

Dramen als small worlds? Netzwerkdaten zur Geschichte und Typologie deutschsprachiger Dramen 1730-1930

,
trilcke@phil.uni-goettingen.de
Georg-August-Universität Göttingen, Deutschland

,
frank.fischer@sub.uni-goettingen.de
Niedersächsische Staats- und Universitätsbibliothek
Göttingen

,
goebel@sub.uni-goettingen.de
Niedersächsische Staats- und Universitätsbibliothek
Göttingen

,
kampkaspar@hab.de
Herzog August Bibliothek Wolfenbüttel

Ansatz

Neben dem ›klassischen‹ strukturalistischen Paradigma, das sich wesentlich an Theoremen der Linguistik orientiert (u. a. Lotman 1972; Titzmann 1977), gibt es in der Literaturwissenschaft bereits seit Jahrzehnten Ansätze zu einer Strukturanalyse, die sich auf die empirische Soziologie – insbesondere auf die *Social Network Analysis* – bezieht und Struktur entsprechend nicht über basale semantische Relationen (etwa als Opposition oder Äquivalenz) definiert, sondern über soziale Interaktionen (Marcus 1973; Stiller et al. 2003; de Nooy 2006; Stiller / Hudson 2005; Elson et al. 2010; Agarwal et al. 2012). Im Kontext der Digital Humanities haben diese Ansätze zu einer literaturwissenschaftlichen Netzwerkanalyse (Trilcke 2013) in den letzten Jahren eine neue Dynamik gewonnen (Moretti 2011; Rydberg-Cox 2011; Park et al. 2013). Aus literaturwissenschaftlicher Sicht versprechen diese Analyseverfahren dabei auf umfangreichen Korpora basierende, von quantitativen Daten gestützte Erkenntnisse über die Literaturgeschichte wie auch über die generischen Eigenarten literarischer Texte. Im Projekt *dlina. Digital Literary Network Analysis* haben wir einen Workflow zur Extraktion, Analyse und Visualisierung von Netzwerkdaten aus dramatischen

Texten mit rudimentärer TEI-Auszeichnung entwickelt (Fischer et al. 2015). Der hier projizierte Vortrag wird Ergebnisse der netzwerkanalytischen Auswertung dieser Daten präsentieren und vor dem Hintergrund etablierter fachwissenschaftlicher Fragestellungen diskutieren.

Datenerhebung und -analyse

Unser derzeitiges Korpus umfasst 465 deutschsprachige Dramen (Zeitraum 1730 bis 1930), die aus dem Textgrid Repository extrahiert wurden. Die für die Netzwerkanalyse relevanten Strukturdaten dieser Dramen (Segmentierung, Figurenidentifikation) wurden in einem regelbasierten Prozess händisch ediert, um OCR- und TEI-Tagging-Fehler zu beheben sowie solchen ›Eigenarten‹ der literarischen Texte zu begegnen, die die Analyseergebnisse verfälschen würden (u. a. unterschiedliche Bezeichnungen identischer Figuren; Bezeichnung von Figurengruppen mit unbestimmten Numeralien wie ›beide‹ oder ›alle‹; etc.). Die edierten Strukturdaten liegen in einem eigens entwickelten Datenformat, dem *dlina*-Format, in Form von XML-Dateien vor. Die Visualisierung der Netzwerke und die Berechnung netzwerkanalytischer Werte erfolgt – mittels Python- und D3-Skripten – automatisiert auf Basis der in den *dlina*-Dateien gespeicherten Strukturdaten. Neben Graphen und basalen Werten, die die Netzwerke global beschreiben (Network Size, Density, Average Degree, Average Path Length), werden dabei auch Zentralitätswerte für sämtliche Figuren eines Dramas erhoben (u. a. Degree, Average Distance, Closeness Centrality, Betweenness Centrality). Die Implementierung weiterer Berechnungsroutinen (u. a. Clustering Coefficient, logarithmierte Degree Distribution-Tabellen) ist für den Winter 2015/16 vorgesehen. Sämtliche Daten und Visualisierungen werden frei verfügbar im Netz publiziert (<https://github.com/dlina> und <https://dlina.github.io/linas/>).

Literaturwissenschaftliche Auswertung 1: Dramengeschichte

Die diachrone Erstreckung unseres Dramenkorpus über ca. 200 Jahre deutscher Literaturgeschichte macht es möglich, größere Entwicklungen im Bereich der strukturellen Komposition von dramatischen Texten zu beobachten (erste Überlegungen dazu haben wir in einem Blogpost skizziert: <https://dlina.github.io/200-Years-of-Literary-Network-Data/>). Neben Werten, die sich auf die Gesamtnetzwerke der einzelnen Dramen beziehen (u. a. Network Size, Density, Average Degree; s. exemplarisch zur Average Path Length, Abbildung 1), werden dabei auch figurenbezogene Werte, v.a. Zentralitätsmaße, einbezogen, die Aufschluss etwa über die Streuung des Personals eines Dramas bzw. dessen Zusammensetzung aus ›zentralen‹ und weniger ›zentralen‹

Figuren gibt. Auf Grundlage dieser Werte sollen im Vortrag einige globale Thesen der Literaturgeschichte diskutiert werden. So werden wir *erstens* diskutieren, inwieweit sich anhand der netzwerkanalytischen Werte eine Ausdifferenzierung der strukturellen Komposition von dramatischen Texten am Ende des 18. Jahrhunderts beobachten lässt: Eine solche Ausdifferenzierung wäre angesichts des Nebeneinanders von ›geschlossenen‹, in der Tradition der Französischen Klassik stehenden Dramen und ›offenen‹ Dramen, die sich u. a. an der Dramatik Shakespeares orientieren, zu erwarten. *Zweitens* werden wir einige geläufige literaturwissenschaftliche Periodisierungshypothesen testen (u. a. aus dem Strukturalismus und der Sozialgeschichte); gefragt werden soll hier, inwieweit die Entwicklung der netzwerkanalytischen Werte mit den von der Forschung vorgeschlagenen Periodisierungen korreliert.

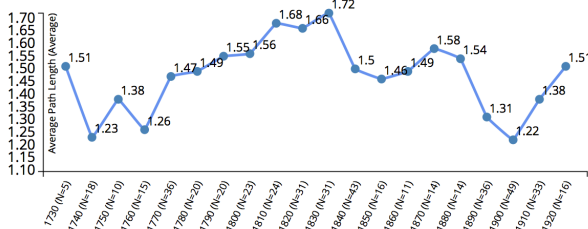
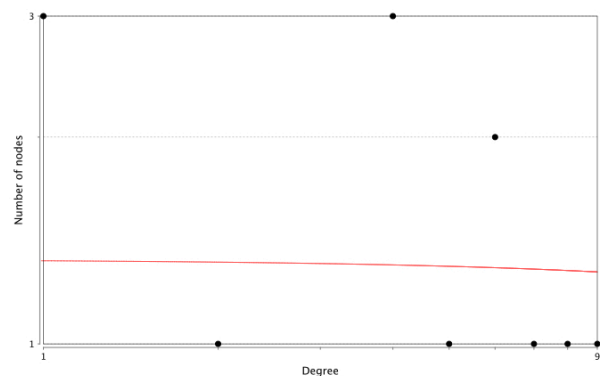
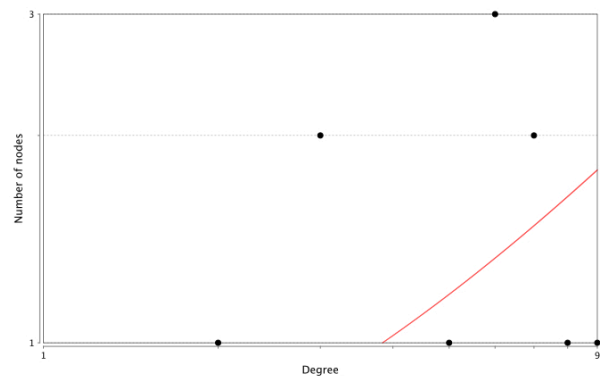


Abb. 1: Average Path Length (Mean; nach Dekaden)

Literaturwissenschaftliche Auswertung 2: Dramentypen

Die von uns bisher erhobenen Werte zeigen, dass Dramen in dem untersuchten Zeitraum auf sehr unterschiedliche Weise strukturiert wurden. In der ›traditionellen‹ Literaturwissenschaft wurden für solche unterschiedlichen ›Bauformen‹ diverse Typologien entwickelt, in der Germanistik am bekanntesten ist Volker Klotz' Unterscheidung in eine ›offene‹ und eine ›geschlossene‹ Dramenform (Klotz 1960). Diesen typologischen Impuls wollen wir aufgreifen und einen Vorschlag unterbreiten, wie sich mittels netzwerkanalytischer Daten bestimmte Typen der strukturellen Komposition von Dramen unterscheiden (und dann wiederum historisch verorten) lassen. Unser Vorschlag greift dabei Überlegungen aus der Forschung zu sog. Small-world-Netzwerken auf. Diese Forschungen setzen bei der Beobachtung an, dass die Werte von empirisch erhobenen Netzwerken nicht selten signifikant von entsprechenden Random-Netzwerken (also z. B. nach dem Erdős-Rényi-Modell erstellten Graphen) abweichen. Abweichungen sind dabei insbesondere beim Clustering Coefficient, bei der Average Path Length sowie bei der Degree Distribution zu beobachten (Albert / Barabási 2002). Für den hier projizierten Vortrag werden wir diese Werte – sowie die Werte für die entsprechenden

Random-Netzwerke – für unser Gesamtkorpus erheben (sowie einen Workflow für die automatisierte Erhebung entwickeln) und diskutieren. Erste Testläufe deuten dabei darauf hin, dass sich auf diese Weise tatsächlich unterschiedliche Typen der strukturellen Komposition von Dramen beschreiben lassen könnten. So zeigen sich z. B. auffällige Unterschiede bei der Degree Distribution (s. exemplarisch die Tabellen für vier Dramen in Abbildung 2); und mit Blick auf den Clustering Coefficient zeigt sich, dass im Vergleich zu Random-Netzwerken signifikant höhere Werte, wie sie bei Small-world-Netzwerken zu erwarten sind, zwar in mehreren Fällen vorkommen, jedoch keineswegs für alle Dramennetzwerke charakteristisch sind (siehe exemplarisch die Werte in Abbildung 3). Im Vortrag werden wir diese Werte für alle Dramen unseres Korpus präsentieren; wir werden diskutieren, inwieweit sich hier – aufbauend auf dem Small-world-Konzept – netzwerkanalytisch basierte Typen der strukturellen Komposition von Dramen unterscheiden lassen und wir werden literarhistorisch fundiert erörtern, welche Eigenschaften der Dramen für die unterschiedlichen Werte verantwortlich sind.



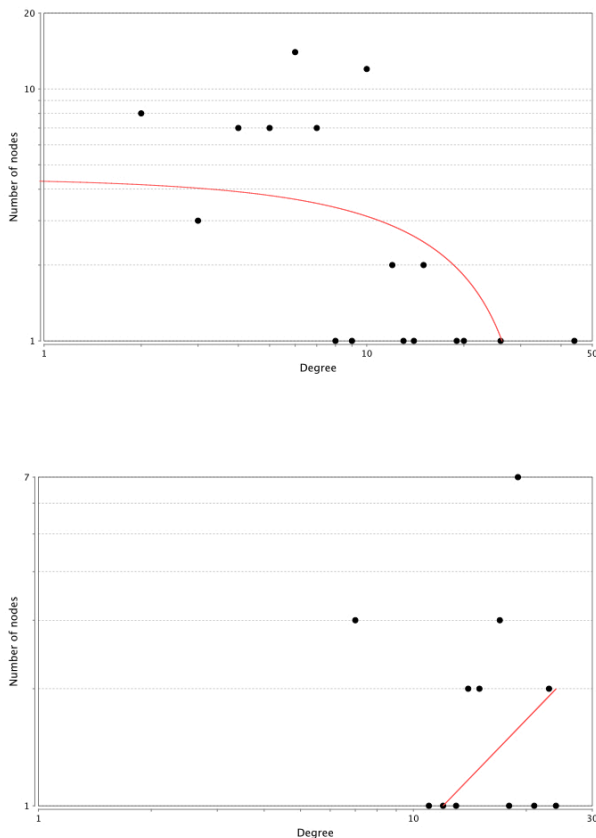


Abb. 2.1 bis 2.4: Node Degree Distribution für »Der sterbende Cato« (1731), »Emilia Galotti« (1772), »Götz von Berlichingen« (1773) und »Die Räuber« (1781)

	Clustering Coefficient	Clustering Coefficient (Random: Erdős-Renyi)	Abw. Clustering Coefficient des Dramennetzwerks vom entsprechenden Random-Netzwerk
Faust II	0,941	0,126	746,83%
Hannibal	0,918	0,352	260,80%
Dantons Tod	0,909	0,419	216,95%
Die Journalisten	0,884	0,753	117,40%
Lucie Woodvil	0,883	0,742	119,00%
Die Räuber	0,867	0,693	125,11%
Götz von Berlichingen	0,852	0,116	734,48%
Der gestiefelte Kater	0,834	0,381	218,90%
Die Jungfrau von Orleans	0,769	0,156	492,95%
Der Hofmeister	0,764	0,261	292,72%
Der sterbende Cato	0,75	0,481	155,93%
Kabale und Liebe	0,745	0,496	150,20%
Canut	0,738	0,619	119,22%
Die Soldaten	0,727	0,296	245,61%
Emilia Galotti	0,517	0,367	140,87%

Abb. 3: Vergleich des Clustering Coefficient des Dramen-Netzwerks mit dem eines jeweils entsprechenden Random-Netzwerks

Bibliographie

Albert, Réka / Barabási, Albert-László (2002): "Statistical mechanics of complex networks", in: *Reviews of Modern Physics* 74: 47–97.

Agarwal, Apoorv / Corvalan, Augusto / Jensen, Jacob / Rambow, Owen (2012): "Social Network Analysis of *Alice in Wonderland*" in: *Proceedings of the Workshop on Computational Linguistics for Literature*. Montréal 88–96.

de Nooy, Wouter (2006): "Stories, Scripts, Roles, and Networks" in: *Structure and Dynamics* 1, 2 <http://escholarship.org/uc/item/8508h946#page-1> [letzter Zugriff 12. Oktober 2015].

Elson, David K. / Dames, Nicholas / McKeown, Kathleen R. (2010): "Extracting Social Networks from Literary Fiction", in: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Uppsala 138–147.

Fischer, Frank / Kampkaspar, Dario / Göbel, Mathias / Trilcke, Peer (2015): "Digital Network Analysis of Dramatic Texts", in: *DH 2015, Sydney, 2. Juli 2015* <https://dlina.github.io/Our-Talk-at-DH2015/> [Skript] und [Slides] [letzter Zugriff 12. Oktober 2015].

Klotz, Volker (1960): *Geschlossene und offene Form im Drama*. München: Hanser.

Lotman, Jurij M. (1972): *Die Struktur literarischer Texte*. München: Wilhelm Fink.

Marcus, Solomon (1973): *Mathematische Poetik*. Frankfurt am Main: Editura Academiei.

Moretti, Franco (2011): "Network Theory, Plot Analysis" in: *Stanford Literary Lab Pamphlets* 2 <http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf> [letzter Zugriff 12. Oktober 2015].

Park, Gyeong-Mi / Kim, Sung-Hwan / Cho, Hwan-Gue (2013): "Structural Analysis on Social Network Constructed from Characters in Literature Texts", in: *Journal of Computers* 8, 9: 2442–2447 <http://ojs.academypublisher.com/index.php/jcp/article/view/jcp080924422447/7672> [letzter Zugriff 12. Oktober 2015].

Rydberg-Cox, Jeff (2011): "Social Networks and the Language of Greek Tragedy", in: *Journal of the Chicago Colloquium on Digital Humanities and Computer Science* 1, 3 <https://letterpress.uchicago.edu/index.php/jdhcs/article/view/86/91> [letzter Zugriff 12. Oktober 2015].

Stiller, James / Nettle, Daniel / Dunbar, Robin I. M. (2003): "The Small World of Shakespeare's Plays", in: *Human Nature* 14: 397–408.

Stiller, James / Hudson, Matthew (2005): "Weak Links and Scene Cliques Within the Small World of Shakespeare", in: *Journal of Cultural and Evolutionary Psychology* 3: 57–73.

TextGrid: TextGrid Repository <https://textgridrep.de> [letzter Zugriff 10. Februar 2016].

Titzmann, Michael (1977): *Strukturelle Textanalyse*. Theorie und Praxis der Interpretation. München: Wilhelm Fink.

Trilcke, Peer (2013): "Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft", in: Ajouri, Philip / Mellmann, Katja / Rauen, Christoph (eds.): *Empirie in der Literaturwissenschaft*. Münster: mentis 201–247.