



Institut für
deutsche Philologie

04-DH-M

Projekt in den Digital Humanities

(04-DH-M-1)

Bitte füllen Sie nur die grau hinterlegten Felder aus.

Dozent:	Dr. Christof Schöch
Semester:	7
Nachname:	Stahl
Vorname:	Jakob
Matrikelnummer:	1818489
Geburtsdatum:	06.07.1990

_____ vom Korrektor auszufüllen _____

Datum	Unterschrift des Korrektors

Note

Prüfungsnummer: 315332

Projektbericht

Projektbeschreibung

Thema: Vergleich spanischer Theaterautoren mithilfe aus
Figurennetzwerken erstellten Merkmalen.

Inhaltsverzeichnis

A) Einleitung.....	3
B) Hauptteil.....	4
1. Thesen und Informationen.....	4
1.1. Der Vergleich von Theaterstücken.....	4
1.2. Das Siglo de Oro.....	5
1.3. Die zu untersuchenden Autoren.....	6
1.4. Thesen.....	9
2. Datengrundlage.....	9
2.1. Deutsche Testkorpus.....	9
2.2. Spanische Hauptkorpus.....	11
3. Methode.....	13
3.1. Methodenübersicht.....	13
3.2. Sprecher und Verbindungen.....	14
3.3. Skriptbeschreibung zur Erstellung der Netzwerke.....	18
3.4. Ermitteln der Merkmale.....	24
4. Auswertung.....	28
4.1. Ergebnisse der Berechnungen.....	28
4.2. Vergleich und Prüfung der Thesen.....	32
4.3. Fazit.....	35
C) Schluss.....	38
D) Literatur.....	40
E) Anhang.....	42

A) Einleitung

Die Ermittlung von Ähnlichkeiten zwischen Autoren ist eine bekannte Forschungsfrage in der Literatur- und Theaterwissenschaft. Dabei wird versucht aus Texten Merkmale zu ermitteln, welche dabei helfen können festzustellen, ob und in welchem Maße eine Ähnlichkeit zwischen zwei oder mehr Autoren nachweisbar ist. Eines der bekanntesten Beispiele hierfür ist die Suche nach der Autorenschaft der sogenannten „*Federalist Papers*“, welche mit verschiedenen stilometrischen Ansätzen, wie jenen von Frederick Mosteller und David Wallace¹ oder John Burrow,² angegangen wurden. Diese Ansätze konzentrieren sich auf Methoden der Statistik, um aus dem Textinhalt eindeutige Unterscheidungsmerkmale zu errechnen, wie zum Beispiel das *Burrows Delta*, das aus den häufigsten im Text vorhandenen Wörtern abgeleitet wird. Ein anderer Ansatz, welcher sich vom eigentlichen Textinhalt distanziert und sich mehr nach der Textstruktur richtet, ist der Versuch ein gesamtes Werk oder einen Teilabschnitt als Netzwerk darzustellen und zu analysieren. Dies ist umso leichter, wenn die Sprecher im Werk deutlich hervorgehoben werden, weswegen für diese Analyse der Fokus auf Dramen gelegt wird. Dieses Vorgehen ist vergleichsweise noch sehr jung und wird hauptsächlich dafür genutzt, um die Figurenkonstellation eines Stückes zu ermitteln.³

Diese literarische Netzwerkanalyse kann als Methodenimport aus den Bereichen Psychologie, Ethnologie und Sozialwissenschaften gesehen werden⁴ und bildet die Beziehung zwischen den Figuren eines literarischen Textes grafisch als Netzwerke ab. Das Netzwerk besteht dabei aus einer Menge an Figuren (Knoten) und den zwischen den Figuren vorhandenen Verbindungen (Kanten). Diese Verbindungen können darin bestehen, dass die Figuren miteinander reden oder eine Figur eindeutig zuordenbar erwähnt wird.⁵ Die Punkte und Kanten eines Netzwerkes werden in der Regel als Matrix gespeichert. Die daraus folgenden Netzwerke können visualisiert werden und zeigen so auf kompakte Weise, welche Figuren im Text mit anderen Figuren in Beziehung stehen, wie diese Beziehungen aussehen und welche Figuren sich im Kern des Netzwerkes oder am Rand aufhalten. Ferner ist ersichtlich, ob eine Figur über andere Figuren theoretisch jede Figur im Text erreichen könnte, also alle Figuren im gleichen Netzwerk angesiedelt sind oder ob voneinander getrennte Figurennetzwerke existieren. Mit der Visualisierung der Gewichtung der Beziehung kann zudem angezeigt werden,

¹ Vgl. Mosteller, F. and D. L. Wallace. *Inference and disputed authorship: The federalist*. Reading, 1964.

² Vgl. Burrows, John, 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship, In: *Literary and Linguistic Computing*. Volume 17 (2002), Issue 3, S. 267 –287.

³ Als Beispiel siehe das Projekt: *Network Analysis of dramatic texts* von Fischer, Frank, Göbel, Mathias, Kampkaspar, Dario und Trilcke, Peer, einsehbar unter (<https://dlina.github.io/>) (Stand September 2017)

⁴Vgl. Kriskke, André und Röpcke, Helge, *Graphen und Netzwerktheorie: Grundlagen – Methoden – Anwendungen*, München, 2015, S. 125-140 und Stegbauer, Christian (Hrsg.), *Netzwerkanalyse und Netzwerktheorie, ein neues Paradigma in den Sozialwissenschaften*, Wiesbaden, 2010, S. 11-19 und Trilcke, Peer. *Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft*, in: Ajouri, Philip, Mellmann, Katja und Rauen, Christoph (Hg.): *Empirie in der Literaturwissenschaft*, Münster 2013, S. 201-247, S. 201.

⁵ Zum Beispiel in der Regieanweisung als stiller Zuhörer oder Beobachter.

welche Figuren zentrale Handlungsträger sind und welche Figuren eher beiläufig die Szenen bereichern. Werden die Netzwerke mehrerer Texte miteinander verglichen, können zudem textspezifische Eigenschaften, wie Gattungscharakteristika oder Zentralitätswerte ermittelt werden.⁶

Die meisten dieser Informationen können aus den Merkmalen eines Netzwerkes generiert werden. Da anzunehmen ist, dass jeder Autor individuelle Werke schafft, wodurch sich wiederum individuelle Merkmale aus den Netzwerken ergeben sollten, soll in dieser Arbeit geprüft werden, ob mit diesen Netzwerkmerkmalen Unterschiede zwischen Autoren feststellbar sind.

Um dieser Annahme nachzugehen, werden für diese Arbeit vier spanische Theaterautoren des Siglo de Oro ausgewählt und anhand ihrer Lebens- und Wirkungsgeschichte eine mutmaßliche Zuordnung der vier Autoren in zwei Gruppen vorgenommen. Inwieweit diese zwei Gruppen zutreffen, wird anschließend mit Hilfe der Netzwerkanalyse überprüft, indem für insgesamt 99 Werke der vier Autoren die Netzwerkmerkmale berechnet und anschließend miteinander verglichen werden.

Das Ziel dieser Arbeit ist zu prüfen, ob aus Theaterstücken genierte Netzwerkmerkmale geeignet sind, um die Autoren von Theaterstücken miteinander vergleichen zu können.

B) Hauptteil

1. These und Informationen

1.1. Der Vergleich von Theaterautoren

Jedes Theater muss in seiner Epoche eine Reihe vielseitiger Aufgaben erfüllen. Das Theaterspiel soll die Zuschauer unterhalten und in bestimmten Fällen belehren, aufklären oder überraschen. Gleichzeitig müssen Bühne und Theatergebäude in Stand gehalten und Schauspieler und Mitarbeiter bezahlt werden. Die Frage, wie ein Theater finanziert und welche Stücke dort aufgeführt werden, kann somit wichtige Einblicke in die Gesellschaft der jeweiligen Epoche liefern. Denn so wie die historische Epoche durch das Theater geprägt wird, wird gleichsam das Theater durch die begleitende Epoche bestimmt.

Eine besondere Rolle nimmt dabei der Autor eines Theaterstückes ein. Wenn dieser nicht auf die Unterstützung wohlgesonnener Mäzenen, ein großes Erbe oder ertragreiche Nebeneinkünfte setzen kann, so muss er hauptsächlich durch den Erfolg seiner Werke leben. In diesem Fall muss er häufig Rücksicht auf die Wünsche und Vorlieben seines Publikums und des Zeitgeistes nehmen und gleichzeitig genug Unerwartetes und Eigenständiges bieten, um das Interesse an seinen Werken langfristig zu erhalten. Die Orientierung bei diesen Entscheidungen erfolgte dabei häufig nach

⁶ Vgl. Vortrag: *Digitale Netzwerkanalyse dramatischer Texte* von Fischer Frank, Kampkaspar, Dario und Trilcke, Peer, Graz, 2015, einsehbar unter <http://gams.uni-graz.at/o:dhd2015.v.040> (Stand September 2017), Folie 14-22 und Jannidis, Fortis, Kohle, Hubertus und Rehbein, Malte, *Digital Humanitis, Eine Einführung*, Stuttgart, 2017, S. 147 ff.

Freunden, Vorbildern oder Konkurrenten des gleichen Gewerbes. Es entstanden damit Netzwerke aus Autoren, welche gemeinsam mit ihrer Arbeit einen Stil, einen Dramentyp oder eine ganze Epoche prägten. Als Beispiel hierfür könnten das spanische Siglo de Oro,⁷ die Weimarer Klassik oder das Elisabethanische Theater herangezogen werden.⁸

Inwieweit sich diese Netzwerke zwischen den Autoren aber auch in den Werken selbst widerspiegeln, ist allerdings sehr schwierig zu erfassen. Zwar existiert zur Ähnlichkeit zwischen Theaterautoren bereits eine reichhaltige Forschung, diese beschränkt sich allerdings noch sehr häufig auf eine bestimmte Schaffungsperiode oder auf eine Handvoll ausgewählter Werke, wobei der Schwerpunkt der Analyse in den meisten Fällen auf inhaltliche Fragen wie Thema oder Schreibstil gelegt wird.⁹

1.2. Das Siglo de Oro

Das Siglo de Oro, das Goldene Jahrhundert, bezeichnet eine kulturelle und politische Blütezeit Spaniens im 16. und 17. Jahrhundert. In dieser Zeit beherrschte die spanische Krone neben der Iberischen Halbinsel sowohl Besitztümer im übrigen Europa als auch Kolonien in Mittel- und Südamerika. Gestützt auf diesen weitreichenden Besitz konnte Spanien für lange Zeit eine Großmachtpolitik führen und großen Einfluss auf die europäische Geschichte ausüben. Doch war dieser Einfluss nicht nur politisch, sondern auch kulturell wahrzunehmen. Die eigene Kultur in Spanien selbst wurde wiederum durch unterschiedlichste Einflüsse, wie die italienische Renaissance, das Barock, die katholische Religion und die spanische Gesellschaftsordnung geprägt. Der Zeitraum des Siglo de Oro wird, abhängig vom Forschungsschwerpunkt, entweder an Regierungszeiten von Königen oder an den Lebensdaten bedeutender Künstler festgemacht. So beginnt das Siglo de Oro in der Forschung, mit Schwerpunkt auf die allgemeine spanische Geschichte, häufig mit dem Ende der Reconquista 1492 oder mit der Krönung Karls V. im Jahr 1516. Der Endpunkt des Siglo de Oro wird dann häufig auf das Todesjahr Karls II. im Jahr 1700 gesetzt. Betrachtet man das Siglo de Oro dagegen vorrangig als literarische Epoche, so lässt sich der Beginn des Zeitalters mit der Veröffentlichung des Werkes *La Celestina* von Fernando de Rojas (1499) und das Ende mit dem Todesjahr von Pedro Calderón de la Barca (1681) festmachen.¹⁰

Als ein besonderes Jahr der Weichenstellung der Epoche ist noch auf das Jahr 1561 hinzuweisen. In diesem Jahr wurde der bisher reisende königliche Hofstaat durch Philipp II. fest in Madrid angesiedelt.

⁷ Vgl. Simson, Ingrid: *Das Siglo de Oro*, Stuttgart, 2001, S. 49 f.

⁸ Vgl. Womack, Peter: *English Renaissance drama*, Malden, 2008, S. 95-124.

⁹ Vgl. Charney, Maurice, *Shakespeare- and the Others*, in: *Shakespeare Quarterly*, Vol. 30, No. 3, 1979, S. 325-342, und Ermatinger, Emil, *Gottfried Keller und Goethe*, in: *PMLA*, Vol. 64, No. 1, 1949, S. 79-97 und Hermand, Jost, *Heine und Brecht: Über die Vergleichbarkeit des Unvergleichlichen*, in: *Monatshefte*, Vol. 73, No 4, 1981, S. 429-441.

¹⁰ Vgl. Simson: *Das Siglo de Oro*, S. 49 f.

Hier und in anderen Städten¹¹ entstand damit zur gleichen Zeit eine frühe Form des Nationaltheaters, mit „professionellen Schauspielertruppen und festen Bühnen [...] die sog. *corrales*“,¹² welche im Umfang, Stilreichtum und der Art der kommerziellen Nutzung für die damalige Zeit in Europa fast einzigartig war.¹³

Dies prägt das moderne Bild des Theaters des Siglo de Oro als vorwiegend kommerzielles Geschäft, welches so überhaupt die Grundlage für die Vielzahl an bekannten Werken der Epoche lieferte. Durch die Ausbeutung der amerikanischen Kolonien gelangte eine große Menge Gold und Silber ins spanische Heimatland. Obgleich dieser Reichtum nicht bei allen Bevölkerungsschichten gleichermaßen ankam, blühten die Städte auf und damit stiegen die Mittel für und das Interesse am Theater. Am Ende des Siglo de Oro dagegen, als die Macht des spanischen Reiches langsam verblasste, sollte das Theater nicht mehr alleine zerstreuen, sondern auch an die glorreiche Vergangenheit erinnern.¹⁴ Es bot sich also in dieser Epoche für viele Menschen die Möglichkeit und der finanzielle Anreiz sich neben-oder hauptberuflich als Autor für Theaterstücke verdient zu machen, ohne direkt von einem Förderer und Mäzen abhängig zu sein. Die folgenden vier Autoren sollten diesem Weg zumindest teilweise folgen.

1.3. Die zu untersuchenden Autoren

Lope de Rueda (1505?-1565)

Einer der ersten bekannten Theaterdichter des Siglo de Oro war Lope de Rueda, welcher nicht nur Bühnenstücke schrieb, sondern auch selbst Schauspieler und Theaterdirektor für reisende Theatergruppen war. Der Kern vieler seiner Stücke war häufig eine Nacherzählung italienischer Renaissancekomödien, deren Akte durch eingeschobene Zwischenspiele, die sogenannten *pasos*, erweitert wurden. Diese Einschübe vor oder zwischen den Akten sollten dabei nicht unbedingt die eigentliche Handlung voranbringen, sondern vielmehr das Publikum erheitern und erfreuen, indem komische oder lächerliche Figuren auftraten, welche wenig mit dem Schicksal der Protagonisten zu tun hatten.

Lope de Rueda wird daher häufig mit der Entwicklung des spanischen Volkstheaters, dem Zwischenspiel und der wandernden Bühnen in Zusammenhang gebracht.¹⁵

¹¹ Neben Madrid besonders in Sevilla und Valencia.

¹² Vgl. Stenzel, Hartmut, *Einführung in die spanische Literaturwissenschaft*, Stuttgart, 2001, S. 136.

¹³ Vgl. Strosetzki, Christoph, *Geschichte der spanischen Literatur*, Tübingen, 1996, S. 161 f.

¹⁴ Vgl. Neuschäfer, Hans-Jörg: *Spanische Literaturgeschichte*, Stuttgart 2011, S. 75-79 und S. 155 f. und S. 163 f. und Simson: *Das Siglo de Oro*, S. 49 ff. und 78.

¹⁵ Vgl. König, Bernhard, *Miguel de Cervantes Saavedra – Entremes*, in: *Das spanische Theater, Vom Mittelalter bis zur Gegenwart*, hg. v. Roloff, Volker und Wentzlaff-Eggebert, Harald, S. 53-69, hier S. 53 f. und Simson: *Das Siglo de Oro*, S. 49 und Strosetzki, *Geschichte der spanischen Literatur*, S. 171 ff.

Miguel de Cervantes Saavedra (1547-1616)

Cervantes ist heute vor allen durch seinen Roman *El ingenioso hidalgo Don Quixote de la Mancha* (*Don Quijote*) bekannt, verfasste allerdings auch eine Vielzahl an Theaterstücken. Hierzu gehörten in den ersten Schaffungsjahren hauptsächlich Tragödien wie *El trato de Argel* oder *El cerco de Numancia*. Diese Tragödien waren dabei im Gegensatz zu den Volksstücken von Lope de Rueda hauptsächlich für das Publikum großer Theaterstädte wie Madrid und Valencia gedacht. Aufgrund der großen Konkurrenz durch seinen Zeitgenossen Lope de Vega zog sich Cervantes jedoch vom klassischen Tragödientheater zurück und verfasste Werke mit eigenen Schwerpunkten, darunter Komödien (*comedias*) und Zwischenspiele (*entremes*), welche häufig als Weiterentwicklung des *pasos* von Lope begriffen werden. Es ist daher oft eine Streitfrage, ob Cervantes Werke zu Lopes *comedia nueva* oder eher zu Ruedas Volkswerken zu zählen sind.¹⁶

Lope de Vega (1562-1635)

Eine besondere Prägung erhielt die spanische Theaterwelt des Siglo de Oro durch die *comedia nueva* von Lope de Vega. In diesem Konzept vereinte Lope de Vega Merkmale von Tragödie und Komödie, brach starre Strukturen auf, erschuf neue Figurenkonstellationen und versuchte ein möglichst breites Publikum aus allen Bevölkerungsschichten anzusprechen. Der Hauptfigur wird dabei häufig der Diener, die Figur des *gracioso*, gegenübergestellt. Während die erste Figur häufig allein der Ehre oder der Liebe verfallen ist, muss sich die zweite Figur mit weltlichen Bedürfnissen wie Hunger oder Schlaf herumschlagen. Der *gracioso* ist oft ein Ratgeber der Hauptfigur und ist andererseits häufig Hauptträger der Komik im Stück. Aufgrund der großen Zahl der Werke von Vega ist es schwierig seine Werke eindeutig einer Gattung zuzuordnen. Eine Möglichkeit der Unterscheidung wäre die Einteilung seiner Stücke in Ehrendramen (*Drama de honor*), in welchen meist ein Feudalherr einer Landbevölkerung Unrecht tut, welches durch den König wieder gut gemacht werden muss, lyrische Dramen, Tragödien, Mantel-und Degenstücke (*Comedia de capa y espada*), Schäferdramen (*Comedia pastoril*) und in historische Dramen (*Comedia histórica*).¹⁷ Als Autor selbst wird Lope de Vega in dieser Arbeit nicht behandelt. Seine Werke und seine *comedia nueva* sollte aber maßgeblich die zwei folgenden Autoren beeinflussen.

Tirso de Molina (1579?-1648)

Tirso de Molina war hauptsächlich Mönch im Mercedarierorden, schrieb allerdings neben seiner Tätigkeit als Chronist und Theologielehrer ebenso Theaterstücke. Den Kern seiner Stücke bildet häufig

¹⁶ Vgl. König, *Miguel de Cervantes Saavedra – Entremes*, hier S. 53-58 und Simson: *Das Siglo de Oro*, S. 62 f. und Strosetzki, *Geschichte der spanischen Literatur*, S. 174 ff.

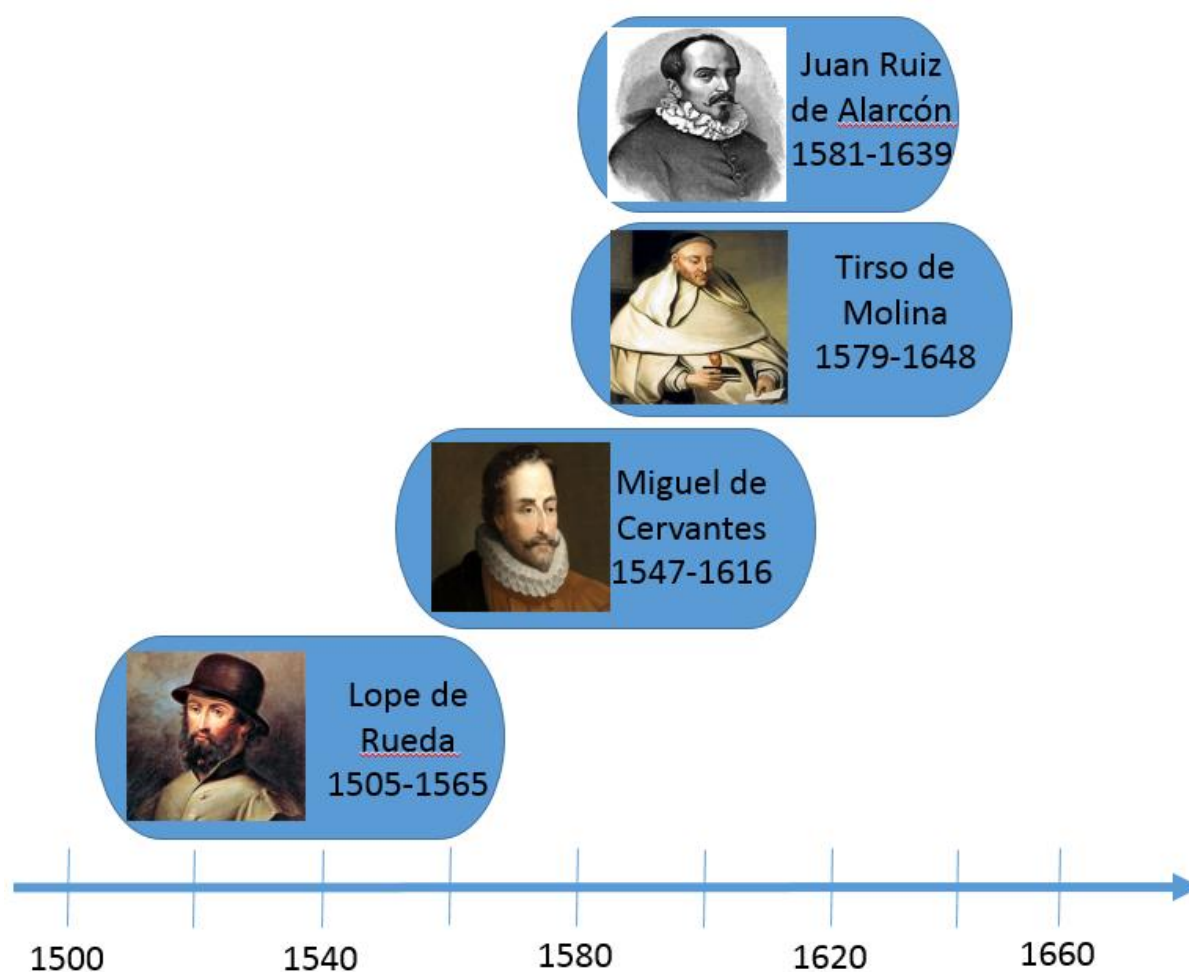
¹⁷ Vgl. Delgado, Mariano, *Das Spanische Jahrhundert (1492-1659)*, Darmstadt, 2016, S. 126 f. und Simson: *Das Siglo de Oro*, S. 59 f. und Stenzel, *Einführung in die spanische Literaturwissenschaft*, S. 136 f. und Strosetzki, *Geschichte der spanischen Literatur*, S. 182-185.

ein religiöses Drama über Moral, Sittenverfall und Sünde. Stilistisch wird ihm eine Nähe zu Vegas *comedia nueva* nachgesagt.¹⁸

Juan Ruiz de Alarcón y Mendoza (1581?-1639)

Juan Ruiz de Alarcón y Mendoza wurde in den amerikanischen Kolonien geboren und bemühte sich nach seiner Übersiedlung nach Spanien zuerst um eine juristische, später um eine administrative Anstellung in der Kolonialverwaltung. Obgleich er nie als hauptberuflicher Schreiber auftrat, sind eine Reihe von Werken mit seiner sicheren Autorenschaft überliefert. Diese Werke werden ähnlich wie die Stücke von Tirso als moralische Lehrstücke betrachtet. Allerdings setzte Ruiz mehr auf eine „Art humanistische Sittenkritik“¹⁹ statt auf religiöse Moral. Moralische Überlegenheit erhielten seine Figuren durch die Ablehnung von Vorurteilen und die Beherrschung der Leidenschaft, während Ruiz wiederum besonders Lüge, bösen Tratsch und Undankbarkeit in seinen Werken als moralisch schlechte Eigenschaften anklagte.²⁰

Die folgende Darstellung verdeutlicht, wie die vier Autoren zeitlich zueinander einzuordnen sind.²¹



¹⁸ Vgl. Simson: *Das Siglo de Oro*, S. 63 ff. und Strosetzki, *Geschichte der spanischen Literatur*, S. 184 f.

¹⁹ Strosetzki, *Geschichte der spanischen Literatur*, S. 185.

²⁰ Vgl. Simson: *Das Siglo de Oro*, S. 65 f. und Strosetzki, *Geschichte der spanischen Literatur*, S. 185 f.

²¹ Bilder entstammen aus <https://www.biografiasyvidas.com/> (Stand August 2017)

1.4. Thesen

Die vorgestellten historischen Persönlichkeiten lebten im relativ nahen Zeitraum zueinander und waren alle als Dramatiker tätig. Doch sie unterschieden sich von Art und Stil der Stücke und des besuchenden Publikums. Während Rueda und Cervantes hauptsächlich für kleine, für die Unterhaltung des Publikums gedachte, Einschübe (*pasos* und *entremes*) bekannt sind, richteten sich Tirso und Ruiz an das von Vega neu definierte Theater der *comedia nueva* und verstanden das Theater unter anderem als Mittel zur Moralkritik. Es wird daher davon ausgegangen, dass zwischen den Autoren mit ähnlicher Intention und Schwerpunkten eine größere Ähnlichkeit als zu den anderen Autoren besteht.

Unterthese I

Da Cervantes nachgesagt wird, dass seine *entremés* eine Fortführung der *pasos* von Rueda seien, sollte die Ähnlichkeit zwischen den Werken von Cervantes und Rueda, trotz der jährlichen Differenz, größer sein, als von Cervantes Werken zu den Werken seiner Zeitgenossen Tirso und Ruiz.

- 1) Die Ähnlichkeit zwischen den *pasos* von Lope und den *entremés* von Cervantes sollte größer sein als die Ähnlichkeit zwischen den *entremés* von Cervantes und den *entremés* von Tirso.
- 2) Die Ähnlichkeit zwischen den *comedias* von Lope und den *comedias* von Cervantes sollte größer sein als die Ähnlichkeit zwischen den *comedias* von Cervantes und den *comedias* von Tirso oder den *comedias* von Cervantes und den *comedias* von Ruiz.

Unterthese II

Aufgrund der Themenähnlichkeit und der Berufung auf Vega sollte die Ähnlichkeit der Werke von Tirso und Ruiz größer sein als die Ähnlichkeit der Werke eines der beiden Autoren zu den Werken von Cervantes.

- 1) Die Ähnlichkeit zwischen den *comedias* von Ruiz und den *comedias* von Tirso sollte größer sein als die Ähnlichkeit zwischen den *comedias* von Cervantes und den *comedias* von Ruiz oder den *comedias* von Cervantes und den *comedias* von Tirso.

Werden also die *comedias* von Ruiz mit den *comedias* von Tirso verglichen, sollte eine größere Ähnlichkeit sichtbar sein, als beim Vergleich der *comedias* von Cervantes mit den *comedias* von Ruiz oder Tirso.

2. Datengrundlage

2.1. Deutsches Testkorpus

Für diese Arbeit wurden zwei verschiedene Korpora für unterschiedliche Zwecke verwendet. Das Testkorpus beinhaltet vier zufällige ausgewählte Texte deutschsprachiger Autoren und dient allein als Testumgebung für später verwendete Methoden. Das Ziel ist, mit einer begrenzten Menge an Texten

die Methoden zu prüfen und Fehler und Schwächen zu identifizieren und beheben zu können. Die Ergebnisse dieses Korpus sind daher für die eigentliche Fragestellung dieser Arbeit nicht relevant und werden daher getrennt von den Ergebnissen des Hauptkorpus nur kurz dargestellt.

Die vier Texte sollten folgende Kriterien erfüllen:

- Die gleiche Sprache für alle Texte, in diesem Fall Deutsch.
- Unterschiedliche Textlänge, um mögliche Fehler der Methoden in Bezug auf Textlänge zu erkennen
- Freie Verfügbarkeit (kein aktives Urheberrecht) und Verfügbarkeit als TEI Datei.

Nach diesen Vorgaben wurden die Werke *Kabale und Liebe – ein bürgerliches Trauerspiel in fünf Aufzügen* und *Dom Karlos – Infant von Spanien* von Friedrich Schiller, *Der zerbrochene Krug – ein Lustspiel* von Heinrich von Kleist und *Komzesse Mizzi oder Der Familientag* von Arthur Schnitzler ausgesucht.

Die folgende Übersicht zeigt die ausgewählten Dramen samt Autor, Ort und Erscheinungsjahr. Die jeweilige Quelle ist als Fußnote verzeichnet.

Titel	Autor	Verlag und Ort	Erscheinungsjahr
Kabale und Liebe – ein bürgerliches Trauerspiel in fünf Aufzügen ²²	Friedrich Schiller	Schwan, Mannheim	1784
Dom Karlos/ Don Carlos – Infant von Spanien ²³	Friedrich Schiller	Göschen, Leipzig	1787
Der zerbrochne Krug – ein Lustspiel ²⁴	Heinrich von Kleist	Realschulbuchhandlung, Berlin	1811
Komtesse Mizzi oder Der Familientag ²⁵	Arthur Schnitzler	Fischer, Berlin	1909

Die Texte wurden von den genannten Adressen als TEI Dateien runtergeladen und die Textauszeichnung anschließend reduziert. Nach dem Vorbild der spanischen Werke erhielt jedes TEI der deutschen Stücke eine Castlist und jede Person darin eine eigene ID. Diese ID wurde anschließend für jeden Sprecherauftritt als Attribut für das <sp> Tag übernommen. Die folgenden Abbildungen

²² TEI: http://www.deutschestextarchiv.de/book/view/schiller_kabale_1784/ (Stand 05.2017).

²³ TEI: http://www.deutschestextarchiv.de/book/show/schiller_domkarlos_1787/ (Stand 05.2017).

²⁴ TEI: http://www.deutschestextarchiv.de/book/view/kleist_krug_1811/ (Stand 05.2017).

²⁵ TEI: https://textgridrep.org/browse/-/browse/v3sw_0 (Stand 05.2017).

zeigen einen Auszug aus der Castlist des Werkes *Kabale und Liebe* mit ID für die Sprecher und einen Auszug aus dem Haupttext mit ID für die Sprecheranteile.

Auszug Castliste

```
<castItem>
  <role xml:id="te0001-04">Wurm, Haussecretär des Präsidenten.</role>
</castItem>
<castItem>
  <role xml:id="te0001-05">Miller, Stadtmusikant oder, wie man sie an einigen Orten nennt, Kunstpfeifer.</role>
</castItem>
<castItem>
  <role xml:id="te0001-06">Dessen Frau.</role>
</castItem>
```

Auszug Haupttext

```
<sp who="#te0001-04">
  <speaker>Wurm.</speaker>
  <p>Meinerseits, meinerseits, Frau Base. Wo eine Kavaliersgnade
    einspricht, kommt mein bürgerliches Vergnügen in gar keine
    Rechnung.</p>
</sp>
<sp who="#te0001-06">
  <speaker>Frau.</speaker>
  <p>Was Sie nicht sagen, Herr Sekertare! Des Herrn Majors von Walter
    hohe Gnaden machen uns wohl je und je das Bläsier, doch verachten
    wir darum niemand.</p>
</sp>
<sp who="#te0001-05">
  <speaker>Miller</speaker>
  <stage>verdrüßlich. </stage>
  <l>Dem Herrn einen Sessel, Frau. Wollens ablegen, Herr Landsmann?</l>
</sp>
```

Nach erfolgreicher Umwandlung besaß das deutsche Testkorpus die gleiche Auszeichnung, wie das spanische Hauptkorpus.

2.2. Spanisches Hauptkorpus

Das Hauptkorpus bilden insgesamt 99 Theaterstücke der vier vorgestellten spanischen Autoren,²⁶ welche ursprünglich aus der digitalen Quelle Teatro Español del Siglo de Oro stammen²⁷ und in vorherigen Projekten in TEI umgewandelt wurden.

²⁶ Eine detaillierte Ansicht über Titel und Typ der einzelnen Texte findet sich im Anhang, S.

²⁷ <http://teso.chadwyck.com/> (Stand September 2017).

	Pasos	Entremés	Comedia
Lope de Rueda (1505?-1565)	7	0	4
Miguel de Cervantes Saavedra (1547-1616)	0	7	9
Tirso de Molina (1579?-1648)	0	9	43
Juan Ruiz de Alarcón y Mendoza (1581?-1639)	0	0	20

Da die Textgattungen *paso/entremé* und *comedia* inhaltlich sehr unterschiedlich sind, werden die Theaterstücke anhand der Textgattung getrennt und nicht vermischt. Damit wird ein Vergleich eines *paso* oder eines *entremé* mit einer *comedia* vermieden. Alle Werke lagen als TEI Dateien vor und waren ähnlich dem deutschen Testkorpus mit einer Castlist versehen. Jeder Sprecher erhielt auch hier eine eigene ID, wobei anzumerken ist, dass in Rahmen vorangegangener Textüberarbeitungen die ID-Vergabe bei einigen Texten nicht eindeutig durchgeführt werden konnte. So war es möglich, dass in der Castlist zwei Figuren mit den gleichen Anfangsbuchstaben auftauchten und die für die Sprecher verwendeten Abkürzungen keine eindeutige Unterscheidung zwischen den Figuren ermöglichte.

Im Folgenden ist ein Auszug aus der Castlist von Tirsos Werk *La prvdencia en la mvger* zu sehen, welche zwei Figuren namens Don Iuan enthält.

```
<castItem><role xml:id="te0497-01">Don Iuan.</role></castItem>
<castItem><role xml:id="te0497-02">Don Diego.</role></castItem>
<castItem><role xml:id="te0497-03">Carrillo criado.</role></castItem>
<castItem><role xml:id="te0497-04">Don Luys.</role></castItem>
<castItem><role xml:id="te0497-05">Vn Mayordomo.</role></castItem>
<castItem><role xml:id="te0497-06">Don Nuño.</role></castItem>
<castItem><role xml:id="te0497-07">Don Melendo.</role></castItem>
<castItem><role xml:id="te0497-08">El Rey de 17. años.</role></castItem>
<castItem><role xml:id="te0497-09">Garrote Pastor.</role></castItem>
<castItem><role xml:id="te0497-10">La Reyna doña Maria.</role></castItem>
<castItem><role xml:id="te0497-11">El Rey Fernando IIII.</role></castItem>
<castItem><role xml:id="te0497-12">Vn criado.</role></castItem>
<castItem><role xml:id="te0497-13">Don Iuan Alonso Caruajal.</role></castItem>
```

Im Werk selbst tauchen mehrere Figuren mit den Kürzungen *Iu.*, *Iua.* und *Iuan.* auf.

```
</sp><sp who="#te0497-13"><speaker>Iu.</speaker>

<l>A quien, que sabello intento?</l>

</sp><sp who="#te0497-11"><speaker>Re.</speaker>

<l>El que està, en esse aposento</l>
<l>os dirà para quien es.</l>

<stage>Vase.</stage>

</sp><sp who="#te0497-13"><speaker>Iuan.</speaker>
```

Eine eindeutige Aussage, welche Abkürzung nun zu welchen der beiden *Don luans* in der Castlist gehört, ist ohne Überprüfung am Originaltext nicht möglich. Zusätzlich musste bei einigen spanischen Theaterstücken die Castliste entweder erweitert werden, wenn Figuren im Text auftraten, die nicht in der Castlist aufgelistet waren oder Figuren, die nicht einer Person zuzuordnen waren, wie *todos* als Angabe, dass nun alle Sprecher der Szene sprachen.

Verwendete 90er IDs in den spanischen Texten

99	Tod./Todos (alle)
98	speaker ana="implicit"
97	(unknown speaker)
96	Las Dos
95	Name, y Name (Zwei Sprecher gleichzeitig)
94	Keine eindeutige Zuordnung möglich
93	otro/otros (andere)

Da diese Fehler jedoch nicht regelmäßig auftauchen, sondern in bestimmten Spezialfällen, werden die betroffenen Werke im Corpus belassen. Es sei aber unter diesen Umständen darauf hinzuweisen, dass die Ergebnisse der Netzwerkberechnungen nicht zu 100% das Original abbilden können. Da aber keine genaue Überprüfung der Autorentexte, sondern der Vergleich der Autoren angestrebt wird, werden diese Fehler wissend in Kenntnis genommen.

3. Methode

3.1. Methodenübersicht

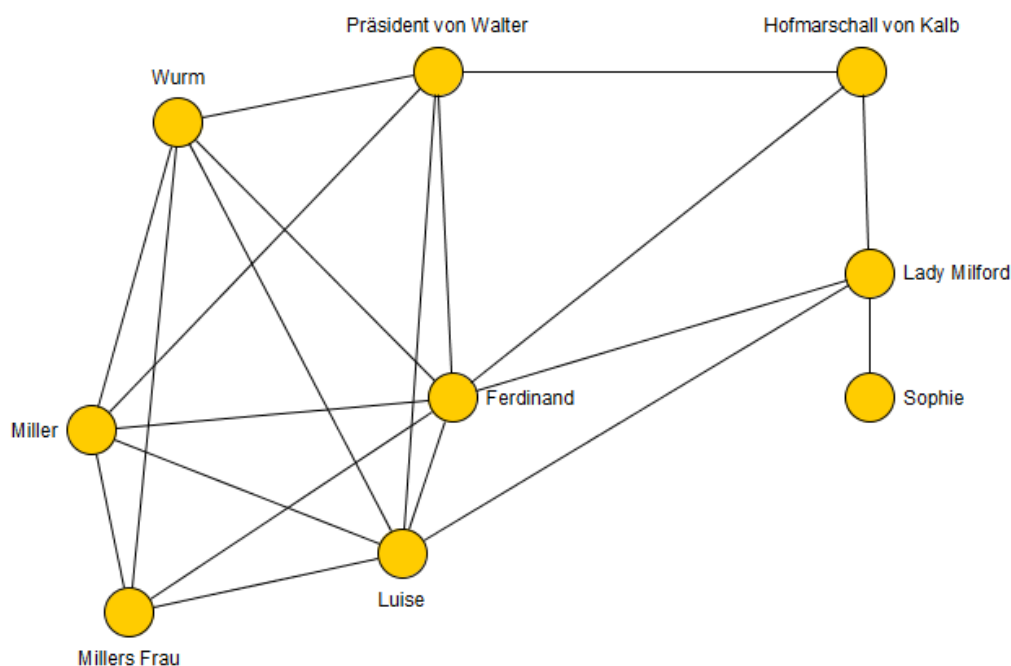
Das Ziel ist es aus den Texten die Daten zu ziehen, welche benötigt werden, um Merkmale von Netzwerken berechnen zu können. Dies sind zum einen die Anzahl der Sprecher, dargestellt als Knoten, und die Anzahl der Verbindungen zwischen den Sprechern, dargestellt als Kanten. Mit Hilfe dieser Informationen können die Figurenkonstellationen der Werke als Netzwerke visualisiert werden. Aus diesen Netzwerken wiederum lassen sich die Merkmale berechnen, welche dabei helfen, die Werke der Autoren nach einem einheitlichen Maß einzuordnen. Der Vorgang kann daher in folgende Abschnitte unterteilt werden.

- I. Einlesen der Theaterstücke mit Hilfe von Python.
- II. Extrahierung von Knoten und Kanten mit Hilfe selbst definierter Funktionen und Formeln.
- III. Erstellen von Netzwerken mit Hilfe der Pythonbibliothek
NetworkX und Berechnung der Merkmale.
- IV. Vergleich der Merkmale.

3.2. Sprecher und Verbindungen

Bedingt durch den Ansatz, dass aus den Texten zuerst die Anzahl der Sprecher und Verbindungen gefiltert und anschließend daraus Netzwerke und Merkmale der Figurenkonstellationen erstellt werden sollen, ist zuerst zu erklären, wie diese Netzwerke aussehen können und wie Sprecher und Sprecherfolge dargestellt werden können. Zudem zeigt eine solche Darstellung auch, welche Probleme zu erwarten sind, welche sich nicht durch die Netzwerkanalyse lösen lassen.

Die folgende Abbildung zeigt ein manuell gezeichnetes, ungerichtetes Figurennetzwerk von Friedrich Schillers *Kabale und Liebe* ohne Gewichtung der Kanten.



Das Netzwerk präsentiert die Figurenkonstellation des gesamten Werkes, es sind also alle Kapitel enthalten. Jede eindeutig zuzuordnende Figur wird dabei als gelber Punkt mit Beschriftung des Sprechernamens dargestellt. Das sind die Knoten des Graphen.²⁸ Zwei Knoten werden dann mit einer Linie verbunden, wenn diese zwei Sprecher nacheinander im Text auftreten und also sehr wahrscheinlich einen Dialog miteinander führen. Diese Verbindungslinien im Graph werden Kanten genannt.²⁹ Der Grad eines Knoten bezeichnet die Verbindung zu anderen Knoten, also die Anzahl der verbundenen Kanten.

²⁸ Manchmal auch Ecken genannt, vgl. Diestel, Reinhard, *Graphentheorie*, Berlin, ³2006, S. 2 f.

²⁹ Vgl. Ebd. S. 2 f.

Dies wir für alle im Text auftauchenden Dialogpaare gemacht, sodass am Ende ein Graph beziehungsweise ein Netzwerk, aller Figuren gezeichnet ist, welche im Werk direkten Kontakt zueinander haben.³⁰

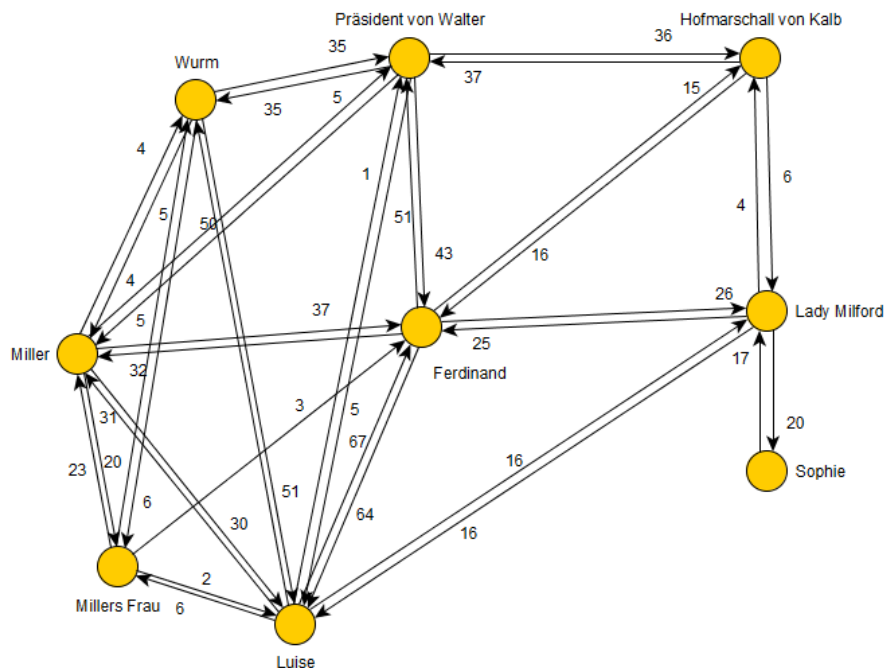
Die Kanten zwischen den Knoten *Lady Milford*, *Luise* und *Ferdinand* geben somit an, dass die Figur *Lady Milford* mindestens einmal auf *Ferdinand* und einmal auf *Luise* folgt, ein Gespräch zwischen *Lady Milford* und den beiden Personen also sehr wahrscheinlich stattfindet. Gleiches gilt für die Figuren *Ferdinand* und *Luise*. Der Knoten *Sophie* hat dagegen nur eine Kante zum Knoten *Lady Milford*, die Figur der *Sophie* tritt daher mit keiner anderen Hauptfigur außer *Lady Milford* in unmittelbaren Kontakt.

Diese Darstellung funktioniert allerdings nur für die Figuren, welche eindeutig identifizierbar sind. Wird zum Beispiel ein Sprecher als Platzhalter für verschiedene Nebenfiguren verwendet, so ist eine eindeutige Zuordnung der Nebenfiguren zu einem Knoten nicht möglich. Im konkreten Beispiel treten im Stück *Kabale und Liebe* mindestens drei oder mehr unterschiedliche Kammerdiener mit unterschiedlichen Funktionen auf. Während die Kammerdiener des Präsidenten und Ferdinands nur als Befehlsempfänger und Meldegänger agieren, hat der Kammerdiener der Lady Milford eine eigene Rolle, indem er die Lady über den Soldatenhandel des Landesfürsten³¹ unterrichtet. Allerdings sind alle Kammerdiener ohne Namen und eine genaue Zuordnung, welcher Diener nun mit dem Präsidenten oder Ferdinand agiert ist nicht möglich. Ähnlich verhält es sich mit den als Nebenfiguren klassifizierten Sprechern, welche allesamt Bediente der Lady Milford sind. Diese sind, genauso wie die Kammerdiener, alle namenlos und treten nur in Erscheinung, um Befehle zu empfangen oder eine Person anzukündigen. Da diese Diener nicht zu unterscheiden sind und damit eine genaue Zuordnung, welcher Diener mit wem agiert, nicht möglich ist, werden sie in diesem Beispielnetzwerk nicht angezeigt.

Erweitert man dieses Netzwerk nun um die Information, welcher Sprecher wie häufig auf welchen Sprecher folgt, erhält man einen gerichteten Graphen mit Gewichten.

³⁰ Abhängig von der Art des Netzwerkes können die Punkte und Verbindungen des Netzwerkes auch anders benannt werden. Vgl. Newman, M., *Networks An Introduction*, Oxford, 2010, S. 109 f.

³¹ Der Verkauf von eigenen Soldaten an die britische Armee für den Kolonialeinsatz im Amerikanischen Unabhängigkeitskrieg.



Diese Darstellung ist eine Erweiterung des ersten Netzwerkes, wobei jede Kante nun eine Richtung und eine Gewichtung erhalten hat.

Die Richtung wird durch einen Pfeil dargestellt, während die Gewichtung durch eine kleine Zahl neben dem Pfeil visualisiert wird. Es ist damit ein Digraph entstanden (*directed graph*), also ein gerichteter Graph bei welchem jede Kante eine Richtung und optional eine Gewichtung besitzt.³²

Bei gerichteten Netzwerken werden Knoten nach Art ihres Grades häufig in Hubs oder Authorities unterteilt. Hubs sind dabei Knoten, von welchen viele Pfeile weggehen. Sie stellen damit wichtige Knotenpunkte im Netzwerk da und helfen andere Bereiche des Netzwerkes zu erreichen. Authorities bezeichnen Knoten, zu welchen verhältnismäßig viele Pfeile hingehen.³³

In diesem Beispiel zeigt die Gewichtung, wie oft welcher Sprecher auf wen folgt. Zum Beispiel folgt die Kammerzofe *Sophie* 20 Mal auf *Lady Milford*, während *Lady Milford* 17 Mal auf *Sophie* folgt.

Doch wie ergibt sich diese Lücke, dass eine Person 20 Mal und die andere dafür nur 17 Mal auf diese folgt? Werden die Auftritte händisch gezählt, ergeben sich bei diesem Beispiel folgende Zahlen: Die Figur *Sophie* hat 22 Auftritte im Werk. Davon folgt sie bei 20 Auftritten direkt auf die Sprecherin *Lady Milford*. Allerdings erscheinen zwischenzeitlich manchmal Diener, wodurch die Folge der Auftritte von *Sophie* und *Lady Milford* unterbrochen wird. Als Beispiel siehe den folgenden Textabschnitt aus dem 2. Akt, 2. Szene von *Kabale und Liebe*.

³² Vgl. Jannidis, Kohle und Rehbein, *Digital Humanitis*, S. 147 ff. und Newman, *Networks*, S. 112 ff.

³³ Vgl. Krischke und Röpcke, *Graphen und Netzwerktheorie*, S. 146 f.


```

<sp who="#te0001-08">
  <speaker>Sophie.</speaker>
  <p> Aber Juwelen wie diese! Hätten Sie nicht Ihre schlechtern nehmen
    können? Nein, wahrlich, Mylady! Es ist Ihnen nicht zu vergeben.</p>
</sp>
<sp who="#te0001-03">
  <speaker>Lady.</speaker>
  <p> Närrisches Mädchen! Dafür werden in einem Augenblick mehr
    Brillanten und Perlen für mich fallen, als zehen Könige in ihren
    Diademen getragen, und schönere -</p>
</sp>
<sp who="#te0001-10">
  <speaker>Bedienter</speaker>
  <stage> kommt zurück. </stage>
  <l> Major von Walter -</l>
</sp>
<sp who="#te0001-08">
  <speaker>Sophie</speaker>

```

Hier spricht zuerst *Sophie*, anschließend *Lady Milford*, anschließend ein *Bedienter* und dann wieder *Sophie*. Die Kette zwischen *Sophie* und *Lady Milford* wird damit kurzzeitig unterbrochen. In der maschinellen Auszählung des Scripts fehlt damit eine Folge von *Sophie* auf *Lady Milford*. Da dieser Fall mehrmals auftritt, ergibt sich damit am Ende der Fall, dass beide Personen unterschiedlich oft unmittelbar nacheinander sprechen.

Ähnlich verhält es sich bei den Gesprächen zwischen dem *Präsidenten* und seinen Sekretär *Wurm* oder den Gesprächen zwischen dem *Präsidenten* und seinem Sohn *Ferdinand*. Immer wieder wird die Sprecherfolge durch einen Diener unterbrochen, welcher Befehle entgegennimmt oder eine Person ankündigt. Der Dialog zwischen den beiden Hauptsprechern wird dabei zwar kurz unterbrochen, aber nicht abgebrochen.

Ein anderer Fall tritt beim Dialog zwischen *Luise* und ihrer Mutter auf. *Frau Miller* tritt im Text insgesamt 35 Mal auf. Sie spricht sechs Mal nach *Luise*, *Luise* allerdings nur zwei Mal nach ihrer Mutter. Das hat jedoch damit zu tun, dass *Luise* und *Frau Miller* nur in Anwesenheit von *Herrn Miller* miteinander sprechen, die drei Sprecher also zusammen einen Dialog führen. Sehr häufig folgt dabei auf *Frau Miller* zuerst *Herr Miller* und dann erst *Luise*.

Diese Beispiele zeigen, dass allein mit dem Zählen der Abfolge der Sprecher nicht alle Dialoge vollständig erfasst werden können. Allerdings müsste der Text von einem Leser vollständig geprüft werden, um alle Dialoge eindeutig bestimmen zu können. Da dies aber besonders bei den spanischen Texten in vollem Umfang nicht möglich ist, ist davon auszugehen, dass abhängig vom Werk eine bestimmte Anzahl an Dialogen nicht erfasst wird. Es ist aber anzunehmen, dass alle Personen, welche im Dialog stehen, mindestens einmal unmittelbar aufeinander folgen und somit zumindest bei ungerichteten Graphen das Ergebnis dadurch nicht negativ beeinflusst wird.

3.3. Skriptbeschreibung zur Erstellung der Netzwerke

Für die Ermittlung der Merkmale wird auf die Programmiersprache Python (Version 3.4) zurückgegriffen. Der Aufbau und die Funktionen des Skriptes werden nun Schritt für Schritt erklärt.

Verwendete Bibliotheken

Für das Skript werden folgende Bibliotheken benötigt.

- `glob` (<https://docs.python.org/3.4/library/glob.html>) (Stand Oktober 2017)
- `os` (<https://docs.python.org/3.4/library/os.html>) (Stand Oktober 2017)
- `re` (<https://docs.python.org/3.4/library/re.html>) (Stand Oktober 2017)
- `pandas` (<https://pandas.pydata.org/pandas-docs/stable/index.html>) (Stand Oktober 2017)
- `collections` (<https://docs.python.org/3.4/library/collections.html>) (Stand Oktober 2017)
- `networkx` (<https://networkx.github.io/>) (Stand Oktober 2017)

Die Bibliotheken *glob* und *os* werden für die Navigation Pythons in den entsprechenden Ordnern benutzt, um für das Pythonscript die benötigten Input Texte und den Ausgabe Ordner zu definieren. Die Bibliothek *re* ermöglicht die Verwendung von Regulären Ausdrücken (*Regular Expressions*) und ist notwendig für das Finden der Sprecher im Volltext.³⁴

Mit Hilfe von *pandas* lassen sich DataFrames erstellen, um in diesen die Ergebnisse der Berechnungen zu speichern und später als CSV-Datei auszugeben.

Die Bibliothek *collections* beinhaltet die Funktion *Counter*, welche die Anzahl von gleichen Elementen in Listen berechnet.

Die Bibliothek *networkx* ist notwendig für die Erstellung der Graphen und Berechnung der Netzwerkmerkmale.

`main`

```
def main(inputpath, outputpath):
    for file in glob.glob(inputpath):
        find_all(file)

main('./input/*.xml', './output/')
```

In der Funktion mit den frei gewählten Namen *main* wird in kurzen Zeilen definiert, das im Ordner *input* (welcher im Unterordner des Ordners des Pythonscripts liegt), alle Dateien mit der Endung *xml* für die Funktion *find_all* verwendet werden sollen. Die Ergebnisse sollen abschließend im Ordner *output* gespeichert werden.

`find_all`

Diese Funktion mit frei gewählten Namen ist die Hauptfunktion, in welchen alle Schritte des Scripts durchgeführt werden.

³⁴ Für eine gute Einführung in Reguläre Ausdrücke siehe Fitzgerald, Michael, *Introducing Regular Expressions*, Sebastopol, 2012.

Nacheinander werden alle folgenden Schritte für jede Datei mit xml Endung im input Ordner durchgeführt.

Namen der Datei finden

```
"""I. Name der Datei"""
file_name = os.path.basename(file)[-4]
print (file_name)
```

Dieser Schritt dient der Identifizierung der Datei. Mit dem Befehl *os.path.basename* wird der Dateiname gelesen, wobei die letzten vier Stellen, welche das .xml bilden, vom Namen abgeschnitten werden. Aus dem Dateinamen Ruiz_15730410_te0421.xml wird so Ruiz_15730410_te0421.

Textinhalt der Datei lesen

```
"""II. Textinhalt der Datei"""
with open(file,"r", encoding = "utf-8") as sourcefile:
    full_text = sourcefile.read()
```

Hier wird die Datei geöffnet und der gesamte Textinhalt (full_text) mit dem Befehl *read()* gelesen.

Rollen erfassen

```
"""III. Rollenliste der Datei"""
roles_ids=re.findall("xml:id=\"te(.*)\">", full_text)
```

Da jede Datei eine Castlist besitzt, in welchen die Sprecher mit folgenden Tag ausgezeichnet sind: <role xml:id="Id">Name</role>, kann mit obigen Regulären Ausdruck eine Liste aller Rollennamen erstellt werden.

Sprecher erfassen

```
"""IV. Speaker_Ids der Datei"""
all_ids_in_text = re.findall("who=\"#te(.*)\">", full_text)
```

Nun sollen alle Sprecher im Haupttext identifiziert werden. Dies ist besonders deshalb wichtig, da einige Sprecher zwar im Text vorkommen, nicht aber in der Castlist, darunter die 90er IDs.

Eine zweite ID Liste

```
"""V. Eine zweite Id Liste anlegen, falls 90er Ids vorhanden sind"""
ids_no_ninety = [item for item in all_ids_in_text if not re.search(r".*-9[0-9]", item)]
```

Da sich die Anzahl der Rollennamen nicht mit denen der Sprecher IDs deckt, wenn 90er IDs vorhanden sind, wird mit dieser Zeile jedes Item aus der Sprecherliste außer die 90er IDs in die neue Liste *ids_no_ninety* übertragen.

Eine zweite Rollenliste

```
"""VI. Eine zweite Rollenliste anlegen, falls 90er Ids vorhanden sind"""
roles_ninety = roles_ids.copy()
for item in all_ids_in_text:
    ninety = re.findall(".*-9[0-9]", item)
    if ninety not in roles_ids:
        roles_ninety.extend(ninety)
```

Da allerdings auch der Fall betrachtet werden soll, was passiert, wenn die 90er IDs mit in die Merkmalsberechnung einfließen, wird eine zweite Rollenliste erstellt, welche um alle 90er IDs aus dem Text erweitert wird. Mit diesen zwei Schritten sind nun die Daten vorhanden, um sowohl die Merkmale mit Einfluss der 90er IDs als auch ohne Einfluss der 90er IDs zu ermitteln.

Die Folge der Sprecher

Nun muss ermittelt werden, welcher Sprecher auf wen folgt. Es wird ein Fenster konstruiert, das von Person zu Person läuft und alle Personen in Zweier-Gruppen einteilt. Dieser Durchlauf wird viermal durchgeführt.

- Einmal für zwei Personen ohne Gewichtung der Kanten und ohne 90er IDs.
- Einmal für zwei Personen ohne Gewichtung der Kanten und mit 90er IDs.
- Einmal für zwei Personen mit Gewichtung der Kanten und ohne 90er IDs.
- Einmal für zwei Personen mit Gewichtung der Kanten und mit 90er IDs.

Für alle vier Fälle folgt die gleiche Berechnung der Sprecherfolge, wobei jedoch unterschiedliche Grunddaten herangezogen werden. Für die Berechnung ohne 90er IDs wird die originale Castlist (*roles_ids*) und die angepasste Sprecherliste (*ids_no_ninety*) verwendet. Für die Berechnung mit 90er IDs dient die angepasste Castlist (*roles_ninety*) und die originale Sprecherliste (*all_ids_in_text*).

Sprecherpaare

```
list_of_teams = []
for item in zip(ids_no_ninety[:-1],ids_no_ninety[1:]):
    list_of_teams.append(item)
GroupofSpeaker=list(set(list_of_teams))
```

Im Falle von zwei Personen ohne Gewichtung und ohne 90er IDs wird nun aus den Items der Liste eine neue Liste erstellt, in welcher jeweils ein Paar aus zwei Sprechern gebildet wird. Im konkreten Fall werden in der Liste (hier *ids_no_ninety*) alle Sprecher nacheinander aufgeführt. Nun wird mit Hilfe einer For-Schleife diese Liste Position für Position abgegangen, wobei das Fenster stetig um ein Listenelement weitergeschoben wird.

```
['0482-00', '0482-00', '0482-01', '0482-00', '0482-01',
```

Die erste Person steht an der Position 0, die zweite an der Position 1 und so fort. Für das erste Element (*item*) in der Liste werden nun die Positionen 0 und 1 mit der Positionsangabe `[:-1]` und `[1:]` ermittelt, was 0482-00 und 0482-00 ergibt. Für das zweite Element geht das Fenster mit `[:-1]` und `[1:]` um einen Schritt weiter, wodurch nun Position 1 und Position 2 erfasst werden, in diesem Fall 0482-00 und 0482-01. Diese wird solange fortgeführt, bis der letzte Sprecher in der Liste erreicht wird.

```
Schritt:1
('0482-00', '0482-00')
Schritt:2
('0482-00', '0482-01')
Schritt:3
('0482-01', '0482-00')
Schritt:4
('0482-00', '0482-01')
Schritt:5
('0482-01', '0482-00')
```

Als Ergebnis erhält man eine Liste der Sprecherpaare. Da uns in diesem Fall nur jede einmalig vorkommende Verbindung interessiert, werden mit dem Befehl `set` die doppelten Verbindungen entfernt.

Da im obigen Falle keine Gewichtung der Kanten vorgenommen wird, muss die Liste nochmals angepasst werden, indem Kanten von Knoten zum gleichen Knoten und doppelte Kanten entfernt werden.

```
"""
- Entfernt Kante vom Knoten zum gleichen Knoten
- Entfernt doppelte Kante zwischen zwei Knoten
"""
for item in GroupofSpeaker:
    if (item[0] == item[1]):
        GroupofSpeaker.remove(item)
simple_edges = set(tuple(sorted(t)) for t in GroupofSpeaker)
```

Dies geschieht mit der obigen If-Schleife und dem Befehl `set`.

Der Fall, dass ein Knoten mit sich selbst verbunden ist, wie im obigen Beispiel der 0482-00 Wert, liegt daran, dass im Originaltext auf dem Sprecher mit der ID 0482-00 an dieser Stelle zuerst die 90er ID 0482-99 folgt und dann wieder die ID 0482-00. Da aber in einem der ersten Schritte für diese Liste alle 90er IDs gelöscht wurden, muss der dadurch entstandene Fehler mit der obigen Schleife entfernt werden, indem jedes Paar entfernt wird, dessen Item an Position 0 und 1 identisch ist.

Als Ergebnis erhält man nun eine Liste der Kanten, die die Knoten verbinden.

```
"""Anzahl der Knoten und der Verbindungen"""
all_nodes = (len(roles_ids))
all_edges = (len(simple_edges))
```

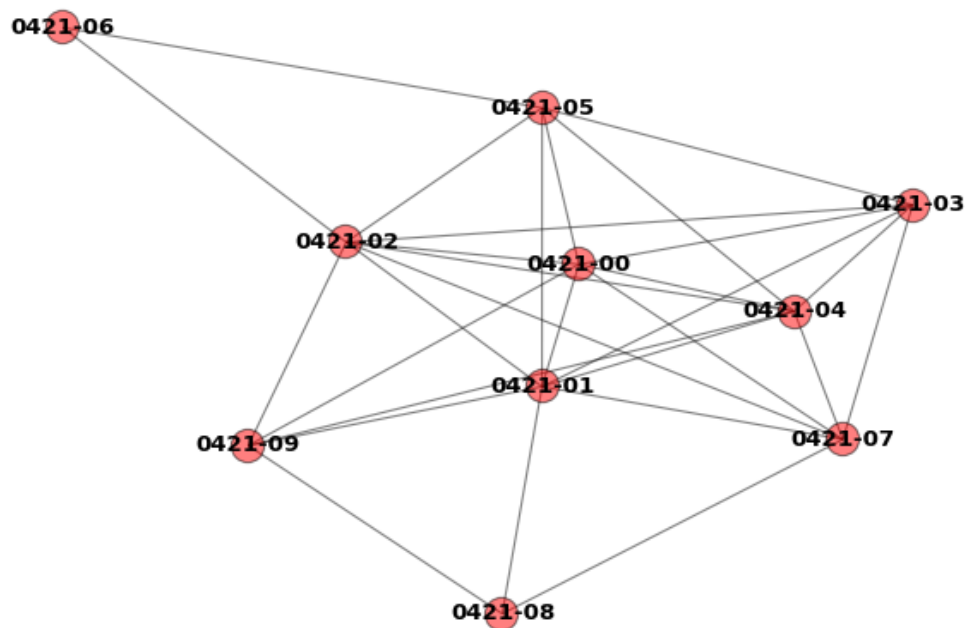
Mit der Länge dieser beiden Listen kann nun die Anzahl der Knoten und der Kanten für ein 2 Personen Fenster ohne Gewichtung und ohne 90er IDs ermittelt werden. Abschließend kann mit der Bibliothek *networkX* der Graph ermittelt werden.

Graph

```
""" Netzwerk """
G1 = nx.Graph()
G1.add_nodes_from(roles_ids)
G1.add_edges_from(simple_edges)
```

Als Knoten dient hierbei die Liste der originalen Rollenids, da diese keine 90er IDs besitzen. Die Kanten werden über die aus den Sprecherpaaren ermittelte Liste gebildet.

Folgendes Beispiel zeigt den Graphen des Netzwerkes des Werkes Ruiz_15730410_te0421 (*La amistad castigada*), ungewichtet und ohne 90er IDs.

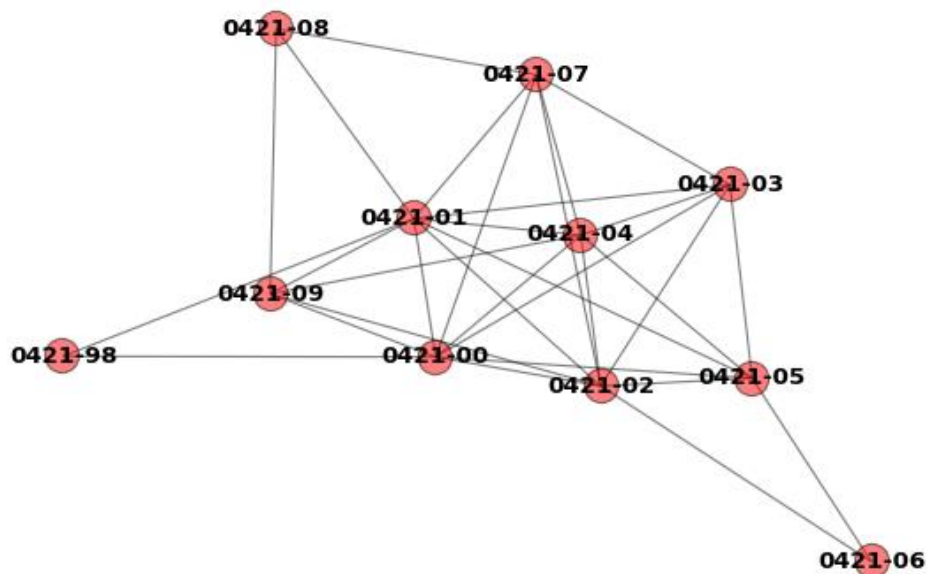


Der Fall mit 90er IDs

Nun ist zu untersuchen, wie der Fall aussieht, wenn 90er IDs berücksichtigt werden. Das Skript folgt dem gleichen Ablauf, verwendet aber statt der originalen Castlist und der veränderten Sprecherliste diesmal die erweiterte Castlist (*roles_ninety*) und die originale Sprecherliste (*all_ids_in_text*).

Die Berechnungen folgen dem gleichen Weg, nur dass diesmal die 90er IDs mitberechnet und im Graphen dargestellt werden.

Folgendes Beispiel zeigt den Graphen des Netzwerkes des Werkes Ruiz_15730410_te0421 (*La amistad castigada*), ungewichtet und diesmal mit 90er IDs (in diesem Fall die 98).



Graph mit gewichteten Kanten

Will man die Graphen nun mit Gewichten versehen, also zählen, wie häufig die Verbindung zwischen zwei Personen vorkommt, ist ein anderer Weg für die Berechnung der Kanten zu gehen.

```
list_of_teams = []
for z in zip(ids_no_ninety[:-1], ids_no_ninety[1:]):
    list_of_teams.append(z)
```

Zuerst wird wieder die Liste aus den Paaren der Sprecher gebildet.

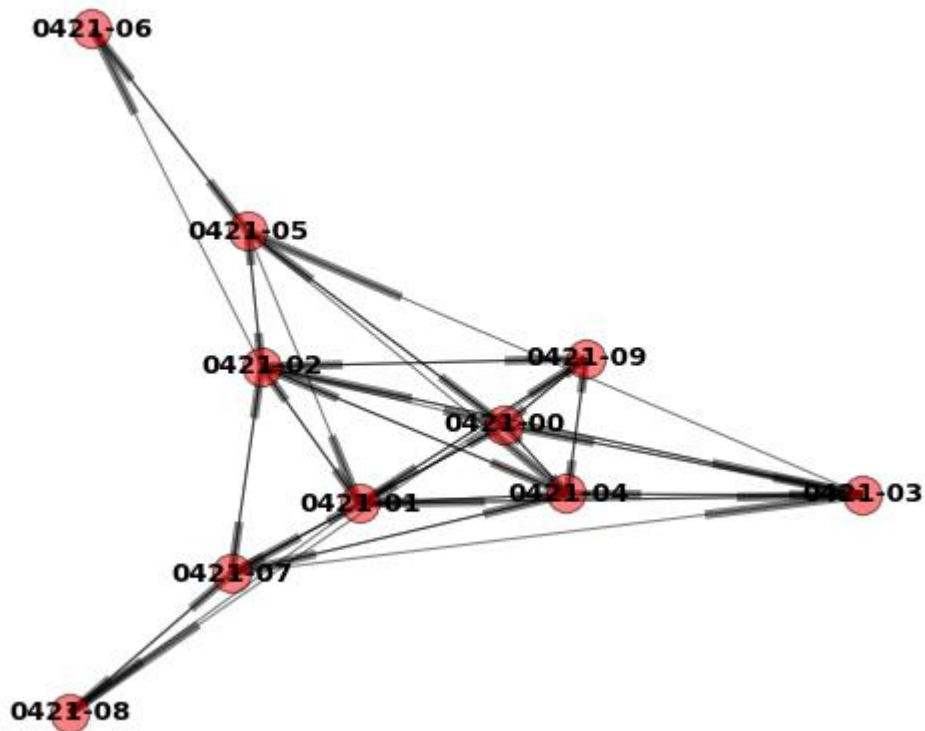
```
"""
- Entfernt Kante vom Knoten zum gleichen Knoten
"""
for item in list_of_teams:
    if (item[0] == item[1]):
        list_of_teams.remove(item)

counter=dict(Counter(list_of_teams))
```

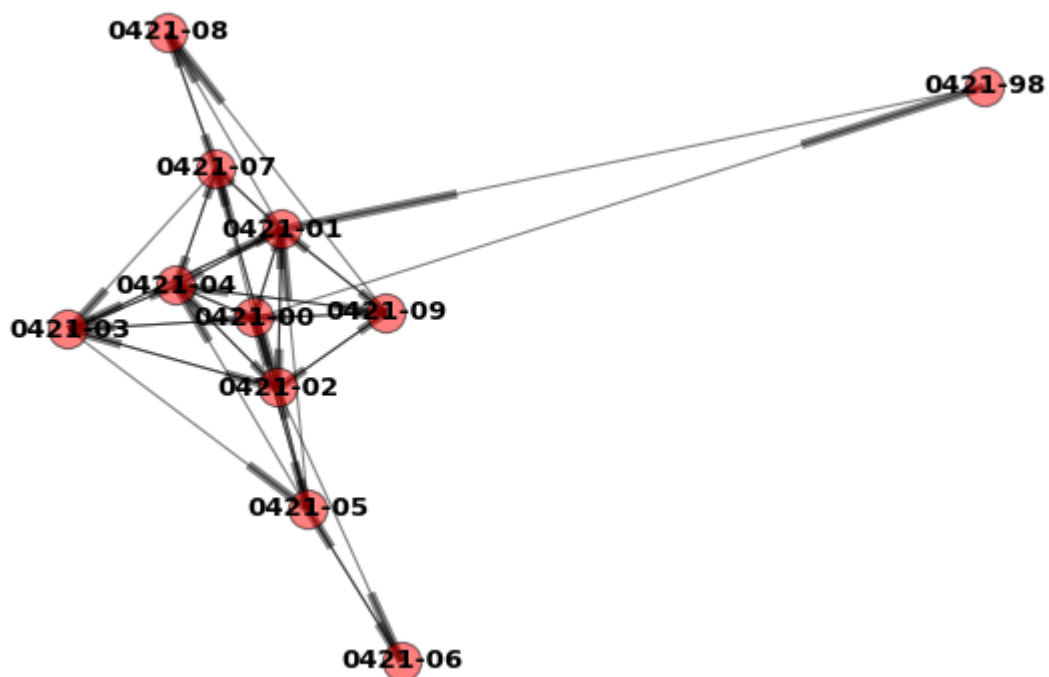
Dieses Mal werden jedoch nicht die Duplikate der Paare entfernt, sondern mit Hilfe der Funktion *Counter* aus *collections* gezählt.

```
""" Netzwerk """
G3 = nx.DiGraph((x, y, {'weight': v}) for (x, y), v in Counter(list_of_teams).items())
```

Daraus kann mit NetworkX ein sogenannter DiGraph erzeugt werden, also ein Graph, dessen Kanten gewichtet sind. In der vereinfachten Network-Darstellung ergibt sich damit folgender Graph vom Werk Ruiz_15730410_te0421 (*La amistad castigada*), mit gewichteten Kanten und ohne 90er IDs.



Ähnlich wird bei der Grapherstellung mit 90er IDs und gewichteten Kanten vorgegangen.



Da nun die Punkte und Kanten bekannt sind, können mit NetworkX die Merkmale berechnet werden.

3.4. Ermitteln der Merkmale

Allgemein ist zuerst die Frage zu klären, welche Merkmale aus den vorliegenden Daten erhoben werden können und ob die Erhebung bestimmter Merkmale für diese Arbeit sinnvoll ist?

Anzahl der Sprecher (n) / Anzahl der Knoten (n)

Vorteile	Nachteile
Einfach zu ermitteln	Nur grober Vergleich möglich
Unabhängig vom Netzwerk	Geringe Aussagekraft
Nötig für weitere Merkmale	
Verdeutlichen die Aussagekraft bestimmter Merkmale	

Das erste Merkmal ist die Anzahl der auftretenden Sprecher und zeigt, wie viele Figuren im Werk als Sprecher auftreten. Dieses Merkmal nimmt eine Sonderrolle ein und ist vorrangig als experimentelles Merkmal zu sehen. Eigentlich gehört die Anzahl der Sprecher nicht zu den Netzwerkmerkmalen und kann direkt aus dem Text abgezählt werden. Allerdings gibt es Vorteile, welche zumindest für eine Betrachtung der Sprecheranzahl sprechen.

Erstens ist die Anzahl der Sprecher sehr einfach und fehlerfrei zu ermitteln, indem die Anzahl einfach aus der Castlist extrahiert wird. Die Anzahl bietet einen groben Überblick, um die Werke der Autoren nach offensichtlichen Faktoren einzuordnen. Zusätzlich ist die Anzahl der Sprecher als Menge der Knoten im Netzwerk Voraussetzung für alle folgenden Merkmale und ist damit bei der Bearbeitung in jedem Fall zu ermitteln.

Zweitens ist die Anzahl der Sprecher notwendig, um andere Merkmale besser verstehen zu können. Zum Beispiel gibt der Degree Centrality Wert an, wie gut ein Knoten mit anderen vernetzt ist. Ein Wert von 1 bedeutet in diesem Fall, dass ein Sprecher im Werk mindestens einmal mit allen anderen Figuren des Werkes in Kontakt tritt. Dies ist nicht sehr ungewöhnlich, wenn das Werk weniger als 10 Figuren besitzt. Wenn allerdings mehr als 10 Figuren auftreten, wird es für jede weitere Figur unwahrscheinlicher, dass ein Degree Centrality Wert von 1 erreicht wird. Ein Degree Centrality Wert von 1 wäre somit bei 25 Personen etwas sehr Besonderes, da alle 25 Personen über eine Person verknüpft wären.

In einer Darstellung ohne Anzahl der Sprecher und Knoten würde dies nicht auffallen.

Werk A	Degree Centrality 1
Werk B	Degree Centrality 1

Daher wird die Anzahl der Sprecher bei den Merkmalen mitangeben.

Werk A	Anzahl der Sprecher: 5	Degree Centrality 1
Werk B	Anzahl der Sprecher: 25	Degree Centrality 1

Maximal $\Delta(G)$ -und Minimalgrad $\delta(G)$

Der Minimalgrad ($\delta(G)$) gibt die Anzahl an Kanten an, welcher der Knoten mit den wenigstens Verbindungen besitzt. Es ist auch möglich, dass ein Knoten isoliert und mit keinen anderen Knoten im Graphen verbunden ist. In diesem Fall hat der Minimalgrad den Wert 0. Der Maximalgrad $\Delta(G)$ dagegen liefert die Anzahl der Kanten, welche der Knoten mit den meisten Verbindungen besitzt.³⁵

```
"""Maximalen und Minimalen Grad"""
degree_sequence=sorted(nx.degree(G1).values(),reverse=True)
maxdegree = degree_sequence[:1]
mindegree = degree_sequence[-1:]
```

Die Grade im Graphen können mit dem Networkbefehl `degree().values()` ermittelt werden und müssen anschließend nur noch sortiert werden. An erster Stelle steht dann der Minimalgrad und an letzter Stelle der Maximalgrad. Beide Grade sind nützlich, um andere Merkmale, wie den Durchschnittsgrad, besser einschätzen zu können. Der Minimalgrad weist zudem auf isolierte Knoten hin, wenn der Wert null beträgt.

Durchschnittsgrad $d(g)$

Aus den Graden aller Knoten lässt sich anschließend der Durchschnittsgrad ($d(G)$) berechnen.³⁶

$$d(G) := \sum_{v \in V} \frac{d(v)}{|V|}$$

Der Durchschnittsgrad zeigt pro Graphen an, wie viele Kanten ein Knoten im Durchschnitt besitzt. Er ist damit ein guter Indikator, um zu ermitteln, ob die Knoten im Netzwerk viele oder wenige Verbindungen zu anderen Knoten besitzen. Wie viele Durchschnittswerte kann der Durchschnittsgrad durch besonders hohe oder niedrige Werte verzerrt werden. Daher sollte bei der Betrachtung des Durchschnittsgrads auch der Minimal-und Maximalgrad betrachtet werden.

```
"""Durchschnittsgrad"""
avdegree = sum(G1.degree().values())/float(len(G1))
```

In dieser Arbeit wurde der Durchschnittsgrad durch die Umsetzung der oberen Formel berechnet, indem die Summe der Gradwerte durch die Anzahl der Grade geteilt wurde.

höchste Degree Centrality C_{Degree}

Das Zentralitätsmaß (degree centrality) spiegelt die Bedeutung eines Knotens innerhalb eines Graphen wieder, indem die Anzahl der Kanten eines Knoten im Verhältnis zur Gesamtanzahl der Knoten gesetzt wird. Im Falle, dass das Zentralitätsmaß im Netzwerk ohne Gewichte den Wert 1 besitzt, bedeutet dies, dass der zugehörige Knoten mit allen anderen Knoten im Netzwerk verbunden ist und eine zentrale Rolle einnimmt.³⁷

³⁵ Vgl. Diestel, *Graphentheorie*, S. 4.

³⁶ Formel entnommen aus: ebd., S. 5.

³⁷ Vgl. Krischke und Röpcke, *Graphen und Netzwerktheorie*, S. 148 ff.

```

"""Zentralitätsmaß (degree centrality)"""
dc = sorted(nx.degree_centrality(G1).values(), reverse=True)
maxdc = (dc[:1])

```

Zur Berechnung der einzelnen Zentralitätsmaße besitzt NetworkX eine eingebaute Funktion. Da für diese Arbeit nur der höchste Degree Centrality Wert nötig ist, werden die Maße der Größe nach sortiert und der letzte Eintrag als höchster Degree Centrality übernommen. Bei Digraphen ist darauf zu achten, dass die Degree Centrality abhängig davon definiert wird, ob die einkommenden oder ausgehenden Verbindungen eines Knotens berücksichtigt werden. In diesem Fall wird von *in_degree_centrality* und *out_degree_centrality* gesprochen.³⁸ Aufgrund der schwierigen Auswertungsmöglichkeit von *in_degree_centrality* und *out_degree_centrality* in dieser Arbeit wird bei gerichteten Graphen das Zentralitätsmaß probeweise berechnet, aber nicht für den Vergleich herangezogen.

Globaler Cluster-Koeffizient

Betrachtet man ein Netzwerk nicht als Ganzes, sondern unterteilt dieses in verschiedene Bereiche, vorzugsweise eine kleine Gruppe verknüpfter Knoten, so stellt sich die Frage, ob bestimmte Bereiche stärker miteinander verknüpft sind als andere. Diese „Stärke“ der Verknüpfung lässt sich mit dem lokalen Cluster-Koeffizienten berechnen.

Der lokale Cluster-Koeffizient gibt dabei die Wahrscheinlich für einen Knoten an, dass die mit diesem Knoten verbundenen Knoten ebenfalls untereinander verknüpft sind. Ist zum Beispiel der Knoten A mit dem Knoten B und dem Knoten C verbunden, gibt der lokale Cluster-Koeffizient die Wahrscheinlichkeit für eine weitere Verbindung zwischen Knoten B und Knoten C wieder.³⁹ Einen guten Eindruck über den lokalen Cluster-Koeffizienten vermittelt die folgende Abbildung aus *Krischke und Röpckes Werk Graphen und Netzwerktheorie*.⁴⁰

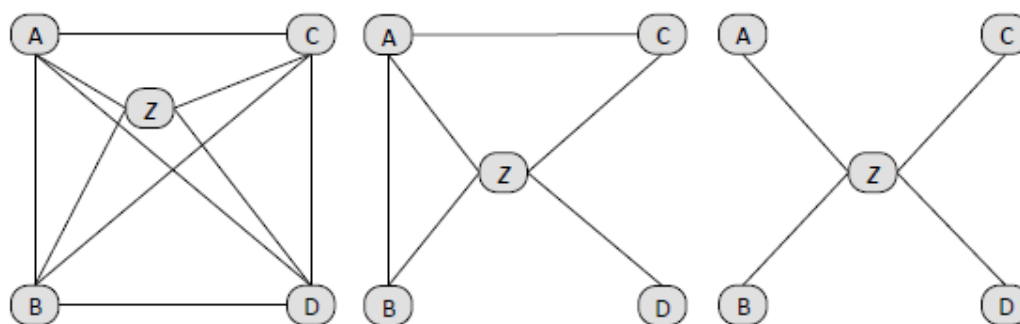


Bild 10.1 Graphen zur Veranschaulichung der Berechnung des lokalen Cluster-Koeffizienten C_i : Der zentrale Knoten Z besitzt den lokalen Cluster-Koeffizienten $C_i = 1$ (links), $C_i = 1/3$ (Mitte) bzw. $C_i = 0$ (rechts).

³⁸ Vgl. Newman, *Networks*, S. 135 f. 475 ff.

³⁹ Vgl. Krischke und Röpcke, *Graphen und Netzwerktheorie*, S. 147 f.

⁴⁰ Abbildung aus ebd., S. 148.

Will man für den ganzen Graphen die Dichte der Cluster berechnen, wird der globale Cluster-Koeffizient benötigt. Dieser wird aus dem Mittelwert der lokalen Cluster-Koeffizienten aller Knoten des Netzwerkes berechnet.⁴¹

$$C = \frac{1}{N} \sum_{1 \leq i \leq N} C_i$$

```
"""Globaler Cluster-Koeffizient"""
ck = nx.average_clustering(G1)
```

In NetworkX ist dies mit der Funktion *average_clustering* möglich. Wichtig hierbei ist zu erwähnen, dass es zwei Wege der Clusterberechnung mit unterschiedlichen Ergebnissen gibt. Die hier verwendete Formel wird verwendet, da sie durch den Computer leichter berechnet werden kann.⁴²

Der globale Cluster-Koeffizient wird in dieser Arbeit nur für ungerichtete Graphen berechnet.

Durchschnittliche Dichte

Die Dichte eines Graphen ist ein Wert, um die Vollständigkeit des Netzwerkes zu bestimmen. Eine Dichte von 1.0 bedeutet, dass alle möglichen Verbindungen zwischen den Knoten vorhanden sind, während eine Dichte von 0.0 ein Netzwerk ohne eine einzige Verbindung zwischen den Knoten wäre.⁴³

```
"""Dichte"""
den = nx.density(G1)
```

In NetworkX lässt sich die Dichte mit der Funktion *density* ermitteln.⁴⁴

4. Auswertung

4.1. Ergebnisse der Berechnungen

Die vom Skript erzeugten Merkmale werden in einer CSV-Datei abgespeichert und später zusammengeführt. Um dem Umfang des Haupttextes nicht über Maßen zu strapazieren, werden hier nur kurze Beispiele gezeigt, um den Aufbau der Ergebnisse zu zeigen.

⁴¹ Formel entnommen aus: Krishcke und Röpcke, *Graphen und Netzwerktheorie*, S. 161.

⁴² Vgl. Newman, M., *The Structure and Function of Complex Networks*, in: SIAM REVIEW, Vol. 45, No. 2, 2003, S. 167-256, hier S. 183 ff.

⁴³ Vgl. Krishcke und Röpcke, *Graphen und Netzwerktheorie*, S. 158 und Newman, *Networks*, S. 134 f.

⁴⁴ Vgl. NetworkX Function density, einsehbar unter (<https://networkx.github.io/documentation/networkx-1.9/reference/generated/networkx.classes.function.density.html>) (Stand Oktober 2017).

Lope de Rueda paso (ungewichtet) ohne 90er IDs								
Werk	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C_{Degree}	C	p(G)
Passo								
primero	3	3	2	2	2	1	1	1
Passo								
quarto	3	3	2	2	2	1	1	1
Passo								
quinto	3	3	2	2	2	1	1	1
Passo								
segundo	2	1	1	1	1	1	0	1
Passo								
septimo	4	6	3	3	3	1	1	1
Passo								
septimo	3	3	2	2	2	1	1	1
Passo								
tercero	4	4	3	1	2	1	0.58333333	0.66666667

Lope de Rueda comedia (ungewichtet) ohne 90er IDs								
Werk	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C_{Degree}	C	p(G)
Armeline	14	35	10	0	5	0.769	0.677	0.385
Los								
engañados	14	45	9	2	6.429	0.692	0.590	0.495
Eufemia	12	30	8	2	5	0.727	0.653	0.455
Medora	18	47	12	0	5.22222222	0.70588235	0.55121052	0.30718954

n=Anzahl der Knoten (Sprecher), m = Anzahl der Kanten, $\Delta(G)$ = Maximalgrad, $\delta(G)$ = Minimalgrad,

C_{Degree} = höchste Degree Centrality, C = globaler Cluster-Koeffizient, p(G) = durchschnittliche Dichte

Aus diesen Zahlen werden jetzt der Mittelwert und der Median des Autors nach Genre getrennt berechnet, was folgende Ergebnisse ergibt.

Zwischenergebnis Lope de Rueda, Mittelwerte und Median
(ungewichtet) ohne 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Paso	3.14	3.29	2.14	1.86	2	1	0.8	0.95
Median	Paso	3	3	2	2	2	1	1	1
Mittelwert	Comedia	14.5	39.25	9.75	1	5.41	0.72	0.62	0.41
Median	Comedia	14	40	9.5	1	5.11	0.72	0.62	0.42

n=Anzahl der Knoten (Sprecher), m = Anzahl der Kanten, $\Delta(G)$ = Maximalgrad, $\delta(G)$ = Minimalgrad,

C_{Degree}= höchste Degree Centrality, C = globaler Cluster-Koeffizient, p(G) = durchschnittliche Dichte

Nun erfolgt die Berechnung alle Zahlen aller Fälle nach obigem Muster, um die Autoren miteinander vergleichen zu können. Da für den Vergleich der Autoren die durchschnittlichen Werte aller Werke benötigt werden, wird nun der Mittelwert nach Gattung getrennt ermittelt. Bei einem Test, bei dem zusätzlich der Median ermittelt wurde, ergab sich, dass der Median des Minimalgrads für die Werke des Autoren Juan Ruiz stets 0 ist, also über die Hälfte der Werke isolierte Sprecher aufweist. Daher wird der Median als Vergleichswert weggelassen und nur der Mittelwert angegeben.

Vergleich der Mittelwerte der *paso* und *entremés* zwischen **Lope de Rueda, Miguel de Cervantes**
Saavedra und Tirso de Molina.

Graph ohne 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Rueda	3.1429	3.2857	2.1429	1.8571	2	1	0.7976	0.9523
Cervantes	11	26	9	1	5	1	1	1
Tirso	9.1111	13.4444	5.6667	1	3.0151	0.8935	0.5647	0.5108

Graph mit 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Rueda	3.1429	3.2857	2.1429	1.8571	2	1	0.7976	0.9523
Cervantes	12.1426	28.1429	9.1429	1	4.5003	0.8460	0.6458	0.4193
Tirso	10.4444	19.5556	6.6667	0.7778	3.5106	0.8292	0.5805	0.4521

Digraph ohne 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Rueda	3.1429	6.4286	83.4286	50.7143	64.3810	-	-	0.9286
Cervantes	10.8571	41.5714	81	2	28.7139	-	-	0.4066
Tirso	9.6667	30.7778	102	3.2222	39.5715	-	-	0.4864

Digraph mit 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	3.1429	6.4286	83.4286	50.7143	64.3810	-	-	0.9286
Cervantes	12.1428	44	81.2857	1.5714	25.7264	-	-	0.3385
Tirso	10.4444	33.2222	102.7778	1.4444	34.8583	-	-	0.3898

n=Anzahl der Knoten (Sprecher), m = Anzahl der Kanten, $\Delta(G)$ = Maximalgrad, $\delta(G)$ = Minimalgrad,

C_{degree}= höchste Degree Centrality, C = globaler Cluster-Koeffizient, p(G) = durchschnittliche Dichte

Vergleich der Mittelwerte der *comedia* zwischen **Lope de Rueda**, **Miguel de Cervantes Saavedra**,
Tirso de Molina und **Juan Ruiz de Alarcón y Mendoza**.

Graph ohne 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	14.5	39.25	9.75	1	5.4127	0.7237	0.6176	0.4102
Cervantes	26.8889	78.5556	15.6667	0.2222	5.7941	0.6336	0.5387	0.2417
Tirso	23.4828	56.3448	13.4826	0.3276	5.1113	0.6702	0.5669	0.2960
Ruiz	15.1	47.1	11.55	0.85	6.2903	0.8379	0.7194	0.4796

Graph mit 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	14.75	39.5	9.75	0.75	5.3550	0.7085	0.6028	0.3959
Cervantes	28.3333	81.7778	16.2222	0.1111	5.7269	0.6165	0.5239	0.2249
Tirso	25.0517	60.7931	14.1897	0.3103	5.1129	0.6535	0.5502	0.2754
Ruiz	16.25	49.65	12.25	0.55	6.1627	0.8152	0.6937	0.4345

Digraph ohne 90er IDs

	N	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	14.5	68.75	191.75	13.5	69.7698	-	-	0.3586
Cervantes	26.8889	130.1111	212.3333	0.2222	45.6914	-	-	0.1989
Tirso	23.4828	95.2759	310.5690	0.8966	69.0384	-	-	0.2585
Ruiz	15.1	81.35	292.8	4.1	92.1981	-	-	0.4194

Digraph mit 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	14.75	69.25	191.75	8.5	68.1929	-	-	0.3463
Cervantes	28.3333	134.3333	212.5556	0.1111	43.6245	-	-	0.1834
Tirso	25.0517	101.1897	311.4310	0.8966	65.4433	-	-	0.2379
Ruiz	16.25	84.75	293.6	1.1	86.4569	-	-	0.3761

4.2. Vergleich und Prüfung der Thesen

Vergleicht man nun die einzelnen Werte, um die aufgestellten Thesen zu prüfen, ergibt sich folgendes Bild.

Die erste These nimmt an, dass die Werte der Werke Cervantes ähnlicher zu den Werten der Werke Ruedas seien als zu den Werken Tirsos und Ruiz. Um dies zu überprüfen werden die Differenzen der Werte Cervantes zu den Werten der anderen Autoren berechnet und jeweils die beiden Werte mit der geringsten Differenz farblich hervorgehoben.

Vergleich der Mittelwerte der *paso* und *entremés* zwischen **Lope de Rueda, Miguel de Cervantes Saavedra** und **Tirso de Molina**.

Graph ohne 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	3.1429	3.2857	2.1429	1.8571	2	1	0.7976	0.9523
Cervantes	11	26	9	1	5	1	1	1
Tirso	9.1111	13.4444	5.6667	1	3.0151	0.8935	0.5647	0.5108

Graph mit 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	3.1429	3.2857	2.1429	1.8571	2	1	0.7976	0.9523
Cervantes	12.1426	28.1429	9.1429	1	4.5003	0.8460	0.6458	0.4193
Tirso	10.4444	19.5556	6.6667	0.7778	3.5106	0.8292	0.5805	0.4521

Digraph ohne 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	3.1429	6.4286	83.4286	50.7143	64.3810	-	-	0.9286
Cervantes	10.8571	41.5714	81	2	28.7139	-	-	0.4066
Tirso	9.6667	30.7778	102	3.2222	39.5715	-	-	0.4864

Digraph mit 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	3.1429	6.4286	83.4286	50.7143	64.3810	-	-	0.9286
Cervantes	12.1428	44	81.2857	1.5714	25.7264	-	-	0.3385
Tirso	10.4444	33.2222	102.7778	1.4444	34.8583	-	-	0.3898

Vergleich der Mittelwerte der *comedia* zwischen **Lope de Rueda, Miguel de Cervantes Saavedra, Tirso de Molina und Juan Ruiz de Alarcón y Mendoza.**

Graph ohne 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	14.5	39.25	9.75	1	5.4127	0.7237	0.6176	0.4102
Cervantes	26.8889	78.5556	15.6667	0.2222	5.7941	0.6336	0.5387	0.2417
Tirso	23.4828	56.3448	13.4826	0.3276	5.1113	0.6702	0.5669	0.2960
Ruiz	15.1	47.1	11.55	0.85	6.2903	0.8379	0.7194	0.4796

Graph mit 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	14.75	39.5	9.75	0.75	5.3550	0.7085	0.6028	0.3959
Cervantes	28.3333	81.7778	16.2222	0.1111	5.7269	0.6165	0.5239	0.2249
Tirso	25.0517	60.7931	14.1897	0.3103	5.1129	0.6535	0.5502	0.2754
Ruiz	16.25	49.65	12.25	0.55	6.1627	0.8152	0.6937	0.4345

Digraph ohne 90er IDs

	N	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	14.5	68.75	191.75	13.5	69.7698	-	-	0.3586
Cervantes	26.8889	130.1111	212.3333	0.2222	45.6914	-	-	0.1989
Tirso	23.4828	95.2759	310.5690	0.8966	69.0384	-	-	0.2585
Ruiz	15.1	81.35	292.8	4.1	92.1981	-	-	0.4194

Digraph mit 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	14.75	69.25	191.75	8.5	68.1929	-	-	0.3463
Cervantes	28.3333	134.3333	212.5556	0.1111	43.6245	-	-	0.1834
Tirso	25.0517	101.1897	311.4310	0.8966	65.4433	-	-	0.2379
Ruiz	16.25	84.75	293.6	1.1	86.4569	-	-	0.3761

Im Falle, dass die 90er IDs ignoriert werden und die Graphen ohne Richtung und Gewichtung miteinander verglichen werden, so sind die *paso* von Rueda und die *entremés* von Cervantes im Hinblick auf die durchschnittliche Dichte, den globalen Cluster-Koeffizienten und das Zentralitätsmaß zueinander ähnlicher als die *entremés* des Cervantes zu den *entremés* des Tirso. Allerdings ähneln sich Cervantes und Tirso stärker als Cervantes und Rueda, wenn es um die absoluten Zahlen von Knoten, Kanten und den Maximal- und Minimalgrad geht.

Betrachtet man nun für den Graphen auch die 90er IDs, so sind Cervantes und Tirso in jeder Hinsicht zueinander ähnlicher, als Cervantes zu Rueda. Dies wird seinen Grund besonders darin haben, da in

den überarbeiteten Werken von Rueda kaum 90er IDs vorhanden sind, während diese bei Cervantes und Tirso im Durchschnitt sehr häufig vorkommen. Für ein Fazit werden daher nur die Ergebnisse ohne 90er IDs verwendet. In diesem Fall scheint, zumindest was die Dichte des Werkes angeht, eine Ähnlichkeit zwischen Rueda und Cervantes zu herrschen. Diese ist aber im Fall eines gerichteten Graphen mit Gewichten nicht mehr gegeben.

Vergleicht man noch die Zahlen der Digraphen miteinander, so ist, unabhängig von der Berücksichtigung der 90er IDs, in fast allen Fällen eine größere Ähnlichkeit zwischen den *entremés* des Cervantes und des Tirso als zwischen den *entremés* des Cervantes und der *paso* von Rueda festzustellen.

Auch im Falle der *comedia* ist der Unterschied zwischen Cervantes und Rueda in den meisten Fällen größer als der Unterschied von Cervantes zu einem der anderen Autoren.

Damit kann gesagt werden, dass die erste Unterthese nicht bestätigt werden kann, beziehungsweise, dass die Ähnlichkeit zwischen Cervantes und Rueda nach den Merkmalen der Figurennetzwerke ihrer Werke nicht größer ist, als die Ähnlichkeit zu den Merkmalen der Netzwerke der anderen Autoren.

Die zweite These nimmt an, dass die Ähnlichkeit zwischen den *comedias* von Ruiz und den *comedias* von Tirso größer sein sollte, als die Ähnlichkeit zu den *comedias* der anderen beiden Autoren. Hierfür wird nun die kleinste Wertedifferenz des Autors Ruiz zu den anderen Autoren berechnet und hervorgehoben.

Vergleich der Mittelwerte der *comedia* zwischen **Lope de Rueda, Miguel de Cervantes Saavedra, Tirso de Molina und Juan Ruiz de Alarcón y Mendoza**

Graph ohne 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	14.5	39.25	9.75	1	5.4127	0.7237	0.6176	0.4102
Cervantes	26.8889	78.5556	15.6667	0.2222	5.7941	0.6336	0.5387	0.2417
Tirso	23.4828	56.3448	13.4826	0.3276	5.1113	0.6702	0.5669	0.2960
Ruiz	15.1	47.1	11.55	0.85	6.2903	0.8379	0.7194	0.4796

Graph mit 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	d(G)	CDegree	C	p(G)
Rueda	14.75	39.5	9.75	0.75	5.3550	0.7085	0.6028	0.3959
Cervantes	28.3333	81.7778	16.2222	0.1111	5.7269	0.6165	0.5239	0.2249
Tirso	25.0517	60.7931	14.1897	0.3103	5.1129	0.6535	0.5502	0.2754
Ruiz	16.25	49.65	12.25	0.55	6.1627	0.8152	0.6937	0.4345

Digraph ohne 90er IDs

	N	m	$\Delta(G)$	$\delta(G)$	$d(G)$	CDegree	C	$p(G)$
Rueda	14.5	68.75	191.75	13.5	69.7698	-	-	0.3586
Cervantes	26.8889	130.1111	212.3333	0.2222	45.6914	-	-	0.1989
Tirso	23.4828	95.2759	310.5690	0.8966	69.0384	-	-	0.2585
Ruiz	15.1	81.35	292.8	4.1	92.1981	-	-	0.4194

Digraph mit 90er IDs

	n	m	$\Delta(G)$	$\delta(G)$	$d(G)$	CDegree	C	$p(G)$
Rueda	14.75	69.25	191.75	8.5	68.1929	-	-	0.3463
Cervantes	28.3333	134.3333	212.5556	0.1111	43.6245	-	-	0.1834
Tirso	25.0517	101.1897	311.4310	0.8966	65.4433	-	-	0.2379
Ruiz	16.25	84.75	293.6	1.1	86.4569	-	-	0.3761

Auch hier trifft die zu überprüfende These nicht zu. Es ist keine größere Ähnlichkeit zwischen den Werken von Ruiz und Tirso festzustellen. Tatsächlich scheint Ruiz trotz der zeitlichen Differenz scheinbar die meisten Ähnlichkeiten zu den *comedias* Ruedas zu haben. Leider kann der Vergleich zwischen Rueda und Tirso nicht vertieft werden, da im Fall Ruiz keine Werke der Gattung *entremés* vorliegen, welche mit den *paseo* Ruedas verglichen werden könnten.

4.3. Fazit

Die Auswertung der Merkmale ermöglichen einen neuen Blick auf die vier Autoren und ihre Werke. Allgemein kann zuerst gesagt werden, dass die gewählte Methode des Vergleichs der Netzwerkmerkmale nicht zum gewünschten Erfolg geführt hat. Nun ist zu ermitteln, welche Ursachen dem zugrunde liegen könnten.

Zum einen ist es natürlich möglich, dass die Annahme zur Autorenähnlichkeit und die am Anfang getroffene Einteilung in zwei Gruppen falsch war und Cervantes, trotz der in der Literatur beschriebenen Nähe zu Rueda, tatsächlich eher einen ähnlichen Stil wie Ruiz und Tirso pflegte. Ebenso kann es möglich sein, dass Ruiz aus dem Grunde, weil er mit seinen Werken beim Publikum eine Veränderung bewirken wollte, in seiner Art sich wissentlich oder unwissentlich dem Vorgänger Rueda angepasst hatte. Allerdings sind die errechneten Werte nicht differenzierbar genug, um in dieser Sache eine klare Aussage zu treffen.

Eine andere Möglichkeit das Ergebnis zu erklären, könnte sein, dass die Netzwerkanalyse für diese Art der Überprüfung kein geeignetes Werkzeug ist. Literarische Texte sind künstlerische Schöpfungen, folgen jedoch häufig auch durch die Gesellschaft vorgegebenen Normen. Obgleich der Schreibstil der vier Autoren sicherlich unterschiedlich war, stützt sich diese Arbeit hauptsächlich auf die mathematischen Werte Sprecheranzahl und die Folge der Figuren aufeinander. Besonders der erste

Wert kann durch herrschende Traditionen, wie die Einbehaltung einer festen Maximalanzahl an Figuren für ein Theaterstück, zu statisch für eine differenzierte Unterscheidung sein.

Ein weiterer Faktor könnte sein, dass die Autoren ihre Art zu schreiben, über die Zeit anpassten und die ermittelten Werte so keinen momentanen „Fingerabdruck“ eines Autors liefern, sondern eher dessen künstlerische Entwicklung wiedergeben.

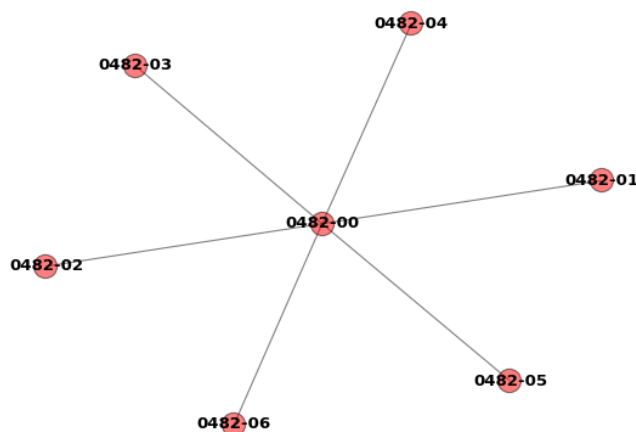
Ein weiterer Faktor könnte die wankende Qualität der bearbeiteten Texte liefern. Während, wie bereits angesprochen, die 90er IDs beim Autor Rueda nur in einem Werk vorkommen, sind die 90er IDs bei den Werken der anderen Autoren sehr häufig vorhanden. Auch IDs, die nicht eindeutig zugeordnet werden konnten, könnte abhängig vom Autor eine Rolle spielen. Zusammen mit weiteren möglichen Fehlerquellen in der Bearbeitung, könnte daher einfach die Textqualität für diese Art des Vergleichs nicht ausreichen.

Jedoch lässt sich auch festhalten, dass sich durch die Netzwerkanalyse eine Vielzahl neuer Erkenntnisse gewinnen lässt, besonders wenn man die Einzelwerke betrachtet. Lope de Rueda lässt sowohl in seinen *paso* als auch in seinen *comedia* nicht viele Figuren auftreten. So sind in seinen *paso* nur 2 bis 4 Personen vorhanden, aber alle miteinander ins Gespräch treten, was sich durch die hohen Werte der Dichte und des Zentralitätsmaßes zeigt. Die Anzahl der auftretenden Sprecher ist allerdings kein Hinweis auf die zu erwartende Zahl an Dialogen. Besonders in seinem Werk *Passo segvndo* mit nur zwei Personen wird 118 Mal das Wort gewechselt, während im Werk *Passo septimo* mit vier Personen durchschnittlich weniger Dialoganteil pro Person zu erkennen ist.

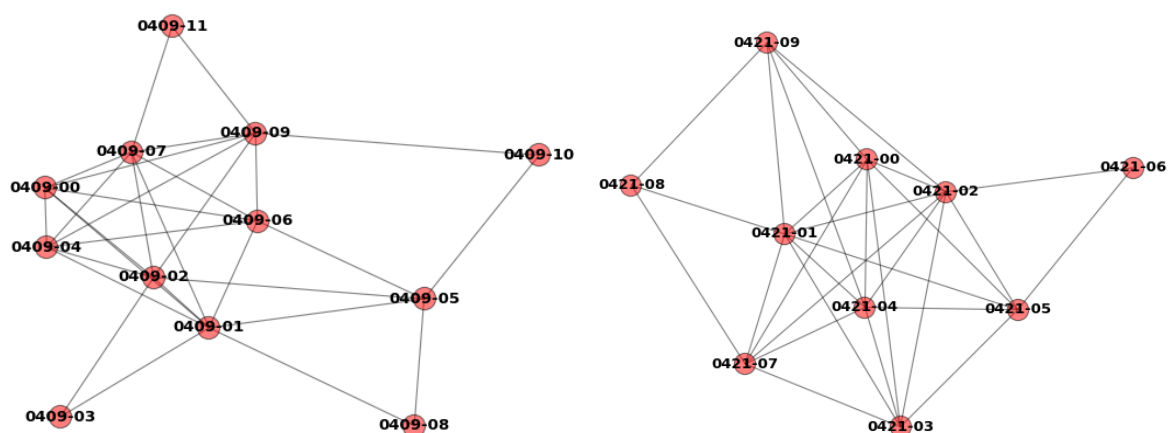
Im Hinblick auf seine *comedia* gibt sich Rueda mit 14 bis 18 Personen relativ bescheiden, während Cervantes und Tirso oft über 20 Personen in einem Stück zu Wort kommen lassen. Ähnlicher zu Rueda sind dagegen die Ergebnisse bei Ruiz mit durchschnittlich 10 bis 20 Personen pro Stück.

Bei den *comedia* von Tirso ist dagegen eine hohe Variation der Sprecheranzahl in den Werken zu beobachten. Während die *comedia Amar por razon de estado* mit sieben Figuren auskommt, sind es in der *comedia Antona García* 41 Sprecher mit der originalen Castlist und 54 Sprecher, wenn die Castlist um die Personen erweitert wird, welche im Text vorkommen, aber keinen Eintrag in die Castlist haben. In Tirsos Werken findet sich zudem ein von der Struktur her unter den Autorenwerken einzigartiges Werk. Im *entemés La malcontenta* verbindet eine Figur alleine alle anderen Figuren im Werk, ohne dass die anderen Figuren untereinander in Kontakt treten. In der Netzwerkanalyse ist dies durch den globalen Cluster-Koeffizienten mit Wert 0 sichtbar geworden (es bildet sich kein einziges geschlossenes Dreieck aus drei Knoten).⁴⁵

⁴⁵ Vgl. Krischke und Röpcke, *Graphen und Netzwerktheorie*, S. 147 f.

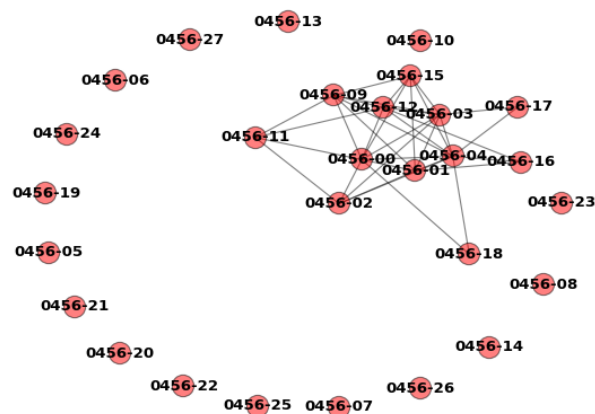


Untersucht man dagegen die unterschiedlichen Zentralitätsmaße, lässt sich erkennen, dass diese Werte bei den meisten Werken sehr hoch sind. Es sind also in den Werken häufig zentrale Sprecher vorhanden, welche im Verhältnis zur Knotenanzahl des Netzwerkes, mit vielen anderen Sprechern verbunden sind. Das Figurennetzwerk fokussiert sich damit häufig mehrere Personen im Mittelpunkt. Die folgenden zwei Abbildungen zeigen den Graphen des Werkes *Eufemia (comedia)* (links) von Rueda und des Werkes *La amistad castigada (comedia)* (rechts) von Ruiz.



Hier lässt sich anhand der Kantenstruktur schnell erkennen, dass in beiden Werken mehrere Figuren innerhalb des Netzes mit hohen Graden positioniert sind, während Nebenfiguren am Rande nur mit einem Grad von 2 oder 3 erscheinen. Nun könnte man die Untersuchung der Texte ausweiten und zum Beispiel erläutern, welche Personen nur mit wenigen Personen verknüpft sind und welche in den Zentren der Netze anzutreffen sind. Ebenso könnte man die Methode umdrehen und versuchen die Netzwerke zu lokalisieren, in welchen zwei oder drei Unternetzwerke ohne zentralen Verbindungsknoten existieren. Dies ist jedoch mit dem verwendeten Korpus schwierig, da wie bereits angesprochen in vielen Werken Sprecher ohne eindeutige Zuweisung auftauchen, wodurch ein geringes maximales Zentralitätsmaß auch eine bedeuten kann, dass viele Knoten isoliert sind.

Als unglückliches Beispiel sei hier auf dem Graphen von Tirso *comedia El amor médico* zu verweisen.



Untersucht man aber in diesem Fall wieder den Originaltext, so stellt sich heraus, dass in der originalen Castlist viele Sprecher doppelt vorhanden sind.⁴⁶ Dies könnte an einem Fehler im Ursprungstext oder einem Fehler bei der spanischen Übertragung in XML liegen. Insofern sind ungewöhnliche Netzwerkmerkmale auch brauchbar, um fehlerhafte Transkriptionen in einer großen Textmenge einfacher zu identifizieren.

C) Schluss

Allgemein scheinen Netzwerke und Netzwerkmerkmale ein gutes Mittel zu sein, um die Beschaffenheit von Theaterstücken zu beobachten und neue Thesen zu prüfen. Um diese Möglichkeiten aufzuzeigen, werden einige Beispiele aufgeführt, wie man die Netzwerkanalyse für weitere Fragen nutzen könnte. Netzwerke könnten sich dazu eignen, die Entwicklung eines Autors selbst abzubilden. Bleibt der Autor einem festen Muster treu und ergeben sich dadurch ähnliche Merkmale oder versucht sich der Autor innerhalb seiner Karriere in verschiedenen Stillrichtungen, was sich in unterschiedlichen Werten für die verschiedenen Merkmale widerspiegeln würde? So scheint Rueda seine Figuren oft eng zusammenzuhalten und im gleichen Verhältnis sprechen zu lassen, während die anderen drei Autoren eine größere Differenz in den Merkmalen Zentralitätsmaß, Cluster-Koeffizienten und Dichte aufweisen und damit wohl häufiger das Verhältnis, wie der Sprechereinsatz auf die einzelnen Figuren aufgeteilt wird, zu wechseln scheinen. Auch können die Merkmale dabei helfen zu zeigen, ob ein Werk auf eine bestimmte Weise heraussticht, wie das oben aufgeführte Beispiel von Tirso's *entemés La malcontenta*. Ein möglicher anderer Ansatz wäre der Versuch Spuren der Geschichte des Theaters selbst innerhalb der Werke zu suchen. Rueda schrieb hauptsächlich für Wanderbühnen mit einer kleinen Schaustellertruppe, während sich in der Zeit der anderen Autoren bereits feste Theater mit einer größeren Menge verfügbarer Künstler etabliert hatten. Vielleicht wäre es daher möglich anhand der Netzwerke zu prüfen, ob die Anzahl der aufgestellten Sprecher im Laufe der Jahre stieg oder sich anderen Faktoren anpasste? Auch die Präsenz eines Chors oder einer Menschenmenge für eine Szene

⁴⁶ 3mal Doña Geronima, 2mal Don Gonçalo, 3mal Don Gaspar, 2mal Don Iñigo, 2mal Doña Estefania ...

könnten hier wichtige Hinweise liefern. Hierfür wären allerdings mehrere Autoren unterschiedlicher Schaffenszeit nötig, um einen größeren Zeitraum abzudecken, und dazu belegbare Informationen, in welchen Jahr welches Stück entstanden ist.

Aber auch die Frage, ob sich unterschiedliche getrennte Gesellschaftsgruppen in einem Werk wiederfinden, könnte geprüft werden. Besonders die von Lope de Vega „erschaffene“ *comedia nueva* soll ihren Reiz unter anderem im Zusammenspiel aus ritterlichen Herrn und einfachen Diener haben. Geht man nun von einer historischen Gesellschaft mit verschiedenen Schichten aus, so wäre es interessant zu überprüfen, ob der ritterliche Herr hauptsächlich im Dialog mit Adeligen und Klerus tritt und der Diener häufiger das Wort mit Bauern und Handwerkern wechselt oder ob eine gesellschaftliche Trennung im Theater gar nicht existiert und sich daher nicht mehrere, nur lose über die Hauptfiguren verknüpfte Teilnetzwerke, sondern ein großes umfangreiches Netzwerk bildet.⁴⁷

Geht man diesen Weg weiter, könnte man die Gesellschaftsfrage auch weiter spezialisieren und als Beispiel die Rolle des Königs oder der Kirche in den Stücken herausarbeiten. Tritt etwa ein König im Werk auf und wenn ja, mit wem, wann und wie? Ist die Figur des Königs eine wiederkehrende Person, welche häufig das Wort ergreift und das ganze Werke über präsent ist oder fungiert der König eher als eine Art „*Deus ex machina*“, eine Richterfigur, welche am Ende eines Werkes auftritt und die Gerechtigkeit wiederherstellt?⁴⁸ Dies könnte man damit überprüfen, indem in Werke die Figur des Königs identifiziert und dessen Position und Verbindung zu anderen Figuren mit verschiedenen Merkmalen herausgearbeitet wird.

Dies sind einige Ideen und zeigen die verschiedensten Möglichkeiten, welche die Netzwerkanalyse bietet und welche neuen Fragen sich aus einer nicht bestätigten These ableiten lassen. Für das Finden von Strukturen in geisteswissenschaftlichen Aufgaben scheint die Netzwerkanalyse damit ein sehr Beispiel dafür zu sein,⁴⁹ wie mit Hilfe moderne Informationstechnologien neue Fragen in den Wissenschaften beantwortet werden können.

⁴⁷ Vgl. als Beispiel für Teilnetzwerke, Krishke und Röpcke, *Graphen und Netzwerktheorie*, S. 150 ff.

⁴⁸ Vgl. Strosetzki, *Geschichte der spanischen Literatur*, S. 182-185.

⁴⁹ Vgl. als weiterführenden Gedanken Jannidis, Kohle und Rehbein, *Digital Humanitis*, S. 13-18 und 147.

D) Literatur:

Aufsätze

Burrows, John. 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship, In: Literary and Linguistic Computing, Volume 17 (2002), Issue 3, S. 267 –287.

Charney, Maurice, *Shakespeare- and the Others*, in: Shakespeare Quarterly, Vol. 30, No. 3, 1979, S. 325-342.

Ermatinger, Emil, *Gottfried Keller und Goethe*, in: PMLA, Vol. 64, No. 1, 1949, S. 79-97.

Hermand, Jost, Heine und Brecht: *Über die Vergleichbarkeit des Unvergleichlichen*, in: Monatshefte, Vol. 73, No 4, 1981, S 429-441.

König, Bernhard, *Miguel de Cervantes Saavedra – Entremes*, in: Das spanische Theater, Vom Mittelalter bis zur Gegenwart, Roloff, Volker und Wentzlaff-Eggebert, Harald (Hg.), S. 53-69.

Mosteller, F. and D. L. Wallace. *Inference and disputed authorship: The federalist*. Reading, 1964.

Newman, M., *The Structure and Function of Complex Networks*, in: SIAM REVIEW, Vol. 45, No. 2, 2003, S. 167-256.

Trilcke, Peer. *Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft*, in: Ajouri, Philip, Mellmann, Katja und Rauen, Christoph (Hg.): *Empirie in der Literaturwissenschaft*, Münster 2013, S. 201-247.

Monographien

Delgado, Mariano, *Das Spanische Jahrhundert (1492-1659)*, Darmstadt, 2016.

Diestel, Reinhard, *Graphentheorie*, Berlin, ³2006.

Fitzgerald, Michael, *Introducing Regular Expressions*, Sebastopol, 2012.

Jannidis, Fortis, Kohle, Hubertus und Rehbein, Malte, *Digital Humanitis, Eine Einführung*, Stuttgart, 2017.

Krischke, André und Röpcke, Helge, *Graphen und Netzwerktheorie: Grundlagen – Methoden – Anwendungen*, München, 2015.

Newman, M., *Networks An Introduction*, Oxford, 2010.

Neuschäfer, Hans-Jörg: *Spanische Literaturgeschichte*, Stuttgart, 2011.

Womack, Peter: *Englisch Renaissance drama*, Malden, 2008.

Simson, Ingrid: *Das Siglo de Oro*, Stuttgart, 2001.

Stegbauer, Christian (Hrsg.), *Netzwerkanalyse und Netzwerktheorie, ein neues Paradigma in den Sozialwissenschaften*, Wiesbaden, ²2010.

Stenzel, Hartmut, *Einführung in die spanische Literaturwissenschaft*, Stuttgart, 2001.

Strosetzki, Christoph, *Geschichte der spanischen Literatur*, Tübingen, 1996.

Internetquellen:

Projekt: *Network Analysis of dramatic texts* von Fischer, Frank, Göbel, Mathias, Kampkaspar, Dario und Trilcke, Peer, einsehbar unter:

<https://dlina.github.io/> (Stand September 2017).

Vortrag: *Digitale Netzwerkanalyse dramatischer Texte* von Fischer, Frank, Kampkaspar, Dario und Trilcke, Peer, Graz, 2015, einsehbar unter:

<http://gams.uni-graz.at/o:dhd2015.v.040> (Stand September 2017).

Bilder der spanischen Autoren:

<https://www.biografiasyvidas.com/> (Stand Juli 2017).

TEI Schiller Kabale und Liebe:

http://www.deutschestextarchiv.de/book/view/schiller_kabale_1784/ (Stand Juli 2017)

TEI Schiller Dom Karlos:

http://www.deutschestextarchiv.de/book/show/schiller_domkarlos_1787/ (Stand Juli.2017).

TEI Kleist Der Zerbrochene Krug:

http://www.deutschestextarchiv.de/book/view/kleist_krug_1811/ (Stand Juli.2017).

TEI Schnitzler Komtesse Mizzi:

https://textgridrep.org/browse/-/browse/v3sw_0 (Stand Juli.2017).

Teatro Español del Siglo de Oro:

<http://teso.chadwyck.com/>

Pythondokumentationen:

glob (<https://docs.python.org/3.4/library/glob.html>) (Stand Oktober 2017)

os (<https://docs.python.org/3.4/library/os.html>) (Stand Oktober 2017)

re (<https://docs.python.org/3.4/library/re.html>) (Stand Oktober 2017)

pandas (<https://pandas.pydata.org/pandas-docs/stable/index.html>) (Stand Oktober 2017)

collcetions (<https://docs.python.org/3.4/library/collections.html>) (Stand Oktober 2017)

networkx (<https://networkx.github.io/>) (Stand Oktober 2017)

NetworkX-Homepage (<https://networkx.github.io/>)(Stand Oktober 2017)

E) Anhang

Spanische Werke

Werke von Lope de Rueda (1505?-1565)

Interne ID	Titel	Genre
412	Passo primero	Paso
413	Passo quarto	Paso
414	Passo quinto	Paso
415	Passo segvndo	Paso
416	Passo septimo	Paso
417	Passo septino	Paso
418	Passo tercero	Paso
406	Armelina	Comedia
408	Los engañados	Comedia
409	Eufemia	Comedia
411	Medora	Comedia

Werke von Miguel de Cervantes Saavedra (1547-1616)

Interne ID	Titel	Genre
258	La cueua de Salamanca	Entremés
259	La elección de los alcaldes de Daganzo	Entremés
263	La guarda cuydadosa	Entremés
267	El retablo de las maravillas	Entremés
268	El rufián viudo, llamado Trampagos	Entremés
270	El viejo zeloso	Entremés
271	El vizcayno fingido	Entremés
256	Los baños de Argel	Comedia
257	La casa de los zelos y seluas de Ardenia	Comedia
260	La entretenida	Comedia
261	El gallardo español	Comedia
262	La gran Svltana	Comedia
264	El juez de los diuorcios	Comedia
265	El laberinto de amor	Comedia
266	Pedro de Urdemalas	Comedia
269	El rvfian dichoso	Comedia

Werke von Tirso de Molina (1579?-1648)

Interne ID	Titel	Genre
451	Los alcaldes	Entremés
465	Los coches de Benavente	Entremés
471	El dvende	Entremés
475	El estvdiente	Entremés
478	El gavacho	Entremés
482	La malcontenta	Entremés
489	El negro	Entremés
510	La venta	Entremés
515	Las vívdas	Entremés
450	Adversa fortvna de don Alvaro de Lvna	Comedia

452	Los amantes de Tervel	Comedia
453	Amar por arte mayor	Comedia
454	Amar por razon de estado	Comedia
455	Amazonas en las Indias	Comedia
456	El amor médico	Comedia
457	El amor, y el amistad	Comedia
458	Amor y zelos hazen discretos	Comedia
459	Antona García	Comedia
460	El Aquiles	Comedia
461	El árbol del mejor fruto	Comedia
462	Averigvelo Vargas	Comedia
463	El castigo del pensequé	Comedia
464	Cavtela contra cavtela	Comedia
466	El condenado por desconfiado	Comedia
467	La Dama del Oliuar	Comedia
468	Del enemigo el primer consejo	Comedia
469	Don Gil de las calzas verdes	Comedia
470	Doña Beatriz de Silva	Comedia
472	La elección por la virtud	Comedia
473	Escarmientos para el cuerdo	Comedia
474	Esto sí que es negociar	Comedia
477	La gallega Mari Hernández	Comedia
479	La huerta de Ivan Fernández	Comedia
480	Los lagos de san Vicente	Comedia
481	La lealtad contra la embidia	Comedia
483	Marta la piadosa	Comedia
484	El mayor desengaño	Comedia
485	La mejor espigadera	Comedia
486	El melancólico	Comedia
487	La mujer que manda en casa	Comedia
488	La mujer por fuerza	Comedia
490	No hay peor sordo	Comedia
491	Palabras y plumas	Comedia
492	La Peña de Francia	Comedia
493	Por el sótano y el torno	Comedia
494	El pretendiente al revés	Comedia
495	Privar contra su gusto	Comedia
496	Próspera fortuna de don Alvaro de Luna	Comedia
497	La providencia en la mujer	Comedia
498	Quien calla otorga	Comedia
499	Quien habló pagó	Comedia
500	Quien no cae no se levanta	Comedia
501	La república al reves	Comedia
502	La reina de los reyes	Comedia
503	La santa Juana	Comedia
504	Santo, y sastre	Comedia
505	Segunda parte de Santa Juana	Comedia
506	Siempre ayuda la verdad	Comedia
507	Tanto es lo demás como lo de menos	Comedia
508	Todo es dar en una cosa	Comedia
509	La venganza de Tamar	Comedia

511	Ventura te dé Dios hijo	Comedia
512	La vida de Herodes	Comedia
513	La villana de la Sagra	Comedia
514	La villana de Vallecas	Comedia
516	Zelos con zelos se curan	Comedia
517	La zelosa de sí misma	Comedia

Werke von Juan Ruiz de Alarcón y Mendoza (1581?-1639)

Interne ID	Titel	Genre
421	La amistad castigada	Comedia
422	El Antichristo	Comedia
423	La crueldad por el honor	Comedia
424	La cueva de Salamanca	Comedia
425	El desdichado en fingir	Comedia
426	El dueño de las estrellas	Comedia
427	Los empeños de un engaño	Comedia
428	Examen de maridos	Comedia
429	Los favores del mundo	Comedia
430	Ganar amigos	Comedia
431	La industria, y la suerte	Comedia
432	La manganilla de Melilla	Comedia
433	Mudarse por mejorarse	Comedia
434	Las paredes oyen	Comedia
435	Los pechos privilegiados	Comedia
436	La prueva de las promesas	Comedia
437	El semejante a sí mismo	Comedia
438	El texedor de Segovia	Comedia
439	Todo es ventura	Comedia
440	La verdad sospechosa	Comedia

Zwischenergebnisse⁵⁰

Lope de Rueda

Zwischenergebnis Lope de Rueda, Mittelwerte und Median

(ungewichtet) ohne 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Paso	3.14	3.29	2.14	1.86	2	1	0.8	0.95
Median	Paso	3	3	2	2	2	1	1	1
Mittelwert	Comedia	14.5	39.25	9.75	1	5.41	0.72	0.62	0.41
Median	Comedia	14	40	9.5	1	5.11	0.72	0.62	0.42

⁵⁰ Für weitere Zahlen siehe die der Arbeit beiliegenden Excel-Tabellen.

Zwischenergebnis Lope de Rueda, Mittelwerte und Median

(ungewichtet) mit 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Paso	3.14	3.29	2.14	1.86	2	1	0.8	0.95
Median	Paso	3	3	2	2	2	1	1	1
Mittelwert	Comedia	14.75	39.5	9.75	0.75	5.36	0.71	0.60	0.40
Median	Comedia	14	40	9.5	0.5	5.11	0.70	0.59	0.40

Zwischenergebnis Lope de Rueda, Mittelwerte und Median

(gewichtet) ohne 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Paso	3.14	6.43	83.43	50.71	64.38	-	-	0.93
Median	Paso	3	6	69	46	58			1
Mittelwert	Comedia	14.5	68.75	191.75	13.5	69.77	-	-	0.36
Median	Comedia	14	69	186	11	68.11			0.36

Zwischenergebnis Lope de Rueda, Mittelwerte und Median

(gewichtet) mit 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Paso	3.14	6.43	83.43	50.71	64.38	-	-	0.93
Median	Paso	3	6	69	46	58			1
Mittelwert	Comedia	14.75	69.25	191.75	8.5	68.19	-	-	0.35
Median	Comedia	14	69	186	1	68.11			0.34

n=Anzahl der Knoten (Sprecher), m = Anzahl der Kanten, $\Delta(G)$ = Maximalgrad, $\delta(G)$ = Minimalgrad,

C_{Degree}= höchste Degree Centrality, C = globaler Cluster-Koeffizient, p(G) = durchschnittliche Dichte

Miguel de Cervantes Saavedra

Zwischenergebnis Cervantes Saavedra, Mittelwerte und Median

(ungewichtet) ohne 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Entremés	11	26	9	1	5	1	1	1
Median	Entremés	10	21	9	1	5	1	1	0
Mittelwert	Comedia	26.88	78.56	15.67	0.22	5.74	0.63	0.54	0.24
Median	Comedia	26	78	16	0	5.23	0.63	0.51	0.21

Zwischenergebnis Cervantes Saavedra, Mittelwerte und Median

(ungewichtet) mit 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Entremés	12.14	28.14	9.14	1	4.50	0.85	0.65	0.42
Median	Entremés	11	23	10	1	4.5	0.83	0.59	0.38
Mittelwert	Comedia	28.33	81.78	16.22	0.1	5.72	0.62	0.52	0.23
Median	Comedia	28	80	17	0	5.28	0.63	0.52	0.20

Zwischenergebnis Cervantes Saavedra, Mittelwerte und Median

(gewichtet) ohne 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Entremés	10.86	41.57	81	2	28.71	-	-	0.41
Median	Entremés	10	33	73	2	27.2			0.37
Mittelwert	Comedia	26.9	130	212	0.22	45.69	-	-	0.20
Median	Comedia	26	124	204	0	42.67			0.19

Zwischenergebnis Cervantes Saavedra, Mittelwerte und Median

(gewichtet) mit 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Entremés	12.14	44	81	1.57	25.72	-	-	0.34
Median	Entremés	11	36	73	2	24.91			0.31
Mittelwert	Comedia	28	134	213	0.11	43.62	-	-	0.18
Median	Comedia	28	129	206	0	40.25			0.17

n=Anzahl der Knoten (Sprecher), m = Anzahl der Kanten, $\Delta(G)$ = Maximalgrad, $\delta(G)$ = Minimalgrad,

C_{Degree}= höchste Degree Centrality, C = globaler Cluster-Koeffizient, p(G) = durchschnittliche Dichte

Tirso de Molina
Zwischenergebnis Tirso de Molina, Mittelwerte und Median
(ungewichtet) ohne 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Entremés	9.11	13.44	5.67	1	3.02	0.89	0.56	0.51
Median	Entremés	7	11	5	1	3.43	1	0.68	0.57
Mittelwert	Comedia	24.48	56.34	13.48	0.33	5.11	0.67	0.57	0.30
Median	Comedia	22	53	13	0	5.22	0.70	0.58	0.24

Zwischenergebnis Tirso de Molina, Mittelwerte und Median
(ungewichtet) mit 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Entremés	10.44	19.56	6.67	0.78	3.51	0.83	0.58	0.45
Median	Entremés	7	12	5	1	3.43	0.83	0.67	0.57
Mittelwert	Comedia	25.05	60.79	14.19	0.31	5.11	0.65	0.55	0.28
Median	Comedia	24	57.5	14	0	5.22	0.70	0.57	0.22

Zwischenergebnis Tirso de Molina, Mittelwerte und Median
(gewichtet) ohne 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Entremés	9.67	30.78	102	3.22	39.57	-	-	0.49
Median	Entremés	7	21	60	2	26			0.5
Mittelwert	Comedia	23.48	95.28	310.60	0.90	69.03	-	-	0.26
Median	Comedia	22	89	298.5	0	65.21			0.20

Zwischenergebnis Tirso de Molina, Mittelwerte und Median
(gewichtet) mit 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Entremés	10.44	33.22	102.78	1.44	34.86	-	-	0.39
Median	Entremés	9	22	62	1	24.57			0.41
Mittelwert	Comedia	25.05	101.19	311.43	0.90	65.44	-	-	0.24
Median	Comedia	24	97	298.5	0	60.51			0.19

n=Anzahl der Knoten (Sprecher), m = Anzahl der Kanten, $\Delta(G)$ = Maximalgrad, $\delta(G)$ = Minimalgrad,
C_{Degree}= höchste Degree Centrality, C = globaler Cluster-Koeffizient, p(G) = durchschnittliche Dichte

Juan Ruiz de Alacrón y Mendoza
Zwischenergebnis Juan Ruiz, Mittelwerte und Median

(ungewichtet) ohne 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Comedia	15.1	47.2	11.55	0.85	6.30	0.84	0.72	0.48
Median	Comedia	14.5	45.5	11	0	6.31	0.85	0.73	0.46

Zwischenergebnis Juan Ruiz, Mittelwerte und Median

(ungewichtet) mit 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Comedia	16.25	49.65	12.25	0.55	6.16	0.82	0.69	0.43
Median	Comedia	15	47.5	11	0	6.08	0.83	0.70	0.42

Zwischenergebnis Juan Ruiz, Mittelwerte und Median

(gewichtet) ohne 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Comedia	15.1	81.35	292.8	4.1	92.2	-	-	0.42
Median	Comedia	14.5	77.5	274	0	88.06			0.42

Zwischenergebnis Juan Ruiz, Mittelwerte und Median

(gewichtet) mit 90er IDs

Gruppe	Genre	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C _{Degree}	C	p(G)
Mittelwert	Comedia	16.25	84.75	293.6	1.1	86.46	-	-	0.38
Median	Comedia	15	80.5	275	0	82.19			0.37

n=Anzahl der Knoten (Sprecher), m = Anzahl der Kanten, $\Delta(G)$ = Maximalgrad, $\delta(G)$ = Minimalgrad,
C_{Degree}= höchste Degree Centrality, C = globaler Cluster-Koeffizient, p(G) = durchschnittliche Dichte

Zwischenergebnis Testkorpora

<i>Testkorpora (ungerichtet)</i>								
Gruppe	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C_{Degree}	C	p(G)
Schiller 1	11	34	9	3	6.18	0,9	0.78	0.62
Schiller 2	22	70	15	0	6.36	0.71	0.60	0.30
Kleist	9	31	8	3	6.89	1	0.94	0.86
Schnitzler	9	23	8	2	5,11	1	0.86	0.64

<i>Testkorpora (gerichtet)</i>								
Gruppe	n	m	$\Delta(G)$	$\delta(G)$	d(G)	C_{InDegree}	$C_{\text{OutDegree}}$	p(G)
Schiller 1	11	60	398	8	167.64	0,8	0.8	0.54
Schiller 2	22	125	654	0	139.18	0.67	0.71	0.27
Kleist	9	56	36	726	263.78	1	1	0.78
Schnitzler	9	43	312	12	125.56	1	1	0.60

n=Anzahl der Knoten (Sprecher), m = Anzahl der Kanten, $\Delta(G)$ = Maximalgrad, $\delta(G)$ = Minimalgrad,

C_{Degree} = höchste Degree Centrality, C = globaler Cluster-Koeffizient, p(G) = durchschnittliche Dichte

Schiller 1 = Kabale und Liebe

Schiller 2 = Don Carlos

Kleist = Der zerbrochene Krug

Schnitzler = Komtesse Mizzi