Einzelfälle aber angepasst oder ergänzt werden muss. Durch seine Modularität und die „Filter“-Schritte soll das Modell für viele Fragestellungen und Untersuchungsgegenstände einsetzbar sein.

## Anhang

Tabelle 4: Textstatistiken zur Verteilung von Figurenreferenzausdrücken und zu Abständen zwischen koreferenten Referenzausdrücken. Berechnungen je Text-Datei und für ausgewählte Figuren. Die Textauszüge und Annotationen entsprechen DROC (Krug et al., 2018).

Text (Textauszug entspricht DROC) und ausgewählte Figuren	Verteilung von Figurenreferenzen nach Art der Referenz (Eigennamen, Appellative, Pronomen), Figurenvorkommen innerhalb und außerhalb von Redewiedergabe-Passagen.							Abstände (in Tokens) von Pronomen zu koreferentem Appellativ oder Eigennamen	
	Gesamt	Eigenna-me	Appellat-iv	Pron-	In Red-e	Außerhal-b Rede	Median	Mittelwer-t	
Kafka_Amerika	341	34	65	242			43	71	
cr1_Karl	186	29	5	152	70	116	45	62	
cr16_Franz	16	3	6	7	8	8	7	6	
Dickens_OTwist	545	119	114	312			24	37	
cr1_Oliver	246	61	17	168	68	178	26	42	
cr314_Jack	72	5	26	41	13	59	24	28	
cr10_Bumble	43	17	6	20	22	21	14	17	
Fontane_GMinde	514	96	70	348			31	55	
cr1_Grete	125	34	10	81	82	43	26	38	
cr2_Gigas	71	8	8	55	39	32	43	81	
cr3_Trud	69	22	5	42	44	25	31	55	
Fontane_MMöhring	426	67	48	311			80	130	
cr1_Hugo	140	13	3	124	111	29	101	161	
cr70_Vater	33	0	11	22	33	0	29	30	
cr5_Rybinski	132	18	5	109	122	10	97	126	
Raabe_Stopfkuchen	293	32	57	204			49	157	
cr1_Eduard	111	5	10	96	26	85	81	218	
cr58_Stopfkuchen	21	12	3	6	7	14	17	15	
Ahlefeld_Erna	803	90	95	618			74	135	
cr1_Alexander	281	22	12	247	16	265	100	160	
cr7_Erna	234	29	8	197	52	182	92	132	
cr12_Arzt	13	0	7	5	1	12	25	42	
Aston_AusdemLeben	436	52	91	293			38	48	
cr1_Oburn	195	26	19	149	58	137	36	44	
cr5_Frau	123	12	23	88	53	70	42	58	
cr163_Buchhalter	16	3	5	7	9	7	36	33	
Balzac_GlanzElen	795	152	194	449			21	37	
cr1_Jakob	65	13	10	42	25	40	20	33	
cr7_Asien	260	43	39	178	89	171	27	37	
cr99_Romette	18	4	5	9	9	9	12	19	
Ebner-Eschenbach_Gemeindekind	896	98	187	611			30	55	
cr1_Pavel	434	64	41	329	161	274	30	60	
cr10_Lehrer	95	4	36	55	22	73	25	118	
cr62_Milada	35	7	14	14	34	1	14	18	
Verne_2000Meilen	264	26	40	198			61	192	
cr1_ich (Eduard)	113	4	8	101	34	79	79	211	
cr_14_Nemo	73	10	16	47	41	32	19	30	
cr_2_Conseil	14	7	1	6	1	13	82	83	

## Fußnoten

1. Auch dort werden aber nur die Aufgaben „Identifikation von Figuren“ (inkl. Koreferenzresolution) und „Identifikation von Relationen“ (z.B. über Kookkurrenz) zur Erstellung von Figurennetzwerken berücksichtigt – Beeinflussungen der Daten durch Erzählebenen, Exkurse/Kommentare etc. bleiben unbeachtet.
2. Andresen und Vauth stellen fest, dass die Verteilungen für verschiedene Figuren des untersuchten Texts relativ stabil sind; dieser Befund ist wohl nicht generalisierbar, da die hier beigefügten Statistiken zeigen, dass etwa die Art der Referenzierung (die Verteilung von Eigennamen/Appellativen/Pronomen) je nach Figur stark variieren kann. Statistiken zu einem englischsprachigen Korpus liefern Bamman et al. (2020), die Verteilungen von Entitätenklassen und verschiedenen Referenzierungstypen sowie Abstände zu Antezedenten berechnen (allerdings nicht figurenspezifisch).
3. Auch die Herausforderungen einer automatischen Szenen-Annotation zeugen von der Komplexität dieser Aufgabe, vgl. Zehe et al., 2021. Zum Begriff des „Segments“ und verschiedenen Segmentierungsansätzen in der DH vgl. Bartsch et al., 2023.
4. Die Netzwerke wurden mit *Gephi* (Bastian et al., 2009) erstellt. Die Größe der Knoten und ihrer Beschriftung richtet sich nach dem Grad, die Dicke der Kanten nach dem Gewicht.
5. Das hier vorgestellte Vorgehensmodell kann auch auf andere Entitätenklassen (z.B. LOC/ORG) übertragen werden. Für Richtlinien zur Entitätenreferenzannotation vgl. Reiter et al., 2023.

## Bibliographie

- Agarwal, Apoorv, Augusto Corvalan, Jacob Jensen und 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. <https://aclanthology.org/W12-2513> (zugegriffen: 19.11.2024).
- Andresen, Melanie und Michael Vauth. 2018. “Added Value of Coreference Annotation for Character Analysis in Narratives.” In *Proceedings of the Workshop on Annotation in Digital Humanities*, 1–6. <http://ceur-ws.org/Vol-2155/andresen.pdf> (zugegriffen: 19.11.2024).

- Bamman, David, Olivia Lewke und Anya Mansoor.** 2020. "An Annotated Dataset of Coreference in English Literature." In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, 44–54. <https://aclanthology.org/2020.lrec-1.6> (zugegriffen: 19.11.2024).
- Bartsch, Sabine, Evelyn Gius, Marcus Müller, Andrea Rapp und Thomas Weitin.** 2023. "Sinn und Segment. Wie die digitale Analysepraxis unsere Begriffe schärft." *Zeitschrift für digitale Geisteswissenschaften* 8 (2023). [https://doi.org/10.17175/2023\\_003](https://doi.org/10.17175/2023_003) (zugegriffen: 19.11.2024).
- Bastian, Mathieu, Sebastien Heymann und Mathieu Jacomy.** 2009. "Gephi: An Open Source Software for Exploring and Manipulating Networks." In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, 361f. <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154> (zugegriffen: 19.11.2024).
- Brunner, Annelen, Stefan Engelberg, Fotis Jannidis, Ngoc Duyen Tanja Tu und Lukas Weimer.** 2020a. "Corpus REDEWIEDERGABE." In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, 796–805. <https://aclanthology.org/2020.lrec-1.100> (zugegriffen: 19.11.2024).
- Brunner, Annelen, Stefan Engelberg, Fotis Jannidis, Ngoc Duyen Tanja Tu und Lukas Weimer.** 2020b. "Redewiedergabe in Heftromanen und Hochliteratur." In *7. Tagung des Verbands Digital Humanities im deutschsprachigen Raum e.V. (DHd 2020)*, 190–194. <https://doi.org/10.5281/zenodo.4621814> (zugegriffen: 19.11.2024).
- Elson, David, Nickolas Dames und Kathleen McKeown.** 2010. "Extracting Social Networks from Literary Fiction." In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*, 138–147. <https://aclanthology.org/P10-1015> (zugegriffen: 19.11.2024).
- Gius, Evelyn, Fotis Jannidis, Markus Krug, Albin Zehe, Andreas Hotho, Frank Puppe, Jonathan Krebs, Nils Reiter, Nathalie Wiedmer und Leonard Konle.** 2019a. "Detection of Scenes in Fiction." In *ADHO Digital Humanities Conference (DH2019)*. <https://doi.org/10.34894/OOY9CE> (zugegriffen: 19.11.2024).
- Gius, Evelyn, Nils Reiter und Markus Willand (Hg.).** 2019b. "A Shared Task for the Digital Humanities: Annotating Narrative Levels." *Journal of Cultural Analytics* 4. <https://culturalanalytics.org/issue/2254-vol-4-issue-3-2019> (zugegriffen: 19.11.2024).
- Gius, Evelyn, Nils Reiter und Markus Willand (Hg.).** 2021. "Narrative Levels: A Shared Task for the Digital Humanities." *Journal of Cultural Analytics* 6. <https://culturalanalytics.org/issue/3772> (zugegriffen: 19.11.2024).
- Jannidis, Fotis, Isabella Reger, Markus Krug, Lukas Weimer, Luisa Macharowsky und Frank Puppe.** 2016. "Comparison of Methods for the Identification of Main Characters in German Novels." In *ADHO Digital Humanities Conference (DH2016)*, 578–582. <https://dh2016.adho.org/abstracts/297> (zugegriffen: 19.11.2024).
- Ketschik, Nora.** 2024. "Figurennetzwerke im mittelhochdeutschen Artusroman. Möglichkeiten und Grenzen der sozialen Netzwerkanalyse als Methode für literaturwissenschaftliche Analysen." Metzler/Springer (Reihe: Digitale Literaturwissenschaft, hg. von Evelyn Gius und Thomas Weitin). Zugl. Dissertationsschrift, Universität Stuttgart. <https://doi.org/10.1007/978-3-662-69984-3> (zugegriffen: 19.11.2024).
- Krautter, Benjamin, Janis Pagel, Nils Reiter und Marcus Willand.** 2020. "»[E]in Vater, dächte ich, ist doch immer ein Vater«. Figurentypen und ihre Operationalisierung." *Zeitschrift für digitale Geisteswissenschaften*. [http://dx.doi.org/10.17175/2020\\_007](http://dx.doi.org/10.17175/2020_007) (zugegriffen: 19.11.2024).
- Krautter, Benjamin und Michael Vauth.** 2023. "Konstellationen kommunikativer Macht. Hypothesengeleitete Netzwerkanalyse in der Literaturwissenschaft." In *Figurenkonstellation und Gesellschaftsentwurf. Annäherung an eine narratologische Kategorie und ihre Deutungspotentiale*, hg. von Henrike Schwab, 205–238. Heidelberg: Winter Verlag.
- Krug, Markus, Frank Puppe, Isabella Reger, Lukas Weimer, Luisa Macharowsky, Stephan Feldhaus und Fotis Jannidis.** 2018. "Description of a Corpus of Character References in German Novels - DROC [Deutsches Roman Corpus]." *DARIAH-DE Working Papers* 27. <http://resolver.sub.uni-goettingen.de/purl/?dariah-2018-2> (zugegriffen: 19.11.2024).
- Krug, Markus, Fotis Jannidis, Isabella Reger, Luisa Macharowsky, Lukas Weimer und Frank Puppe.** 2016. "Attribuierung direkter Reden in deutschen Romanen des 18.–20. Jahrhunderts. Methoden zur Bestimmung des Sprechers und des Angesprochenen." In *3. Tagung des Verbands Digital Humanities im deutschsprachigen Raum e.V. (DHd 2016)*, 124–130.
- Reiter, Nils, Andre Blessing, Nora Ketschik, Sandra Murr, Maximilian Overbeck und Axel Pichler.** 2023. "Annotation Guidelines for Entity Reference Annotation (CRETA)." Zenodo (v1.1). <https://doi.org/10.5281/zenodo.7684268> (zugegriffen: 19.11.2024).
- Rochat, Yannick und Frédéric Kaplan.** 2014. "Character Networks in Les Confessions from Jean-Jacques Rousseau." In *Proceedings of Texas Digital Humanities Conference*.
- Szemes, Botond und Bence Vida.** 2024. "Tragic and Comical Networks. Clustering Dramatic Genres According to Structural Properties." In *Computational Drama Analysis. Reflecting on Methods and Interpretations*, hg. von Melanie Andresen und Nils Reiter, 167–188. Berlin/Boston: De Gruyter. <https://doi.org/10.1515/9783111071824> (zugegriffen: 19.11.2024).
- Smeets, Roel.** 2021. "Character Constellations: Representations of Social Groups in Present-Day Dutch Literary Fiction." Leuven: Leuven University Press. <https://doi.org/10.2307/j.ctv21wj5cb> (zugegriffen: 19.11.2024).

**Trilcke, Peer.** 2013. "Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft." In *Empirie in der Literaturwissenschaft*, hg. von Philip Ajouri, Christoph Rauen und Katja Mellmann, 201–247. Münster: Mentis Verlag GmbH.

**Trilcke, Peer, Evgeniya Ustinova, Ingo Börner, Frank Fischer und Carsten Milling.** 2024. "Detecting Small Worlds in a Corpus of Thousands of Theater Plays." In *Computational Drama Analysis. Reflecting on Methods and Interpretations*, hg. von Melanie Andresen und Nils Reiter, 7–33. Berlin/Boston: De Gruyter. <https://doi.org/10.1515/9783111071824> (zugegriffen: 19.11.2024).

**Vauth, Michael.** 2023. "Eine digitale Narratologie der Binnenerzählung. Untersuchungen zu den Dramen und Novellen Heinrich von Kleists." Metzler/Springer (Reihe: Digitale Literaturwissenschaft, hg. von Evelyn Gius und Thomas Weitin). Zugl. Dissertationsschrift, Universität Darmstadt. <https://doi.org/10.1007/978-3-662-67036-1> (zugegriffen: 19.11.2024).

**Vauth, Michael.** 2019. "Netzwerkanalyse narrativer Informationsvermittlung in Dramen." In *6. Tagung des Verbands Digital Humanities im deutschsprachigen Raum e.V. (DHd 2019)*, 184–188.

**Viehhauser, Gabriel.** 2023. "Figurenkonstellationen, Gesellschaftsentwürfe und Netzwerkanalyse. Digitale Zugänge zu Strukturmustern im Fastnachtspiel." In *Figurenkonstellation und Gesellschaftsentwurf. Annäherung an eine narratologische Kategorie und ihre Deutungspotentiale*, hg. von Henrike Schwab, 183–204. Heidelberg: Winter Verlag.

**Weimer, Anna Mareike, Florian Barth, Tillmann Döncke, Luisa Gödeke, Hanna Varachkina, Anke Holler, Caroline Sporleder und Benjamin Gittel.** 2022. "The (In-)Consistency of Literary Concepts – Formalising, Annotating and Detecting Literary Comment." In *1st Annual Conference of Computational Literary Studies. Conference Reader*. <https://doi.org/10.48694/jcls.90> (zugegriffen: 19.11.2024).

**Zehe, Albin, Leonard Konle, Lea Katharina Dümpelmann, Evelyn Gius, Andreas Hotho, Fotis Jannidis, Lucas Kaufmann, Markus Krug, Frank Puppe, Nils Reiter, Anneke Schreiber und Nathalie Wiedmer.** 2021. "Detecting Scenes in Fiction: A new Segmentation Task." In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, 3167–3177. <https://doi.org/10.18653/v1/2021.eacl-main.276> (zugegriffen: 19.11.2024).