

Methodenbeitrag: Netzwerkanalyse

Mareike Schumacher  ¹

1. Universität Regensburg

forTEXT

Thema:	Netzwerkanalyse	DOI:	10.48694/fortext.3759
Jahrgang:	1	Ausgabe:	6
Erscheinungsdatum:	2024-08-30	Erstveröffentlichung:	2018-11-12 auf fortext.net
Lizenz:			open  access

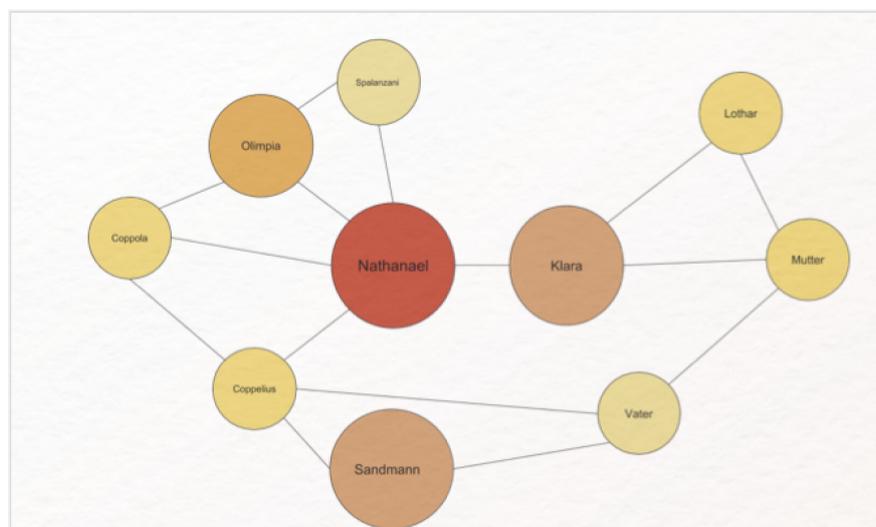
Allgemeiner Hinweis: Rot dargestellte Begriffe werden im Glossar am Ende des Beitrags erläutert. Alle externen Links sind auch am Ende des Beitrags aufgeführt.

1. Definition

In der Netzwerkanalyse werden zuvor bestimmte Größen (z. B. Figuren, Autoren, Orte) in ihrer Beziehung zueinander als Netzwerk aus Knotenpunkten (auch als Ecke oder mit dem englischen Ausdruck *Node* bezeichnet) und Verbindungslinien (auch Kanten, Relationen oder auf Englisch *Edges* genannt) untersucht. Die Verbindungslinien können gerichtet (meist in Form von Pfeilen dargestellt) oder ungerichtet (Linien) sein. Gerade in der digitalen Netzwerkanalyse werden zur Netzwerkanalyse häufig Visualisierungen (Horstmann und Stange 2024) genutzt. Dadurch treten zunächst quantitative Aspekte des relationalen Systems, in erster Linie die Anzahl von Knoten, Verbindungslinien und Verknüpfungen, deutlich hervor, die als Basis einer qualitativen Analyse dienen können. Netzwerkanalysen werden durch digitale Programme erleichtert, können aber auch manuell durchgeführt werden.

2. Anwendungsbeispiel

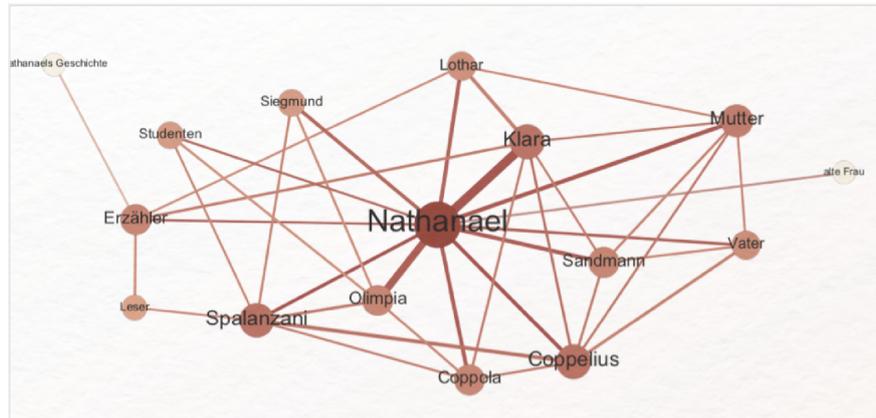
Sie wollen die Figurenkonstellation in E.T.A. Hoffmanns Erzähltext *Der Sandmann* analysieren. Die Methode der Netzwerkanalyse kann Sie dabei unterstützen. Ihre Hypothese ist, dass die Figur der Klara besonders stark mit Figuren des familiären Umfeldes vernetzt ist und dass die Verbindung dieses Unternetzwerkes zur Hauptfigur Nathanael hauptsächlich über Klara hergestellt wird. Nathanael hingegen scheint – so Ihre intuitive Beobachtung – geradezu in einem anderen Unternetzwerk gefangen zu sein: dem von den Figuren Sandmann/Coppelius/Coppola/Professor/Olimpia gebildeten, das Nathanael von anderen, positiven Sozialkontakten weitgehend isoliert. Das Gesamtnetzwerk, das Ihre Ausgangshypothese zur Figurenkonstellation der *Der Sandmann* abbildet, sieht damit in etwa so aus:



Netzwerkanalyse ist eine Methode, die analog und digital durchgeführt werden kann.

Die Verbindungen zwischen den Figuren, die hier als Kanten visualisiert sind, wären hier definiert als deren Interaktionen (miteinander sprechen, einander treffen, einander schreiben usw.). Der Text liegt Ihnen in digitaler Form (vgl. OCR) vor und Sie haben bereits eine Liste aller vorkommenden Figuren erstellt und markiert (vgl.

Annotation; Tagset), wo Figuren miteinander in Interaktion treten. Diese Daten visualisieren Sie nun mit einem Netzwerkanalysetool wie z. B. Gephi (Schumacher 2024):



Digital mit Gephi erstelltes Figuren-Netzwerk

In der Analyse des Netzwerkes wird deutlich, dass Nathanael – im Gegensatz zu Ihrer Ausgangshypothese – zwar einerseits durchaus mit allen Figuren direkt vernetzt ist. Neben der Beziehung zu Klara ist vor allem die zu seiner Mutter von großer Bedeutung, aber auch mit Lothar, seinem Freund und Klaras Bruder, steht er in relativ intensivem Kontakt. Im Netzwerk zeigt sich andererseits jedoch auch deutlich, dass die Verbindung Nathanaels zu Klara die am stärksten ausgeprägte in der gesamten Erzählung ist. Aus den Vorarbeiten – unter anderem dem Versuch, ein gerichtetes Netzwerk zu erstellen – wissen Sie, dass die Stärke dieser Verbindung vor allem darauf beruht, dass es sich hierbei um eine gegenseitige Beziehung handelt. Nathanaels empfundene Liebe zu Olympia wird zwar häufiger erwähnt als seine Gefühle für Klara, Olympia kann aber aufgrund der Tatsache, dass sie eine Puppe ist, natürlich nur wenig zurückgeben. Bei dem Netzwerk der Figuren in *Der Sandmann* von E.T.A. Hoffmann handelt es sich also um ein stark auf die Hauptfigur zentriertes Netzwerk. Die Grundannahme Ihrer Analyse, dass Nathanael stärker sozial isoliert ist als Klara, ist somit zunächst falsifiziert. Darüber hinaus werden Sie auf weitere Phänomene aufmerksam, die Sie in Anschlussuntersuchungen betrachten möchten. Welche Rolle spielen z. B. der Erzähler und der implizite Leser für das Figurennetzwerk? Inwiefern wirkt hier der Erzählanlass auf das narrative Gefüge ein? Eine umfassendere Netzwerkanalyse sämtlicher figuraler Instanzen kann also auch einen explorativen Zugang ermöglichen und interessantere Perspektiven eröffnen als das eher schematisch-hypothetische, manuell erstellte Netzwerk, mit dem Sie Ihre Fallstudie begonnen haben.

3. Literaturwissenschaftliche Tradition

Der Begriff des Netzwerkes findet sich bereits im Grimmschen Wörterbuch aus dem 18. Jahrhundert belegt. Er wurde damals allerdings als Bezeichnung für netzartige Verzerrungen z. B. von Mode oder Einrichtungsgegenständen gebraucht. Netzwerke in dem heute gebräuchlichen Sinn eines Konstrukts aus Verbindungen zwischen einzelnen Punkten sind in der Wissenschaftsgeschichte in unterschiedlichen Disziplinen belegt. Über Technik, Ökonomie, Sozialwissenschaft und schließlich Soziolinguistik gelangte der Begriff auch in die Literaturwissenschaften (Sauder 2003, 167f.).

Die literarische Netzwerkanalyse ist eine Methodik, die in unterschiedlichen Traditionen der Literaturwissenschaft verankert ist, von denen hier vier skizziert seien. Netzwerke können

1. Figurenkonstellationen darstellen (Moretti 2011),
2. Korrespondenzen von Autoren verbildlichen (Hildenbrandt und Kamzelak 2014),
3. intertextuelle Bezüge aufzeigen (Bartalesi u. a. 2015) oder
4. Ähnlichkeiten literarischer Stilistik visualisieren (Eder 2017).

Auch in stärker thematisch oder kontextuell orientierter literaturwissenschaftlicher Forschung werden Netzwerke entworfen und zur Analyse herangezogen, auch wenn diese nicht immer als solche benannt werden. So untersucht beispielsweise Rieger (1985) die Machtstrukturen in Kleists Erzählungen anhand mehrerer (gerichteter) Netzwerke, die er nicht als solche, sondern als Abbildung der Familienstrukturen bezeichnet. In den letzten fünfzig Jahren wurde eine Zunahme im Einsatz von Netzwerkanalysen in zahlreichen Disziplinen u. a. in den Geisteswissenschaften beobachtet (Jannidis 2017, 147).

1. Die Analyse der Figurenkonstellation gehört zu den grundlegenden Techniken der Literaturwissenschaft, da Figuren immer/meistens in Relation zu anderen Figuren stehen und diese Wechselbeziehungen mithilfe

einer Netzwerkanalyse ermittelt sowie visualisiert werden können. Diese kann grundsätzlich sowohl manuell als auch mit digitalen Hilfsmitteln erstellt werden, bestimmte Aspekte, wie z. B. die Gewichtung bestimmter Figuren oder Relationen auf Basis quantitativer Aspekte, sind allerdings leichter mit digitalen Tools zu ermitteln (Moretti 2011). Meist wird die Analyse von Figurenkonstellationen entweder anhand von Kollokationen (vgl. **Kollokation**) durchgeführt oder mittels qualitativer Kriterien der Relationen zwischen einzelnen Figuren wie z. B. Liebesbeziehungen oder Rivalität (Schneider 2003, 21). Die Analyse der Figurenkonstellation dient dazu, die Kompositionsstruktur eines Werkes besser zu verstehen (vgl. ebd., 22). Nicht immer werden zur Analyse der Figurenkonstellation Grafiken genutzt; wenn dies der Fall ist, haben sie aber häufig die Form eines ungerichteten (z. B. im Falle der Kollokationsanalyse) oder gerichteten (z. B. im Falle der Analyse von Liebesbeziehungen) Netzwerkes. Auch in interdisziplinären Forschungsansätzen konnten Netzwerke literarischer Figuren gewinnbringend eingesetzt werden. So analysieren Schweizer und Schnegg (1998) Netzwerke in Ingo Schulzes Roman *Simple Stories* aus ethnologischer Perspektive. In dieser interdisziplinär angelegten Studie wird unter anderem deutlich, dass sich die Figurenkonstellation im betrachteten literarischen Werk von tatsächlichen sozialen Systemen dadurch unterscheidet, dass die Figuren untereinander sehr viel stärker vernetzt sind als tatsächliche Personen (vgl. Schweizer und Schnegg 1998).

2. Fast noch naheliegender als bei der Figurenkonstellation ist das Erstellen von Netzwerken als Analysemethode bei Korrespondenzen. Briefe sind in soziale Kommunikationsprozesse eingebunden und haben darum eine inhärente netzwerkbildende Funktion (Hildenbrandt und Kamzelak 2014, 175). Als Knotenpunkte eines Briefnetzwerkes werden die miteinander kommunizierenden Personen, die Orte, mit denen sie verbunden sind (vgl. ebd., 175) oder auch Institutionen zum Gegenstand der Betrachtung. In vielen Untersuchungen werden mehrere Aspekte gleichzeitig erforscht. Zum Beispiel untersucht Siegert sowohl Korrespondenznetzwerke als auch die Bedeutung der Postämter, die die Korrespondenz leiten (Siegert 1990). Häufig werden kommunikative Netzwerke ausgehend von einer zentralen Figur, also als sog. egozentrierte Netzwerke (mehr zum Begriff des egozentrierten Netzwerkes z. B. in Wolf 2010, 471–483), angelegt und untersucht (Widawska 2011, 129). Die Netzwerke, die sich durch Briefkorrespondenz ergeben, sind komplex, da sie unterschiedliche relationale Ebenen abbilden. In der Netzwerkanalyse werden diese Ebenen – z. B. die soziale, thematische oder räumliche Ebene – einzeln oder gemeinsam betrachtet (Hildenbrandt und Kamzelak 2014, 176). Briefnetzwerke einzelner Schriftsteller zeigen deren soziale und geografische Position und welche Funktion ihnen jeweils innerhalb eines Gefüges literarischer Kontakte zukommt. In der Netzwerkanalyse von Briefkorrespondenzen werden quantitative Aspekte in Verbindung gebracht mit Textanalysen von Briefinhalten und Selbstaussagen des Autors und dadurch mehrdeutig interpretiert (Sauder 2003). In der Literatursoziologie werden Knotenpunkte und Verbindungen eines Netzwerkes kontextuell bzw. in einem gesellschaftlichen Systemzusammenhang betrachtet. In Siegerts (1990) Analyse von Briefnetzwerken, die gänzlich ohne Visualisierung des Netzwerkes auskommt, werden z. B. Postämter zu Knotenpunkten, zwischen denen nicht nur Briefe hin- und hergeschickt werden können bzw. durch die Monopolisierung auch müssen, sondern die gleichzeitig auch die Transportwege anderer Akteure, nämlich von Personen, sind (Siegert 1990, 537). Eine weitere Besonderheit von Briefnetzwerken ist, dass die Relationen nicht nur die neutralen Werte gerichtet oder ungerichtet oder qualitative Auszeichnungen wie „freundschaftlich“, „feindschaftlich“, „tröstend“, „bittend“ und viele weitere haben können, sondern dass sie Teil eines Wissenssystems sind (vgl. ebd., 550) und zum Kulturtransfer beitragen können (Widawska 2011, 130). Gerade diese Komplexität ist es, die sich auch in der Methode der Netzwerkanalyse widerspiegelt: Der systemische Charakter kann immer mitgedacht werden.
3. In der Intertextualitätsforschung gibt es die Basisannahme, dass Texte nicht für sich allein, als abgeschlossene Systeme wirken, sondern dass sie in Bezug zu einem referentiellen Wissenssystem von Autor*innen auf der einen und von Leser*innen auf der anderen Seite stehen (Helbling 1995, 8), wobei der Leserschaft über die Interpretation der Intertextualitäten eine besondere Rolle zukommt (Pfister 1985, 20). Darüber hinaus wurde Bachtins (1979) Idee der sog. *Dialogizität* in dieser literaturwissenschaftlichen Traditionslinie weitergedacht; von einer Kommunikation zwischen zwei Texten hin zu einer Vielstimmigkeit, die Texte nicht nur untereinander sondern auch mit der Gesellschaft verbindet (Pfister 1985). Auch hier gibt es also die Basiskomponenten eines Netzwerkes: Knotenpunkte in Form kultureller Werke (Texte) und Beziehungen dieser Werke untereinander und/oder zu gesellschaftlichen Phänomenen, die innerhalb der Intertextualitätsforschung zum Teil ebenfalls als „Text“ verstanden werden (Pfister 1985, 7). Intertextuelle Bezüge können – ähnlich wie Autorenkorrespondenzen – als egozentrierte Netzwerke modelliert werden, in deren Fokus ein bestimmtes Werk steht. Die Referenzen zu vorhergehenden Texten der Literaturgeschichte oder eigenen Werken (intratextuell) und auch innerhalb des Textes (inner- oder intratextuell) bilden ein System, das seinerseits eingebettet ist in einen übergeordneten Diskurs (Gray 2017, 127). Intertextuelle Bezüge können im literarischen Text sowohl über die Sprache als auch intermedial z. B. über Abbildungen aufgerufen werden (vgl. ebd., 128). Grundsätzlich ist es sehr schwierig, vielleicht sogar unmöglich, Intertextualität so zu operationalisieren, dass sie exakt messbar wird (Pfister 1985, 26). Jegliche Analyse, auch die mit Hilfe von Netzwerken, ist dadurch stark von heuristischen Interpretationen abhängig (vgl. ebd., 25–26).

4. In der digitalen Forschung zum Autor*innenstil werden dagegen ganz konkrete textuelle Merkmale statistisch ausgewertet. Anhand von Kriterien wie Worthäufigkeiten können dann Texte unterschiedlicher Autoren miteinander verglichen und klassifiziert werden (Weitin, Gilli und Kunkel 2016, 107). Hier wird meist eher der Versuch unternommen, Autoren voneinander zu unterscheiden (Barth 2018, 95), statt sie wie in der Intertextualitätsforschung miteinander in Verbindung zu setzen. Allerdings können die ermittelten Zahlenwerte auch dazu dienen, Netzwerkvisualisierungen von Autoren zu erstellen, die aufzeigen, welche Autoren sich stilistisch nahe stehen. Diese können dann wiederum zum Ausgangspunkt einer Netzwerkanalyse werden (Eder 2017).

4. Diskussion

Jede Form der Netzwerkanalyse, sei sie analog oder digital unterstützt durchgeführt, bietet zunächst den Vorteil, Strukturen im Text fassbar zu machen, die nicht auf den ersten Blick sichtbar sind (vgl. Abschnitt 3). Darüber hinaus können komplexe Kontexte mit Hilfe dieser Methode besonders gut berücksichtigt werden. Grundsätzlich birgt die Methode bzw. ihre Anwendung in der Literaturwissenschaft allerdings auch nachteilige Aspekte wie z. B. eine relativ hohe Abstraktion vom Text und die Konzentration auf einzelne Aspekte desselben.

Speziell die digitale Netzwerkanalyse bietet den Vorteil, dass eine große Datenbasis berücksichtigt werden kann und dass mathematische Verfahren die Auswertung dieser Daten unterstützen (vgl. *Text Mining*). Auf dieser Grundlage können Berechnungen darüber in die Analyse einfließen, welcher Art die Verbindungen zwischen zwei Figuren sind. Es kann z. B. berücksichtigt werden, dass Nathanael in einem Verhältnis der Angst zur Figur des Sandmannes steht, dass er hingegen für Olympia Liebe empfindet und dass die Automate Olympia zwar in einem Schöpfungsverhältnis zum Professor steht und häufig mit Nathanael kopräsent ist, dass von ihr aber keinerlei Emotionen ausgehen. Darüber hinaus könnte abgebildet werden, dass Nathanaels Angst vor dem Sandmann größer ist als die vor dem Professor. Es können also einerseits skalare Größen mit einbezogen und andererseits komplexe Verhältnisse der Figuren untereinander abgebildet werden.

Henning zeigt anhand ihrer Analyse von Thomas Manns *Zauberberg* auf, dass Netzwerkanalyse versteckte soziale Strukturen in Narrativen aufzeigt und dass diese Methode dabei eine vergleichsweise hohe Präzision ermöglicht (Henning 2006, 466). Auch Moretti (2011, 4) verweist darauf, dass die vergleichsweise hohe Abstraktion einer Netzwerkanalyse besonders geeignet ist, um verborgene Strukturen sichtbar zu machen, und damit auch Vorteile hat. Wie das Anwendungsbeispiel *Der Sandmann* gezeigt hat, kann die Möglichkeit, eine große Datenmenge zu berücksichtigen, auch einen explorativen Zugang zu Texten ermöglichen.

Einer der größten Vorteile der digitalen Netzwerkanalyse ist die Möglichkeit, sehr komplexe Zusammenhänge und/oder sehr viele Parameter abbilden zu können. Dies birgt aber umgekehrt auch die Gefahr, dass das Netzwerk unübersichtlich und als Hilfsmittel zur Textinterpretation unbrauchbar wird. Es ist also auch für die digitale Netzwerkanalyse zunächst eine Komplexitätsreduktion und eine klare Definition dessen erforderlich, was in den Fokus der Betrachtung gerückt werden soll. Mit digitalen Hilfsmitteln erstellte Netzwerkvisualisierungen wirken selbst bei vorangegangener Konzentration auf Kernaspekte häufig noch sehr komplex; eine Tatsache, die verschleiern, dass nicht sämtliche Aspekte eines Forschungsgegenstandes wie z. B. der Figurenkonstellation des *Sandmann* abgebildet werden.

Der größte Nachteil der digitalen Netzwerkanalyse in der Literaturwissenschaft ist aber, dass sie stets nur eine Art der Verbindung und nur eine im mathematischen Sinne positive anzeigen kann. In unserem Beispiel *Der Sandmann* hat Nathanael eine ungerichtete Verbindung zu seinem Vater (Vater – Sohn). Außerdem erwähnt Nathanael seinen Vater gegenüber Lothar, was als gerichtete Verbindung visualisiert werden könnte. Da aber stets nur eine Art der Verbindung in einer Netzwerkvisualisierung angezeigt werden kann, muss hier eine Auswahl getroffen werden. Im oben genannten Beispiel fiel diese auf die ungerichtete Variante, unter der dann auch gerichtete Arten der Kommunikation subsumiert wurden. Dieser Kompromiss ist für dieses Beispiel hinnehmbar, dass aber etwas, das nicht stattfindet, auch nicht oder nur unzureichend in die Visualisierung einbezogen werden kann, ist hier von entscheidender Bedeutung. Ein zentraler Bestandteil des Erzähltextes ist, dass Nathanael nach dem ersten Zusammentreffen mit Coppelius nicht mit seinem Vater darüber spricht. Diese nicht stattfindende Kommunikation bedingt, dass die Zusammenarbeit zwischen Coppelius und Nathanaels Vater mystifiziert bleibt. In der Netzwerkvisualisierung bleibt dieses Nicht-Gespräch (das im Falle unseres Beispiels als negativer Wert einbezogen wurde, der in die Berechnung der Verbindungsstärke eingeflossen ist und somit die Verbindung zwischen Vater und Sohn geschwächt hat) verborgen.

In der digitalen Netzwerkanalyse treten quantitativ bedeutsame Phänomene besonders hervor und werden dadurch mit Bedeutung versehen. Dies bewirkt nicht nur, wie oben bereits ausgeführt, dass nicht stattfindendes nicht oder nur unzureichend einbezogen werden kann, sondern auch, dass selten stattfindende Ereignisse weniger betont werden. Im oben genannten Fallbeispiel haben Coppelius, der Sandmann und Coppola eine von Nathanael unterstellte Verbindung, die gar nicht stärker sein könnte – er glaubt, sie seien eine Person. Der einzige Hinweis darauf, dass diese Vermutung richtig sein könnte, ist, dass die drei Figuren nie gemeinsam in Aktion treten. Dieses nicht zusammen Agieren, das auf die stärkste mögliche Verbindung hindeutet, kann aber im digital erstellten Netzwerk nicht angezeigt werden. Was sichtbar wird, ist die relativ selten von Nathanael geäußerte Vermutung, dass es sich um eine einzige Person handeln könnte, eine nicht besonders auffällig

erscheinende Verbindung unter anderen.

In Verbindung mit anderen digitalen Textanalysetools kann Netzwerkanalyse als reine **Distant Reading**-Methode eingesetzt werden. In diesem Falle können vollautomatisch Kanten und Verbindungen berechnet werden. Hier müssen die Parameter sehr bewusst ausgewählt und möglichst in der Analyse auch offengelegt werden. Geschieht dies nicht, so kann es dazu kommen, dass das resultierende Netzwerk zwar objektiv erscheint, dass es aber undurchsichtig bleibt, da nicht klar wird, welche Aspekte von Texten einbezogen und auf welche verzichtet wurde. Die Methode wird dann zu einer sog. Black Box, von der unklar ist, ob sie es tatsächlich vermag, Texte zu entschlüsseln, oder ob sie nicht eher Gegenteiliges bewirkt.

5. Technische Grundlagen

Die digital unterstützte Form der Netzwerkanalyse ist grundsätzlich ohne tiefgreifende technische Kenntnisse durchführbar. Bei der Erstellung einer Netzwerkanalyse gibt es unterschiedliche Verfahren, die grob in drei Ansätze unterteilt werden können.

1. Nutzung einer Netzwerk-Visualisierungs-Software (wie z. B. Gephi (Schumacher 2024)), die als Eingabemaske für im **Close Reading**-Verfahren am Text festgestellte Merkmale verwendet wird. Der Computer rechnet dann lediglich die Häufigkeiten der Eingaben oder die selbst vergebenen Gewichtungen zusammen.
2. Nutzung eines semi-automatischen Verfahrens, bei dem mit Hilfe einer anderen Software bestimmte Parameter wie z. B. Kollokation/Kopräsenz untersucht wurden, die dann manuell in ein Netzwerkanalysetool (wie z. B. Ezlinavis (Flüh 2024)) übertragen werden. Auch hier können noch Gewichtungen vorgenommen werden.
3. Übertragung von statistischen Daten, die mit einem anderen Tool (wie z. B. des Stilometrie-Tools Stylo (Horstmann 2024a; Horstmann 2024b)) generiert wurden, in das Netzwerkanalyse-Tool, das diese Daten dann für die Visualisierung (Horstmann und Stange 2024) auswertet.

Gängige Tools stellen grafische Benutzeroberflächen (vgl. **GUI**), Bedienungsanleitungen und Tutorials bereit und machen es möglich, Netzwerkvisualisierungen zu erstellen, ohne Programmcodes nutzen zu müssen. Allerdings sind gewisse mathematische und statistische Verfahren grundlegend, deren Kenntnis bei der Interpretation des Netzwerkes sehr hilfreich ist. Eine umfassende Einführung in diese Grundlagen finden Sie bei Jannidis (2017, 147–161). Im Folgenden seien nur einige zentrale Aspekte davon aufgeführt:

- **Graphentheorie:** Ein Netzwerk ist ein Graph, der aus Knoten und Verbindungen besteht. Außer diesen beiden sind zunächst keine weiteren Parameter mit Bedeutung versehen. Es können allerdings sowohl bei den Knoten- als auch bei den Kantenelementen Bedeutungen hinzugefügt werden, dazu gehören:
 - *Verbindungen:* Richtung, Dicke, Länge und Farbe (um z. B. die Verbindungsstärke oder -qualität zwischen zwei Figuren zu zeigen)
 - *Knoten:* Größe und Farbe des Knotens, um z. B. die (quantitative) Bedeutung einer Figur innerhalb einer Figurenkonstellation hervorzuheben
- **Zentralitätsmaß:** Vor allem bei egozentrierten Netzwerken ist es wichtig offenzulegen, welche Parameter einen Knoten ins Zentrum rücken. Ein wichtiges Maß für Zentralität ist z. B. die Menge der Verbindungen eines Knotens. Auch die Position zwischen zwei Teilbereichen eines Netzwerkes kann einen Knoten ins Zentrum rücken (wie in unserem Beispiel Nathanael). Nutzt man digitale Tools mit vorgegebenen Layouts, ist es wichtig zu wissen, welcher Art das zu Grunde liegende Zentralitätsmaß ist bzw. ob das Layout ein solches überhaupt verwendet (dies ist z. B. der Fall beim sog. *Force Directed Network Graph*).
- **Abstandsmessungen:** Bei automatischen Verfahren zur Erstellung der Datengrundlage eines Netzwerkes werden häufig Algorithmen benutzt, die ermitteln, welche Knotenpunkte wie nah beieinander liegen. Dazu wird zuerst ermittelt, zu welchem anderen Knotenpunkt ein Ausgangsknoten besonders dicht steht (*nearest neighbours*) Dann wird einbezogen, mit welchen Knotenpunkten der Ausgangsknoten über einen anderen Knoten verbunden ist (*second neighbours*) usw. Die Abstandsmessungen können beliebig viele Ausgangsknoten berücksichtigen und dementsprechend komplex werden. Bei der vollautomatischen Erstellung von Netzwerkvisualisierungen kann oft ausgewählt werden, welche Relationsabstände berücksichtigt werden sollen (z. B. nur *nearest neighbours*).
- **Manuelle Anpassung der Visualisierung:** Es gehört zu den gängigen Funktionen von Netzwerkanalysetools, dass eine Visualisierung zuerst statistisch erstellt und dann manuell nachbearbeitet wird. Dabei ist es nicht nur möglich, die unter dem Punkt Graphentheorie aufgeführten Bedeutungsebenen einzuarbeiten, sondern auch Zentralität und Abstand zu verändern. Dieser Arbeitsschritt ist essentiell für die Methode, sollte aber in dem Bewusstsein durchgeführt werden, dass damit bestimmte Aspekte der vorangegangenen Berechnungen verändert werden. Ebenso wie bei der Aufbereitung der Daten, die wie oben beschrieben in unterschiedlicher Weise durchgeführt werden kann, handelt es sich hierbei um einen interpretatorischen Arbeitsschritt. Am Ende verdeutlicht also auch das digital erstellte Netzwerk das

Ergebnis einer Interpretation und macht diese nachvollziehbar (zu der nur vermeintlichen Objektivität von Visualisierungen vgl. Textvisualisierung (Horstmann und Stange 2024)).

Bibliographie

- Bachtin, Michail. 1979. *Die Ästhetik des Wortes*. Frankfurt: Suhrkamp.
- Bartalesi, Valentina, Carlo Megini, Paola Andriani und Mirko Tavoni. 2015. Towards a Semantic Network of Dante's Works and Their Contextual Knowledge. *Digital Scholarship in the Humanities* 30, Nr. 1: i28–i35. doi: 10.1093/llc/fqv044.
- Barth, Florian. 2018. Zwischen Elisabeth Hauptmann und Bertolt Brecht: Stilometrische Studien einer Zusammenarbeit. In: *Quantitative Ansätze in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven*, hg. von Toni Bernhard, Markus Willand, Sandra Richter, und Andrea Albrecht, 95–120. Berlin, Boston: de Gruyter. <https://doi.org/10.1515/9783110523300-005> (zugegriffen: 8. November 2018).
- Eder, Maciej. 2017. Visualization in stylometry: Cluster analysis using networks. *Digital Scholarship in the Humanities* 32, Nr. 1: 50–64. doi: 10.1093/llc/fqv061, (zugegriffen: 13. November 2018).
- Flüh, Marie. 2024. Toolbeitrag: Ezlinavis. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 6. Netzwerkanalyse (30. August). doi: 10.48694/fortext.3784, <https://fortext.net/tools/tools/ezlinavis>.
- Gamper, Markus, Linda Reschke und Michael Schönhuth. 2012. *Knoten und Kanten 2.0. Soziale Netzwerkanalyse in Medienforschung und Kulturanthropologie*. Bielefeld: Transcript.
- Giessmann, Sebastian. 2006. *Netze und Netzwerke. Archäologie einer Kulturtechnik 1740-1840*. Bielefeld: Transcript.
- Gray, Richard T. 2017. Intertextualität/Vernetzung. In: *W. G. Sebald Handbuch*, hg. von Claudia Öhlschläger und Michael Niehaus, 122–129. Stuttgart: Springer.
- Helbling, Brigitte. 1995. *Vernetzte Texte. Ein literarisches Verfahren von Weltenbau. Mit den Fallbeispielen Ingeborg Bachmann, Uwe Johnson und einer Digression zum Comic Strip Doonesbury*. Würzburg: Königshausen & Neumann.
- Henning, Martina. 2006. Netzwerkanalyse literarischer Texte – am Beispiel Thomas Manns „Der Zauberberg“. In: *Qualitative Netzwerkanalyse*, hg. von Martina Hollstein und Florian Strauss, 465–480. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hildenbrandt, Vera und Roland S. Kamzelak. 2014. Im Exil erweitert sich die Welt. Neue Zugangswege zu Korrespondenzen durch Visualisierung. *Internationales Jahrbuch für Editionswissenschaft* 28, Nr. 1: 175–192. doi: 10.1515/editio-2014-011, (zugegriffen: 12. November 2018).
- Horstmann, Jan. 2024b. Methodenbeitrag: Stilometrie. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 1. Stilometrie (26. Februar). doi: 10.48694/fortext.3769, <https://fortext.net/routinen/methoden/stilometrie>.
- . 2024a. Toolbeitrag: Stylo. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 1. Stilometrie (26. Februar). doi: 10.48694/fortext.3770, <https://fortext.net/tools/tools/stylo>.
- Horstmann, Jan und Jan-Erik Stange. 2024. Methodenbeitrag: Textvisualisierung. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 5. Textvisualisierung (7. August). doi: 10.48694/fortext.3772, <https://fortext.net/routinen/methoden/textvisualisierung>.
- Jannidis, Fotis. 2017. Netzwerkanalyse. In: *Digital Humanities. Eine Einführung*, hg. von Fotis Jannidis, Hubertus Kohle, und Malte Rehbein, 147–161. Stuttgart: Metzler.
- Moretti, Franco. 2011. Network Theory, Plot Analysis. *Stanford Literary Lab Pamphlets 2*. <https://litlab.stanford.edu/LiteraryLabPamphlet2.pdf> (zugegriffen: 10. Oktober 2018).
- Pfister, Manfred. 1985. Konzepte der Intertextualität. In: *Intertextualität. Formen, Funktionen, anglistische Fallstudien*, hg. von Ulrich Broich und Manfred Pfister, 1–30. Tübingen: Niemeyer.
- Rieger, Bernhard. 1985. *Geschlechterrollen und Familienstrukturen in den Erzählungen Heinrich von Kleists*. Frankfurt am Main: Lang.
- Sauder, Gerhard. 2003. Netzwerk der Aufklärung: Mercks Briefe. In: *Netzwerk der Aufklärung. Neue Lektüren zu Johann Heinrich Merck*, hg. von Ulrike Leuschner und Matthias Luserke-Jaqui, 167–182. Berlin: de Gruyter.
- Schneider, Jost. 2003. *Einführung in die Roman-Analyse*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Schumacher, Mareike. 2024. Toolbeitrag: Gephi. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 6. Netzwerkanalyse (30. August). doi: 10.48694/fortext.3783, <https://fortext.net/tools/tools/gephi>.
- Siebert, Bernhard. 1990. Netzwerke der Regimentalität Harsdörfers Teutscher Secretarius und die Schicklichkeit der Briefe im 17. Jahrhundert. *Modern Language Notes* 105, Nr. 3: 536–562. <https://www.jstor.org/stable/pdf/2905073.pdf> (zugegriffen: 2. November 2018).
- Spoerhase, Carlos. 2014. „Manuscript für Freunde“, die materielle Textualität literarischer Netzwerke, 1760-1830 (Gleim, Klopstock, Lavatar, Fichte, Reinhold, Goethe). *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 88, Nr. 2: 172–205. doi: 10.1007/BF03374726, <https://link.springer.com/article/10.1007/BF03374726> (zugegriffen: 5. November 2018).
- Trilcke, Peer. 2013. Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft. In: *Empirie in der Literaturwissenschaft*, hg. von Philip Ajouri, Katja Mellmann, und Christoph Rauen, 201–247. Brill | mentis, 1. Januar. doi: 10.30965/9783957439710_012, https://www.mentis.de/view/book/edcoll/9783957439710/B9783957439710_s012.xml (zugegriffen: 8. September 2021).

- Weitin, Thomas, Thomas Gilli und Nico Kunkel. 2016. Auslegen und Ausrechnen. Zum Verhältnis hermeneutischer und quantitativer Verfahren in den Literaturwissenschaften. *Zeitschrift für Literaturwissenschaft und Linguistik* 46, Nr. 1: 103–115. doi: 10.1007/s41244-016-0004-8, <https://link.springer.com/article/10.1007/s41244-016-0004-8> (zugegriffen: 8. November 2018).
- Widawska, Barbara. 2011. Informationstransfer in den deutsch-polnischen Korrespondenznetzwerken der zweiten Hälfte des 19. Jahrhunderts. Zum Briefwerk von Albert Zipper (1855-1936). *Studia Germanica Gedanensia* 25: 128–142. https://fil.ug.edu.pl/sites/default/files/_nodes/strona-filologiczny/14508/files/ssg_25drukarnia.pdf#page=125 (zugegriffen: 5. November 2018).

Glossar

- Annotation** Annotation beschreibt die manuelle oder automatische Hinzufügung von Zusatzinformationen zu einem Text. Die manuelle Annotation wird händisch durchgeführt, während die (teil-)automatisierte Annotation durch **Machine-Learning-Verfahren** durchgeführt wird. Ein klassisches Beispiel ist das automatisierte **PoS-Tagging** (Part-of-Speech-Tagging), welches oftmals als Grundlage (**Preprocessing**) für weitere Analysen wie Named Entity Recognition (NER) nötig ist. Annotationen können zudem deskriptiv oder analytisch sein.
- Browser** Mit Browser ist in der Regel ein Webbrowser gemeint, also ein Computerprogramm, mit dem das Anschauen, Navigieren auf, und Interagieren mit Webseiten möglich wird. Am häufigsten genutzt werden dafür Chrome, Firefox, Safari oder der Internet Explorer.
- Close Reading** Close Reading bezeichnet die sorgfältige Lektüre und Interpretation eines einzelnen oder weniger Texte. Close Reading ist in der digitalen Literaturwissenschaft außerdem mit der manuellen **Annotation** textueller Phänomene verbunden (vgl. auch **Distant Reading** als Gegenbegriff).
- Commandline** Die Commandline (engl. *command line interface* (CLI)), auch Kommandozeile, Konsole, Terminal oder Eingabeaufforderung genannt, ist die direkteste Methode zur Interaktion eines Menschen mit einem Computer. Programme ohne eine grafische Benutzeroberfläche (**GUI**) werden i. d. R. durch Texteingabe in die Commandline gesteuert. Um die Commandline zu öffnen, klicken Sie auf Ihrem Mac „cmd“ + „space“, geben „Terminal“ ein und doppelklicken auf das Suchergebnis. Bei Windows klicken Sie die Windowstaste + „R“, geben „cmd.exe“ ein und klicken Enter.
- Data Mining** Data Mining gehört zum Fachbereich **Information Retrieval** und bezieht sich auf die systematische Anwendung computergestützter Methoden, die darauf abzielt, in vorhandenen Datenbeständen Muster, Trends oder Zusammenhänge zu erkennen. Textbasierte Formen des Data Minings sind u. a. **Text Mining**, **Web Mining** und **Opinion Mining**.
- Distant Reading** Distant Reading ist ein Ansatz aus den digitalen Literaturwissenschaften, bei dem computationale Verfahren auf häufig große Mengen an Textdaten angewandt werden, ohne dass die Texte selber gelesen werden. Meist stehen hier quantitative Analysen im Vordergrund, es lassen sich jedoch auch qualitative **Metadaten** quantitativ vergleichen. Als Gegenbegriff zu **Close Reading** wurde der Begriff insbesondere von Franco Moretti (2000) geprägt.
- GUI** GUI steht für *Graphical User Interface* und bezeichnet eine grafische Benutzeroberfläche. Ein GUI ermöglicht es, Tools mithilfe von grafischen Schaltflächen zu bedienen, um somit beispielsweise den Umgang mit der **Commandline** zu umgehen.
- HTML** HTML steht für *Hypertext Markup Language* und ist eine textbasierte Auszeichnungssprache zur Strukturierung elektronischer Dokumente. HTML-Dokumente werden von **Webbrowsern** dargestellt und geben die Struktur und Online-Darstellung eines Textes vor. HTML-Dateien können außerdem zusätzliche **Metainformationen** enthalten, die auf einer Webseite selbst nicht ersichtlich sind.
- Information Retrieval** Die Teildisziplin der Informatik, das Information Retrieval, beschäftigt sich mit der computergestützten Suche und Erschließung komplexer Informationen in meist unstrukturierten Datensammlungen.
- Kollokation** Als Kollokation bezeichnet man das häufige, gemeinsame Auftreten von Wörtern oder Wortpaaren in einem vordefinierten Textabschnitt.
- Lemmatisieren** Die Lemmatisierung von Textdaten gehört zu den wichtigen **Preprocessing**-Schritten in der Textverarbeitung. Dabei werden alle Wörter (**Token**) eines Textes auf ihre Grundform zurückgeführt. So werden beispielsweise Flexionsformen wie „schneller“ und „schnelle“ dem Lemma „schnell“ zugeordnet.
- Machine Learning** Machine Learning, bzw. maschinelles Lernen im Deutschen, ist ein Teilbereich der künstlichen Intelligenz. Auf Grundlage möglichst vieler (Text-)Daten erkennt und erlernt ein Computer die häufig sehr komplexen Muster und Gesetzmäßigkeiten bestimmter Phänomene. Daraufhin können die aus den Daten gewonnen Erkenntnisse verallgemeinert werden und für neue Problemlösungen oder für die Analyse von bisher unbekanntem Daten verwendet werden.

- Markup Language** Markup Language bezeichnet eine maschinenlesbare Auszeichnungssprache, wie z. B. **HTML**, zur Formatierung und Gliederung von Texten und anderen Daten. So werden beispielsweise auch **Annotationen** durch ihre Digitalisierung oder ihre digitale Erstellung zu Markup, indem sie den Inhalt eines Dokumentes strukturieren.
- Metadaten** Metadaten oder Metainformationen sind strukturierte Daten, die andere Daten beschreiben. Dabei kann zwischen administrativen (z. B. Zugriffsrechte, Lizenzierung), deskriptiven (z. B. Textsorte), strukturellen (z. B. Absätze oder Kapitel eines Textes) und technischen (z. B. digitale Auflösung, Material) Metadaten unterschieden werden. Auch **Annotationen** bzw. **Markup** sind Metadaten, da sie Daten/Informationen sind, die den eigentlichen Textdaten hinzugefügt werden und Informationen über die Merkmale der beschriebenen Daten liefern.
- Named Entities** Eine Named Entity (NE) ist eine Entität, oft ein Eigenname, die meist in Form einer Nominalphrase zu identifizieren ist. Named Entities können beispielsweise Personen wie „Nils Holgerson“, Organisationen wie „WHO“ oder Orte wie „New York“ sein. Named Entities können durch das Verfahren der Named Entity Recognition (NER) automatisiert ermittelt werden.
- OCR** OCR steht für *Optical Character Recognition* und bezeichnet die automatische Texterkennung von gedruckten Texten, d. h. ein Computer „liest“ ein eingescanntes Dokument, erkennt und erfasst den Text darin und generiert daraufhin eine elektronische Version.
- Opinion Mining** Unter Opinion Mining, oder Sentiment Analysis, versteht man die Analyse von Stimmungen oder Haltungen gegenüber einem Thema, durch die Analyse natürlicher Sprache. Das Opinion Mining gehört zu den Verfahren des **Text Minings**.
- POS** PoS steht für *Part of Speech*, oder „Wortart“ auf Deutsch. Das PoS- **Tagging** beschreibt die (automatische) Erfassung und Kennzeichnung von Wortarten in einem Text und ist of ein wichtiger **Preprocessing**-Schritt, beispielsweise für die Analyse von **Named Entities**.
- Preprocessing** Für viele digitale Methoden müssen die zu analysierenden Texte vorab „bereinigt“ oder „vorbereitet“ werden. Für statistische Zwecke werden Texte bspw. häufig in gleich große Segmente unterteilt (*chunking*), Großbuchstaben werden in Kleinbuchstaben verwandelt oder Wörter werden **lemmatisiert**.
- Tagset** Ein Tagset definiert die Taxonomie, anhand derer **Annotationen** in einem Projekt erstellt werden. Ein Tagset beinhaltet immer mehrere Tags und ggf. auch Subtags. Ähnlich der **Type/Token**-Differenz in der Linguistik sind Tags deskriptive Kategorien, wohingegen Annotationen die einzelnen Vorkommnisse dieser Kategorien im Text sind.
- Text Mining** Das Text Mining ist eine textbasierte Form des **Data Minings**. Prozesse & Methoden, computer-gestützt und automatisch Informationen bzw. Wissen aus unstrukturierten Textdaten zu extrahieren, werden als Text Mining zusammengefasst.
- Type/Token** Das Begriffspaar „Type/Token“ wird grundsätzlich zur Unterscheidung von einzelnen Vorkommnissen (Token) und Typen (Types) von Wörtern oder Äußerungen in Texten genutzt. Ein Token ist also ein konkretes Exemplar eines bestimmten Typs, während ein Typ eine im Prinzip unbegrenzte Menge von Exemplaren (Token) umfasst.
Es gibt allerdings etwas divergierende Definitionen zur Type-Token-Unterscheidung. Eine präzise Definition ist daher immer erstrebenswert. Der Satz „Ein Bär ist ein Bär.“ beinhaltet beispielsweise fünf Worttoken („Ein“, „Bär“, „ist“, „ein“, „Bär“) und drei Types, nämlich: „ein“, „Bär“, „ist“. Allerdings könnten auch vier Types, „Ein“, „ein“, „Bär“ und „ist“, als solche identifiziert werden, wenn Großbuchstaben beachtet werden.
- Web Mining** Unter Web Mining versteht man die Anwendung von Techniken des **Data Mining** zur Extraktion von Informationen aus dem World Wide Web. Das Web Mining ist ein Teilbereich des Data Minings und zählt zu einem der wichtigsten Anwendungsgebiete für das **Text Mining**.