
Eine Netzanalyse der Organisationsstrukturen im Bereich Geldwäsche

Masterarbeit

zur Erlangung des Grades eines Master of Science
im Studiengang Wirtschaftsinformatik

Vorgelegt von

Nicholas Sartor

Immatrikulationsnummer: 208110667

E-Mail: nicholassartor@uni-koblenz.de

Fachbereich 4: Informatik

Institut für Wirtschafts- und Verwaltungsinformatik

Universität Koblenz-Landau

Erstgutachter:

Prof. Dr. Klaus G. Troitzsch

Zweitgutachter:

Dr. Michael Möhring

Koblenz, Dezember 2014

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

	Ja	Nein
Mit der Einstellung dieser Arbeit in die Bibliothek bin ich einverstanden.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Der Veröffentlichung dieser Arbeit im Internet stimme ich zu.	<input checked="" type="checkbox"/>	<input type="checkbox"/>

(Ort, Datum)

(Unterschrift)

Abstrakt

Diese Arbeit analysiert die Möglichkeiten der Text- und Netzwerkanalyse kriminologischer Daten. Dabei werden diese Daten unter dem Aspekt der Geldwäsche betrachtet.

Zum Zweck der Analyse wurden aus dem Text die wichtigsten Konzepte in elf verschiedene Klassen aufgeteilt. Die Zusammenhänge zwischen den identifizierten Konzepten wurden ebenfalls untersucht.

Einige Aussagen über Geldwäsche lassen sich durch die Interpretation der Ergebnisse validieren. Es werden bestimmte Konzepte, wie Banken, Unternehmen und ausländische Tochtergesellschaften als wichtige Konzepte identifiziert. Das bestätigt zusammen mit dem in Kapitel 1.4.3 dargestellten zyklischen Ablauf von Geldwäschevorgängen die Vermutung, dass bestimmte Organisationen und Personen, innerhalb des betrachteten Falles, Gelder über Banken, Gesellschaften und Investitionen in den legalen Finanzmarkt einfließen lassen.

Ebenfalls wird innerhalb dieser Arbeit der Nutzen der betrachteten Tools für den hier zugrundeliegenden Forschungsprozess analysiert und eine Einschätzung der Eignung von ORA und Automap für diese Art von Forschung geliefert.

Abstract

This Thesis conducts a text and network analysis of criminological files. The specific focus during the research is the field money laundering.

The analysis showed the most important concepts present in the text which were classified in eleven different classes. The relationships of those concepts were analysed using ego networks, key entity identification and clustering.

Some of the statements given about money laundering could be validated by the findings of this analysis and their interpretation. Specific concepts like banks and organizations as well as foreign subsidiaries were identified. Aggregating these concepts with the statements in chapter 1.4.3 on the circular process of money laundering it can be stated that different organizations and individuals, present in the criminological files, were placing money through different banks, organizations and investments in the legal financial market.

At last this thesis tries to validate the benefits of the used tools for the kind of conducted research process. An estimation on ORA's and Automap's applicability for this kind of research is given in the end.

Inhaltsverzeichnis

Erklärung	iii
Abstrakt	v
Abstract	vi
Inhaltsverzeichnis	vii
Abbildungsverzeichnis	ix
Tabellenverzeichnis	x
Anhangsverzeichnis	x
1 Einleitung	1
1.1 Forschungsziel	2
1.2 Forschungsfragen	2
1.3 Die Datengrundlage und Herkunft	3
1.4 Definitionen und Hintergrund	4
1.4.1 Organisiertes Verbrechen	4
1.4.2 Extortion Racket Systems	5
1.4.3 Geldwäsche	7
1.4.4 Ontologie	10
1.5 Struktur der Arbeit	11
2 Forschungsdesign	12
2.1 Toolauswahl	13
2.1.1 Betrachtete Tools	13
2.1.1.1 RapidMiner	13
2.1.1.2 Knime	13
2.1.1.3 KHCoder	14
2.1.1.4 Automap	14
2.1.1.5 ORA	14
2.1.1.6 Pajek	15
2.1.1.7 Gate	15
2.1.2 Analyseschema	16
2.2 Methodische Beschreibung	20
2.2.1 Textanalyse	20
2.2.1.1 Manuelle Vorbearbeitung	23
2.2.1.2 Textanalyse mit Automap	25
2.2.1.2.1 Textvorbereitung	26
2.2.1.2.2 Weitere Beschreibungen	29

	2.2.1.2.3	Thesaurus Dateien	30
	2.2.1.2.4	Konzeptklassen in ORA und Automap.....	32
	2.2.1.2.5	Das DyNetML Format	34
2.2.2		Netzwerkanalyse	35
	2.2.2.1	Theoretische Grundlagen	35
	2.2.2.2	Explizites Analysekonzept.....	37
	2.2.2.3	Die Erstellung des Netzwerkes	38
	2.2.2.3.1	Workaround für alle Automap Klassen	41
	2.2.2.4	Die Netzwerkanalyse in ORA	43
	2.2.2.4.1	Maßzahlen innerhalb der Netzwerkanalyse	45
	2.2.2.4.2	Analyse der Netzwerkstruktur	47
	2.2.2.4.3	Identifikation der Key-Entities	50
	2.2.2.4.4	Analyse der Egonetzwerke	56
	2.2.2.4.5	Cliquenanalyse	64
3		Zusammenfassung der Ergebnisse.....	67
3.1		Ergebnisse der Textanalyse.....	67
	3.1.1	Forschungsfrage 1	67
	3.1.2	Forschungsfrage 2	68
	3.1.3	Forschungsfrage 3	69
3.2		Ergebnisse Netzwerkanalyse	70
	3.2.1	Forschungsfrage 2	70
	3.2.2	Forschungsfrage 3	71
	3.2.3	Forschungsfrage 4	71
3.3		Ergebnisse der Ontologiebildung.....	72
4		Fazit und Ausblick.....	73
	4.1	Inhaltliches Fazit.....	73
	4.2	Restriktionen der Arbeit.....	73
	4.3	Wertung der genutzten Tools	74
	4.4	Ausblick	75
5		Literaturverzeichnis.....	76
		Anhang.....	79

Abbildungsverzeichnis

Abbildung 1.1: Verwendung der Erlöse aus illegalen Geschäften (vgl.: Altenkirch 2002, S. 8)	7
Abbildung 1.2: Das typische Geldwäscheschema	9
Abbildung 2.1: Vorgehensmodell.....	12
Abbildung 2.2: Text- und Netzanalyse	20
Abbildung 2.3: Abgrenzung von Text Mining zu bekannten Forschungsgebieten.....	21
Abbildung 2.4: Beispielanzeige Automap.....	26
Abbildung 2.5: Beispielanzeige Automap nach Vorbereitung	28
Abbildung 2.6: Auszug einer Konzeptliste des Reuters-Beispiels	29
Abbildung 2.7: Ansicht des Thesaurus Editors	31
Abbildung 2.8: Dynamische Netzwerke in DyNetML	34
Abbildung 2.9: Vorgehen der Netzwerkanalyse.....	37
Abbildung 2.10: Thesaurusprüfung in ORA.....	38
Abbildung 2.11: Thesaurusprüfung in Automap	39
Abbildung 2.12: Workaround für alle Automap Klassen, Schritt 1	42
Abbildung 2.13: Workaround für alle Automap Klassen, Schritt 2	42
Abbildung 2.14: Das Netzwerk im ORA Visualizer.....	44
Abbildung 2.15: Zusammenführungsfehler in ORA.....	45
Abbildung 2.16: Wichtige Schlüsselpersonen	53
Abbildung 2.17: Wichtige Organisationen	54
Abbildung 2.18: Beispiel eines Egonetzwerkes	56
Abbildung 2.19: Egonetzwerk (Location x Task) von Vereinbarung.....	57
Abbildung 2.20: Egonetz Geldwäsche x Agent-344.....	60
Abbildung 2.21: Egonetzwerk der Schlüsselgelder (Location x Organization x Gelder)	61
Abbildung 2.22: Egonetzwerk der verschiedenen Gesellschaftstypen	62
Abbildung 2.23: Egonetz Beratung x Off_Shore_Beratung.....	63
Abbildung 3.1: Anforderungen an Text- und Netzanalysetools.....	68
Abbildung 3.2: Grundlagenklassifizierung im Bereich Geldwäsche	72

Tabellenverzeichnis

Tabelle 1-1: Forschungsfragen und -ziele.....	3
Tabelle 2-1: Analyseschema der betrachteten Tools	16
Tabelle 2-2: Angaben zur Textdatei (Gloders Quelldatei)	29
Tabelle 2-3: Mehrfachnennung innerhalb eines Thesaurus.....	32
Tabelle 2-4: Zuweisung der Konzeptklassen	43
Tabelle 2-5: Auszug des Context Reports	49
Tabelle 2-6: Context Report Ergebnisse	49
Tabelle 2-7: Beispiel Key Entities (Agent)	51
Tabelle 2-8: Auländisch und ähnliche Konzepte.....	58
Tabelle 2-9: Ähnliche Konzepte (Attribute-Klasse).....	58
Tabelle 2-10: Beispielinhalt (Clique 1)	65
Tabelle 3-1: Zuweisung der Thesauruseinträge zu den Ontologieklassen von Automap	69

Anhangsverzeichnis

Anhang 1: Top 5 Key Entities der relevanten Knotenklassen.....	79
Anhang 2: Context Report (Zusammengefasst)	80
Anhang 3: Newman Cluster.....	82
Anhang 4: Newman Cluster CSV	82
Anhang 5: Verbindungen von Geldern und Agenten	83
Anhang 6: Verbindungen von Geldern und Organisation 31	84

1 Einleitung

Organisiertes Verbrechen ist in vielen Staaten ein Phänomen. Innerhalb von Deutschland liegt die Zahl der Strafverfahren bezüglich organisiertem Verbrechen im Zeitraum von 2004 bis 2013 bei durchschnittlich 599 jährlich. Davon entfallen im Durchschnitt 304 Verfahren jährlich auf Erstmeldungen. 2013 wurden 282 Ermittlungsverfahren abgeschlossen, im Jahr 2012 waren es 274 Verfahren. In 35% der Fälle handelt es sich um Verfahren aus dem Bereich von Geldwäscheaktivitäten (Bundeskriminalamt 2013).

In anderen Ländern, wie beispielsweise Italien, gehören Konzepte wie die Mafia seit Jahren zu den bekannten Formen des organisierten Verbrechens. Innerhalb dieser Verbrechensorganisationen bestehen feste Strukturen, die sich auszeichnen durch Attribute wie "Familie", "Macht", "Respekt" und "Territorium". Die älteste Form der Mafia in Italien ist die Cosa Nostra, eine sizilianische Form der Mafia, die hauptsächlich in Richtung Nordamerika expandiert. Einige der "fähigen" Geldwäscher der Cosa Nostra haben es geschafft, legitime Organisationsstrukturen innerhalb von Ländern wie Südafrika, Kanada und den USA zu etablieren (EUROPOL 2013, S. 15).

Jedoch existieren ebenfalls Formen des organisierten Verbrechens, deren Strukturen nicht bzw. weniger hierarchisch aufgebaut sind. Aufgrund der flacheren Hierarchie dieser Organisationen, in denen es keine eindeutige Weisungsbefugnis geben muss, kann man von einem Personennetzwerk sprechen.

Der technische Fortschritt ermöglicht es nicht nur Analysten und Kriminologen, gegen Verbrechen vorzugehen, sondern erleichtert ebenfalls das Taggeschäft von Kriminellen (EUROPOL 2013). Durch das Internet und weitere Entwicklungen (bspw. Soziale Netzwerke etc.) kann sich die Dynamik krimineller Gruppierungen erhöhen (UNODC, 2011, S. 8). Dies bedeutet, dass es neue Angriffsmöglichkeiten für Kriminelle gibt, bspw. Datenraub aus sozialen Netzen, aber auch eine bessere Absprachemöglichkeit innerhalb einer Gruppierung.

Das, durch das 7. Rahmenprogramm der EU geförderte, Forschungs- & Entwicklungsprojekt GLODERS (Global Dynamics of Extortion Racket Systems) analysiert diese Dynamiken. Dazu werden bestehende Datensätze genutzt und analysiert, um so Vorhersagen zu Strukturen und Dynamiken von Extortion Racket Systemen (kurz: ERS) treffen zu können.

Insbesondere möchte das GLODERS-Team Informationssysteme und Modelle anhand von Textanalysen erstellen. Mit diesen Modellen sollen Regeln bzw. Muster geformt werden um die Entstehung oben genannter Strukturen zu unterbinden. Man interessiert sich hier vor allem für die kulturellen, ökonomischen und sozialen Umfeldler in denen diese Systeme entstehen (GLODERS.eu 2012). Die identifizierten Muster, Regeln, Modelle und Vorhersagen sollen für weiterführende Simulationen genutzt werden.

Innerhalb der Forschungsarbeit liegt der Fokus der Analyse auf dem Themengebiet der Geldwäsche. Es soll ermittelt werden, wie Gelder, die durch die illegalen Geschäfte des organisierten Verbrechens eingenommen werden, in den "legalen" Markt eingeschleust werden können.

1.1 Forschungsziel

Das Ziel der Arbeit ist der Versuch, Netzwerke aus einer bestehenden Datengrundlage zu extrahieren und zu analysieren. Dabei geht es um die Textanalyse der vorliegenden Datengrundlage zur Identifikation relevanter Netzwerkkarten und deren Attribute.

Es soll bewertet werden, inwieweit manuelle Vorarbeit und automatische Analyse miteinander in Verbindung stehen, d.h., es wird geprüft, welche Arbeiten ein Forscher manuell erledigen muss, damit ein Analysetool die Daten sinnvoll nutzen und verstehen kann. Es geht in erster Linie um die Betrachtung dieses Zusammenspiels und in zweiter Linie um die Erkenntnisse der eigentlichen Analyse.

Die Ergebnisse dieser Arbeit dienen zu weiteren Forschungszwecken des GLODERS Projektes. Die identifizierten Netzwerke können so in weiteren Schritten als Grundlage für Simulationsmodelle genutzt werden.

Daher gibt es in dieser Arbeit drei Zielsetzungen:

- Ziel 1: Das Bilden einer geeigneten Ontologie im Bereich Geldwäsche
- Ziel 2: Eine Bewertung der automatischen Text- und Netzanalyse anhand der erforderlichen manuellen Vorbereitungsschritte
- Ziel 3: Die Analyse verschiedener Netzwerke innerhalb des Datensatzes und deren Zusammenspiel

Ziel 1 dient dabei als Grundlage für weitergehende Forschung. Es soll eine Grundlagenontologie gebildet werden, die wesentliche Konzepte der zugrundeliegenden Daten erfasst und kategorisiert. Ebenfalls hängt dieses Ziel wesentlich von den Erkenntnissen der Netzwerkanalyse ab.

1.2 Forschungsfragen

Um die oben genannten Ziele zu erreichen wurden folgende Forschungsfragen aufgestellt:

- FF1: Welche Konstellation an (halb-)automatischen Analysewerkzeugen ermöglicht eine inhaltliche Text- und Netzwerkanalyse?
- FF2: Wie können frei zugängliche Programme das gewünschte Ziel allein oder mit Zusatz von Eigenprogrammierung lösen?
- FF3: Welche Einheiten lassen sich aus dem Text extrahieren?
- FF4: Wie stehen diese Einheiten miteinander im Verhältnis?

Zum besseren Verständnis werden die Forschungsfragen nun den jeweiligen Forschungszielen zugeordnet und mit den dafür erforderlichen Forschungsmethoden verbunden. Die Forschungsmethoden werden in Kapitel 2 weitergehend ausgeführt.

Tabelle 1-1: Forschungsfragen und -ziele

Ziel	Forschungsfrage	Forschungsmethoden
Ziel 1: Bilden einer geeigneten Ontologie	FF3: Welche Einheiten lassen sich aus dem Text extrahieren?	Textanalyse
Ziel 3: Netzwerkanalyse des Datensatzes	FF4: Wie stehen diese Einheiten miteinander in Verbindung?	Netzanalyse
Ziel 2: Bewertung der automatischen Text- Netzanalyse	FF1: Welche Konstellation an (halb-) automatischen Analysewerkzeugen ermöglicht eine inhaltliche Text- und Netzanalyse?	Bewertungsschema
	FF2: Wie können frei zugängliche Programme das gewünschte Ziel allein oder mit Zusatz von Eigenprogrammierung lösen?	

1.3 Die Datengrundlage und Herkunft

Die Daten, die dieser Arbeit zugrunde liegen, sind Dokumente der niederländischen Polizei. Es handelt sich bei den Inhalten um Interviewabschriften und Ermittlungsnotizen der Polizei. Innerhalb des Falles wurden Zeugen zu Mitgliedern und Aktivitäten des organisierten Verbrechens befragt. Die Inhalte der Befragungen sollen dazu genutzt werden, die Aktivitäten einzelner Personen und Organisationen zu verstehen und miteinander zu verknüpfen, sodass im späteren Verlauf ein gemischtes Netzwerk aus verschiedenen Klassen entsteht.

Aus Gründen des Datenschutzes werden innerhalb dieser Arbeit weder Namen der Organisationen noch der Personen genannt. Die Ergebnisse werden auf einer Metaebene präsentiert, damit die Zusammenhänge für den Leser deutlich werden, aber keine genannte Einzelperson negativ betroffen wird.

Die Daten wurden durch das Team des GLODERS-Projektes bereitgestellt.

Herausfordernd ist hierbei, dass eine Analyse der Daten nur im Wechselspiel von manueller und automatischer Bearbeitung möglich ist und es sich, um die Analyse von personenbezogenen Daten handeln kann.

Ein wichtiger Faktor dieser Arbeit ist die Einschätzung eventuell notwendiger Vorverarbeitungsschritte. In diesem Zusammenhang werden in der Arbeit verschiedene Data Mining Tools angewandt, um

einen guten Kompromiss zwischen manueller Datenvorbereitung und automatischer Datenanalyse zu finden.

Diese Arbeit fokussiert sich auf den Teilaspekt der Geldwäsche. Die Daten beinhalten Angaben sowohl zu Personen, Geldbeträgen, Zeiträumen, Institutionen als auch zu weiteren Einheiten. Die genauen Einheiten und deren Herleitung werden in Kapitel 3 näher beschrieben. In Kapitel 1.4.4 wird geschildert, was innerhalb dieser Arbeit als Ontologie verstanden wird.

Eine Netzwerkanalyse wurde deshalb gewählt, um die erhebliche Datenmenge nach Möglichkeit in der Komplexität zu reduzieren. Die Forschungsfragen sollen dabei helfen das Gebiet der Geldwäsche in die einzelnen Bestandteile aufteilen zu können und bestimmte Teilbereiche zu betrachten. So könnten beispielsweise Netzwerke zwischen Geldeinheiten und Personen oder Organisationen geprüft werden, um den Aspekt der Verschleierung zu ermitteln.

Um die Komplexität der Daten handhaben zu können, ist daher interessant welche Tools einem Nutzer einen schnellen Überblick über verschiedene Eigenschaften der Daten liefern können. Dies könnten bestimmte Textpassagen oder einzelne Wörter mit besonderer Bedeutung oder eine Klassifizierung einzelner Wörter in die zugehörigen Wortgruppen sein (Subjekt, Prädikat oder Objekt).

Netzwerkanalysertools können dahingehend bei einer solchen Betrachtung helfen, als dass sie die Beziehungen einzelner Wörter untereinander visualisieren und analysieren können. Dem Nutzer ist es daher meist möglich einen grafischen oder analytischen Ansatz zu wählen.

1.4 Definitionen und Hintergrund

Innerhalb dieses Abschnittes werden zuerst Begriffserklärungen gegeben und anschließend erläutert, wie bestimmte Begriffe innerhalb dieser Forschungsarbeit genutzt werden. Zu den thematischen Begriffen zählen: "Organisiertes Verbrechen", "Extortion Racket Systems" und "Geldwäsche". Daran anschließend wird die Nutzung des methodischen Begriffs Ontologie erklärt.

1.4.1 Organisiertes Verbrechen

Die erste Definition stammt vom Bundeskriminalamt und soll einen Einblick darauf geben, was die deutsche Polizei und Justiz unter dem Begriff des organisierten Verbrechens versteht.

„Organisierte Kriminalität ist die von Gewinn- oder Machtstreben bestimmte planmäßige Begehung von Straftaten, die einzeln oder in ihrer Gesamtheit von erheblicher Bedeutung sind, wenn mehr als zwei Beteiligte auf längere oder unbestimmte Dauer arbeitsteilig

a)unter Verwendung gewerblicher oder geschäftsähnlicher Strukturen,

b)unter Anwendung von Gewalt oder anderer zur Einschüchterung geeigneter Mittel oder

c)unter Einflussnahme auf Politik, Medien, öffentliche Verwaltung, Justiz oder Wirtschaft zusammenwirken“¹ (Bundeskriminalamt 2013, S. 9).

Dabei wird angemerkt, dass zwischen den Aktivitäten des Terrorismus und der organisierten Kriminalität unterschieden wird. Erstere werden nicht mit in die obige Definition eingeschlossen.

In einem Bericht des Schweizer Bundesrates an das Parlament wurde folgende Definition von organisierter Kriminalität gegeben.

"Organisiertes Verbrechen liegt dort vor, wo Organisationen in Annäherung an die Funktionsweise internationaler Unternehmen hochgradig arbeitsteilig, stark abgeschottet, planmäßig und auf Dauer angelegt sind und durch Begehung von Delikten sowie durch Teilnahme an der legalen Wirtschaft möglichst hohe Gewinne anstreben. Die Organisation bedient sich dabei der Mittel der Gewalt, Einschüchterung, Einflussnahme auf Politik und Wirtschaft. Sie weist regelmäßig einen stark hierarchischen Aufbau auf und verfügt über wirksame Durchsetzungsmechanismen für interne Gruppennormen. Ihre Akteure sind dabei weitgehend austauschbar" (Schweizer Bundesrat 2002, S. 1841).

Hierbei wird jedoch darauf hingewiesen, dass die Strukturen eine derart hohe Dynamik haben, dass eine abschließende Definition im Gesetz nicht gewährleistet werden kann. Dies soll speziell für Gruppierungen gelten, die keine strikte Hierarchie ähnlich der italienischen Mafia besitzen, sondern aus mehreren Teilgruppen bestehende Netzwerke sind, welche zwar für bestimmte "Operationen" zusammen agieren, ansonsten aber autonom handeln (Schweizer Bundesrat 2002).

Für diese Arbeit wird organisiertes Verbrechen daher als "die Zusammenarbeit mehrerer Personen über einen längeren Zeitraum, um durch Delikte Geld oder Macht zu erlangen" verstanden. Dabei sind in erster Linie nicht die Hierarchie, sondern Schlüsselpersonen bzw. -gruppen relevant.

1.4.2 Extortion Racket Systems

Extortion Racket² wird übersetzt mit [Schutzgeld-]Erpressung bzw. verbrecherischer Erpressung. Daher handelt es sich bei Extortion Racket Systemen um einen Zustand, welcher es einer Einzelperson bzw. einer Gruppe erlaubt, durch verbrecherische Erpressung Dritte durch Drohung zu einer bestimmten Tat zu bringen.

¹ Arbeitsdefinition „Organisierte Kriminalität“, die im Mai 1990 von der GAG Justiz/Polizei verabschiedet wurde

² Extortion racket wird übersetzt mit "[Schutzgeld]erpressung" bzw. "verbrecherische Erpressung" (Pons.de, <http://de.pons.com/%C3%BCbersetzung?q=Extortion+Racket&l=deen&in=&lf=>, letzter Zugriff: 08.10.2014)

Im deutschen Strafgesetzbuch (StGB) werden in §253 Erpressung und in §255 Räuberische Erpressung als Straftatbestände erfasst. Der deutsche Gesetzgeber definiert in §253 StGB den Straftatbestand der Erpressung und dessen Strafbarkeit wie folgt:

- "(1) Wer einen Menschen rechtswidrig mit Gewalt oder durch Drohung mit einem empfindlichen Übel zu einer Handlung, Duldung oder Unterlassung nötigt und dadurch dem Vermögen des Genötigten oder eines anderen Nachteil zufügt, um sich oder einen Dritten zu Unrecht zu bereichern, wird mit Freiheitsstrafe bis zu fünf Jahren oder mit Geldstrafe bestraft.
- (2) Rechtswidrig ist die Tat, wenn die Anwendung der Gewalt oder die Androhung des Übels zu dem angestrebten Zweck als verwerflich anzusehen ist.
- (3) Der Versuch ist strafbar.
- (4) In besonders schweren Fällen ist die Strafe Freiheitsstrafe nicht unter einem Jahr. Ein besonders schwerer Fall liegt in der Regel vor, wenn der Täter gewerbsmäßig oder als Mitglied einer Bande handelt, die sich zur fortgesetzten Begehung einer Erpressung verbunden hat" (Gesetze im Internet - Strafgesetzbuch, § 253).

In §255 StGB heißt es weiter:

"Wird die Erpressung durch Gewalt gegen eine Person oder unter Anwendung von Drohungen mit gegenwärtiger Gefahr für Leib oder Leben begangen, so ist der Täter gleich einem Räuber zu bestrafen" (Gesetze im Internet - Strafgesetzbuch, § 255).

In einer Studie des Joint Research Centre on Transnational Crime wurde zwischen zwei verschiedenen Formen des Extortion Racketeering unterschieden. Demnach wird zwischen einer gelegentlichen und einer systematischen Form unterschieden. Diese Formen würden sich vor allem durch folgende Attribute unterscheiden und definieren:

1. die Organisationsstruktur der kriminellen Gruppe, welche die Erpressung ausübt
2. deren starke Präsenz innerhalb der Territorien
3. das Verhältnis zwischen Täter und Opfer

Weiterhin wird das Verhältnis der Variablen als reziprok bezeichnet, also beeinflussen sich die Variablen gegenseitig (TRANSCRIME 2009).

Im Cambridge Dictionary wird extortion racket definiert als " an organized illegal activity in which a person or group tries to get money from someone by using force or threats" (Cambridge Dictionaries Online).

Aus den oben genannten Definitionen bzw. Erläuterungen ergibt sich für diese Arbeit das Verständnis, dass es sich bei Extortion Racket Systems um die Existenz einer oder mehrerer Personen handelt, welche durch Drohung Mittel oder Handlungen von Dritten erzwingt. In den zugrundeliegenden Daten handelt es sich vorwiegend um Erpressung von "Kriminellen" durch "Kriminelle".

1.4.3 Geldwäsche

Wie in Abschnitt 1.4.1 erwähnt wurde, handelt es sich dann um organisiertes Verbrechen, wenn über einen längeren Zeitraum mehrere Personen durch illegale Maßnahmen zu Geld oder Macht kommen. In diesem Abschnitt wird eine besondere Form dieser illegalen Geschäfte beleuchtet - die Geldwäsche. Ebenfalls bestehen für diese Arbeit Zusammenhänge zu dem vorherigen Thema des Extortion Racketeering, da die Gelder, welche "gewaschen" werden, aus Erpressungen stammen und über Investitionen oder ähnliches in den "legalen" Markt eingeführt werden. Dieser Zusammenhang ist im betrachteten Fall von bedeutender, kritisch zu beurteilender, Tragweite, da die erpressten Personen, Kriminelle sind. Für den Übergang der illegalen Gelder in den legalen Markt bedeutet das, dass sowohl legale als auch illegale Gewinne von Person A (Krimineller und erpresste Person) als illegaler Erlös (Erpresste Summe) für Person B (Erpresser) in Betracht kommen.

Altenkirch stellt die Verwendung illegaler Erlöse grafisch wie folgt dar:

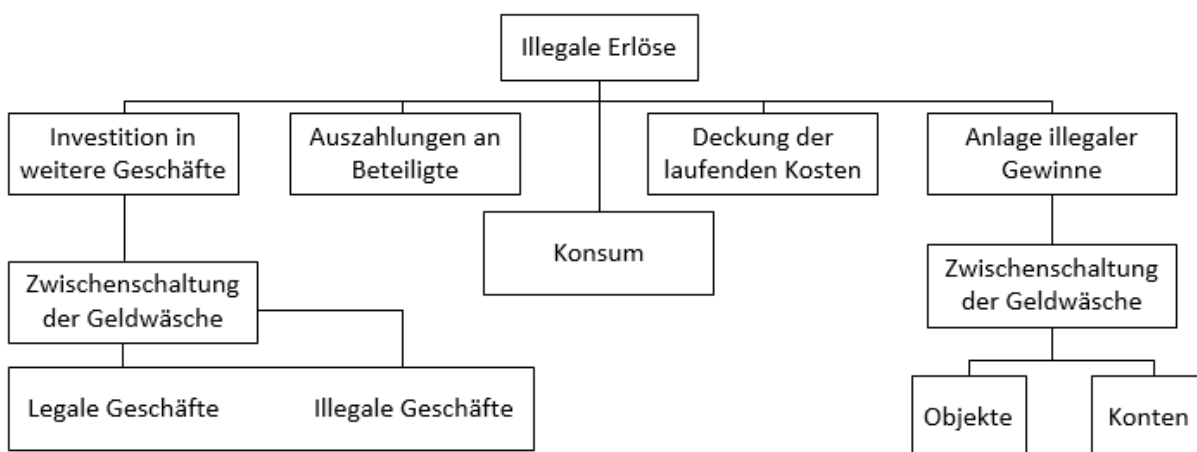


Abbildung 1.1: Verwendung der Erlöse aus illegalen Geschäften (vgl.: Altenkirch 2002, S. 8)³

Illegale Erlöse werden also unter anderem dazu genutzt, Beteiligte auszuzahlen, das Tagesgeschäft zu finanzieren (Konsum & Deckung laufender Kosten) sowie in weitere Geschäfte bzw. Objekte und Konten anzulegen. Daher erkennen Schneider et al., dass eine "Legalisierung" des Vermögens lediglich dann verhindert bzw. aufgedeckt werden kann, wenn die kriminellen Aktivitäten, aus denen das Geld stammt, erkannt und nachverfolgt werden (Schneider et al. 2006, S. 12).

³ Aufgrund höherer Bildqualität nachgezeichnet.

In §261 des deutschen Strafgesetzbuches wird der Straftatbestand der Geldwäsche und dessen Strafbarkeit wie folgt gefasst:

"(1) Wer einen Gegenstand, der aus einer in Satz 2 genannten rechtswidrigen Tat herrührt, verbirgt, dessen Herkunft verschleiert oder die Ermittlung der Herkunft, das Auffinden, den Verfall, die Einziehung oder die Sicherstellung eines solchen Gegenstandes vereitelt oder gefährdet, wird mit Freiheitsstrafe von drei Monaten bis zu fünf Jahren bestraft" (Gesetze im Internet - Strafgesetzbuch, §261)

Gegenstand einer strafbaren Geldwäsche können Gegenstände sein, die aus einem Verbrechen oder einem der in § 261 StGB genannten Vergehen herrühren. Für die vorliegende Arbeit ist relevant, dass zu den In § 261 StGB genannten Vergehen ebenfalls jene aus §253 gehören, "... die gewerbsmäßig oder von einem Mitglied einer Bande, die sich zur fortgesetzten Begehung solcher Taten verbunden hat, begangen worden sind..." (Gesetze im Internet - Strafgesetzbuch, § 261).

Schneider et al. erläutern, dass "Gegenstand der Geldwäsche [...] alle Arten von Vermögensgegenständen sein [können], die [...] aus einer Straftat stammen." Allerdings seien dies oft Gelder, die aus dem illegalen Drogenhandel stammen. Der kritische Punkt der Geldwäsche ist, dargestellt in Abbildung 1.1, der Übergang von "schmutzigen" Geldern in das legale Finanzsystem (Schneider et al. 2006, S. 16).

Es werden vier Ziele der Geldwäsche aufgezeigt.

- Erstes Ziel sei es, die Einziehung der Geldmittel durch Behörden zu verhindern.
- Zweites Ziel ist, den Zugriff auf das Geld dennoch möglich zu erhalten.
- Drittens soll das Geld, in das legale Finanzsystem überführt werden.
- Viertens sollen Täter vor Strafverfolgung bzw. Verurteilung geschützt werden.

In der Schweiz wurde zur Bekämpfung der Geldwäsche das "Bundesgesetz über die Bekämpfung der Geldwäscherei und der Terrorismusfinanzierung im Finanzsektor" (Geldwäschereigesetz, GwG) erlassen. Durch dieses Gesetz werden die Sorgfaltspflichten der Finanzintermediäre definiert. Daraus lässt sich einerseits ableiten, dass die Notwendigkeit besteht, bestimmte Berufszweige besonders zu kontrollieren. Zu den Finanzintermediären zählen laut GwG ⁴: "Banken", "Investmentgesellschaften", "Versicherungseinrichtungen"

⁴ Für genaue Eingrenzung der Gruppen vgl. GwG Art. 2 Geltungsbereich



Abbildung 1.2: Das typische Geldwäscheschema
(United Nations Office on Drugs and Crime)

Laut Abbildung 1.2 teilt sich die Geldwäsche üblicherweise in drei Phasen ein. Die Phasen werden "Placement", "Layering" und "Integration" genannt. Die erste Phase besteht darin, das Geld aus illegalen Geschäften auf Bankkonten einzuzahlen, sodass es sich im Finanzsystem befindet.

Innerhalb des Finanzsystems folgt die Phase "Layering". Sie beinhaltet die Überweisung des Geldbetrages zwischen verschiedenen Unternehmen und Banken unter (teils) vorgetäuschten Rechnungsangaben. Der Schritt der Überweisung über internationale Grenzen wiederholt sich viele Male, sodass ein Rückschluss auf das Ursprungsgeschäft, das illegale Geschäft, nicht möglich ist. Ist diese Verschleierung gelungen, tritt die Phase "Integration" ein, und das Geld wird dazu genutzt, um aus dem Geld der Unternehmen Güter zu erwerben. Diese gehören dann rechtmäßig dem Unternehmen und können gewinnbringend veräußert oder als Geldanlage weiter genutzt werden (United Nations Office on Drugs and Crime). Aus Abbildung 1.2 lassen sich erste Konzeptklassen erstellen, die in der späteren Grundlagenontologie genutzt werden können, so beispielsweise offshore (für Banken oder Unternehmen in einem offeneren Rechtsbereich), Banken, Unternehmen und Geldanlagen.

Die Schwierigkeit bei der Bekämpfung bzw. Aufdeckung von Geldwäschegeschäften ist so die Zuordnung der Phase, in der sich der Untersuchungsfall befindet, und die Erstellung von nachvollziehbaren Geld- / Wertströmen.

Der Begriff der Geldwäsche beschreibt aus Sicht dieser Arbeit alle Personen, Organisationen, Ressourcen, Aufgaben und weiteren Konzepte, welche die Verschleierung oder Verbreitung illegaler Erlöse beeinflussen.

1.4.4 Ontologie

Ontologien werden verwendet, um bestimmte Aspekte der Welt in einem Begriffssystem zu sammeln. Dabei geht es darum, Konzepte miteinander so zu verbinden, dass eine Person schrittweise den Teilbereich, für den die Ontologie entwickelt wurde, verstehen kann. Cocchiarella sagt, dass die Qualität von (formalen) Ontologien davon abhängt, inwieweit die gebildeten Kategorien die Welt abbilden und inwieweit sie mit "gesundem Menschenverstand" nachvollziehbar sind (Cocchiarella 2007).

Smith gibt sieben Eigenschaften an, die eine Ontologie möglichst erfüllen sollte (Smith in: Bennett und Fellbaum 2006, S. 15–19).

- Als erstes sollte eine Ontologie, die für den Gebrauch durch Dritte kreiert wurde, so gestaltet sein, dass erfahrene Personen aus dem Bereich der Ontologierstellung sie mit geringem Aufwand nachvollziehen können sollten.
- Zweitens sollte eine Ontologie für Änderungen offen sein, es sei denn, es handelt sich um essentielle Änderungen, dann sollte eine neue Ontologie erstellt werden.
- Um änderbar bzw. erweiterbar zu sein, muss eine Ontologie daher so geschildert werden, dass sie einem Menschen verständlich ist und es Computern ermöglichen, mit dieser Ontologie zu arbeiten. Eine Ontologie sollte den Zweck erfüllen, sich über den Inhalt mit Dritten auszutauschen, sie sollte daher nicht erstellt werden, nur um erstellt zu werden.
- Viertens sollte man sich bei der Erstellung einer Ontologie an bereits existierenden Teilbereichen orientieren und diese eventuell nutzen.
- Außerdem sollte man nicht zu viele neue Konzepte erstellen, sondern, wenn möglich, so nah wie möglich bei existierenden Begriffen des Fachgebiets bleiben, es sei denn, die Aussagen sind so wichtig, um eine Ontologie inhaltlich zu komplettieren und logisch zu verbinden.
- Die in der Ontologie genutzten Begriffe sollten, den Menschen verständlich und möglichst durch Standardsoftware weiter verwendbar sein.
- Als letztes sagt Smith, dass ein Begriff innerhalb der Ontologie eindeutig in dem Sinne sein soll, dass, egal aus welcher Perspektive man den Begriff betrachtet, er immer die gleiche Aussage behält.

Gruber beschreibt Ontologien als eine explizite Spezifikation einer Konzeptualisierung. Wenn das Wissen eines Fachbereichs in beschreibender Form dargestellt wird, so bezeichnet man die Menge an darstellbaren Objekten als Gegenstandsbereich. Er sagt weiterhin, dass Ontologien daher als eine Sammlung von repräsentativen Texteinheiten definiert werden (Gruber 1993).

Ähnlich zu Smith beschreibt Gruber, dass die Textpassagen für einen Menschen verständlich festgehalten werden müssen und es eine Form von Nutzungsbedingungen für die erstellten Konzepte gibt. Er unterscheidet fünf Kriterien für Ontologien (Gruber 1993, S. 908–909).

- Erstens sollten die benutzten Begriffe möglichst eindeutig und objektiv gewählt sein, als auch möglichst komplett beschrieben werden.
- Zweitens sollte die Ontologie kohärent sein, also in sich logisch verknüpft. Es dürfen keine Aussagen dazu führen, dass andere Aussagen nicht mehr stimmen.
- Weiterhin sollte eine Ontologie erweiterbar sein, ohne ein aufwendiges Studium der existierenden Begriffe vorauszusetzen.
- Es sollten keine Kodierungsformen genutzt werden, die nur dazu dienen, bestimmte Programme oder Systeme zu nutzen.
- Als letztes drückt Gruber aus, dass die Ontologien möglichst für Erweiterungen offen sind, die es späteren Anwendern erlaubt, bestimmte Teilbereiche weiter zu erläutern .

Innerhalb dieser Arbeit soll eine Grundlage für eine Ontologie im Bereich des organisierten Verbrechens, genauer im Teilbereich der Geldwäsche, gelegt werden. Dabei werden innerhalb einer Netzwerkanalyse verschiedene Konzeptklassen identifiziert und miteinander in Verbindung gebracht. Daher ist im Rahmen dieser Arbeit eine Ontologie als die Identifikation und Verfeinerung von gültigen Konzeptklassen im Fokusbereich (hier: Geldwäsche) verstanden worden.

1.5 Struktur der Arbeit

Die Arbeit besteht aus vier Kapiteln. Im ersten Kapitel werden die Hintergründe der Arbeit und deren Motivation sowie die Ziele der Arbeit und die Forschungsfragen geschildert. Ebenfalls werden erste Definitionen der, für diese Arbeit notwendigen, Begriffe geliefert.

Im nächsten Kapitel wird das Forschungsvorgehen im Zuge der Arbeit grafisch dargestellt und schrittweise erläutert. Das dritte Kapitel präsentiert die Ergebnisse der Arbeit und zeigt daher potentielle Netzwerke sowie deren Zusammenspiel und Aussagefähigkeit.

Ein Fazit wird im letzten Kapitel der Arbeit gezogen, in welchem ebenfalls Limitationen der Arbeit sowie ein möglicher weiterer Forschungsbedarf erläutert werden.

Kapitel 1: Einführung und Hintergrund

Kapitel 2: Forschungsdesign

Kapitel 3: Forschungsergebnisse

Kapitel 4: Fazit

2 Forschungsdesign

Innerhalb dieses Kapitels wird das der Arbeit zugrundeliegende Forschungsvorgehen beschrieben. Wie in Abbildung 2.1 zu sehen ist, beginnt die Arbeit mit der Auswahl der benutzten Tools. Dafür wurde ein Raster erstellt, welches die notwendigen Attribute an die Tools festlegt. Verschiedene Tools wurden dafür ausgewählt und vom Autor getestet. Dabei spielen ebenfalls subjektive Attribute wie die Einfachheit der Bedienung und die Mächtigkeit der Programme eine Rolle. In Kapitel 2.1.1.4 wird das Analyseschema näher beschrieben und die Auswahl der Tools vorgestellt. Am Ende des Kapitels wird die Wahl der letztendlich genutzten Tools erläutert.

Im darauffolgenden Unterkapitel 2.2 werden die Methoden der Text- und Netzwerkanalyse eingeführt und die für die Arbeit relevanten Teilgebiete eingegrenzt. In Kapitel 3 werden die erzielten Ergebnisse vorgestellt und deren Bedeutung für den Kontext der Arbeit erläutert, ebenfalls werden die Forschungsfragen, die zu Beginn der Arbeit aufgestellt wurden beantwortet.

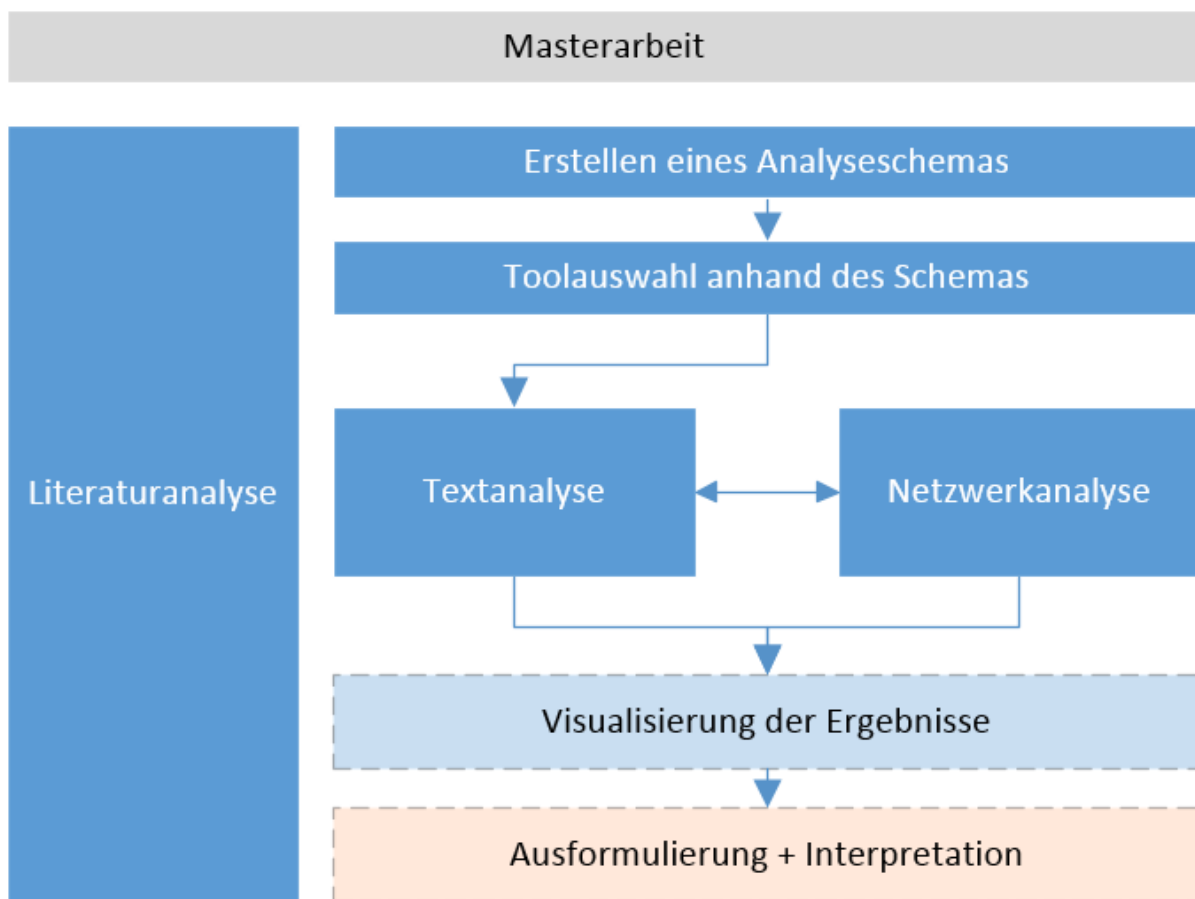


Abbildung 2.1: Vorgehensmodell

2.1 Toolauswahl

Es wurden zur Auswahl der Tools sowohl objektive Attribute wie die Verfügbarkeit bestimmter Funktionen, sowie subjektive Attribute wie eine intuitive Benutzbarkeit der Tools einbezogen. Das Analyseschema gibt ebenfalls das primäre Einsatzgebiet der vorliegenden Tools an. Dies ist vor allem für jene Tools relevant, die zwar Funktionen der Text- bzw. Netzwerkanalyse bieten, allerdings in keinem zufriedenstellendem Umfang. Es folgt die Aufzählung und kurze Beschreibung der Tools und im Abschnitt 2.1.1.4 wird das Analyseschema anhand der einzelnen Attribute erläutert und dargestellt.

2.1.1 Betrachtete Tools

Es wurden sowohl Programme zur reinen Textanalyse, zur reinen Netzanalyse und Mischtools betrachtet. Folgende Programme wurden für die Nutzung der Arbeit getestet: RapidMiner, Knime, KHCoder, ORA, Automap, Pajek und GATE.

2.1.1.1 RapidMiner

RapidMiner ist eine Data Mining Tool der Firma RapidMiner. Rapidminer, bietet dem Nutzer ein grafisches Interface, auf welchem er Prozesse bilden kann. Einzelne Arbeitsschritte werden in eine Art Prozesskette aneinandergereiht und dann sequentiell ausgeführt. Es ist möglich, parallele Arbeitsschritte zu erstellen. In der Version RapidMiner Studio 5.3.015 sind Text-Analyse Methoden nur über Plugins verfügbar. Mit RapidMiner ist es möglich, Textdokumente zu laden, ein bag of words zu erstellen und mithilfe von Marktanalyseverfahren Wörter in einen Zusammenhang zu bringen. Dabei werden Regeln erstellt, wie wahrscheinlich es ist, dass zwei Wörter zusammen genannt werden.

2.1.1.2 Knime

KNIME wird entwickelt und betreut von der KNIME GmbH. Betrachtet wird die Version 2.9.4. Ähnlich zu RapidMiner ist KNIME ein Tool aus dem Bereich des Data Mining. Es ist ebenfalls möglich, Prozessketten zu bilden. Ebenfalls müssen TextMining Mechaniken über Plugins hinzugefügt werden. Dies führt dazu, dass die Prozesse schnell an Größe gewinnen und an Übersichtlichkeit verlieren. Ein Ampelsystem der einzelnen Prozessschritte zeigt an, welche Funktionen durchliefen, welche eventuell noch angepasst werden müssen und welche zu einem Fehler führten. Dafür erhalten die Plugins für KNIME mehr Funktionen für Textanalysen. Es ist beispielsweise möglich, Stemmer anzuwenden, Termhäufigkeiten auszurechnen und die Co-Occurence von Wörtern berechnen zu lassen. Ebenfalls können Stopwortlisten auf die Daten und Vorverarbeitungsschritte wie die Entfernung von Satzzeichen angewandt werden.

2.1.1.3 KHCoder

KH Coder⁵ ist ein quantitatives Inhaltsanalyse bzw. Text Mining Tool. Entwickelt wird es von Koichi Higuchi. Es handelt sich hierbei um ein javabasiertes Tool mit einer minimalistischen grafischen Oberfläche. Die Menüführung ist nicht intuitiv. Es bietet Möglichkeiten der Clusteranalyse und der Textbearbeitung. Allerdings fiel aufgrund der schlechten Dokumentation und der Unübersichtlichkeit der Menüs und Arbeitsschritte die Entscheidung gegen dieses Tool.

2.1.1.4 Automap

Automap ist ein Text Mining tool, welches an der Carnegie Mellon Universität in Pittsburgh vom "Center for Computational Analysis of Social and Organizational Systems" (kurz CASOS) entwickelt wird.

Es ist möglich, einen oder mehrere Textdaten, sowie Blog, Twitter oder andere Daten aus sozialen Netzwerken als Input für die Analysen auswählen. Das Tool arbeitet überwiegend mit Daten im UTF-8 Format. Es kann Duplikate von Texten erkennen und hat viele Vorverarbeitungsmethoden zur Textbereinigung. Es kann automatisch Konzepte erkennen und diese im besten Fall klassifizieren.

Dabei ist es möglich, Konzepte aus dem Text in neue Konzepte zu überführen und ihnen Metaontologien zuzuweisen.

Dem Benutzer ist es möglich, selbst Thesauri Dateien zu erstellen, die als Grundlage für eine spätere Erstellung von Netzwerken dienen. Sind die Texte bereinigt und die Thesauri erstellt, kann Automap DyNetML Daten erstellen. Bei DyNetML handelt es sich um eine XML basierte Sprache zur Darstellung von Knoten, Metadaten und Beziehungen der Knoten. Prozesse lassen sich ebenfalls erstellen, allerdings nicht intuitiv über eine Prozesskette wie bei KNIME oder RapidMiner, sondern über ein Scriptrunner. Weitere Ausführungen zu den einzelnen Funktionen von Automap folgen in Kapitel 2.2.1.

2.1.1.5 ORA

ORA ist ein Tool zur Analyse von dynamischen Metanetzwerken und steht für "Organisation Risk Analyser". Es wird ebenso wie Automap von CASOS an der Carnegie Mellon Universität in Pittsburgh entwickelt. Netzwerke lassen sich über XML Netzwerk Daten importieren oder aus Datengrundlagen wie Excel-Dateien, Pajek Netzwerken und vielen weiteren. ORA bietet sich vor allem durch die hohe Anbindung an Automap als Analysetool für Netzwerke an. Vom selben Entwickler stammend, ist es in der Lage, die Daten, die zur Erstellung der Netzwerke genutzt werden, komplett zu nutzen.

⁵ <http://khc.sourceforge.net/en/>

Es ist möglich, die erstellten Daten von Automap in ORA sogar soweit anzupassen, dass man Gruppierungen ändern und Konzepte somit in eine Analyse ein- oder ausbeziehen kann. Betrachtet wird die Version 3.0.9.9j.

ORA erkennt auf Basis der erstellten Thesaurus Daten Netzwerke zwischen verschiedenen Konzeptgruppen. So können beispielsweise Netzwerke zwischen Personen und Organisationen, sowie zwischen Personen und Aktivitäten betrachtet werden. Es lassen sich alle Klassen zuschalten, sodass man auch ein gesamtes Netzwerk aus allen existierenden Klassen betrachten kann.

Die Ausgaben der Analysen können als Text, HTML, CSV, PowerPoint oder PDF ausgegeben werden, sodass sich dieses Tool ebenfalls für Reports eignet. Weiteres über die genauen Metriken und Methoden dieses Tools folgen in Kapitel 2.2.2.

2.1.1.6 Pajek

Pajek ist ein Netzwerkanalysetool für Windows und wurde an der Universität Ljubljana entwickelt. Sein Zweck ist es, besonders große Netzwerke zu untersuchen. Es kann beispielsweise Soziogramme darstellen und Unternetze aus existierenden Netzwerken extrahieren. Hierbei können in Menüs die spezifischen Wünsche gewählt werden. Beispielsweise lassen sich nur bidirektionale Verbindungen betrachten, also solche Teilnetze, in denen Personen in beide Richtungen agieren. Person A spricht Person B an und Person B reagiert auf Person A.

Pajek kann Netzwerkoperationen ausführen, wie eine Reduktion auf Unternetze, sowie Erkennung von Triads oder anderen Aspekten der Netzwerke. Ebenfalls sind Clusterbildung und -analysen möglich. Reports öffnen sich in Pajek in einem neuen Fenster und lassen sich in Textformat abspeichern. Ebenfalls ist es möglich, Vektoren zu bilden, mit denen sich wiederum beispielsweise geografische Daten repräsentieren lassen. Mehr zu Pajek erfährt man im zugehörigen Wiki⁶, in welchen man ebenfalls die Dokumentation der wichtigsten Methoden findet oder in Nooy et al. 2005. In ihrer Arbeit erklären Nooy et al. schrittweise die Benutzung von Pajek unter Zuhilfenahme mehrerer Testdatensätze.

2.1.1.7 Gate

Gate ist ein Textanalysetool basierend auf Java. Es wurde 1995 an der Universität Sheffield entwickelt und bietet besonders im Bereich der natürlich sprachlichen Bearbeitung und der Informationsextraktion umfangreiche Anwendungsmöglichkeiten. In jedem Zeitpunkt kann man eine Übersicht über die betrachteten Text erhalten.

6 <http://pajek.imfm.si/doku.php> (Letzer Zugriff: 12.10.2014)

In Gate basieren die Textanalysen auf einem Corpus bzw. einer Corpus-Pipeline, dem Nutzer ist es so möglich, ein oder mehrere Dokumente für die Analyse zur Verfügung zu stellen.

Der zweite Schritt in Gate ist es, eine Anwendung zu erstellen, indem man verschiedene Prozesse hinzufügt und in eine Reihenfolge bringt. Jede Methode lässt sich individuell einstellen. Um eine Analyse auszuführen, gibt der Nutzer an, welchen Corpus bzw. Corpus-Pipeline eine Anwendung nutzen soll.

Die Ergebnisse der Anwendung kann der Nutzer dann ansehen und Textpassagen mit möglichen Klassen taggen.

2.1.2 Analyseschema

Das Analyseschema besteht aus objektiven und subjektiven Attributen. Die objektiven Attribute sind vor allem bezogen auf die Existenz bestimmter Funktionen oder anderer Daten, die für alle Betrachtungen (im Zeitraum dieser Masterarbeit) gleichbleibend sind.

Tabelle 2-1: Analyseschema der betrachteten Tools

Attribute	Rapid Miner	Knime	KHCoder	ORA	Automap	Pajek	Gate
Textbezug	+/-	+/-	+	-	+	-	+
Ontologien	+/-	+/-	+/-	+ (+/-)	+	+/-	+
Prozessfluss	+	+	-	-	+	-	+
Dokumentation	+/-	+/-	-	+/-	+/-	+	+
Intuitivität	+	+	-	+	+	+/-	-
Anbindung von Netzwerk & Textanalyse	+/-	+/-	+	+	+	-	-
Vielfalt an Textanalysemethoden	-	- (+/-)	+	-	+	-	+
Vielfalt an Netzanalysemethoden	+/-	+/-	+	+	-	+	-

Der Vorteil von RapidMiner ist, dass man einen Prozess sehr schnell erstellen und für viele Textdaten nutzen kann. Jeder einzelne Knoten der Prozesskette ist anpassbar, sodass trotz eines vorgegebenen Prozesses noch Feineinstellungen getätigt werden können. Ein weiterer Vorteil ist, dass die einzelnen Prozessschritte in der grafischen Oberfläche anzeigen, ob sie erfolgreich durchliefen oder ob innerhalb des Prozesses ein Fehler auftrat. Dies macht es grade bei Subprozessen einfacher, die fehlerhafte Einstellung in der richtigen Methode zu finden.

Allerdings gibt es auch einige Nachteile des RapidMiners. Angefangen damit, dass die Methoden zur Textanalyse nur über Plugins verfügbar sind und somit kein direkter Support von Seiten der Hersteller im Sinne einer kompletten Dokumentation verfügbar ist.

Ein zweiter Nachteil ist, dass der Bezug zur Textebene verloren geht. Die Oberfläche gestattet es zwar, die Prozesse zu betrachten, aber Änderungen, die bei der Vorverarbeitung getätigt wurden, sind nicht direkt auffindbar.

RapidMiner wurde daher als ungeeignet eingeschätzt, da die Netzwerkanalyse über die Methoden der Marktanalyse nur bedingt sinnvoll erschien. Ebenfalls unterstützt RapidMiner in der betrachteten Version keine Form von Ontologien.

Für Knime gilt, dass die Übersichtlichkeit nicht so gut wie bei RapidMiner ist, da keine Funktion für Subprozesse gefunden wurde. Ebenfalls geht auch in KNIME der Textbezug verloren.

Automap bietet eine grafische Oberfläche, die auf den ersten Blick sehr intuitiv ist. Die automatische Zuweisung von Konzepten funktioniert in englischen Texten sehr gut, allerdings sinkt die Qualität dieser bei anderssprachigen Textquellen. Die einzigen Nachteile von Automap scheinen die teilweise nicht ausführliche Dokumentation, die allerdings durch einen Support über ein Online Forum⁷ und die schnelle Reaktionszeit der Entwickler wieder gut gemacht wird, und ein Fokus des Tools auf den englischen Sprachraum zu sein; lediglich einzelne Funktionen wie Stemming Algorithmen beinhalten die Auswahl mehrerer Sprachen.

Gate besitzt lediglich eine nicht intuitive Menüführung und die fehlende Aussagekraft einzelner Methoden erschweren die Arbeit mit diesem Tool. Vor allem sind die Ergebnisse innerhalb von Gate für den Zweck einer nachfolgenden Netzwerkanalyse nicht primär geeignet, da es kein Format gibt in dem Netzwerke sinnvoll gespeichert werden könnten, dies wäre nur über Eigenprogrammierung in Java möglich.

Das Attribut Textbezug drückt aus, ob es dem Nutzer während der Textanalyse möglich ist nachzuvollziehen, wie das Programm den Text verwendet. Werden dem Nutzer alle Schritte und Änderungen angezeigt, wird dieses Attribut positiv (+) bewertet, sind die Textansichten nur teilweise ersichtlich, beispielsweise über das Zwischenspeichern in eine Textdatei, wird dieses Attribut durchschnittlich (+ / -) bewertet. Sollte es hingegen keine Möglichkeit geben, den Text nach Einspeisen in einen Prozess zu sehen, so wird dieses Attribut negativ (-) bewertet. In Abschnitt 2.1 wurde ersichtlich, dass hier ledig-

⁷ <https://groups.google.com/forum/#!forum/ora-google-group> (Letzter Zugriff: 13.11.2014)

lich die Tools KHCoder, Automap und Gate in der Lage sind, den Text komplett darzustellen. In Automap kann man nach jedem Textsäuberungsschritt die Änderung nachverfolgen.

Für das Attribut Ontologien wurde eine positive Wertung gegeben, wenn es innerhalb des Tools möglich ist, Klassen für bestimmte Textpassagen zu definieren. Hier wurden vor allem Gate, Automap und ORA positiv gewertet. Wobei ORA hier lediglich aufgeführt wird, da es in der Lage ist, die in Automap erstellten Metaontologien zu editieren und zu erweitern.

Ein Prozessfluss im Sinne des Analyseschemas ist die Möglichkeit, einen automatisch ausführbaren Prozess zu definieren. Positive Wertungen erhielten die Tools, die entweder die Analyseschritte direkt in eine Prozesskette knüpfen oder diejenigen, die Möglichkeiten bieten, die ausgeführten Schritte in einer Art Skript zu speichern. Positive Wertungen erhielten vor allem RapidMiner und Knime, da die Textanalyse hauptsächlich als Prozesskette dargestellt werden konnte. Außerdem wurden noch Gate und Automap positiv bewertet, da sie die Möglichkeit bieten, Prozesse (bzw. Skripte) zu speichern und zu einem späteren Zeitpunkt zu laden.

Für die Dokumentation wurden dann positive Wertungen gegeben, wenn die programmeigenen Methoden komplett und eindeutig beschrieben werden. Die Tools RapidMiner, Knime, Automap und ORA haben hier lediglich eine durchschnittliche Wertung erhalten. Für RapidMiner und Knime ist diese Einordnung darauf zurückzuführen, dass die Methoden zur Textanalyse lediglich durch Plugins erhältlich und diese nur zum Teil hinreichend beschrieben sind. ORA und Automap haben zwar ausführliche Beschreibungen, Tutorials und Dokumentation, allerdings ist diese an einigen Stellen nicht komplett bzw. widersprüchlich. Pajek und Gate hingegen haben eine sehr gute Dokumentation, insbesondere durch die Popularität der Tools auch Beschreibung in Fachliteratur.

Die Intuitivität wurde daran gemessen, wie schnell sich in der Menüführung zurechtgefunden wurde. Ebenfalls hierfür relevant sind eine eindeutige Benennung der Methoden und die inhaltliche Verknüpfung dieser. Es sollte dem Nutzer ersichtlich sein, welche Konsequenzen die einzelnen Methoden haben. Dieses Attribut ist ein subjektiv wahrgenommenes und ist daher auf die Erfahrung des Nutzers mit den betrachteten Tools zurückzuführen. KNIME und RapidMiner haben eine intuitive Menüführung und man gelangt schnell zu lauffähigen Prozessen. Automap und ORA besitzen ebenfalls klar strukturierte Menüs, lediglich an wenigen Stellen gab es Abstriche bei der Ungenauigkeit der Zwecke bestimmter Methoden.

Die Anbindung von Netzwerk- und Textanalyse klassifiziert, wie gut eine Integration von einer Textanalyse in eine Netzwerkanalyse möglich ist. Hier schneiden besonders die Programme KHCoder, ORA und Automap gut ab. KHCoder vereint beide Analysewerkzeuge in sich. Automap und ORA eignen sich hier besonders, da beide Tools vom selben Institut entwickelt werden. Eine Integration der Ergebnisse der Textanalyse in die Netzwerkanalyse ist vor allem insoweit positiv zu betrachten, als dass das

Textanalysetool Netzwerke erstellen kann, welche folgend in ORA betrachtet werden können. Andererseits besitzt ORA, die Möglichkeit, die Attribute von Automap, welche zur Erstellung der Netzwerke genutzt werden, zu lesen und zu editieren, ohne dafür in Automap zurückkehren zu müssen.

Die Vielfalt der Analysemethoden ergeben sich hauptsächlich aus dem Kernbereich der verwendeten Tools. KNIME und RapidMiner als eher klassischere Data Mining Tools bieten zwar Methoden zur Clusterbildung und Analyse von Knoten, allerdings sind die einschlägigen Textanalysemethoden nur über Plugins zugänglich. KHCoder vereint als einziges Programm Methoden beider Schwerpunkte in sich. ORA und Pajek umfassen beide ausführliche Netzwerkanalysemethoden, wohingegen GATE und Automap die Dominanz im Bereich der Textanalysemethoden bieten.

Zusammenfassend an dieser Stelle sei gesagt, dass die Toolauswahl nicht Kernpunkt, aber entscheidender Bestandteil dieser Arbeit ist. Das Analyseschema beruht auf den Erfahrungen, die der Nutzer mit den Programmen innerhalb einer Einarbeitungszeit gemacht hat und sich die Entscheidung für die letztlich verwendeten Tools aus zwei Faktoren zusammensetzt.

Der erste Faktor ist, dass so gut wie alle Tools gute Wertungen erhielten. Es ist mit den meisten der Tools möglich, bis zu einem bestimmten Punkt die Analyse, die dieser Arbeit zugrunde liegt, durchzuführen. Der zweite und für diese Arbeit entscheidende Faktor ist die Intuitivität der Tools.

ORA und Automap werden für diese Arbeit als Netzwerk- bzw. Textanalysetools verwendet, da die gegenseitige Integration ineinander und die Übersichtlichkeit der Methoden auf den jeweiligen Fachgebieten den entscheidenden Vorteil gegenüber den anderen Tools gegeben hat.

2.2 Methodische Beschreibung

Dieses Kapitel leitet in die Grundzüge des Text Mining und der Netzwerkanalyse ein. Es wird erklärt, welche Komponenten bei den Analysen eine Rolle spielen und mit welchen Verfahren die Ergebnisse erzielt werden. Weiterhin wird erklärt, wie Text Mining und die Netzwerkanalyse in dieser Arbeit zusammenhängen.

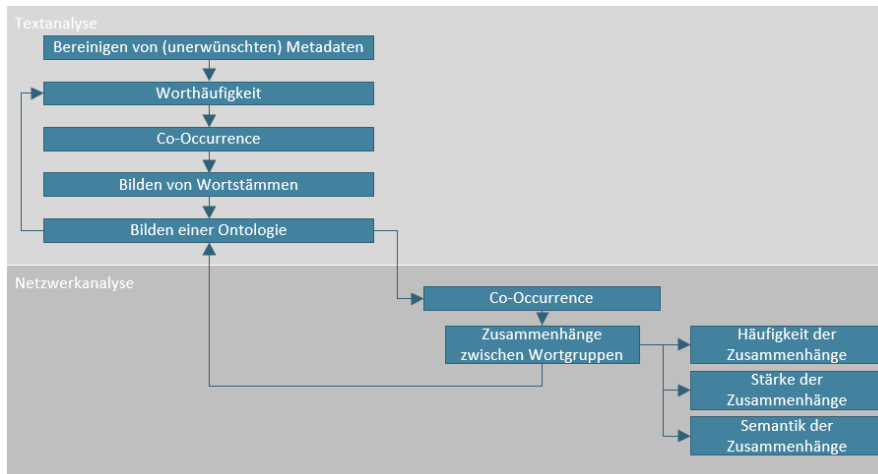


Abbildung 2.2: Text- und Netzanalyse

In Abbildung 2.2 ist zu sehen, aus welchen Arbeitsschritten sich die jeweiligen Analyseschritte zusammensetzen. Nachfolgend werden Textanalyse und Netzwerkanalyse thematisch beschrieben. Am Ende der jeweiligen Kapitel findet sich eine Eingrenzung der genutzten Arbeitsschritte. Diese wurden wiederum zur Erstellung / Erweiterung des oben vorgestellten Analyseschemas in Kapitel 2.1.2 genutzt.

2.2.1 Textanalyse

Eine Textanalyse kann je nach Schwerpunkt des Forschungsansatzes in unterschiedlichen wissenschaftlichen Kategorien eingeordnet werden. Textanalysen bzw. Text Mining bedient sich laut Abbildung 2.3 Methoden des "Natural Language Processing" (NLP), des "Knowledge Discovery in Databases" (KDD), des "Knowledge Discovery in Texts" (KDT), der Statistik und des "Machine Learning" (ML).

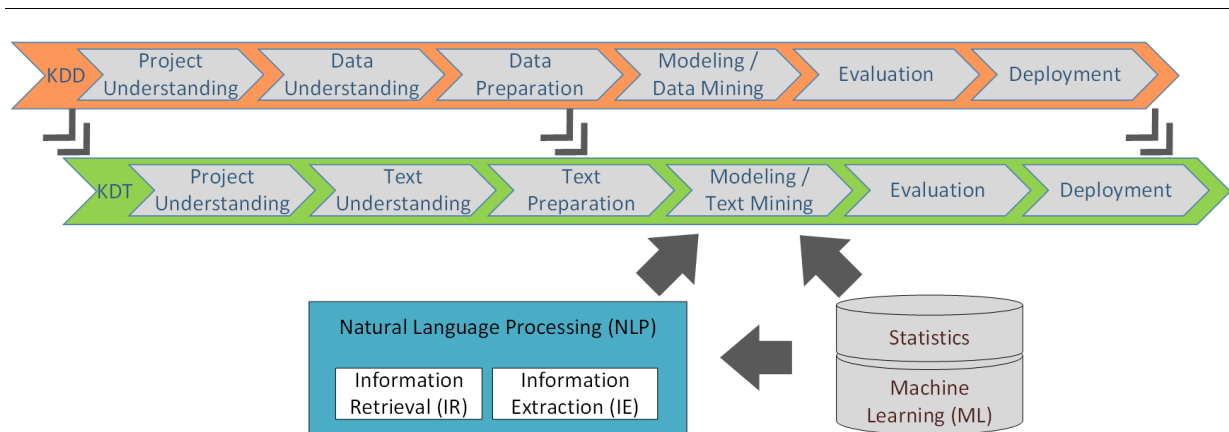


Abbildung 2.3: Abgrenzung von Text Mining zu bekannten Forschungsgebieten
(angelehnt an: Krukow 2013, S. 6 & Chapman et al. 2000)

Jiang schreibt, dass Information Extraction ein großer Bestandteil von Text Mining und Natural Language Processing ist. Er geht dabei vor allem auf Information Extraction ein, indem er sagt, dass Suchsysteme die extrahierten Daten nutzen können, um Relationen zwischen den identifizierten Konzepten zu entwickeln bzw. visualisieren (Jiang 2012). Weiter merkt er an, dass es allerdings Szenarien gibt, die nicht gut strukturiert sind, sodass Informationen vor allem aus großen Textquellen nicht leicht zu erhalten sind, und dass die Studien, die sich mit der Extraction von Konzepten aus unstrukturierten Texten befassen, alle in ihrem Erfolg limitiert sind (Jiang 2012, S. 14).

Laut Nenkova und McKeown wird Machine Learning häufig dazu verwendet, um dem Forscher / Nutzer zu zeigen, welche Teile eines Textes für den gewünschten Inhalt relevant sind. Allerdings sagen sie auch, dass die meisten Methoden des Machine Learnings auf Training angewiesen sind. Dies bedeutet, dass es Texte oder Textstellen geben muss, die von einem Menschen markiert wurden, sodass dieser prüfen kann, ob die Zuordnung durch die Machine Learning Methode korrekt durchgeführt wurde oder nicht (Nenkova und McKeown 2012, S. 63).

Um die spätere Netzwerkanalyse zu ermöglichen, muss der vorhandene Text zuerst untersucht werden. Aggarwal und Zhai schreiben, dass Text Mining dazu genutzt werden kann, um Einheiten wie Personen, Organisationen, Orte und deren Relationen zueinander aufzudecken und zu untersuchen. Gleichzeitig drücken sie aber aus, dass die Forschungsfortschritte im Gebiet der "Verarbeitung natürlicher Sprache" nicht weit genug sind, um "... akkurate semantische Repräsentationen von Texten ..." zu erstellen (Aggarwal und Zhai 2012, S. 3).

Ihrer Meinung nach sind viele Text Mining Ansätze daher noch auf simplere Darstellungen von Textdaten wie dem "bag of words" beschränkt. Unter "bag of words" versteht man eine Ansammlung aller in Textdateien vorkommenden Wörter. Dadurch gehen in einem ersten Schritt die Positionen der Wör-

ter, Sätze und deren Zusammenhänge verloren, allerdings ermöglicht dieser Ansatz auch weitere Möglichkeiten, wie das Erkennen von scheinbar relevanten Texteinheiten.

Diesner schreibt, dass es verschiedene Möglichkeiten gibt Textdaten und Netzwerke miteinander in Verbindung zu bringen. Es gibt einerseits Methoden, die existierende Netzwerke beschreiben und erweitern, und andererseits gibt es Fälle, innerhalb welcher Textdaten als einzige Grundlage zur Bildung von Netzwerken genutzt werden. Sie unterscheidet drei Fälle. Innerhalb eines ersten Falles werden Konzepte aus Texten extrahiert, um Schlüsseltermine aus dem zugrundeliegenden Text zu erhalten. Dies wird vor allem zur Nutzung von "one-mode" Netzwerken genutzt, mehr zu dieser Art von Netzwerk steht in Kapitel 2.2.2. Eine weitere Art ist es, mehrere Arten von Konzepten aus dem Netz zu extrahieren, um bspw. Verbindungen zwischen Personen und Organisationen zu erstellen und zu analysieren. Diese Art von Netzwerk wird ebenfalls in dem Kapitel der Netzwerkanalyse erläutert. Als letzte Art erläutert Diesner, dass es ebenfalls möglich ist, Dokumente selbst als Grundlage für Netzwerke genutzt werden können, um Personen zu bestimmten Dokumenten in Relation zu setzen (Diesner 2012).

Anhand des obigen Abschnittes ist zu erkennen, dass die Bereiche Text- und Netzwerkanalyse sehr weit ineinandergreifen und man sich bereits zum Zeitpunkt der Textanalyse über die Art der Netzwerkanalysen bewusst sein sollte. Dabei unterscheiden sich vor allem die Metaebenen der analysierten Gegenstände. Für die hier durchgeführte Textanalyse kommt vor allem der zweite Ansatz von Diesner in Frage, da innerhalb dieser Arbeit verschiedene Konzepte und deren Relationen untereinander betrachtet werden. Für die Textanalyse bedeutet dies, dass vor allem diejenigen Methoden herangezogen werden können, die verschiedene Wortgruppen voneinander unterscheiden können, wie bspw. "Part of Speech" oder "Thesauri".

Innerhalb dieser Arbeit wird sich vor allem auf den Teilbereich des Information Extraction beschränkt, da Konzepte erkannt und klassifiziert werden. Die eingebauten Methoden von Automap führen durch das Erstellen eines Metanetzwerkes diese identifizierten Konzepte zusammen und machen so die Relationen zwischen ihnen sichtbar.

2.2.1.1 Manuelle Vorbereitung

Bei der Bearbeitung von Textdaten ist oft ein Teil manueller Vorarbeit zu leisten. Dies ist einerseits abhängig von den gewünschten Analysen und deren Ergebnissen und andererseits von der Qualität der Daten. Die für diese Arbeit vorliegenden Daten sind Berichte / Akten, die zuerst eingescannt und daraufhin mit einem sogenannten OCR-Tool eingelesen wurden.

Diese Art von Tools ermöglicht die visuelle Erkennung von Schrift und liefert daher eine Textdatei, die man für weitere Schritte nutzen kann. Allerdings haben diese Tools ebenfalls einen Nachteil, da nicht immer alle Buchstaben und Zeichen richtig erkannt werden. Dies führt wiederum dazu, dass nachgeschaltete Schritte wie eine Textanalyse nur dann möglich sind, wenn die entstandenen Fehler in den relevanten Passagen der Texte behoben werden.

Innerhalb dieser Arbeit geht es vor allem um die Erkennung von Personen, Institutionen, Geldbeträgen und deren Zusammenhängen. Die Problematik, die sich daraus für diese Arbeit ergibt ist, dass einige Ziffern und Buchstaben nicht richtig erkannt wurden und so beispielsweise aus der Ziffer "5" der Buchstabe "S" oder aus der Ziffer "0" der Groß- oder Kleinbuchstabe "O" bzw. "o" wurde. Aus dem Buchstaben "l" (kleines L) wurde häufig ein "/". Laut Mori et al. ist die Qualität der Eingabedaten wichtig für den OCR Prozess. Er gibt viele Beispiele, wie Buchstaben ursprünglich aussehen und woraus sich die Probleme der Tools ergeben, wie bspw. alte Dokumente, deren Buchstaben nicht durchgängig gedruckt sind oder über die Zeit verblassen (Mori et al. 1999, S. 7–10). Da dem GLODERS Projektpartnern nur die Daten nach der OCR Anwendung zur Verfügung stehen, ist über die ursprüngliche Textqualität keine Aussage zu treffen. Allerdings ergibt sich für das Projekt und insbesondere für diese Arbeit die Notwendigkeit, die relevanten Textpassagen nach Möglichkeit zu korrigieren. Es gibt allerdings wenige Passagen, die ohne Zuhilfenahme der ursprünglichen physischen Dokumente bzw. der Scans dieser nicht möglich sind.

Ein weiterer Punkt ist, dass eine Übersetzung der Daten zusätzlich zu den existierenden Fehlerquellen des OCR weitere Fehler durch Fehlübersetzung und Verschiebung der Satzstruktur hinzugefügt hat. Zu Illustrationszwecken wird hier ein fiktives englisches Beispiel aufgezeigt. Angenommen der ursprüngliche Satz im Dokument hieß: "Bruce Wayne had a difficult childhood and is now Batman", so wurde in der Übersetzung daraus "Bruce hatte eine Wayne hart Kindheit und ist jetzt Batman"⁸.

Als Beispiel für eine Fehlübersetzung innerhalb der zugrundeliegenden Daten wurde aus dem niederländischen "betaling" ein nicht existierendes "Beta-Zellen". Für die Vorbereitung bedeutet dies, dass

⁸ Aus Gründen des Datenschutzes dürfen in dieser Arbeit keine Klarnamen aus den Textpassagen wiedergegeben werden. Das Beispiel basiert daher auf einer fiktiven Figur einer bekannten Comicbuchreihe. Es wurde aus Zwecken des allgemeineren Verständnisses ein englischer Beispieltext gewählt.

sowohl die OCR als auch Übersetzungsfehler identifiziert und wenn möglich korrigiert werden müssen. Von besonderer Bedeutung sind diejenigen Übersetzungsfehler, die die Namen von Personen oder Organisationen betreffen. Da in der Textdatei sowohl richtige als auch falsche Übersetzungen vorkommen, werden für Organisationen und Personen die korrekten Formen der niederländischen Datei zum Vergleich herangezogen. Dies ist, bei Organisationen besonders wichtig, da durch die Übersetzungsprogramme Abkürzungen der Gesellschaftsformen zum Teil ebenfalls auftreten. Da es in dieser Arbeit um Geldwäsche geht und ein Indikator für Geldwäsche ein Unternehmen mit vielen Tochtergesellschaften in unterschiedlichen Ländern sein kann, darf diese Art von Fehlern nicht in den Texten auftreten.

Da diese Vorbereitung notwendig ist, wird der Umfang der Vorarbeit erweitert, indem Namen, sowohl für Personen als auch Organisationen, Geldbeträge und Geldzeichen durch Unterstriche ("_") verbunden werden. Für die nachgehende Analyse hat dies vor allem im Bereich der Geldbeträge den Vorteil, dass selbst wenn zuerst, ein "bag of words" erstellt wird, die Einzelbeträge in ihrer jeweiligen Währung erhalten bleiben, was eine mögliche Analyse von Geldwäsche im internationalen Raum unterstützt.

Bei nicht klarer Nennung der Währungsart in den Originaldaten wird das Wort "Währung" vor den Betrag gesetzt, sodass sich ein Betrag von bspw. "Währung_100" ergibt. Dies hat vor allem für den Schritt der Ontologiebildung erhebliche Vorteile, da so zwischen Währungen, Terminen, Aktenzeichen und Kontonummern unterschieden werden kann. Für Währungen ergibt sich eine weitere Erschwerung, die mit folgenden Textbereinigungsmaßnahmen einhergeht. Die Währungen werden überwiegend in der jeweiligen Landesdarstellung angezeigt. Für USD ergibt sich so eine Bezeichnung mit Tausenderkommata und Nullpunkt wie bspw. "10,000.00 \$", für EUR hingegen wird die Schreibweise "10.000,00 €" verwendet. Um eine einheitliche Darstellung zu gewährleisten, werden daher alle Tausendertrenner entfernt und die Nullstelle mit einem Semikolon versehen, was zu dem Ergebnis "10000_00" für alle Währungen führt. Dies dient nicht nur der Einheitlichkeit, sondern ebenfalls der Anwendung durch späteres Entfernen der Satzzeichen, mehr dazu folgt in Kapitel 2.2.1.2.

Ebenfalls Namen und Pseudonyme, welche innerhalb der Akten verwendet werden, sind durch eine solche Zusammenführung in ihrer Eindeutigkeit gefördert. Das bedeutet, dass vor allem Personen, welche denselben Vor- oder Nachnamen tragen, in den folgenden Analyseschritten genauer voneinander abgrenzbar sind. Für die Visualisierung der Ergebnisse in dieser Arbeit ist dieser Schritt zwar nicht sichtbar, da alle Namen von Institutionen und Personen anonymisiert werden, allerdings für die Folgearbeit im Rahmen des GLODERS- oder eines Folgeprojektes können diese Schritte von Bedeutung sein.

Die Dauer der Vorbearbeitung wird vor allem durch die eben genannte Doppelung von Namen beeinträchtigt, da man die jeweils adressierte Person identifizieren muss.

Dies wird dadurch erschwert, dass Pseudonyme bereits in den zur Verfügung gestellten Dokumenten eingeführt wurden. Um dies zu veranschaulichen, wird das vorherige fiktive Beispiel wieder aufgegriffen.

Angenommen eine Person A heißt "Bruce Wayne" und eine Person B heißt "Martha Wayne". In den Akten wurde dann der Nachname Wayne durch ein Pseudonym ersetzt wie bspw. "P1". Dies führt dazu, dass Personen, die ebenfalls zufälligerweise den Nachnamen "Wayne" tragen, ebenfalls mit diesem Pseudonym versehen sind. Da jedoch nur das Wort "Wayne" durch das Pseudonym "P1" ersetzt wurde und nicht, durch oben genannte OCR-Fehler, Variationen des Wortes wie "Wagne" oder "Wapne", konnten einige Klarnamen auf Personen zurückgeführt werden. Da allerdings zu einem späteren Zeitpunkt der Analyse selbst definierte Pseudonyme zum Einsatz kommen, wird dieser Fehler in der Arbeit nicht weiter auffallen.

Daraus entsteht das Problem, dass im Rahmen der Vorarbeit sichergestellt werden muss, ob es sich um eine Person handelt, die evtl. wirklich "Wapne" heißt oder ob sich diese Nennung auf die ursprünglich gemeinten "Bruce" oder "Martha" bezieht.

Sind alle genannten Vorbearbeitungsschritte durchgeführt, bleibt als letztes die Eliminierung von Metadaten, die nicht zur Analyse der Inhalte notwendig sind. Hierzu werden die Daten vor dem Einlesen in eine Textdatei mit geeignetem Zeichenformat (UTF-8) überführt. Dieser Schritt eliminiert Metadaten von Microsoft Word Dokumenten. Andere Metadaten wie die Seitenzahlen der ursprünglichen Akten oder Referenzen auf weitere Akten, die dem GLODERS Projektteam nicht zur Verfügung stehen, werden ebenfalls entfernt, da diese in keinem für diese Arbeit relevanten Zusammenhang mit dem Inhalt des organisierten Verbrechens bzw. der Geldwäsche stehen. Dieser Schritt ergibt eine Auseinandersetzung mit dem Text im Sinne der Geldwäsche. Es sind Inhalte zu identifizieren, die zwar Polizeiarbeit und Anmerkungen von Ermittlern beinhalten, welche sich aber durchaus auf die Geldwäsche bzw. anderen Verbrechen beziehen. Von diesen Inhalten sind diejenigen Inhalte abzugrenzen, die sich auf die Polizeiarbeit mit anderen Akten und Einlesen in neue Inhalte beziehen, welche in dem Textdokument jedoch nicht weiter ersichtlich sind.

Sind all diese Schritte der manuellen Vorarbeit geleistet, wird die Textquelle in Automap geladen.

2.2.1.2 Textanalyse mit Automap

Ist eine Textdatei in Automap geladen, ergibt sich die Darstellung des Textes, wie in Abbildung 2.4 zu sehen. Automap bietet nun Textbereinigungsmaßnahmen für verschiedene Zwecke. Der Menüpunkt

"Edit" ermöglicht es dem Nutzer, existierende Tags anzuzeigen oder nach bestimmten Wörtern zu suchen. Ebenfalls können bekannte Begriffe aus existierenden Thesauri ausgeblendet werden, um sich auf "neuen" Inhalt zu fokussieren. Ebenfalls kann der Nutzer Einstellung zu Schriftart, -größe und programmspezifische Präferenzen durchführen.

Im Menüpunkt "CEMap" können Daten aus sozialen Netzwerken, wie Twitter oder Facebook, oder aus E-Mail Programmen und RSS-Feeds geladen werden. "Extractors" erlaubt es dem Nutzer, Daten von Internetseiten oder aus bestehenden HTML Dokumenten zu importieren und in eine reine Textform zu überführen.

2.2.1.2.1 Textvorbereitung

Da in dieser Arbeit lediglich eine Textdatei betrachtet wird, sind diese Funktionen jedoch nicht relevant. Der erste relevante Menüpunkt ist "Preprocess" innerhalb dieses Menüs kann der Nutzer Funktionen nutzen, die zur Säuberung des Textes genutzt werden können.

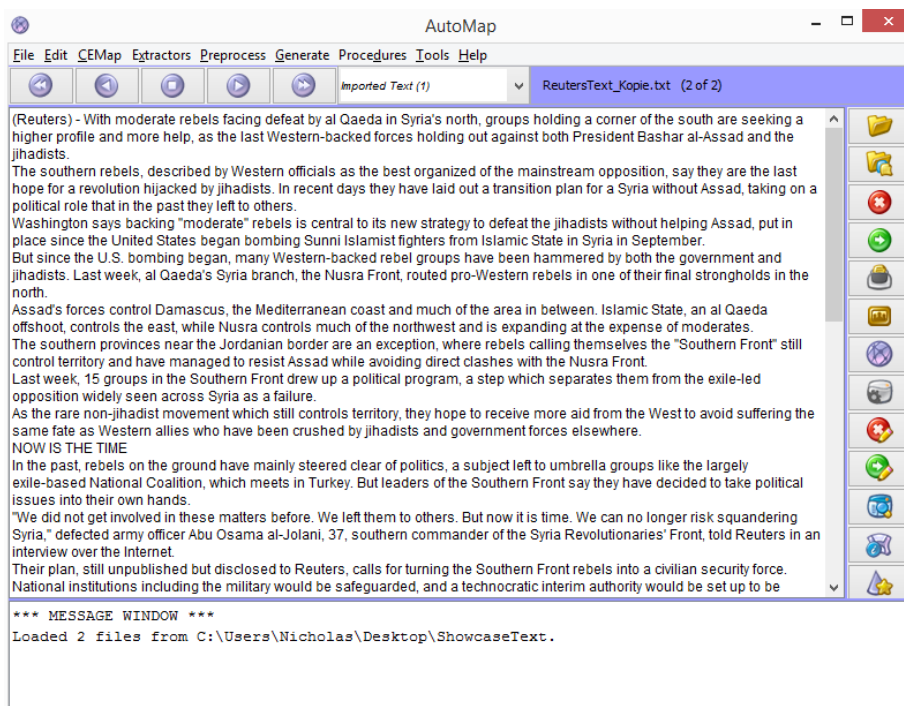


Abbildung 2.4: Beispielanzeige Automap⁹

⁹ Der gezeigte Text wurde von Reuters.com zur Zwecken der Demonstration entnommen (<http://www.reuters.com/article/2014/11/13/us-mideast-crisis-syria-rebels-idUSKCN0IX11I20141113> Letzter Zugriff: 13.11.2014)

Das Textcleaning erlaubt es dem Nutzer, überflüssige Leerzeichen zu entfernen, britisches Englisch in amerikanisches Englisch zu überführen, gängige Rechtschreibfehler zu korrigieren, sowie gängige Abkürzungen auszuschreiben und Pronomen aufzulösen.

Ebenfalls können Einzelbuchstaben, Zahlen und Tages bzw. Monatsangaben entfernt werden. Da diese Funktionen überwiegend auf englische Texte ausgelegt sind und keine Funktion zur Einbettung eines deutschen Wörterbuchs ersichtlich ist, ergibt sich für diese Arbeit, dass lediglich folgende Funktionen genutzt werden:

1. Die Entfernung von Satzzeichen

Satzzeichen werden entfernt, da Wörter, die am Ende eines Satzes oder vor einem Komma stehen, berücksichtigt werden. Dies ist vor allem für die Erstellung einer Konzeptliste relevant. Nutzt ein Benutzer diese Funktion kann er entscheiden, ob er die Satzzeichen mit einem Leerzeichen ersetzen möchte oder nicht. Empfehlenswert ist es, ein Leerzeichen anstelle der Satzzeichen zu setzen, da so vermieden werden kann, dass Wörter, die nur durch ein Satzzeichen voneinander getrennt sind, zusammengefasst werden.

2. Entfernung von Nutzer Symbolen

Ein Nutzer kann bestimmen, welche Symbole er behalten möchte, wie beispielsweise Währungszeichen, indem er eine Liste mit Symbolen erstellt, die er löschen möchte. In diesem Falle werden nur diese Symbole entfernt.

3. Vereinheitlichung von Groß- und Kleinbuchstaben

Alle Wörter einer Datei sollten entweder in Groß- oder Kleinbuchstaben an weitere Analysemethoden oder -tools übergeben werden. Dies führt dazu, dass, sobald ein bag of words erstellt wird, alle Varianten eines Wortes zusammengefasst werden, egal ob sie am Anfang eines Satzes stehen oder in der Mitte.

4. Die Entfernung von überflüssigen Leerzeichen

Mit dieser Funktion entfernt das Programm überflüssige Leerzeichen. Dieser Schritt sollte erst nach dem Entfernen der Satzzeichen genutzt werden, um zu gewährleisten, dass überflüssige Leerzeichen, die durch Punkt 1 entstehen können, wieder gelöscht werden.

5. Anwenden einer Delete List

Eine Delete List in Automap ist eine Datei, welche Wörter beinhaltet, die das Programm nicht für weitere Methoden berücksichtigen soll. Für englischsprachige Texte existieren mehrere vordefinierte Listen, welche vor allem Stoppwörter¹⁰ beinhalten. Stoppwörter treten in ihrer Sprache be-

¹⁰ Für die zugrundeliegende Arbeit wurde eine Stoppwortliste von <http://www.ehtio.de/deutsche-stoppswortliste-2152/> verwendet (Letzer Zugriff: 14.11.2014).

sonders häufig auf und sind deshalb für die meisten Analysen nicht aussagekräftig. Die Entfernung von solchen Stoppwörtern führt ebenfalls in einigen Fällen zu einer Aufwandsreduktion nachgelagerter Methoden. Eine Delete List kann in Automap erstellt werden oder in einem externen Texteditor. Die Ersteller von Automap empfehlen, zwischen einer universellen Delete List und einer themenspezifischen Delete List zu unterscheiden. Erste sollte Stoppwörter enthalten, da sie im Prinzip für alle Texte einer Sprache gelten und Zweite kann mit Wörtern versehen werden, die für einen bestimmten Anwendungsfall keine Relevanz haben.

Der Nutzer kann beim Anwenden einer Delete List zwischen "direct" und "rhetorical" wählen. Die erste Variante löscht alle Wörter der Delete List aus dem Text, und die zweite Variante ersetzt die Wörter mit "xxx". Diese Ersetzung hat den Vorteil gegenüber der ersten Variante, dass die Entfernung der Wörter zueinander in diesem Falle nicht verändert wird und die Struktur des Originaltextes erhalten bleibt.

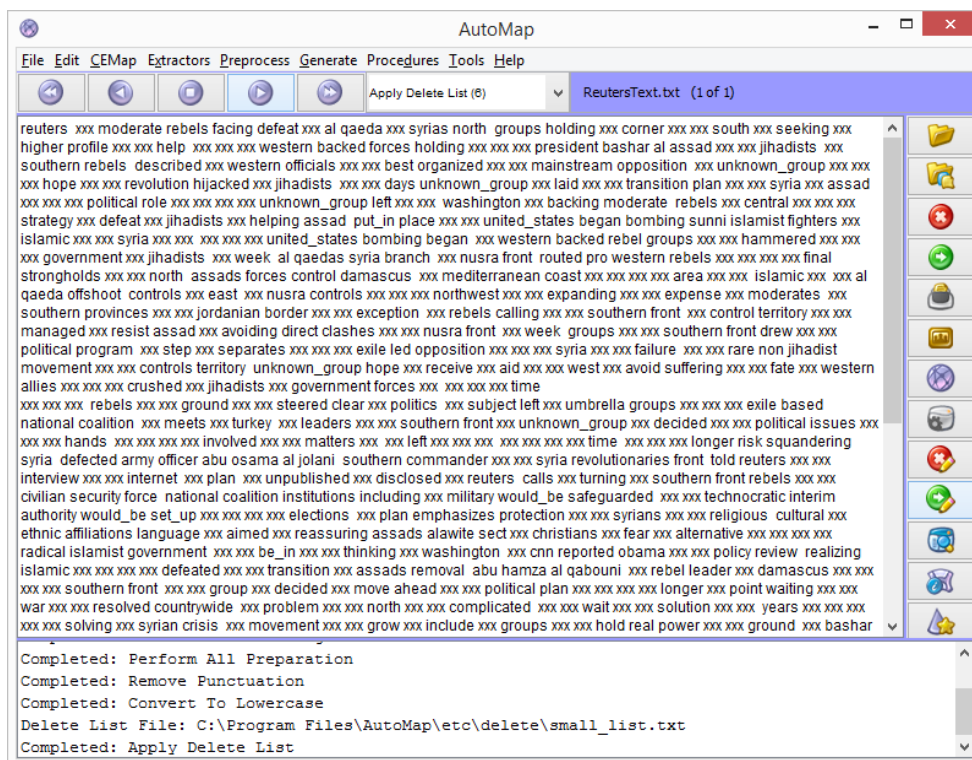


Abbildung 2.5: Beispielanzeige Automap nach Vorbearbeitung

In Abbildung 2.5 erhält man einen Überblick, wie eine Datei aussieht, wenn alle oben beschriebenen Maßnahmen angewandt wurden. An diesem Punkt unterscheidet sich der weitere Verlauf der Analyse je nach Art der Forschung. Wurde schon vorher mit Automap gearbeitet und Thesauri im betrachteten Themenbereich erstellt, vereinfacht dies die folgende Arbeit. Im Falle dieser Forschungsarbeit ist es die erste Arbeit, die innerhalb von Automap stattfindet.

2.2.1.2.2 Weitere Beschreibungen

Um sich an diesem Punkt einen ersten Überblick über die Daten zu verschaffen, eignen sich die Methoden des Menüpunktes "Generate". Den ersten Überblick über die Datei erhält man mit der Funktion "Text Properties". Diese Methode erstellt eine csv Datei mit generellen Angaben über den Text.

In Tabelle 2-2 sieht man, dass man so einen Überblick über die Anzahl an Wörtern und die Größe des Textes erhält.

Tabelle 2-2: Angaben zur Textdatei (Gloders Quelldatei)

Document	...Dateipfad\GLOD_DE.txt	URI
Number of Characters	1145078	Number
Number of Clauses	24463	Number
Number of Sentences	13865	Number
Number of Words	152334	Number
Average Words per Sentence	10,99	Number

Der nächste Schritt ist die Erstellung einer Konzeptliste. Diese erstellt aus jedem Wort, welches in der Datei gefunden wird, ein Konzept und zugehörige Angaben. Aus Datenschutzgründen wird auch hier zu Illustrationszwecken eine Konzeptliste gezeigt, die aus dem Beispieltext von Reuters in Abbildung 2.4 erstellt wurde.

	A	B	C	D	E	F	G	H
1	concept	frequency	relative_frequency-within_text	gram_type	number_of_texts	Evaluation	Potency	Action
2	rebels	15	1.0	single	1			
3	southern	13	0.8666667	single	1			
4	front	9	0.6	single	1			
5	al	8	0.53333336	single	1			
6	nusra	8	0.53333336	single	1			
7	south	8	0.53333336	single	1			
8	groups	7	0.46666667	single	1			
9	syria	7	0.46666667	single	1			
10	western	7	0.46666667	single	1			
11	jihadists	6	0.4	single	1			
12	north	6	0.4	single	1			
13	political	6	0.4	single	1			
14	unknown_group	5	0.33333334	ngram	1			
15	army	5	0.33333334	single	1			
16	assad	5	0.33333334	single	1			
17	forces	5	0.33333334	single	1			
18	ground	5	0.33333334	single	1			
19	syrian	5	0.33333334	single	1			
20	assads	4	0.26666668	single	1	0.74	2.1	1.0
21	coalition	4	0.26666668	single	1			

Abbildung 2.6: Auszug einer Konzeptliste des Reuters-Beispiels

Die Konzeptliste in Automap beinhaltet, wie in Abbildung 2.6 zu sehen ist, 8 Attribute. Die letzten drei Attribute "Potency", "Action" und "Evaluation" sind, wie zu sehen ist, nicht bei allen Konzepten zu er-

kennen. In der Dokumentation des Programmes sind diese Attribute nicht zu finden und werden daher hier auch nicht weiter erläutert.

Die erste Spalte einer Konzeptliste enthält die Konzepte. Ein Konzept kann sowohl ein einzelnes Wort oder eine Folge an Wörtern sein. Eine Folge an zusammengehörigen Wörtern wird n-gram genannt und wird meist durch einen Unterstrich miteinander verbunden. "Frequency" gibt die Häufigkeit eines gefunden Wortes an und "relative Frequency" berechnet die Häufigkeit der Konzepte untereinander. Die Angaben werden jedoch eher für Analysen benötigt, in denen mehrere Dokumente betrachtet werden. In der vorliegenden Forschungsarbeit ist vor allem die erste Spalte relevant, denn diese wurde dazu genutzt, Thesaurus Dateien zu erstellen.

2.2.1.2.3 Thesaurus Dateien

In Automap werden sogenannte Thesaurus¹¹ Dateien zur Bildung von Netzen, als DeleteList¹² oder zur Generalisierung verwendet. Wörter werden je nach Anwendungsfall listenartig aufgeführt. Zur Erstellung von Thesauri können entweder Text- oder Tabellenverarbeitungsprogramme genutzt werden oder der in Automap integrierte Thesaurus Editor. Der Thesaurus Editor benötigt allerdings eine geladene .csv Datei, um zu arbeiten, allerdings ermöglicht er auch einen schnellen Vergleich mehrerer Thesauri miteinander.

¹¹ Ein Thesaurus wird im Sinne des Dudens als eine "Sammlung von Wörtern eines bestimmten [Fach]bereichs" verstanden (<http://www.duden.de/rechtschreibung/Thesaurus> Letzter Zugriff: 15.11.2014).

¹² Eine DeleteList bezeichnet in Automap eine Liste von Wörtern, welche vor der Analyse aus dem Text entfernt werden. Das bekannteste Beispiel einer DeleteList ist eine Stoppwortliste.

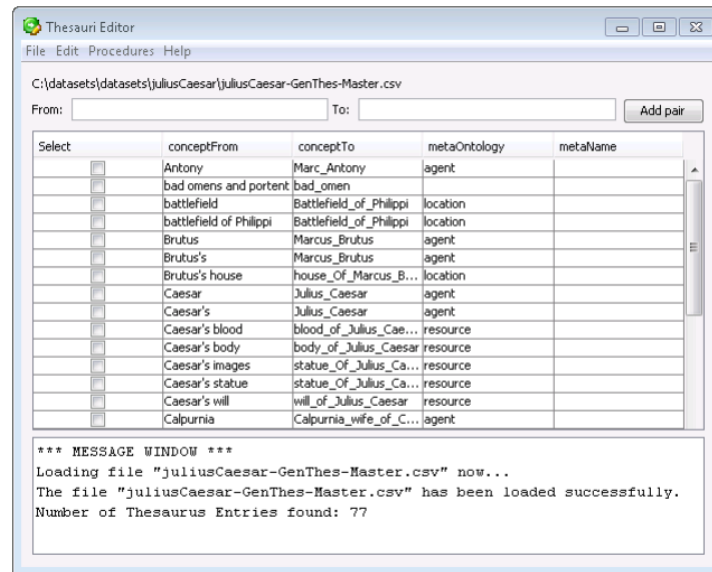


Abbildung 2.7: Ansicht des Thesaurus Editors
(entnommen aus: Carley et al. 2013a, S. 140)

Es folgt eine Unterscheidung der verschiedenen Thesaurusarten in Automap. Es wird unterschieden nach Standard Format und Master Format.

Das Standardformat enthält zwei Spalten. Die erste Spalte enthält die Textangaben, sowie sie innerhalb der Quelldatei zu finden sind, und die zweite Spalte wird genutzt, um einheitliche Konzepte zu erschaffen. Wird das oben erwähnte Illustrationsmaterial (Reuters) herangezogen, so enthält bspw. die erste Spalte den Term "Assad" und die zweite Spalte den "Baschar_al-Assad". Diese Datei wird als .csv Datei gespeichert. Diese Dateien sind Textdateien, innerhalb derer verschiedene Spalten durch Kommata, Semikola oder Zeilenumbrüche voneinander getrennt werden.

Das Masterformat erweitert die Struktur des Standardformats, indem es zwei neue Spalten einführt. Ebenfalls werden im Masterformat nun zur Unterscheidung der Spalten Überschriften herangezogen.

Wie in Abbildung 2.7 zu sehen ist, besteht das Masterformat aus den Spalten "conceptFrom", "conceptTo", "metaOntology" und "metaName". In der Spalte "metaOntology" werden die Konzepte zu verschiedenen in Automap und ORA bekannten Konzeptklassen zugeteilt. Mehr zu den Konzeptklassen in Kapitel 2.2.1.2.4. Innerhalb der Spalte "metaName" können die Klassen weiter spezifiziert werden. Diese Form der Spezialisierung wird in dieser Forschungsarbeit angewandt, um während und nach der Netzwerkanalyse die Ontologie einzuführen.

Alle Einträge in der ersten Spalte der Thesauri müssen einzigartig sein, in dem Sinne, dass jede Mehrfachnennung von Konzepten in der Klassifizierung des erstgenannten resultiert. Die folgende Tabelle enthält eine solche Mehrfachnennung.

Tabelle 2-3: Mehrfachnennung innerhalb eines Thesaurus

der	der_Baum
deR	der_Hund
dEr	der_Platz
DE	Deutschland
Germany	Deutschland

Alle Nennungen von "der" aus der Tabelle resultieren in dem Konzept "der_Baum". Dabei ist die Groß- oder Kleinschreibung der Konzepte der ersten Spalte egal. Die zweite Spalte kann allerdings denselben Begriff mehrfach enthalten, sodass man unterschiedliche Varianten von einem Wort zum selben Konzept überführen kann. So zu sehen an dem Beispiel, dass sowohl die Abkürzung "DE" und das Wort "Germany" im Konzept Deutschland resultieren werden. Für den Inhalt der Konzeptliste ist der Benutzer verantwortlich. Unterschiedliche Bedeutungen von Konzepten sollten daher geprüft werden. Im Beispiel der Tabelle 2-3 muss der Nutzer gewährleisten, dass alle Formen von "DE" im Text sich auf Deutschland beziehen und nicht etwa für das französische Wort "von" stehen.

Bei der Anwendung eines Thesaurus kann der Nutzer entscheiden, ob nur Wörter beachtet werden sollen, die innerhalb des Thesaurus enthalten sind. Dies hat den Vorteil, dass ein Thesaurus, der alle für eine Forschungsarbeit relevanten Einträge beinhaltet, angewandt werden kann, ohne Aufwand in den restlichen Text zu investieren (Carley et al. 2013a, S. 150). Dieser Vorteil kann allerdings nur dann existieren, wenn gewährleistet werden kann, dass "alle" relevanten Einträge innerhalb eines Thesaurus enthalten sind oder in Situationen, die darauf abzielen, nur bestimmte im Thesaurus abgebildete Sachverhalte zu analysieren.

2.2.1.2.4 Konzeptklassen in ORA und Automap

Innerhalb des Nutzerhandbuchs 2013 von Automap werden folgende zwölf Meta-Netzwerk-Kategorien unterstützt: "action", "agent", "attribute", "event", "group", "knowledge", "location", "organization", "resource", "role", "task" und "when" (Carley et al. 2013a, S. 134–135).

- *Action* bezeichnet dabei alle Handlungen, die als Verben im Text aufgeführt werden.
- *Agent* kann jede Person, Institution oder Gruppe sein, die in der Lage ist Informationen aufzunehmen und zu verarbeiten.
- *Attribute* werden Agenten zugeteilt. Sie beschreiben Eigenschaften, die die meisten Agenten gemein haben, wenn auch die Ausprägung anders sein kann. Beispiele dafür sind Augenfarben oder innere Eigenschaften wie die religiöse Einstellung (Carley et al. 2013a, S. 134–135).

- Events sind Ereignisse, die mehrere Agenten gemeinsam ausführen. Dies können Meetings, Feiertage oder sonstige Veranstaltungen sein.
- Knowledge wird laut Carley et al. als Wissensangaben, die durch Studium oder durch Erfahrung erlernt worden sind, verstanden.
- Location bezeichnet jedwede Angabe von einem physischen Raum. Es können Planeten, Kontinente, Länder, Städte oder spezifische Adressangaben in dieser Gruppe eingeordnet werden.
- Organization beinhaltet Netze von Agenten, die auf ein gemeinsames Ziel hinarbeiten. Dies können spezifische Unternehmen oder auch Vereine oder sonstige Gruppierungen sein (Carley et al. 2013a, S. 134–135).
- Als Resource lässt sich jedes physische oder nicht-physische Gut betrachten, welches genutzt werden kann, um bestimmte Aufträge zu erledigen. Es können sowohl nicht materielle Dinge wie Zahlungsströme oder Guthaben auf einem Bankkonto gemeint sein, wie auch ein Auto, Boot oder sonstiges.
- Eine Rolle wird durch einen Agenten bekleidet. Dies können spezielle Berufsangaben sein, wie "Unternehmensberater" oder unklarere Angaben wie "Strohmann" oder "Mittelsmann".
- Carley et al. definieren einen task als einen "Teil von Aktionen zur Erfüllung eines Jobs, Problems oder einer Aufgabe". Dabei merken sie an, dass der Ausdruck "task" als Synonym für "Aktivitäten" herangezogen werden kann, wobei Aktivitäten auf einen längeren Zeitraum ausgelegt sind (Carley et al. 2013a, S. 134–135).
- When bezeichnet zeitliche Nennungen, hier können Monats, Jahres oder datumsspezifische Angaben eingeordnet werden.

Für die hier zugrundeliegende Forschungsarbeit werden die oben genannten Klassen als erste Kategorisierung für die Ontologie verwendet. Dabei unterscheidet sich jedoch das Verständnis der Klassen gegenüber den eben genannten Definitionen.

- Als Agenten werden nur natürliche Personen verstanden, und Attribute werden verstanden als jede Eigenschaft einer Person oder eines Gegenstands.
- Alle Unternehmen oder Unternehmensgruppen werden in einem ersten Schritt in der Klasse "organizations" aufgeführt. Einzelne Abteilungen eines Unternehmens hingegen werden als Gruppe eingeordnet, genau wie Familien- oder Freundeskreise, die spezifisch genannt werden.
- Als "resource" werden alle wertschöpfenden Mittel klassifiziert, dies beinhaltet Geldströme, Objekte sowie Konten. Dies ist besonders relevant im Hinblick auf Adressangaben im Text, diese werden sowohl als location, als auch als resource definiert.
- Als Task werden in dieser Arbeit alle kurz- oder längerfristigen Aufgaben aufgefasst, welche von Personen und / oder Organisationen durchgeführt werden.

- Da "group" im Nutzerhandbuch nicht näher definiert wurde und es gegenüber von Organisationen abzugrenzen ist, wird eine Gruppe als Netz aus Agenten angesehen, die allerdings im Gegensatz zu Organisationen kein gemeinsames Ziel verfolgen müssen.
- Da die Gruppe when in ORA nicht existiert, werden alle zeitlichen Angaben als Event klassifiziert, dies beinhaltet Datumsangaben und Termine wie Meetings oder Ähnliches.

2.2.1.2.5 Das DyNetML Format

"DyNetML¹³ ist eine auf XML basierende Sprache zum Austausch von relationalen und Netzwerkdaten. Es beinhaltet Metadaten, Knoten, Verbindungen und deren Attribute. Es wurde als Verbindung von Analyse und Visualisierungs-Programmen entwickelt und soll reichhaltige Daten von sozialen Netzwerken darstellen und mit Prozessangaben oder anthropologischen Daten erweitern" (CASOS).

Da es von Automap genutzt wird um die Netzwerke für ORA zu erstellen wird hier eine kurze Erläuterung geliefert. Das Format kann somit als Schnittstelle der Text- und Netzwerkanalyse angesehen werden.

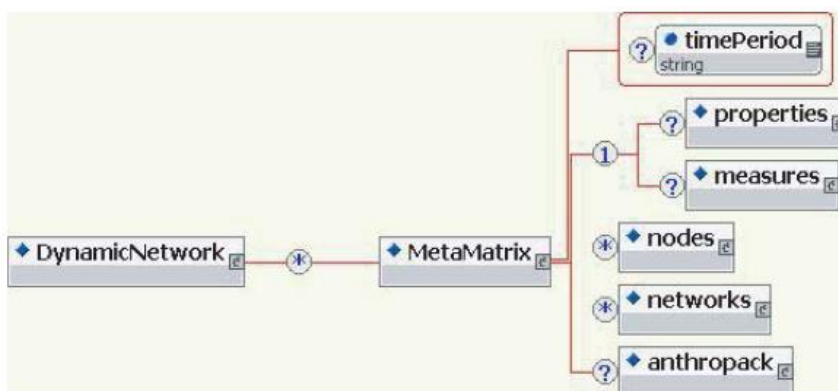


Abbildung 2.8: Dynamische Netzwerke in DyNetML
(Aus: Tsvetovat et al. 2004)

In ihrer Arbeit beschreiben Tsvetovat et al. das DyNetML Format als Möglichkeit, innerhalb eines Graphen mehrere Knotenpunkte unabhängig von ihrer Klasse miteinander zu verbinden. Sie unterscheiden dabei nach den folgenden Knotenklassen: "agent", "organization", "knowledge", "resource", "task", "location" und "graph" (Tsvetovat et al. 2004).

13 Die Spezifikation von DyNetML kann unter: http://www.casos.cs.cmu.edu/dynetml/dynetml_2_0-schema.xml (Letzte Prüfung: 27.11.2014) aufgerufen werden.

Sage und Rouse beschreiben, dass das Format, Daten als eine Momentaufnahme eines Netzwerkes in DyNetML erfasst werden. Für jede Momentaufnahme entsteht ein eigenes Metanetzwerk (Sage und Rouse 2009). Auf diese Weise ist es den Nutzern möglich, Änderungen innerhalb eines Netzes über einen beliebig großen Zeitraum zu erfassen und nachzuvollziehen.

2.2.2 Netzwerkanalyse

Innerhalb dieses Abschnittes werden zuerst die theoretischen Grundlagen zu Netzwerkanalysen geliefert, darauf aufbauen die explizite Vorgehensweise in dieser Forschungsarbeit und letztendlich die Vorgehensweise unter Berücksichtigung des hier verwendeten Tools.

2.2.2.1 Theoretische Grundlagen

Der zentrale Punkt dieser Arbeit ist die Netzwerkanalyse, welche die Beziehungen zwischen Personen, Organisationen und weiteren Klassen beschreiben soll.

Wassermann und Faust haben in ihrem Werk "*Social Network Analysis*" Methoden und Grundlagen zur Analyse von Netzen gegeben. Für die Netzwerkanalyse gibt es verschiedene grundlegende Konzepte, welche Wassermann & Faust vorstellen. Die Kernkonzepte sind Akteure, Verbindungen, Dyade, Triaden, Gruppen, Beziehungen und Netzen.

Dabei werden Akteure als *soziale Einheiten* in einem größeren System angesehen, wie beispielsweise ein Institut in einer Lehreinrichtung. Sollten alle Akteure denselben Typ besitzen, spricht man von "one-mode networks" (Wasserman und Faust 1994, S. 17).

Dyaden nennt sich eine Untergruppe des Netzes, bestehend aus zwei Akteuren. Diese Einheit wird dann untersucht, wenn man feststellen möchte, ob reziproke oder einseitige Verbindungen existieren.

Die Erweiterung um einen dritten Akteur ermöglicht weitere Analysen, die das Verhältnis von Akteur 1 zu Akteur 3 analysieren, wenn es bereits Aussagen zu dem Verhältnis von Akteur 1 zu Akteur 2 und von Akteur 1 zu Akteur 3 gibt. In größeren Netzen können so beispielsweise indirekte Verbindungen zwischen mehreren Akteuren existieren. Dyaden und Triaden lassen sich um weitere Akteure erweitern und werden so zu Gruppen bzw. Untergruppen. In ihrem Werk definieren Wassermann und Faust Gruppen als "... Sammlung aller Akteure, in welcher Verbindungen gemessen werden können" (Wasserman und Faust 1994, S. 19).

An dieser Stelle wird die *Soziale Netzwerkanalyse* von der hier durchgeführten *semantischen Netzwerkanalyse* abgegrenzt.

Reichenberg beschreibt, dass *semantische Netze* jede beliebige Anzahl an Objekten und Eigenschaften besitzen können. Objekte werden untereinander verbunden und so entstehen inhaltliche Zusammen-

hänge. Ebenso wird erwähnt, dass ein Objekt mehrere Namen, bzw. Beschreibungen besitzen kann (Reichenberger 2010). So können mehrere Varianten desselben Objekts in einem Programm bzw. in einem Netz auf denselben Ursprungsknoten deuten. Ein semantisches Netz kann einem Nutzer weiterführende Informationen liefern als ein reines soziales Netzwerk. Durch die Verknüpfung von mehreren Objektklassen zueinander und der Nähe zueinander lassen sich Konzepte, die relevant sind, einfach finden. Reichenberg gibt dies an folgendem Beispiel wieder:

"Es ist wie im Kaufhaus: Wissen Sie, wie die kleinen Dinger aus Messing heißen, mit denen man dicke Briefe und kleine Päckchen verschließt (...) Sie heißen Musterbeutelklammern (...) Auch wenn Sie das nicht wissen (ich wusste es nicht), werden Sie trotzdem keine Schwierigkeiten haben, diesen Artikel in einem größeren Kaufhaus zu beschaffen." (Reichenberger 2010, S. 12–13)

Er sagt weiter, dass es bei semantischen Netzen ähnlich funktioniert. Wenn ein Nutzer weiß, um welche Klasse es sich bei dem gesuchten Objekt handeln könnte, kann er dieses oft schnell finden.

Lietz zeigt in einem kurzen Paper, wie semantische Netzwerkanalyse anhand der europäischen Verfassung genutzt werden kann. Er schreibt, dass "Nichttriviale semantische Aussagen (...) aus dem Zusammenspiel von Wörtern (entstehen). Sätze und Texte können also als Wortnetzwerke betrachtet werden" (Lietz 2007).

Weiter sagt er, dass die Kernaussagen der betrachteten Texte durch das Zusammenspiel aller Wörter und der Visualisierung der Wortbeziehungen entsteht. So sagt er, dass die Aussage des Satzes "Das Wetter ist schlecht", erst durch die Verbindung der Wörter "Wetter" und "schlecht" ersichtlich wird (Lietz 2007, S. 2).

Bei der Analyse von semantischen Netzwerken ist es also möglich, eine Beziehung zwischen bestimmten Konzepten zu identifizieren und zu interpretieren.

Diesner stellt fest, dass innerhalb von Konzeptnetzwerken die Verbindungen zwischen Personen, Organisationen, Events und weiteren Einheiten analysiert werden können. Sie schlägt vor, für Untersuchungen an Organisations- und Personennetzwerken eine Meta-Matrix einzuführen, welche einen höheren Level an Abstraktion bietet. Man führt somit eine Ontologie ein, welche ausdrücken, wie die Gruppen der Konzepte zueinander stehen (Diesner und Carley 2008, S. 1).

Innerhalb der Analysen stehen die Knoten entweder für Personen oder Organisationen, und die Kanten stehen für die Verbindung zwischen den Einheiten. Carley, Diesner und Tambayong geben an, dass Daten die zur Konstruktion dieser Netzwerke genutzt werden können, unter anderem folgende sind: Mitschriften von Gerichtsverfahren, Kommunikationsdaten (E-Mails etc.) und Daten aus online Quellen (Diesner et al. 2012, S. 2).

Für diese Arbeit wird der Unterschied zwischen sozialer und semantischer Netzwerkanalyse vor allem durch die Analyse von verschiedenen Knotenklassen und deren Beziehungen zueinander verstanden. Dabei gibt es in diesem Analysefall verschiedenste Arten an Verbindungen, von "einfacher" Bekanntheit über "Zahlungen von ... an ..." und weiteren Formen.

2.2.2.2 Explizites Analysekonzept

Die Netzwerkanalyse dieser Arbeit ist in Abbildung 2.9 dargestellt. Es sollen sowohl der Aufbau des gesamten Netzwerkes, als auch die Beziehungen innerhalb des Netzes analysiert werden. Die Strukturanalyse soll zeigen, um welche Art von Netzwerk es sich handelt.

Inhaltlich beginnt die Analyse mit der Identifikation von Schlüsselkonzepten, in diesem Schritt sollen die bedeutsamsten Konzepte einer Klasse im Netzwerk identifiziert und weiter analysiert werden.



Abbildung 2.9: Vorgehen der Netzwerkanalyse

Sind die wichtigsten Konzepte erkannt, werden deren Egonetzwerke betrachtet. Auf diese Weise werden die Anknüpfstellen der Schlüsselkonzepte identifiziert. Eine solche Betrachtung soll zeigen, in welchen Kreisen sich die Schlüsselkonzepte bewegen. Auf Personen bezogen, könnten interessante Fragestellungen folgendermaßen aussehen:

- Mit welchen anderen Personen hat Person 1 zu tun?
- Welche Beträge erbringt Person 1?
- Zu welchen Organisationen hat Person 1 Beziehungen?

Der letzte Schritt ist eine Cliques- bzw. Gruppenanalyse. Diese soll zeigen, welche Konzepte inhaltlich zusammengehören, also welche Unternetzwerke innerhalb des Metanetzes noch enthalten sind. Wie genau diese Punkte erreicht werden, ist in Kapitel 2.2.2.4 erklärt. Dort werden die verschiedenen Analysemethoden von ORA erläutert, welche genutzt werden, um die in Abbildung 2.9 beschriebenen Aufgaben zu erfüllen.

2.2.2.3 Die Erstellung des Netzwerkes

In Kapitel 2.2.1.2.5 wurde eine grobe Einführung in das DyNetML Format gegeben, welches Automap bei der Erstellung von Metanetzwerken nutzt. Bevor die eigentliche Erstellung des Netzwerkes beginnt, ist es ratsam, die Prüfmechanismen von ORA und Automap zu nutzen. Diese Mechanismen können Thesaurusdateien analysieren und dem Nutzer Doppelzuweisungen, Zyklen innerhalb des Thesaurus und weitere Fehlerquellen aufzeigen.

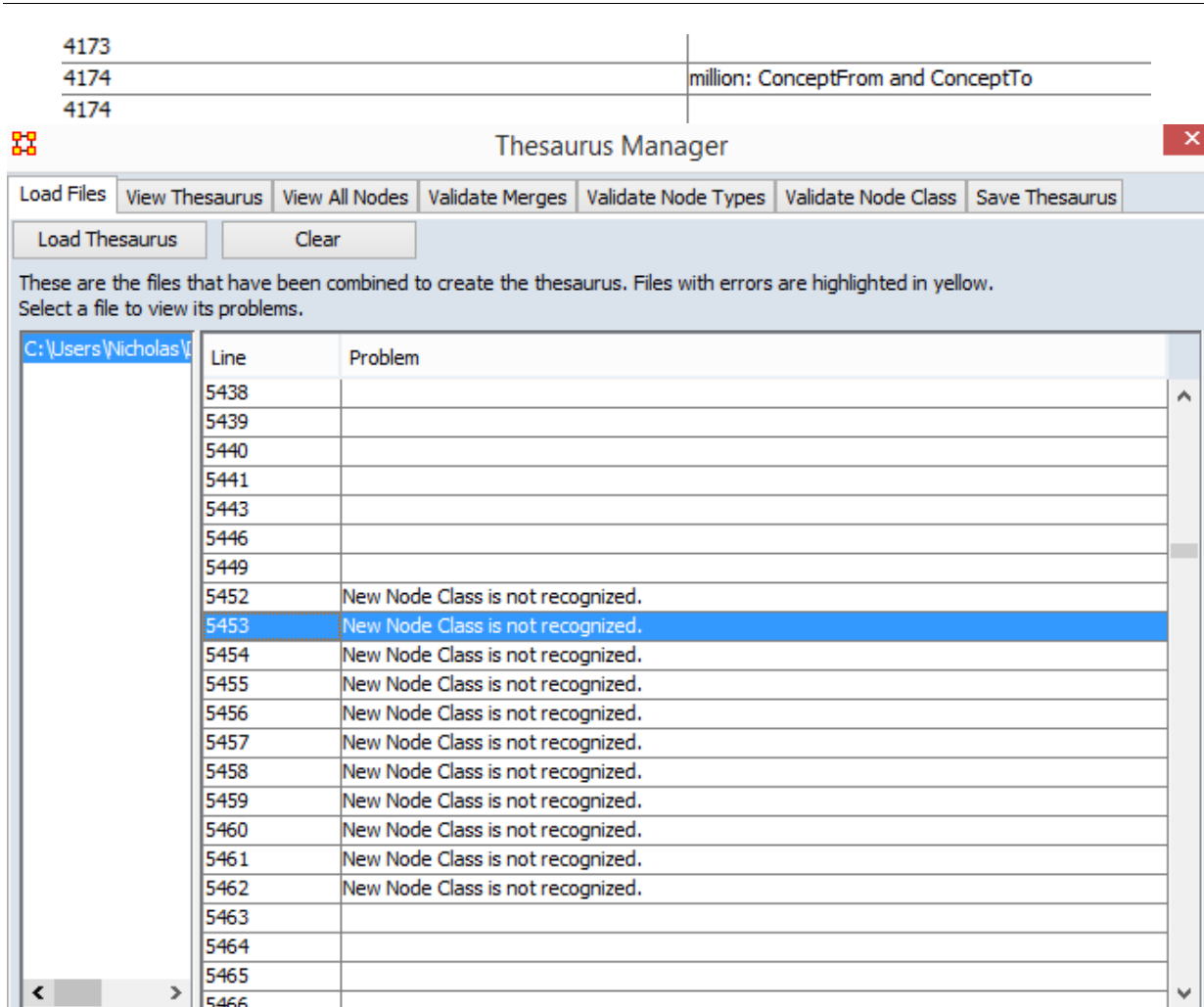


Abbildung 2.10: Thesaurusprüfung in ORA

In ORA gibt es diesen Prüfungsmechanismus als Thesaurus Manager. Dieser ermöglicht es dem Nutzer, die genutzten Thesauri auf Richtigkeit zu prüfen. In Abbildung 2.10 werden zwei Beispiele für mögliche Fehler innerhalb einer Thesaurusdatei erläutert. Der obere Teil des Bildes zeigt die logische Verknüpfung von ConceptTo und ConceptFrom. Der abgebildete Fehler besagt, dass eine Zeile im Thesaurus ein Konzept auf "million" zuordnet und eine andere Zeile das Konzept "million" auf ein weiteres Konzept ordnet. In diesem Fall wird "million" auf "millionen" verwiesen. Das führt dazu, dass Zeile 4174

auf "million" und dann in einer späteren Spalte auf "millionen" zugeordnet wird. Was wiederum den Schluss zulässt, dass der Nutzer Zeile 4174 direkt auf "millionen" zuweisen kann bzw. sollte.

Der zweite Teil des Bildes zeigt, dass ORA nicht alle Konzeptklassen erkennt, die im Userguide von Automap geschildert werden. Dabei kann es sich entweder um Schreibfehler handeln, so wurde in dieser Arbeit für mehrere Zeilen "tas" anstatt "task" verwendet, worauf der Thesaurus Manager hinwies. Anders bei den Klassen "Action", "Attribute" und "Group", die im Nutzerhandbuch aufgelistet werden (Carley et al. 2013a). Diese Klassen werden zwar von Automap als unterstützt angegeben, ORA kennt diese Klassen allerdings nicht direkt. Eine Möglichkeit, diese Zuweisung dennoch hinzubekommen, wird in Kapitel 2.2.2.3.1 gezeigt.

Nicht nur ORA unterstützt Methoden zur Thesaurus Prüfung, sondern auch Automap. In Automap kann der Menüpunkt "Thesaurus Procedures" mehrere Prüfmöglichkeiten durchführen.

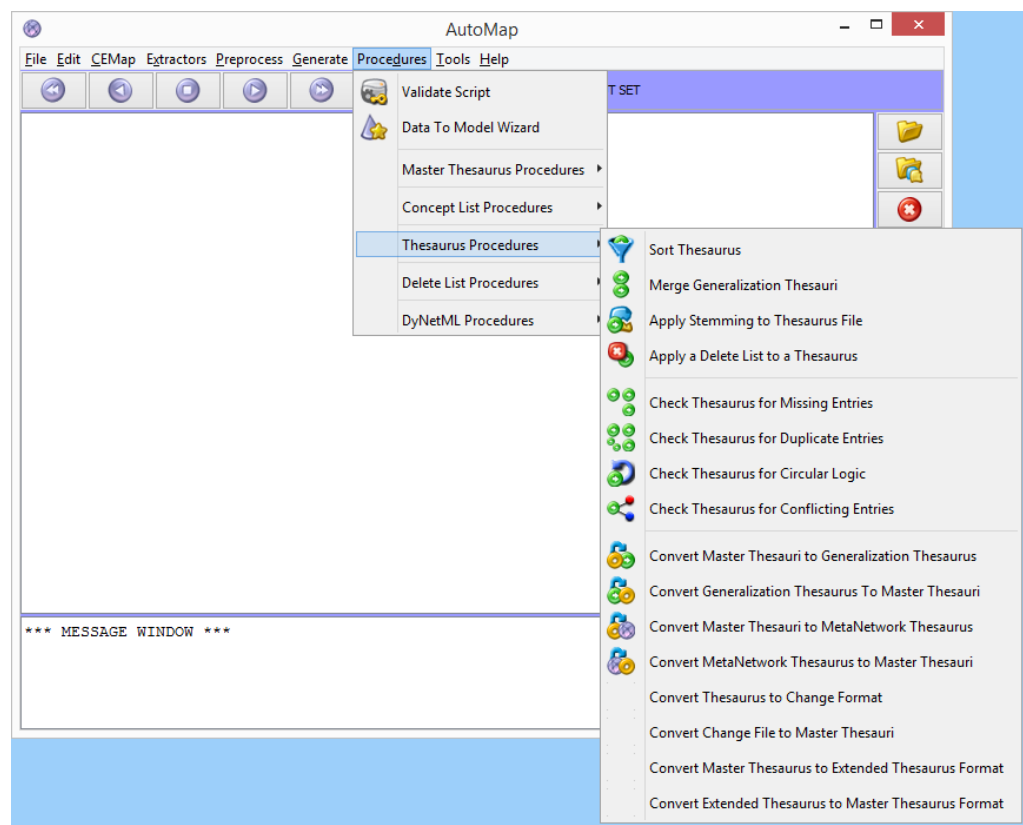


Abbildung 2.11: Thesaurusprüfung in Automap

In Abbildung 2.11 ist zu sehen, dass der Nutzer den Thesaurus nach doppelten Einträgen, nach zyklischer Logik und nach Konflikteinträgen durchsuchen kann. Es werden dabei alle Einträge im Ausgabe-fenster (Message Window) angegeben, sodass der Nutzer die betroffenen Zeilen im Thesaurus identifizieren kann. Lediglich die Funktion "Check Thesaurus for Conflicting Entries" funktionierte während

der Durchführung dieser Arbeit nicht. Das Programm beendete die Funktion mit der Fehlermeldung "Error: Program exited with exit code 1".

In der Arbeit hat sich gezeigt, dass beide Programme zur Thesaurusprüfung genutzt werden sollten, da man so das bestmögliche Ergebnis für die Netzwerkerstellung erzielen kann. Doch selbst wenn beide Programme keine Fehler mehr ausgeben, zeigt ORA den Thesaurus weiterhin als fehlerhaft an. Eine Erklärung hierfür ist während der Arbeit nicht ersichtlich.

Ebenfalls ein wichtiger Punkt vor der Erstellung des Netzwerkes ist die Sortierung des Thesaurus. Carley et al. schreiben im Automap Benutzerhandbuch, dass es "in bestimmten Situationen wichtig ist, den Thesaurus sortieren zu lassen..." (Carley et al. 2013a, S. 89). Dabei wird der Thesaurus vom längsten Konzept zum kürzesten sortiert, damit eine Zuweisung aller Konzepte möglich ist.

Ist der Thesaurus sortiert und geprüft, kann der Nutzer das Metanetzwerk durch Automap erstellen lassen. Hierzu wird der Text in Automap geladen und die notwendigen Bereinigungsschritte durchgeführt. Falls, wie in dieser Arbeit, nur ein einzelner Thesaurus existiert, empfiehlt es sich den Master Thesaurus durch Automap in einen Generalisations-Thesaurus umzuwandeln. Dieser kann dann in der Vorverarbeitung angewandt werden. Nachdem der Thesaurus ausgewählt ist, fragt das Programm, ob lediglich der Inhalt des Thesaurus berücksichtigt werden soll. Ist man sich an diesem Zeitpunkt sicher, dass der Thesaurus alle Konzepte beinhaltet, die man betrachten möchte, kann man an dieser Stelle ja auswählen. Dies führt dazu, dass lediglich nur jene Konzepte im Text verbleiben, die auch im Thesaurus auftauchen, der Rest des Textes wird in die Form "xxx", wie in Abbildung 2.5, überführt, um die Abstände der Wörter zueinander zu erhalten.

Das Netzwerk wird letztendlich über die Funktion "Generate MetaNetwork DyNetML (Per Text)" durchgeführt. Als Input wird wieder der Thesaurus, diesmal aber in der Ursprungsform, welche er vor der Umwandlung zum Generalisation Thesaurus besaß, eingefügt.

Der Nutzer kann für die Erstellung des Netzwerkes verschiedene Einstellungen vornehmen, wie die Fenstergröße zu setzen und Bereiche zu definieren, die Automap für die Berechnung der Zusammenhänge berücksichtigt. Die Fenstergröße gibt dabei an in welcher Entfernung zwei Konzepte im Text erscheinen dürfen um eine Verbindung miteinander zu erhalten. Der Eingrenzungsbereich den der Nutzer wählen kann, gibt ein zusätzliches Stoppkriterium für dieses Vorgehen. Wählt ein Nutzer bspw. "Absatz" oder "Satz" werden nur Wörter innerhalb eines Abschnitts oder eines Satzes miteinander verbunden und nicht mit Wörtern aus vorherigen Abschnitten.

In dieser Arbeit wurde daher "Use all words as size" gesetzt, damit Automap möglichst alle Konzepte miteinander verbindet die zu finden sind.

Durch die zuvor beschriebene Reduktion der Wörter wird die Erstellung des Netzwerkes erheblich beschleunigt. Die Erfahrungswerte dieser Arbeit liegen bei ca. 15-25 Minuten zur Erstellung eines Netzwerkes ohne diese Reduktion gegenüber ca. 1-5 Minuten mit dem kleineren Wörterumfang.

2.2.2.3.1 Workaround für alle Automap Klassen

Wie weiter oben beschrieben, ist es möglich, in einem Netzwerk alle Klassen, die Automap laut Nutzerhandbuch kennen soll, einzubeziehen. Allerdings scheint Automap nicht alle Klassen richtig zu setzen, bzw. ORA diese nicht richtig zu erkennen. In der Arbeit wurden somit nur die "Grundklassen" die ORA kennt erstellt. Es scheint also eine Ungenauigkeit der Klassen zwischen Automap und ORA zu geben. Um alle Klassen, die es laut dem Automap Handbuch geben sollte, nutzen zu können wurde in dieser Arbeit folgender "Workaround" genutzt.

Dazu nutzt man im Thesaurus das Feld "metaOntology" für die , von ORA, unterstützen Gruppen¹⁴, beispielsweise "Task" für "Action", "Resource" für "Attribute" und "Organization" für "Group". Die weiteren Klassen, sofern man sie nutzen möchte, können in der Zeile "metaName" gelistet werden. So ist es dem Nutzer möglich, alle Konzepte zu einer Klasse zuzuordnen und dennoch in ORA eine Zuweisung zu den neuen Klassen zu ermöglichen.

Abbildung 2.12 zeigt, dass in der Spalte "NodeType" die vom Nutzer gewünschten Gruppen ersichtlich sind. Er kann alle Knoten suchen, die einen bestimmten Wert haben, bspw. "Equals: Action", und diese Knoten mit dem Menüpunkt "Move selected nodes" in eine existierende Knotenklasse oder eine neu zu erstellende Klasse bewegen.

14 Dies sind in der betrachteten ORA Version: Agent, Organization, Knowledge, Resource, Belief, Event, Task, Location, Role, Action und Unknown

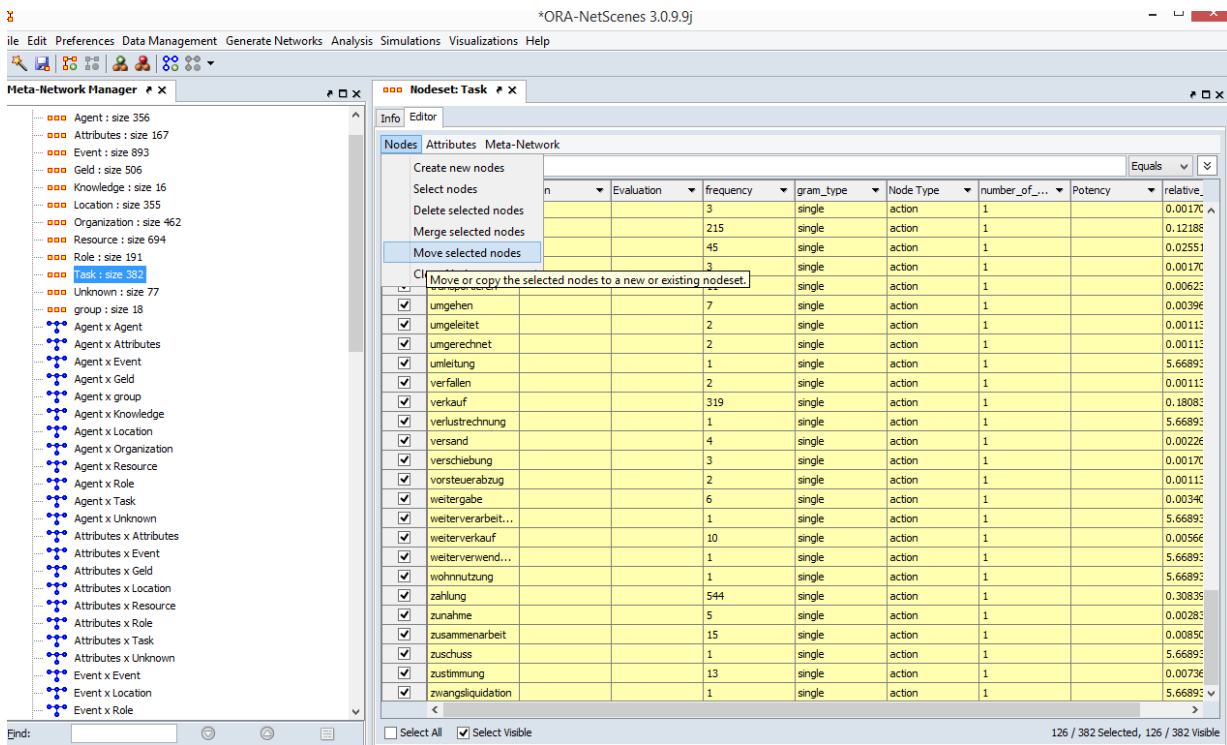


Abbildung 2.12: Workaround für alle Automap Klassen, Schritt 1

In der nächsten Abbildung ist zu sehen, wie die Verschiebung der einzelnen Knoten aussieht. Der Nutzer entscheidet, welcher "Grundklasse" die neu erstellte Klasse am ehesten unterzuordnen ist und wählt diese aus (Abbildung 2.13 links). Danach gibt er der neuen Klasse einen Namen. Alle ausgewählten Knoten und ihre Relationen werden daraufhin in die neu erstellte Klasse übertragen. Dies hat den Vorteil, dass ORA automatisch neue Netzwerke erstellt, die zur erzeugten Klasse gehören, sodass keine Informationen durch eine solche Umsortierung verloren gehen (Abbildung 2.13, rechts).

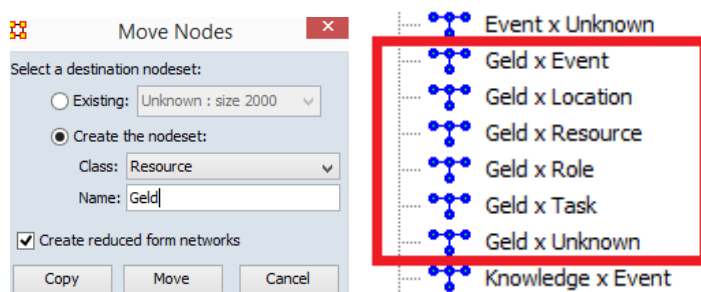


Abbildung 2.13: Workaround für alle Automap Klassen, Schritt 2

2.2.2.4 Die Netzwerkanalyse in ORA

Nachdem das Netzwerk wie in den weiter oben beschriebenen Schritten erstellt und in ORA geladen wurde, beginnt die Netzwerkanalyse ebenfalls mit dem Schritt der Vorverarbeitung. Der erste Verarbeitungsschritt für diese Forschungsarbeit wurde in Kapitel 2.2.2.3.1 beschrieben. Die Konzepte werden zu den Konzeptklassen hinzugefügt, welche der Nutzer zu Beginn analysieren möchte.

Automap hat eine XML Datei erstellt, deren Einlesen in ORA folgende Zuweisung der Klassen zur Folge hatten:

Tabelle 2-4: Zuweisung der Konzeptklassen

Knotenklasse	Beinhaltete Konzepte	Konzepte nach Neuzuweisung
Agent	353	357
Event	842	892
Knowledge	16	16
Location	340	353
Organization	478	472
Resource	785	485
Role	150	190
Task	274	380
Unknown	892	72
Attribute	-	173
Geld	-	717
Group	-	16

Tabelle 2-4 zeigt, dass Automap die Konzepte trotz der Zuweisung, die innerhalb des Thesaurus vorgenommen wurde, für die Erstellung des Metanetzwerkes nicht komplett genutzt hat. Eine Fehlfunktion ist daran zu merken, dass der Thesaurus innerhalb der Spalte "conceptTo" 4059 unterschiedliche Konzepte beinhaltet. Diese Zahl steht daher für die maximale Anzahl an Konzepten, die aus einem Text erstellt werden kann, wenn lediglich die Inhalte des Thesaurus für die Erstellung des Metanetzwerkes genutzt werden dürfen. Summiert man allerdings die Konzeptangaben aus Tabelle 2-4, so gelangt man zu dem Schluss, dass 71 Konzepte zu viel im Netzwerk enthalten sind.

Nach näherer Betrachtung der Klasse "unknown" lässt sich nachvollziehen, dass einige Konzepte, die im Thesaurus nicht in der Spalte "conceptTo" beinhaltet sind, als Konzept in das Metanetzwerk übernommen wurden. Nach weiterer Betrachtung des Thesaurus sind diese Konzepte selbst in der Spalte "conceptFrom" nicht auszumachen, was dafür spricht, dass die Beschränkung auf lediglich die Konzepte des Thesaurus nicht korrekt ausgeführt wurde.

Allerdings ließen sich durch manuelle Arbeit die Klassen weiter spezifizieren, was ebenfalls in Tabelle 2-4 zu sehen ist. Die rechte Spalte zeigt die Zuweisung der Konzepte zu Konzeptklassen, wie sie nach der manuellen Anpassung gegeben ist. Ebenso zeigt sich, dass die Erstellung der fehlenden Klassen innerhalb dieses Schrittes geschieht.

Visualisiert man das Netzwerk mit Hilfe der Funktion "visualize", so erhält man ein Netzwerk, wie es in Abbildung 2.14 zu sehen ist. Es besteht aus 4123, bzw. ohne die Klasse unknown aus 4051, Knoten mit 62114, bzw. 61177, Verbindungen zwischen ihnen.

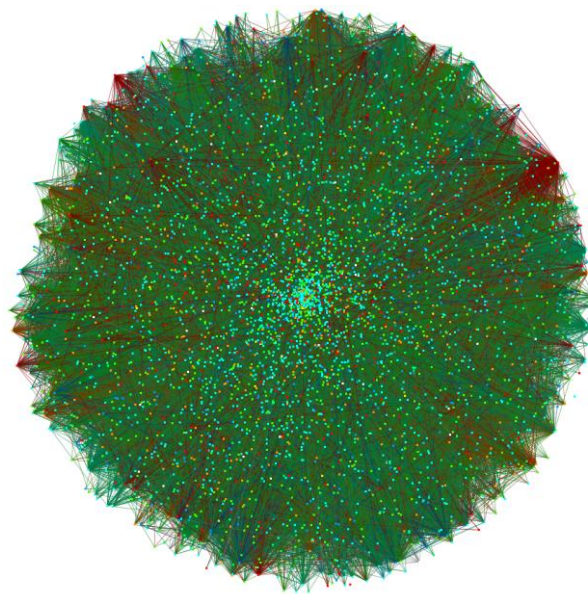


Abbildung 2.14: Das Netzwerk im ORA Visualizer

Da allerdings auffällt, dass die Komplexität des Netzwerkes für einen Nutzer händisch nicht handhabbar ist, sind weitere Analysen und vor allem die Reduktion auf aussagekräftige Unternetze erforderlich.

Diejenigen Konzepte, die entweder durch einen Fehler von Automap oder durch eine Ungenauigkeit innerhalb der Thesaurusdatei nicht zusammengefasst wurden, können innerhalb von Automap über den Dateneditor selektiert und über den Menüpunkt "Nodes" - "Merge selected nodes" zusammengefasst werden. Folgende Grafik zeigt, dass jedoch auch an dieser Stelle Fehler innerhalb des Programmes auftreten. Die Fehlermeldung suggeriert, dass ein unerwarteter Fehler innerhalb der Zusammenführung mehrerer Zeilen stattfand. Allerdings scheint innerhalb dieser Fälle eine Zusammenführung zumindest zum Teil zu funktionieren, da der Wert in der Spalte "Frequency", welcher angibt, wie häufig ein Konzept innerhalb des Netzwerkes bzw. innerhalb des zugrundeliegenden Textes auftaucht, inkrementiert wird. Trotz der Erhöhung des Wertes existierten die ursprünglichen Zeilen teilweise den-

noch. Es ist für den Nutzer daher nicht nachvollziehbar, an welchem Punkt des Zusammenführens der Prozess geendet hat. Innerhalb der Hauptanalyse dieser Arbeit wurde daher auf das Zusammenführen von inhaltlich gleichen bzw. ähnlichen Zeilen verzichtet, da bei einem Fehlschlag nur durch neues Laden des Netzwerkes die Ursprungsdaten wiederhergestellt werden können. Dabei gehen alle vorangegangenen Schritte ebenfalls verloren, hat man keine Sicherungskopie direkt vor dem Zusammenführungsschritt erstellt.

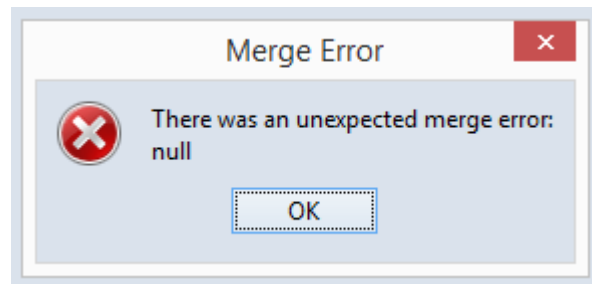


Abbildung 2.15: Zusammenführungsfehler in ORA

Die Vorbereitung der Analyse endet mit dem Schritt der Anonymisierung. Über den Menüpunkt "Data Management" - "Meta-Network Anonymize..." kann man die zu untersuchenden Daten mit einem zusätzlichen Attribut versehen. Dieses Attribut ersetzt bei allen Ausgaben des Tools, wie Reports, den Namen des Knoten, sodass es selbst für personenbezogene Daten für den Nutzer möglich ist, seine Erkenntnisse der Analysen an Dritte weiterzugeben.

Es ist möglich, die Anonymisierung der Daten durch das Bereitstellen einer Datei einzuschränken. Innerhalb der Datei können "friendly names" definiert werden, also solche Angaben, die innerhalb von Reports auftauchen dürfen. Für Datensätze aus demselben Gebiet bzw. mit gleichen Inhalten ist dies eine Erleichterung, für eine einmalige Arbeit mit dem Tool allerdings ein weiterer manueller Arbeitsschritt in dem zu automatisierenden Analyseprozess. Spezifiziert der Nutzer keine solche Datei, benennt ORA alle Daten automatisch mit dem Namen der Knotenklasse und einer ansteigenden Kennzahl, bspw. agent-1 bis agent-357.

Der Nutzer kann auf diese Weise sicher sein, dass vorerst alle Daten in anonymer Form vorliegen. Möchte der Nutzer nun bestimmte Zeilen benannt und einige anonym in die Reports integrieren, kann er manuell entweder für eine komplette Knotenklasse das Anonymisierungsattribut löschen oder für die gewünschten Zeilen den Wert aus der Spalte des Attributes entfernen.

2.2.2.4.1 Maßzahlen innerhalb der Netzwerkanalyse

ORAs Analyseschritte nutzen in den erstellten Reports folgende Maßzahlen "Degree Centrality", "Hub Centrality", "Betweenness", "Eigenvektor" & "Closeness". Innerhalb dieses Abschnittes werden diese Maßzahlen kurz vorgestellt. Sowohl in der Literatur, als auch in ORA, stehen weitere Maßzahlen zur

Netzanalyse zur Verfügung. Die in diesem Kapitel vorgestellten Maße stellen lediglich die inhaltlich passendsten Werte dar, da sie die relevantesten Eigenschaften der zu untersuchenden Elemente beschreiben.

1. Degree Centrality

Laut Jansen wird die Degree Zentralität für einen Akteur ermittelt, "(...) indem man die Zahl seiner Beziehungen zu anderen Akteuren zählt" (Jansen 2003). Gemeint sind alle ausgehenden Verbindungen eines Akteurs zu anderen Akteuren. Für andere Konzeptarten gilt dies ebenso. Sie unterscheidet hierbei die Berechnung der Zentralität in gerichteten oder symmetrischen Netzwerken. Innerhalb symmetrischer Netze gibt es nur eine Art von Beziehung, also keine Unterscheidung nach eingehenden oder ausgehenden Knoten. Daher definiert sich die Degree Zentralität innerhalb symmetrischer Netzwerke durch die Anzahl an Beziehungen zu anderen Konzepten. Innerhalb von gerichteten Netzwerken entspricht sie allerdings nur der Anzahl der ausgehenden Beziehungen (Jansen 2003, S. 132–133).

2. Closeness

Laut Jansen unterscheidet sich die nähenbasierte Zentralität durch die Einbeziehung von indirekten Verbindungen. Es wird also der Wert gemessen, den ein Konzept besitzt, wenn man direkte und indirekte Distanzwerte addiert und den Kehrwert bildet, wobei indirekte Beziehungen "schwächer" als direkte Beziehungen gewertet werden (Jansen 2003, S. 133–134). Sie macht deutlich, dass dieses Maß lediglich dann Sinn ergibt, wenn alle Akteure eine Verbindung besitzen. Sollte es andererseits Konzepte geben, die nicht mit anderen verbunden sind, so dürfen diese Konzepte bei der Berechnung des Wertes nicht berücksichtigt werden, da der Knoten nie erreichbar wäre.

3. Betweenness

Die Betweenness unterscheidet sich in der Art der Berechnung gegenüber der beiden Vorgänger dadurch, dass bei der Berechnung der Betweenness in jedem Schritt 3 Konzepte betrachtet werden. Ist ein Konzept zwischen zwei anderen Konzepten zu finden, handelt es sich laut Jansen um einen "Mittler". Dieses Konzept verbindet die anderen Konzepte miteinander. Je häufiger ein Konzept an der Stelle eines "Mittlers" zu finden ist, desto höher ist der Betweenness Wert. Errechnet wird für jeden Knoten i , wie hoch die Wahrscheinlichkeit ist, dass eine Kommunikation zwischen zwei anderen Knoten x & y über i läuft. Dies ist vor allem dann relevant, wenn es mehrere Verbindungen zwischen x & y gibt. Jede Konstellation wird nur einmal berücksichtigt (Jansen 2003, S. 134–137).

4. Eigenvektor

Jansen schreibt, dass jede Matrix ein Eigensystem besitzt. Innerhalb dieser Systeme existieren Eigenwerte und Eigenvektoren. Dabei sind Eigenvektoren all die Vektoren, die vom Nullvektor unterschiedlich sind und durch die Multiplikation mit einem Eigenwert auf einen bestimmten Punkt im Koordinatensystem zeigen. Jansen schreibt weiter:

"Eine $n \times n$ -Matrix hat maximal n verschiedene Eigenwerte und n Eigenvektoren. Die Eigenwerte einer $n \times n$ Matrix M sind die Werte λ_i für die man von Null verschiedene Vektoren x_i , bestimmen kann, so dass gilt $M * x_i = \lambda_i * x_i$ " (Jansen 2003, S. 150–151).

Die Berechnung des Eigenvektors bzw. der Eigenwerte wird in der Netzanalyse laut Jansen zur Ermittlung der Prestigewerte eines Konzeptes innerhalb des Netzes genutzt.

5. Hub- / Authority Centrality

Laut Carley et al. handelt es sich bei der Hub Zentralität um eine "(...) Generalisierung der Eigenvektor Zentralität (...) und berücksichtigen direkte Verbindungen" (Carley et al. 2013b). Die Werte Hub und Authority sind besonders bekannt innerhalb des Information Retrieval. Ein Hub ist dabei eine Seite (bzw. für diese Arbeit ein Konzept), die auf viele andere Seiten verlinkt und eine Authority ist eine Seite, auf die von vielen anderen Seiten verlinkt wird (Manning et al. 2008, S. 474–478).

Überträgt man dieses Konzept nun auf eine Netzwerkanalyse, so sind Hubs all die Knoten, welche ausgehende Verbindungen zu vielen anderen Knoten innerhalb des Netzes besitzen und Authorities die Knoten, die eingehende Verbindungen von vielen anderen Knoten besitzen.

2.2.2.4.2 Analyse der Netzwerkstruktur

Wie in Abbildung 2.9 gezeigt wurde, beginnt die Netzwerkanalyse mit der Netzstrukturanalyse. Innerhalb dieser Analyse werden die Eigenschaften des in dieser Arbeit erstellten Netzwerkes mit den Eigenschaften von verschiedenen Netzwerktopologien verglichen. In ORA werden diese als "Stylized Networks" bezeichnet. Die in ORA verwendeten Netzwerke sind "Erdős-Rényi", "Core-Periphery", "Scale-Free", "Cellular", "Lattice", "Small-World" & "Fixed Degree-Distribution".

Die Strukturanalyse innerhalb von ORA kann über das Menü "Analysis" - "Generate Reports" - "Statistical Procedures and Diagnostics" - "Context" erreicht werden.

Der Context Report vergleicht alle quadratischen Netze (agent x agent, location x location etc.) mit Cellular, Core-Periphery und Erdős-Rényi Netzen. Innerhalb der Analyse wird für jedes quadratische Netzwerk anhand der Maßzahlen Betweenness, Closeness und Inverse Closeness berechnet, ob es ei-

ner der grade genannten Topologien ähnelt. Auffällig an dieser Stelle ist, dass der Nutzer zwar "Scale-Free" und "Small-World" Netzwerke erstellen kann, diese aber im Context Report nicht als Vergleichswert herangezogen werden können.

Die grade genannten, unterstützten Topologien werden hier kurz erläutert:

1. Erdős-Rényi

Ein Erdős-Rényi Netzwerk ist das klassische Zufallsnetzwerk. In ihrer Arbeit beschreiben Erdős und Rényi, wie ein Graph sich unabhängig von bereits existierenden Kanten entwickelt. In jedem Knoten n besteht die Wahrscheinlichkeit p , dass eine Verbindung zu einem der anderen $n-1$ Knoten existiert. Dies bedeutet, dass es keine Selbstreferenz von Knoten gibt und dass alle Knoten dieselbe Wahrscheinlichkeit $0 < p < 1$ besitzen, miteinander verbunden zu sein (Kun 2013).

2. Core Periphery

Ein Core Periphery Netzwerk drückt sich laut Carley et al. dadurch aus, dass sowohl Knoten als auch Kanten in zwei Gruppen aufgeteilt werden. Diese Gruppen werden als Kern (Core) und als Umfeld (Periphery) bezeichnet. Konzepte, die innerhalb eines Kerns angesiedelt sind, haben eine hohe Anzahl an Verbindungen zueinander und teilen gemeinsame Events. Das Umfeld besteht dagegen aus Konzepten, die nicht gut untereinander verlinkt sind, also nur wenige Verbindungen besitzen und gemeinsame Events teilen.

Weiter drücken Carley et al. aus, dass alle Konzepte in einem Kern, im Gegensatz zum Umfeld, die Möglichkeit besitzen, ihre Handlungen untereinander zu koordinieren. Da die Knoten, welche zum Umfeld gehören, aber nur wenige Verbindungen zum Kern besitzen und zu sonst keinen Knoten, sind diese nicht in der Lage, eine solche Kommunikation hervorzurufen (Carley et al. 2013b, S. 351).

3. Cellular

Zellenartige (Cellular) Netzwerke beschreiben Carley et al. als Netze, die aus einer kleinen Anzahl an dichten Clustern bestehen, welche untereinander nur wenige Verbindungen besitzen. Dabei bezeichnet der durchschnittliche Zellwert die Größe der Cluster.

Tabelle 2-5: Auszug des Context Reports

Topology Cellular (meanCellSize=18)						
Measure	Input network	2014_12_05	Random	T-Value	SL	Same
Centrality, Betweenness	Geld x Geld	0.002	0	2.349	1432	no
Centrality, Betweenness	Attribute x Attribute	0.603	0	4.428	344	no
Centrality, Betweenness	Knowledge x Knowledge	0	0	0	30	yes
Centrality, Betweenness	Agent x Agent	0.376	0	5.118	712	no
Centrality, Betweenness	Event x Event	0.007	0	5.111	1782	no
Centrality, Betweenness	Role x Role	0.392	0	5.058	378	no
Centrality, Betweenness	Organization x Organization	0.368	0	7.041	942	no
Centrality, Betweenness	Task x Task	0.310	0	6.507	758	no

Tabelle 2-5 zeigt einen Ausschnitt aus dem Context Report¹⁵ für ein Cellular Netzwerk mit der durchschnittlichen Zellgröße von 18. Ein solches Netzwerk besteht laut Carley et al. aus Clustern mit vielen internen Verbindungen, aber nur wenigen Verbindungen untereinander. Die Zellengröße gibt dabei an, wie viele Knoten durchschnittlich zu einem Cluster gehören (Carley et al. 2013b, S. 350).

Zu erkennen ist, dass für jedes one-mode Netzwerke die Wahrscheinlichkeit der Zufallsverteilung (2014_12_05) ein T-Wert und ein Signifikanzlevel (SL) errechnet werden. Für jeweils die Maßzahlen "Centrality, Betweenness", "Centrality, Closeness" und "Centrality, Inverse Closeness" wird anhand der errechneten Werte entschieden ob es sich bei dem Netzwerk um eine Zufallsverteilung handelt oder ob ein bestimmtes Muster (eine Topologie) zu erkennen ist.

Der Report in Tabelle 2-6 ergibt, dass lediglich die Netzwerke Location x Location, Knowledge x Knowledge, Geld x Geld und Group x Group zu den zu vergleichenden Netzwerken zugeordnet werden können.

Tabelle 2-6: Context Report Ergebnisse

Topologie Netzwerk	Erdős-Rényi	Core-Periphery	Cellular (m = 4)	Cellular (m = 10)	Cellular (m = 18)	Cellular (m = 25)
Location x Location	1	-	1	-	-	-
Knowledge x Knowledge	2	2	2	2	2	2
Geld x Geld	1	1	-	-	-	-
Group x Group	2	2	2	2	2	2
Unknown x Unknown	1	1	2	2	2	2

¹⁵ Eine Zusammenfassung des vollständigen Reports ist als Anhang eingefügt (siehe Anhang 2)

Diese Werte werden aus der Summe der Spalte "Same", wie in Tabelle 2-5 zu sehen, errechnet. Auf den ersten Blick ist zu sehen, dass eine eindeutige Unterscheidung aus diesem Report nicht ersichtlich ist, da einige Netzwerke scheinbar allen Topologien gleich ähneln. Außerdem können sieben Knotenklassen nicht zu einer Topologie zugeordnet werden, da für alle Maßzahlen jeder Topology der Wert "Same" mit "no" gesetzt ist ¹⁶. Diese Spalte drückt aus, ob es sich laut der errechneten Werte um dieselbe Art von Netzwerk handelt. Stimmt daher der errechnete T-Wert (siehe Tabelle 2-5) mit dem erwarteten Wert für die jeweilige Topologie überein, so wird der Wert für "Same" auf "yes" gesetzt. Eine Übersicht der Werte, welche ORA für die jeweiligen Topologien erwartet, ist allerdings nicht verfügbar.

Tabelle 2-6 zeigt, dass alle Netzwerke, welche Ergebnisse hervorgebracht haben, zu einem Erdős-Rényi Netzwerk zugeordnet werden können. Für die Knotenklassen "Group" und "Knowledge" kommen laut des Kontextreports aber auch alle Formen eines zellenartigen Netzwerkes in Frage. Das Netzwerk "Location x Location" lässt sich ebenfalls auf ein Erdős-Rényi bzw. ein zellenartiges Netzwerk, allerdings nur mit einer Zellgröße von durchschnittlich 4 Knoten pro Cluster, zuweisen.

Daraus ergibt sich, dass die fünf (bzw. vier relevanten) Knotenklassen aus Tabelle 2-6 einer der Topologien ähneln. Für alle anderen Knotenklassen kann lediglich ausgeschlossen werden, dass sie einer der drei Zufallsnetz Topologien angehören. Aus den durch ORA zur Verfügung stehenden Reports ergibt sich also die Annahme, dass keine erkennbare Verteilung innerhalb der Netzwerke existiert.

2.2.2.4.3 Identifikation der Key-Entities

Genau wie auch die Strukturanalyse des Netzwerkes nutzt die Key Entities Analyse die in Kapitel 2.2.2.4.1 beschriebenen Maßzahlen zur Errechnung der jeweiligen Listen. Wie in dem Kapitel erwähnt, drücken die verschiedenen Maßzahlen unterschiedliche Dinge aus. Für diese Analyse sind alle oben erklärten Zentralitätsmaße relevant.

Key Entities werden in allen Knotengruppen gesucht, sodass jeweils die wichtigsten Konzepte jeder Gruppe analysiert werden können. Der Report findet sich in ORA unter "Analysis" - "Generate Report" - "Locate Key Entities" - "Key Entity". Die Reports werden jeweils in einer anonymisierten Form und einer Klartextform durchgeführt. Auf diese Weise ist es dem Nutzer möglich, die beiden Ergebnisse miteinander zu vergleichen und deren Übereinstimmung zu validieren.

¹⁶ Es gibt zwölf Knotenklassen, von denen lediglich 5 Klassen in der Zuordnung auftauchen.

Für die weitere Analyse wurden vor allem die Key Entities aus den Gruppen "Agent", "Attribute", "Location", "Organization", "Resource", "Geld" und "Task" herangezogen. Die Gruppen "Knowledge" und "Group" wurden aufgrund ihrer Größen und aufgrund des Inhaltes für Geldwäsche als nicht relevant eingeordnet und die Gruppe "Event" wurde ebenfalls als nicht aussagekräftig eingestuft. Die meisten Einträge in der Gruppe Event bestehen aus konkreten Datumsangaben, sodass der Mehrwert für diese Analyse nicht ersichtlich ist. Analysen, die den zugrundeliegenden Fall im Ganzen weiter aufklären sollen, können diese Klasse jedoch eventuell für die konkrete zeitliche Abfolge der Geschehnisse nutzen.

Die Key Entities der Gruppe "Roles" konnten durch ORA nicht erstellt werden, da laut des Reports alle Knoten dieser Klasse denselben Wert aufweisen. Die Einträge dieser Klasse werden daher erst in den weiteren Analysen, wie der Ego-Netzanalyse betrachtet.

In Anhang 1 sieht man eine Excel Tabelle, innerhalb welcher für die Zentralitätsmaße "Total Degree Centrality", "Hub Centrality", "Betweenness", "Eigenvektor" und "Closeness" und für jede Knotenklasse, die eben beschrieben wurde, die ersten 5 Einträge findet. Diese Tabelle wurde als Indikator für die folgende Ego-Netzwerkanalyse herangezogen.

Tabelle 2-7: Beispiel Key Entities (Agent)

	Agent				
Platzierung Maßzahl	1	2	3	4	5
Total Degree	Agent-344	Agent-346	Agent-174	Agent-2	Agent-188
Hub	Agent-174	Agent-346	Agent-2	Agent-344	Agent-188
Betweenness	Agent-344	Agent-346	Agent-2	Agent-174	Agent-134
Eigenvektor	Agent-344	Agent-174	Agent-346	Agent-2	Agent-188
Closeness	Agent-344	Agent-181	Agent-346	Agent-119	Agent-231

Tabelle 2-7 zeigt eine interne Auflistung der Key Entities für die Knotenklasse "agent". Es wurden für diese fünf Zentralitätsmaße je die ersten fünf Konzepte eingetragen. Aus diesen 25 Werten wurden die häufigsten ausgewählt und für die folgende Ego-Netzwerkanalyse genutzt.

Für diese Arbeit von besonderer Bedeutung könnten jedoch jene Konzepte sein, die einen hohen Wert bei der Hub und der Betweenness Zentralität besitzen, da diese, wie weiter oben beschrieben, am ehesten für Mittelsmänner bzw. Strohmänner sprechen.

Ein Hub ist, eine Person (ein Konzept), welche viele ausgehende Verbindungen besitzt, also potentiell viele Personen oder andere Konzepte erreichen kann. Die Betweenness sagt aus, wie häufig ein Konzept als Mittler zwischen anderen Konzepten dient.

Es ist fraglich ob eine Unterscheidung zwischen Mittelsmann und Strohmännchen inhaltlich möglich ist. Ein Mittelsmann könnte von der Tragweite durchaus wichtiger sein als ein Strohmännchen, da ein Mittelsmann mehrere Gruppen bzw. Konzepte verbindet. Diese Konzepte würden also einen hohen Betweenness-Wert aufweisen. Entfernt man diese Person aus dem Netzwerk, so würde es möglicherweise zur Reduktion von Cliquen bzw. Gruppen innerhalb des Netzes führen. Da allerdings ein Strohmännchen eine Person A gegenüber einer Partei B vertritt steht dieser ebenfalls zwischen zwei Parteien. Da die inhaltlichen Unterschiede, also die Einsetzung eines Strohmännchens zur Verschleierung einer Person bzw. die Einbindung eines Mittelsmanns als zentrale Anlaufstelle, durch die hier durchgeführte Netzwerkanalyse nicht unterscheidbar sind, wird innerhalb dieser Arbeit ein Strohmännchen und ein Mittelsmann als gleichwertig bezüglich ihrer Eigenschaften angesehen. Nicht unterscheidbar sind sie, da lediglich eingehende und ausgehende Verbindungen und die daraus resultierenden Maßzahlen betrachtet werden. Gäbe es eine weitere inhaltliche Ebene, die den Zweck dieser Konzepte bestimmt, könnte in späteren Analysen eine Unterscheidung möglich sein.

Diese beiden Werte drücken somit die Eigenschaften aus, die einen klassischen Mittelsmann ausmachen. Da allerdings im Bereich der Geldwäsche viele verschiedene Organisationen und Personen beteiligt sind, ist es wahrscheinlich, dass es viele "kleinere" Strohmännchen gibt, welche für verschiedene Aufgabenbereiche zuständig sind. Dies spricht dafür, dass eventuell diejenigen Personen Strohmännchen sein könnten, die eingehende Verbindungen der Top-Hubs besitzen¹⁷.

Wie in Tabelle 2-7 gezeigt wurde, gibt es einige Schlüsselpersonen, die innerhalb der hier untersuchten Datengrundlage als besonders wichtig einzuordnen sind. In Abbildung 2.16 wird dies innerhalb eines Balkendiagramms noch weiter veranschaulicht. Zu erkennen ist, dass es vier Personen gibt, welche in ihrer Wichtigkeit die anderen Personen weit übertreffen. Percent of measures bezeichnet dabei den Durchschnittswert den ein Agent für alle betrachteten Maßzahlen besitzt.

¹⁷ Ein Strohmännchen wird laut Berwanger definiert als: "Vorgeschobene Person, die im eigenen Namen für jemand anders handelt und die Treuhänderstellung verheimlicht. Häufig bei der Gründung von Gesellschaften (um gesetzlich geforderte Anzahl der Gründer herzustellen)" (Berwanger).

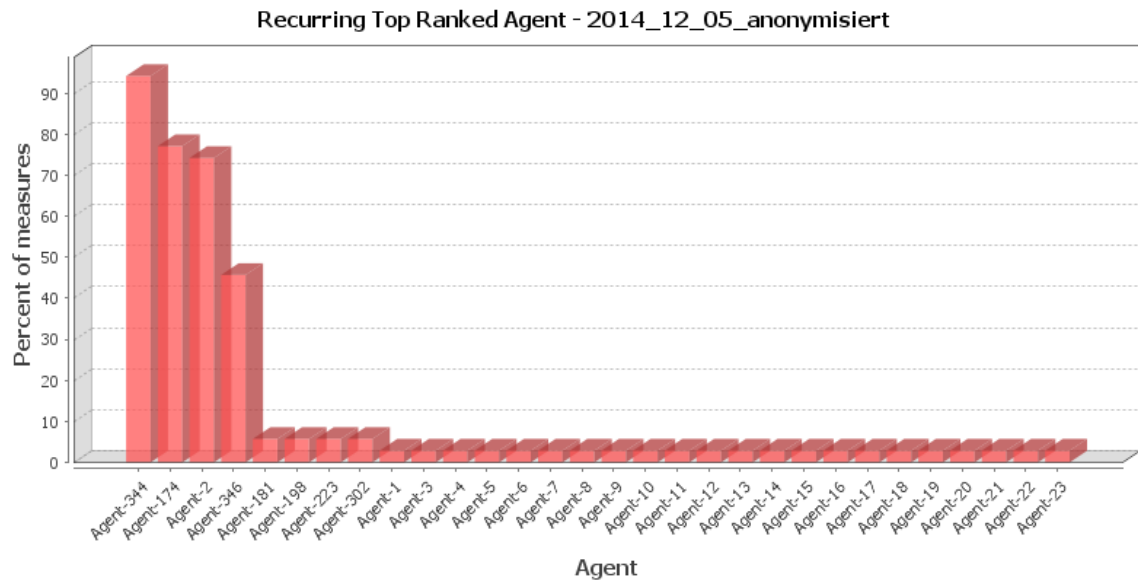


Abbildung 2.16: Wichtige Schlüsselpersonen

An dieser Stelle muss darauf hingewiesen werden, dass sich die Polizeiarbeit auf bestimmte Personen und einen bestimmten Sachverhalt fokussiert hat. Die Wichtigkeit einzelner Konzepte kann daher "systematisch" erhöht worden sein. Alle Interviews, Zeugenaussagen und Schriftstücke, die durch die Polizei zur Verfügung gestellt wurden, können daher ein Störfaktor¹⁸ für die Aussagekraft der Analyse sein. Um diesen Faktor zu verringern oder auszuhebeln, müssten weitere Dokumente zur Verfügung gestellt werden. Die Netzwerkanalyse bezieht sich somit auf die sich aus den Daten ergebenden Häufigkeiten, Verbindungen und Aussagen.

Innerhalb dieses Kapitels werden nur die aussagekräftigsten Reports dargestellt. Aussagekräftig bedeutet an dieser Stelle, dass lediglich diese Reports aufgeführt werden, die klare Tendenzen zeigen, wie die folgende Abbildung.

¹⁸ Störfaktor bezieht sich hier auf die Anwendbarkeit der Ergebnisse auf den real existierenden Polizeifall. Da während der Analysen lediglich die Informationen zur Verfügung standen, welche die Polizei besitzt, kann nicht gewährleistet werden, dass alle beteiligten Personen und Organisationen erkannt wurden.

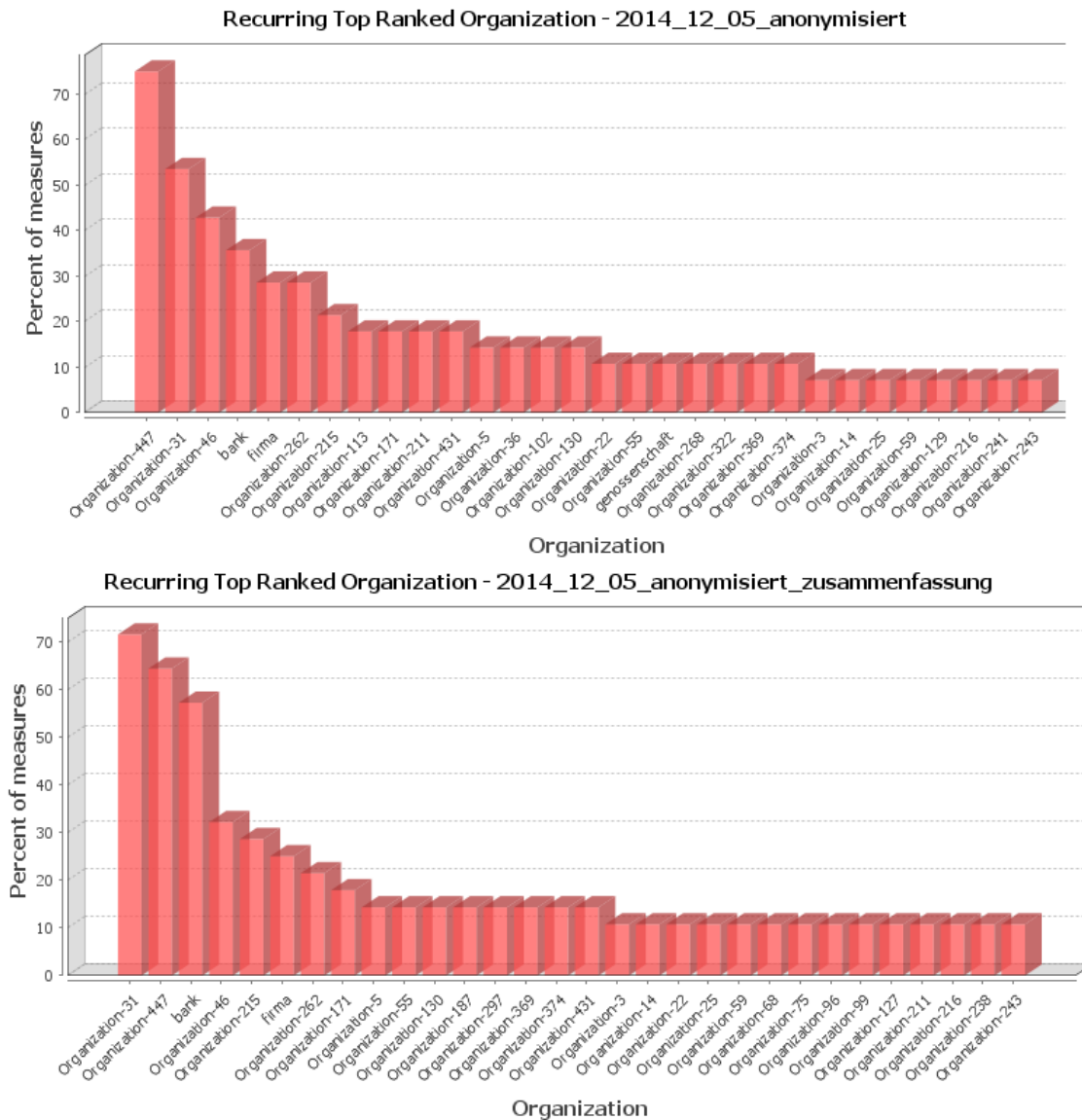


Abbildung 2.17: Wichtige Organisationen

Auf dieser Grafik (Abbildung 2.17) sind die wichtigsten Unternehmen gezeigt, jedoch in zwei verschiedenen Varianten. Der obere Teil des Bildes zeigt die Verteilung der Organisationen, wie sie im Text auftauchen und der untere Teil des Bildes zeigt eine verschobene Verteilung, die sich aus der Zusammenfassung der Konzepte ergibt. Es wird deutlich, dass die beiden ersten Unternehmen eine bedeutende Rolle innerhalb des Falles spielen und dass durch die Hinzunahme der Tochtergesellschaften zu einem Konzept eine noch kritischere Rolle innerhalb des Netzwerkes erreicht werden kann. Dieser Fakt spricht ebenfalls dafür, dass die Bildung von Tochter- oder Schwestergesellschaften eine Rolle innerhalb der Geldanalyse spielen kann. Denn ebenso, wie die Konzepte die im ersten Teil des Bildes

eventuell nicht mehr mit aufgeführt wurden¹⁹, kann dies eine Aussage über die Nachverfolgbarkeit von Transaktionen ausmachen. An dieser Stelle ist ebenfalls anzumerken, dass hier lediglich diejenigen Konzepte zusammengefasst wurden, die namentlich zusammengehörten. Dies ist nur in der nicht anonymisierten Version des Textes bzw. der Daten nachvollziehbar.

Durch das Zusammenführen aller Banken zum allgemeinen Konzept Bank ist der Stellenwert dieses Konzeptes im unteren Teil des Bildes ebenfalls erheblich deutlicher zu erkennen. Neben den beiden Firmen 31 und 447 ist es das drittgrößte Konzept. Führt man dies auf die Inhalte von Kapitel 1.4.3 und insbesondere auf Abbildung 1.2 zurück, unterstützt die Wichtigkeit des Konzeptes Bank die Aussagen über Geldwäsche.

Vergleicht man an dieser Stelle die Abbildung der Personen mit der Abbildung der Organisationen, fällt auf, dass wesentlich mehr Organisationen relevant zu sein scheinen als Personen, da die Steilheit der Kurve in Abbildung 2.16 höher ist als die der Abbildung 2.17. Dies könnte dafür sprechen, dass die vier Personen wesentlich in die Geldwäsche involviert sein könnten. Um diese Hypothese zu stützen bzw. zu prüfen, werden in der folgenden Analyse der Egonetzwerke, sowohl die wichtigsten Agenten, als auch die wichtigsten Organisationen betrachtet.

¹⁹ Auf den Reports sind lediglich die 30 wichtigsten Konzepte aufgeführt.

2.2.2.4.4 Analyse der Egonetzwerke

Egonetzwerke stammen aus der sozialen Netzwerkanalyse. Betrachtet werden die Netzwerke von den im Fokus stehenden Konzepten, welche man innerhalb dieser Analysen als Ego bezeichnet. Dem Ego entgegen stehen die "alters", welche alle Kontakte bezeichnet, die das Ego kennt bzw. zu denen das Ego Verbindungen besitzt (Mcauley und Leskovec 2014). Erweitert man diesen Ansatz auf die hier durchgeführte semantische Netzanalyse, bezeichnet das Ego nicht mehr nur Personen, sondern Konzepte aller bisher vorgestellten Knotenklassen. Auf diesem Weg lässt sich für die im letzten Kapitel identifizierten Schlüsselkonzepte eine genauere Analyse aufziehen. In Abbildung 2.18 sieht man das Egonetzwerk der Person "u" und die verschiedenen Kreise, in denen sie verkehrt.

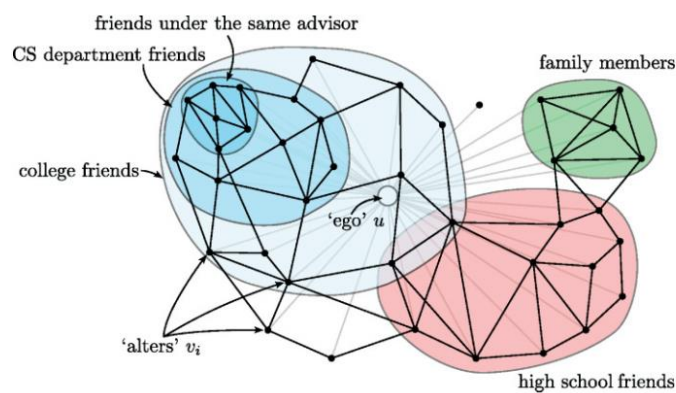


Abbildung 2.18: Beispiel eines Egonetzwerkes
(aus: Mcauley und Leskovec 2014)

Um in ORA solche Egonetzwerke bzw. Ego Reports zu erstellen, kann man die Funktion "View Role and Sphere of Influence" nutzen. Man kann aus allen Knotenklassen diejenigen Knoten anwählen, welche man für die hier genannte Analyse betrachten möchte.

Der Key-Entity Report dient zur Identifikation der für diese Analyse relevanten Knoten. Innerhalb dieses Kapitels werden die interessantesten Erkenntnisse des Egonetzwerkes aufgezeigt.

Weiterhin gliedert sich dieser Arbeitsschritt in zwei Teile. Zuerst werden die Reports, die oben bereits erwähnt wurden, erstellt und analysiert. Aus den Inhalten der Reports werden die interessantesten in diesem Kapitel wiedergegeben. Anknüpfend an die Reports werden diejenigen Egonetzwerke grafisch dargestellt, die aus den Reports oder aus der Themenvorgabe (Geldwäsche) eine wichtige Rolle zu spielen scheinen. Diese Ergebnisse werden ebenfalls festgehalten und interpretiert.

2.2.2.4.4.1 Die Reports

Innerhalb der Reports werden die Schlüsselkonzepte der Daten weiter analysiert. Für jedes Konzept werden die Beziehungen des Egos zu den anderen Knotenklassen visualisiert und verschiedene Maßzahlen berechnet. Ein erster Überblick über die Maßzahlen zeigt, dass zwischen exklusiven, gemeinsamen und indirekten Links unterschieden wird.

Ein exklusiver Link stellt eine Beziehung dar, die lediglich das betrachtete Ego in seiner Knotenklasse besitzt. Beispielsweise besitzt das Konzept "Vereinbarung" als einziger Task eine (ausgehende) Verbindung zum Ort "Location-270". Das ortsbezogene Egonetz von "Vereinbarung" zeigt, dass verschiedene Staaten in den Fall involviert sind.

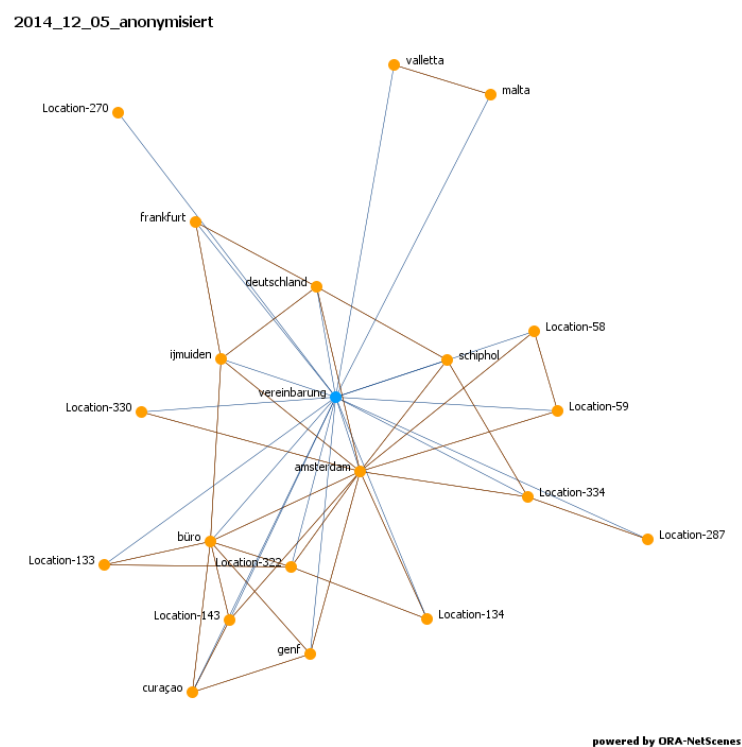


Abbildung 2.19: Egonetzwerk (Location x Task) von Vereinbarung

In Abbildung 2.19 sieht man beispielsweise, dass "Curaçao", "Genf", "Amsterdam", "Schiphol", "Deutschland", "Frankfurt", "Malta", "Valetta" und einige spezifischere²⁰ Ortsangaben eine Verbindung zu Vereinbarung besitzen. Dies legt den Schluss nahe, dass Geschäftsverbindungen international

²⁰ Alle Ortsangaben, die spezifischer als Ortsnamen sind, wurden in dieser Arbeit anonymisiert. Vor allem bezieht sich die Anonymisierung auf konkrete Adressangaben.

ler Art bestehen. Für die Geldwäsche kann dies insbesondere bedeuten, dass es Tochtergesellschaften geben könnte, die in unterschiedlichen Ländern miteinander agieren.

Der Report des Konzeptes "ausländisch" zeigt, dass es Ähnlichkeiten zu einigen anderen Knoten der Klasse Attribute besitzt. In der folgenden Tabelle sieht man, dass zwei inhaltlich naheliegende Konzepte "international" und "offshore" zu 60,3% bzw. 53,8% dem Konzept ausländisch ähneln. Interessant ist jedoch, dass ebenfalls "Unternehmen", "Struktur", "falsch" und "Spezifikation" eine Ähnlichkeit von über 50% besitzen. Diese Angaben sind daher interessant, da sie Aufschluss darüber geben, welche Konzepte innerhalb der internationalen Geldwäsche von Bedeutung sind.

Tabelle 2-8: Auländisch und ähnliche Konzepte

Rank	Attribute	Value
1	struktur	0.805
2	unternehmen	0.604
3	international	0.603
4	falsch	0.582
5	offshore	0.538
6	spezifikation	0.516
7	beschlagnahm	0.499
8	unternehmensstruktur	0.492
9	zugunsten	0.476
10	abgesichert	0.467

Es folgt eine Tabelle, die alle Konzepte zeigt, die einem der Schlüsselkonzepte der Klasse Attribute über 50% ähneln. Diese Konzepte können weitere Indikatoren für Geldwäsche sein und sollten daher entweder weiter einzeln betrachtet werden oder ihr Auftreten in anderen Reports sollte von Wichtigkeit sein.

Tabelle 2-9: Ähnliche Konzepte (Attribute-Klasse)

Schlüsselkonzept	Ähnliche Konzepte
Ausländisch	Struktur, Unternehmen, international, falsch, offshore, Spezifikation
Bedingung	Strafzinsen, Gesellschaft
Betrag	Unternehmen, falsch, Strafzinsen, Gesellschaft, finanziell, Wert, commercial
Signatur	falsch, strafzinsen, bedingungslos, Note, Struktur, beschlagnahm
Struktur	ausländisch, falsch, Unternehmen, international, Stundensatz, beschlagnahm, Spezifikation, abgesichert, Signatur, zugunsten
Unternehmen	falsch, Struktur, Betrag, international, ausländisch, finanziell, regelmäßig, beschlagnahm, zugunsten, Unternehmensstruktur
Wert	wertlos, finanziell, Betrag

Da aus der Datenquelle bekannt ist, dass Immobiliengeschäfte mit zu den Einnahmequellen der Geldwäsche gehört, ist die Betrachtung vom Konzept "Kauf" und "Verkauf" sinnvoll, da sie die verschiedenen Stationen zur Legalisierung der illegalen Gelder darstellen könnten.

Aus der Betrachtung ergibt sich, dass 22 Orte, 36 Personen und 76 Organisationen direkte Verbindungen zum Konzept "Kauf" und 71 Orte, 58 Personen und 101 Organisationen direkte Verbindungen zum Konzept "Verkauf" aufweisen. Dabei sind beim Konzept "Kauf" lediglich 38 verschiedene Geldbeträge direkt verbunden und beim Konzept "Verkauf" lediglich 61. Dies könnte ein Hinweis sein, dass dieselben Gelder über mehrere Personen, Orte bzw. Organisationen geflossen sind, was ebenfalls ein weiterer Indikator für Geldwäsche sein könnte.

Ein weiterer interessanter Faktor zeigt, wie Organisation 31 mit Geldern zusammenhängt. Zu sehen ist diese Grafik in Anhang 6. Gelder scheinen untereinander viele Verbindungen zu besitzen und das Unternehmen zu vielen der einzelnen Geldbeträge. Ein weiterer Indikator, dass Geld über Unternehmen in den legalen Markt transferiert wurde. Anhand dieser Grafik wurde die Entscheidung getroffen, die Schlüsselgelder zu betrachten und deren Verbindungen mit Unternehmen und Personen zu betrachten.

In der Grafik Anhang 5 ist zu erkennen, dass Agent-2 eine Verbindung zu "Währung_15000" (15.000) besitzt, welches wiederum Verbindungen zu mehr als 20 anderen Beträgen aufweist. Interessanter jedoch sind die Währungsangaben "Währung_10000000" (10.000.000) und "Währung_1500000" (1.500.000), da diese beiden Beträge Verbindungen zu mehr als einer Person besitzen, bzw. mehr als eine Person Verbindungen zu diesem Betrag. "Währung_1500000" ist verbunden mit den Agenten 51, 270, 321 und 346.

Der zweistellige Millionenbetrag (10.000.000) besitzt Verbindungen zu elf Personen, darunter ebenfalls Agenten 346 und 321, welche schon Verbindungen zu dem anderen Betrag besaßen. Dies kann für Geldwäsche sprechen, da viele Personen mit den Geldern in Kontakt gekommen sind.

auch gemeinsame, Verbindungen zu diversen Orten besitzen. Diese Orte sind: "Amstel", "Florida", "Türkei", "Belgrad", "Frankreich", "Utrecht", "Luxemburg", "Vaduz", "Lugano", "Amstelland", "Skandinavien", "Niederrhein", "Karibik", "England", "Frankfurt", "Südafrika", "Amsterdam", "Krumnau", "Genf", "Fehmarn", "Sardinien", "Flevoland" und "Schiphol". Allerdings gibt es auch allgemeinere Ortsangaben wie "Spielhalle", "Flughafen" und "Standort".

Dies spricht unter dem Aspekt der Geldwäsche ebenfalls dafür, dass diese Beträge durch verschiedene Unternehmen, an oben genannten Orten, flossen. Diese Aussage unterstützt ebenfalls die Aussagen der Geldwäsche aus Kapitel 1.4.3.

Eine Erweiterung durch die Agentenklasse zeigt, dass 16 Agenten Beziehungen zu den Schlüsselkonzepten besitzen. Darunter befinden sich ebenfalls folgende Schlüsselkonzepte der Agentenklassen:

Agent 2, 188, 344 und 346. Die anderen Agenten sind keine Schlüsselkonzepte und könnten daher ebenfalls kleine Stroh- bzw. Mittelsmänner sein. Die komplette Grafik, die die Beziehung zwischen Agenten und den Schlüsselgeldern zeigt, befindet sich im Anhang.

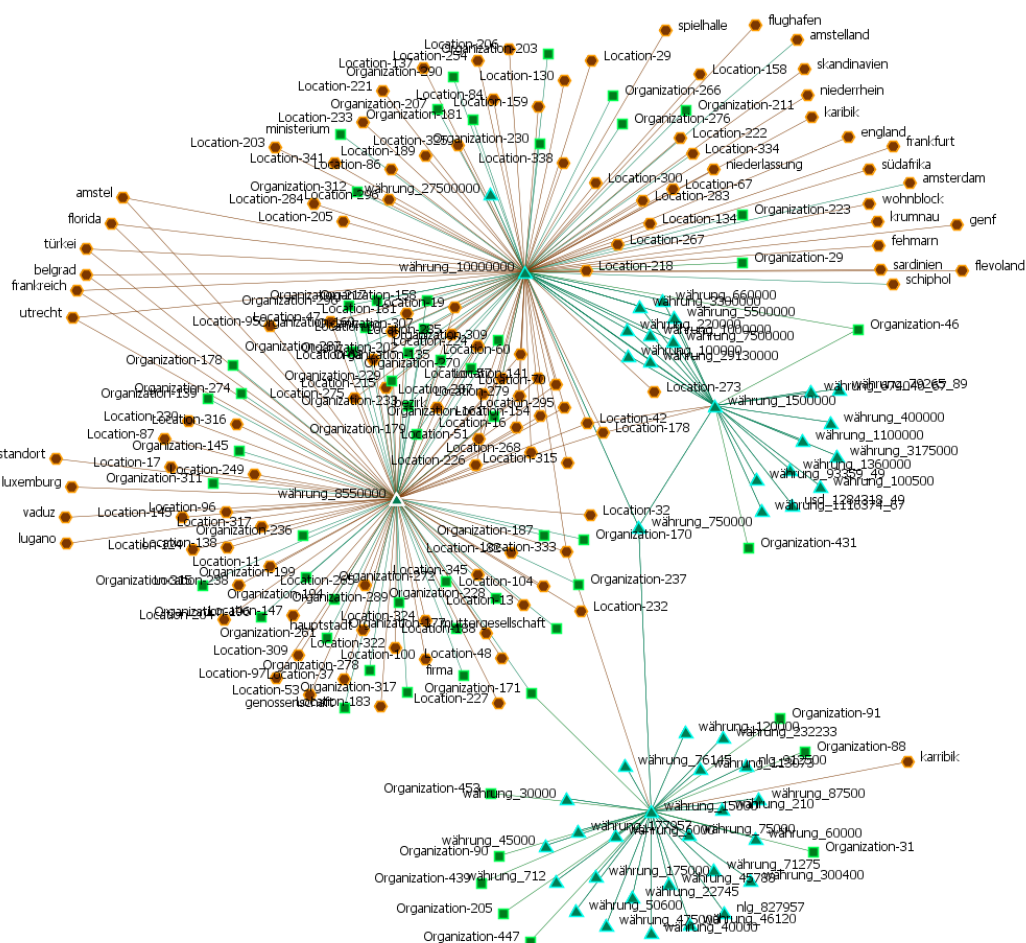


Abbildung 2.21: Egonetzwerk der Schlüsselgelder (Location x Organization x Gelder)

Die Schlüsselkonzepte der Organisationen, die Verbindungen zu den Geldern haben, sind:

Organisation 31, 46, 447 und das allgemeine Konzept "firma". Nimmt man diese Erkenntnisse zusammen, liegt die Vermutung nah, dass sowohl einige der Schlüsselpersonen, als auch einige der Schlüssellorganisationen in Geldwäschegeschäften involviert sind.

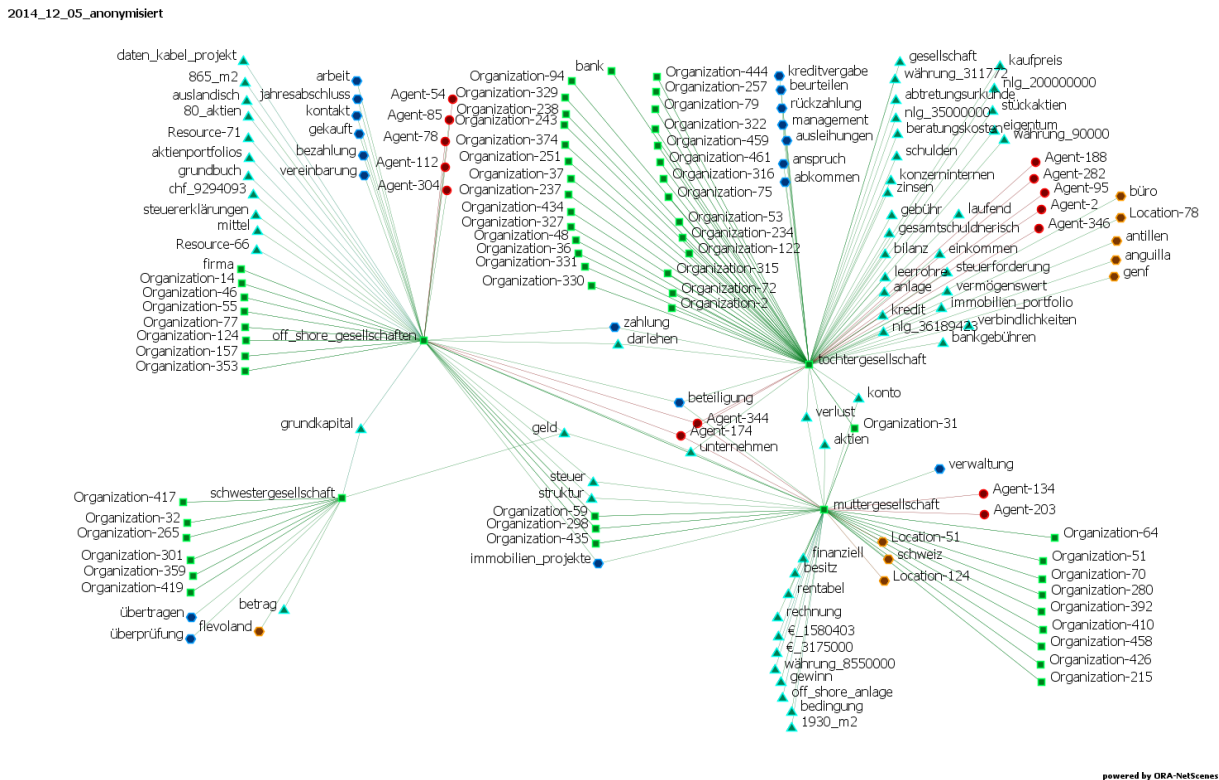


Abbildung 2.22: Egonetzwerk der verschiedenen Gesellschaftstypen

Abbildung 2.22 zeigt die Wichtigkeit der Konzepte "Muttergesellschaft", "Tochtergesellschaft", "Off_shore_Gesellschaft" und "Schwestergesellschaft". Sie beinhaltet die Knotenklassen "Agenten", "Attribute", "Location", "Organization", "Geld", "Resource" und "Task".

Zu sehen ist, dass es viele Offshore Gesellschaften gibt und viele Tochter- und Muttergesellschaften. Die Agenten 344 und 174 besitzen Verbindungen zu allen Gesellschaftsformen außer "Schwestergesellschaft". Dies ist ebenfalls ein weiterer Indikator dafür, dass sie Geldwäsche betreiben bzw. in einer Form an den Geldwäscheaktivitäten beteiligt sind. Ebenfalls fällt auf, dass Offshore Gesellschaft und Muttergesellschaft direkt miteinander verbunden sind. In der Mitte befinden sich Konzepte, welche Verbindungen zu allen Gesellschaftstypen (i.S.v. Mutter, Tochter, etc.) besitzen.

Auffällig sind dabei die Konzepte "Geld", "Unternehmen" und "Beteiligung", welche jeweils zu drei Typen Verbindungen besitzen und die Konzepte "Konto", "Verlust", "Aktien", "Grundkapital", "Steuer",

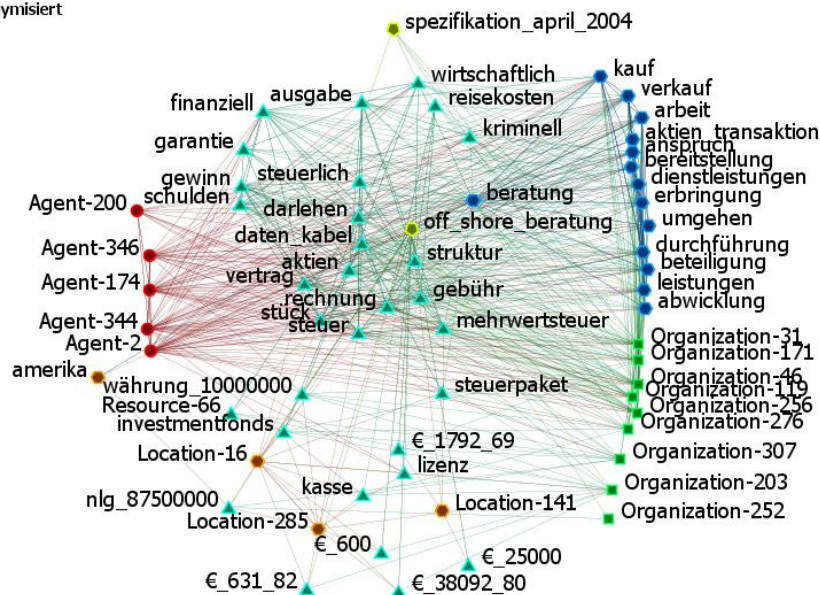
"Struktur", "Immobilien_Projekte" und einige Organisationen, welche zu je zwei Typen Verbindungen besitzen. Ebenfalls eine interessante Tatsache sind die verschiedenen Agenten, die zu den Gesellschaftstypen Beziehungen besitzen, möglicherweise könnten diese Ansprechpartner bei den Unternehmen gewesen sein, welche die Geldwäsche ermöglichten.

Weitere Konzepte, die interessante Aussagen ermöglichen, sind "Urkundenfälschung" und "ausländisch" (in ORA ausländisch). Ausländisch zeigt Verbindungen zu den Agenten 2, 344 und 174, den Orten Lugano, Curaçao und Schweiz, was ebenfalls wieder die Beteiligung an internationalen Geschäften aufzeigt. Ebenso besitzt es Verbindungen zum Konzept "off_shore_gesellschaften", "bankkonten", "geld" und "geldwäsche".

Urkundenfälschung ist sowohl verbunden mit Agent 174 und Agent 344, als auch mit einigen Geldbeträgen, dem Konzept "refinanzierung", "geldwäsche" und "ausländisch". Ein weiterer Indikator dafür, dass Gelder unter fingierten Umständen transferiert wurden.

Betrachtet man sich das Konzept "Immobilien_projekte" genauer, welches aus Abbildung 2.22 bekannt ist, fällt auf, dass es "nur" Verbindungen zu "off_shore_gesellschaften", "muttergesellschaft", "beteiligung" und "Organization_298" besitzt. Daher liegt die Vermutung nahe, dass innerhalb der Organisationen, die sich im Besitz der Geldwäscher befinden, Immobilien weiterverkauft bzw. weitergereicht wurden. Dies kann ein Indikator für den Wechsel von "illegalem" Geld in "legalen" Besitz sein.

2014_12_05_anonymisiert



powered by ORA-NetScenes

Abbildung 2.23: Egonetz Beratung x Off_Shore_Beratung

Auf Abbildung 2.23 kann man erkennen, dass für die Beratungen im Wesentlichen wirtschaftliche Konzepte relevant sind. Im Speziellen scheinen die wichtigen Themen Steuern, Darlehen und Schulden bzw. Gebühren zu sein. Verwandte Tasks scheinen Kauf, Verkauf und sonstige finanzielle Beteiligungen zu sein. Daher liegt hier nahe, dass sich die Geldwäscher durchaus finanziell beraten lassen haben.

Betrachtet man das Konzept "Notar", welches aus der inhaltlichen Auseinandersetzung mit dem Text als relevant erscheint, fällt auf, dass es viele Verbindungen zu Transaktionsaufgaben besitzt. Ebenfalls tauchen in seinem Netz Konzepte wie "Genehmigung", "Vollmacht", "Hypothek", "Kaufvertrag" und weitere Dokumente auf. Ebenfalls besitzt es eine schwache Verbindung zum Konzept "kriminell". Nimmt man diese Verbindungen zusammen, liegt der Schluss nahe, dass zumindest ein Notar in gewissem Umfang an den Geldwäscheaktivitäten beteiligt ist. Durch ihn könnten Beglaubigungen erstellt worden sein, welche Ansprüche oder Nachweise legitimieren sollen.

2.2.2.4.5 Cliquenanalyse

Um einen schnellen Überblick über Gruppierungen zu erhalten, empfiehlt sich die Anwendung des Newman Algorithmus zur Erstellung von Clustern. Innerhalb dieser Arbeit konnte der Newman Algorithmus 115 verschiedene Gruppen identifizieren. Allerdings kann man einstellen, dass er vor der Erstellung der Cluster Einzelfälle, also Werte die keine Verbindung zu andern Werten besitzen, entfernt. Dadurch entsteht eine Reduktion auf 14 Cluster, die mehr als ein Konzept besitzen.

Allerdings fällt hier auf, dass, wenn man sich den Report als HTML und als CSV Dateien ausgeben lässt, die Reports nicht miteinander übereinstimmen. Vergleicht man die Inhalte von Anhang 3 und Anhang 4, lässt sich erkennen, dass in der CSV Datei keine Konzepte gibt, die zu Gruppen 12 und 13 zugeordnet werden und dass Gruppe 14 mehr als 2 Konzepte besitzt.

Daher wurde für diese Arbeit eine Cliquenanalyse zur Betrachtung inhaltlich relevanter Gruppen angewandt.

Eine Cliquenanalyse ist in ORA durchführbar, indem man den Menüpunkt "Analysis" - "Generate Reports" - "Characterize Groups and Networks" - "Locate Groups" wählt und dann den Algorithmus "Locate Patterns" anwendet. Dieser errechnet je nach Einstellungen alle Cliquen einer durch den Benutzer festgelegten Minimalgröße.

Eine Cliquenanalyse der Mindestgröße 30 und 50 zeigte jedoch dieselben Ergebnisse, was verwunderlich ist. In beiden Cliquen-Reports zeigen sich 67867 verschiedene Gruppen, die ORA identifiziert. In je beiden Reports wird die Aussage getroffen, dass jeweils die kleinste Gruppe aus 3 Konzepten besteht und die größte Gruppe aus 39 Konzepten. Bei einer Mindestgröße von 30 wäre jedoch anzunehmen, dass nur diejenigen Gruppen zwischen 30 und 39 Konzepten berücksichtigt werden, und bei einer

Mindestgröße von 50 wären, aufgrund der höchsten identifizierten Gruppe von 39, gar keine Gruppen zu erwarten.

Nachdem in der ORA Gruppe auf diesen Fehler hingewiesen wurde²¹, erklärten die Entwickler, dass es ein Fehler im Programm ist. Eine Möglichkeit dennoch relevante Gruppen zu erstellen ist es sowohl die erwartete Cliquengröße als auch die "hidden links" auf denselben Wert zu setzen. Für diese Arbeit wurden als Input für die Cliquenanalyse folgende Knotenklassen verwendet: "Agent", "Attribute", "Geld", "Location", "Oranization", "Resource" und "Task". Dabei wurden die quadratischen Netze (Klasse1 x Klasse1) von "Location", "Attribute", "Geld", "Resource" und "Task" ausgegrenzt, das lediglich die Interaktion zwischen Personen und zwischen Organisationen relevant erscheinen.

Ein Wert von 15 Konzepten als Minimalgröße der Cliquen führt zu 30 Cliquen der Größe 15 und eine Clique der Größe 18. Folgende Tabelle zeigt die identifizierten Cliquen und die Aufteilung auf die untersuchten Knotenklassen. Extrahiert man die Daten in Excel und Summiert die Zuweisungen zu den Cliquen der einzelnen Konzepte ist ersichtlich, dass lediglich 45 unterschiedliche Konzepte zu diesen Cliquen zugeordnet wurden.

Tabelle 2-10: Beispielinhalt (Clique 1)

organization-426
organization-428
€_1600000
kreditinstituten
organization-166
organization-170
organization-201
organization-224
organization-255
organization-265
organization-280
organization-300
organization-305
organization-313
organization-322
organization-406
organization-416
organization-423

²¹ <https://groups.google.com/forum/#!topic/ora-google-group/OK5SFokjLrU> (Letzter Zugriff: 15.12.2014)

Innerhalb dieser Cliques befinden sich 21 Unternehmen (davon 4 Kreditinstitute), 5 Agenten, 13 Ressourcen (Geld und Aktien), die Ressource "daten_kabel", das Konzept "Übertragungsgeschwindigkeit" und 4 Ortsangaben (Adressen).

Auffällig dabei ist, dass Organisation-426 in alle 31 Cliques eingeordnet wurde und Organisationen 57, 84, 256 und 442 in 30 Cliques, alle anderen Unternehmen sind lediglich Clique 1 zugewiesen. Zusätzlich dazu taucht der Geldbetrag "€_9000" in 30 Cliques auf und die Konzepte "€_1728_06" und "mehrere_millionen_gulden" in 26 der Cliques. Zusammengenommen könnten diese Informationen ein Indiz dafür sein, dass zumindest die oben genannten Unternehmen eine Verbindung zu den Geldbeträgen besitzen. Da dies, wie zu sehen ist, keine runden Beträge sondern sehr spezifische Nennungen sind, könnte dies die Vermutung untermauern, dass dieselben Beträge durch alle Unternehmen geflossen ist.

Clique 1 bildet in dieser Hinsicht die auffälligste Clique, da in ihr 17 Unternehmen ("Kreditinstituten" liegt in der Klasse "organization") und der Geldbetrag "€_1600000" auftauchen. Unter dem Gesichtspunkt des Geldflusses könnte dies ein Anhaltspunkt für Geldwäsche sein.

Clique 2 beinhaltet einige Unternehmen, den Agenten 331, die Geldbeträge "€_33333_33" und "€_9000", "599_aktien", das Konzept "daten_kabel" mehrere Adressen und das Konzept "Übertragungsgeschwindigkeit". Ebenfalls ein Indikator dafür, dass hier Geld und Aktien innerhalb einer bestimmten Zeit transferiert wurden.

Das Konzept "daten_kabel" taucht ebenfalls in 30 der 31 Cliques auf und könnte somit eine besonders wichtige Rolle gespielt haben. Allerdings muss dann nicht mehr Geldwäsche eine mögliche Ursache der Transaktionen sein, sondern die Fertigstellung eines "daten_kabel" Projektes könnte oben genannte Gelder erfordert haben.

Alle Cliques nach Clique 1 unterscheiden sich lediglich in der Zusammenstellung der Gelder und Aktien. In 30 Fällen sind, die im zweiten Abschnitt dieser Seite genannten, Unternehmen und das Konzept "daten_kabel" der wesentliche Inhalt der Gruppen. Ansonsten unterscheiden sich die Gruppen lediglich nach den Geldbeträgen und den Adressen.

Fazit dieser Cliquenanalyse ist somit, dass der Zusammenhang zwischen den Unternehmen und Geldern entweder auf die Geldwäsche oder auf eine "daten_kabel" Projekt zurückzuführen ist. Dieses Projekt könnte allerdings auch zur Geldwäsche genutzt worden sein.

3 Zusammenfassung der Ergebnisse

Innerhalb dieses Kapitels werden die Ergebnisse der einzelnen Forschungsschritte dargestellt, dabei werden zuerst die Ergebnisse der Textanalyse, dann die Ergebnisse der Netzwerkanalyse und zuletzt die der Ontologiebildung erläutert.

3.1 Ergebnisse der Textanalyse

An dieser Stelle werden die Forschungsfragen und Ziele von Seite 3 herangezogen. Die Forschungsfragen, welche die Textanalyse betreffen sind:

- FF1: Welche Konstellation an (halb-)automatischen Analysewerkzeugen ermöglicht eine inhaltliche Text- und Netzwerkanalyse?
- FF2: Wie können frei zugängliche Programme das gewünschte Ziel allein oder mit Zusatz von Eigenprogrammierung lösen?
- FF3: Welche Einheiten lassen sich aus dem Text extrahieren?

3.1.1 Forschungsfrage 1

Wie in Kapitel 2.1 geschildert, lassen sich viele verschiedene Tools zur Analyse von Texten nutzen. Die Einbeziehung der Tools ist vor allem abhängig durch die Präferenzen und den Ansatz des Forschungsthemas. In dieser Arbeit standen eine enge Bindung zum aktuellen Text und eine Vielfalt an Textbereinigungsmöglichkeiten im Vordergrund.

Dieser Fokus ließ von den betrachteten Tools nur Automap, KHCoder und Gate übrig. Wie bereits erwähnt, waren die betrachteten Tools auf Funktionsumfang ähnlich stark, sodass letztendlich subjektive Präferenzen die Auswahl der Tools entschieden haben. Dabei gab die Integration von Text- und Netzwerkanalyse im Fall von ORA und Automap den entscheidenden Vorsprung gegenüber den anderen Tools.

Eine gemeinsame Text- und Netzwerkanalyse ist möglich, wenn ein Textanalysetool geeignete Texteinheiten identifizieren und besonders kenntlich machen kann. Im Fall von Automap kann dies durch Taggen oder durch Erstellen einer Konzeptliste geschehen. Eine Konzeptliste enthält, alle Texteinheiten (Wörter oder n-grams) eines Textes nach den durchgeführten Bereinigungsschritten. Diese Texteinheiten müssen weiter verfeinert oder verallgemeinert werden können.

Für die hier betrachtete Forschungsarbeit ist der Ansatz die Analyse auf einer Metaebene durchzuführen. So werden zwar die genauen Daten aus der Datenquelle als Input geliefert und für eine Netzwerkanalyse genutzt, die Konzepte werden allerdings soweit verallgemeinert, dass generelle Aussagen

möglich sind. Ein Tool muss für diese Arbeit daher die Möglichkeit der Anonymisierung und der Klassenbildung unterstützen.

Um die Text- und Netzwerkanalyse miteinander zu verbinden, muss es möglich sein, alle Erkenntnisse der Textanalyse an das Netzwerkanalysetool zu übergeben. Dies beinhaltet die Konzepte, deren Relationen miteinander und eventuelle weitere Anmerkungen.

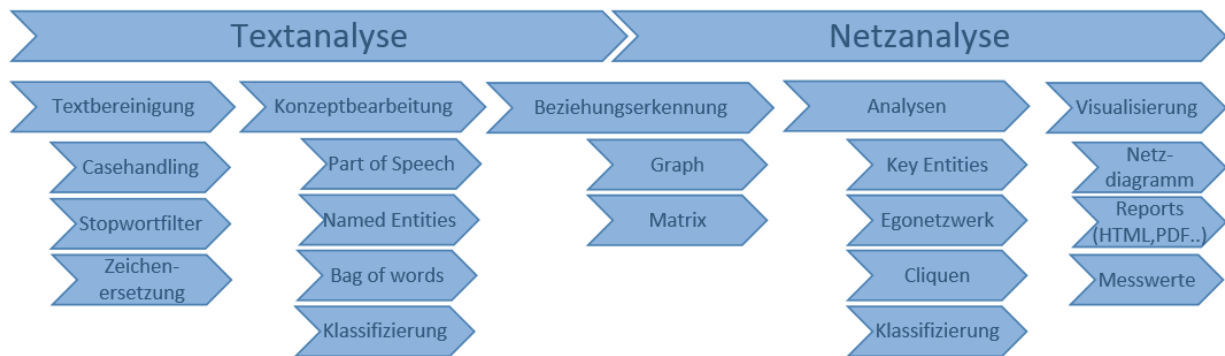


Abbildung 3.1: Anforderungen an Text- und Netzwerkanalysetools

Diese Arbeit sieht Anforderungen an Text- und Netzwerkanalysetools gemäß Abbildung 3.1. Ein Textanalysetool muss in der Lage sein, den Text in gewünschter Form zu bereinigen, Konzepte zu erkennen und diese Konzepte weiter zu bearbeiten, zu verfeinern oder zu generalisieren. Der wichtigste Punkt jedoch ist, dass das Textanalysetool eine Funktion zur Verfügung stellen muss, um nicht nur die Konzepte, sondern auch deren Relation miteinander, an das Netzwerkanalysetool zu übergeben, sei dies in Form einer Matrix, eines XML Formates oder eines Graphen.

3.1.2 Forschungsfrage 2

Die im vorherigen Unterkapitel genannten Funktionen werden in dieser Forschungsarbeit durch Automap und ORA erfüllt. Durch die Erstellung von Thesaurusdateien werden Konzepte mit Meta-Netzwerk Klassen versehen und anhand dieser Klassen werden die Konzepte miteinander in Verbindung gebracht.

Automap und ORA erfüllen die Übergabe der Konzepte und der daraus resultierenden Netze durch die DyNetML-Erweiterung von XML, wie in Kapitel 2.2.1.2.5 beschrieben. Automap unterstützt ebenfalls die gängigsten Methoden, die für eine Textanalyse erforderlich sind, dies jedoch nur in Englisch im vollen Umfang. In deutschen Texten, insbesondere in der betrachteten Datenquelle, erstellt Automap zwar Dateien mit "Part of Speech" Tags oder "Named Entities", diese sind jedoch nicht sonderlich hilfreich, da die sehr hohe Fehlerrate durch bloßes Auge auffällt.

Ein großer Vorteil von Automap im Rahmen der Textanalyse ist, dass man nach jedem durchgeführten Prozess eine Übersicht über den Text hat, sodass man die Auswirkung der einzelnen Methoden nachverfolgen kann. Ebenfalls kann man einen oder mehrere Schritte zurückspringen, falls ein Fehler aufgetreten ist.

Im Text werden verschiedene Unternehmen, Personen und Verben erkannt und für die Analyse genutzt. Allerdings werden die Ergebnisse auf den eingeführten Klassen der Konzepte dargestellt, sodass ein Textanalysetool für diese Arbeit eine Unterstützung für Klassifizierung bzw. Taggen von Texteinheiten bieten muss.

3.1.3 Forschungsfrage 3

Welche Texteinheiten genau identifiziert wurden, wird in dieser Sektion besprochen. Wie in Tabelle 2-2 erkennbar, besteht der hier untersuchte Text, nach der manuellen Vorbearbeitung, aus 152.334 Wörtern. Eliminiert man die Doppelnennungen und erstellt wie weiter oben beschrieben die Konzeptliste, erhält man 12.478 Konzepte. Die relevanten Konzepte dieser Analyse wurden zusammen mit den Experten des GLODERS Projektes iterativ den in Automap unterstützen Ontologieklassen zugeordnet:

Tabelle 3-1: Zuweisung der Thesauruseinträge zu den Ontologieklassen von Automap

Anzahl der Einträge im Thesaurus	Ontologieklasse
286	Action
469	Agent
290	Attribute
906	Event
23	Group
18	Knowledge
488	Location
651	Organization
1374	Resource
225	Role
315	Task
9	Time

Diese Einträge beziehen sich jeweils auf die Zeileneinträge in der Thesaurusdatei und können daher doppelte Konzepte beinhalten. Eine Übersicht über die extrahierten Konzepte, die aus dieser Thesaurusdatei erstellt wurden, findet sich in Kapitel 2.2.2.4.

Die Ontologiekategorie Unknown wird in der Netzwerkanalyse außen vor gelassen, da die relevanten Konzeptklassen für Geldwäsche in Zusammenarbeit mit Experten des GLODERS-Projektes identifiziert bzw. sich auf einen für die Analyse genügenden Umfang an Konzepten geeinigt wurde. Diese Konzept-

zuweisung wird zwar an Automap übergeben, ist aber nicht gleichzustellen mit der Zuweisung, die im Netzwerk entsteht. In dieser Arbeit wurden nie alle Konzepte so klassifiziert, wie sie laut Thesaurus zu klassifizieren sind.

3.2 Ergebnisse Netzwerkanalyse

Die Forschungsfragen, die den Bereich der Netzwerkanalyse betreffen, sind folgende:

- FF2: Wie können frei zugängliche Programme das gewünschte Ziel allein oder mit Zusatz von Eigenprogrammierung lösen?
- FF3: Welche Einheiten lassen sich aus dem Text extrahieren?
- FF4: Wie stehen diese Einheiten miteinander im Verhältnis?

3.2.1 Forschungsfrage 2

ORA beinhaltet viele verschiedene Schnittstellen zu existierenden Programmen bzw. Formaten, in denen Netzwerke dargestellt werden können. Zu den Programmen gehören beispielsweise Pajek, UCINET, Analyst Notebook, PenLink, Path Finder, Thing Finder und einige weitere.

Ebenfalls ist es möglich, Netze über einfache Tabellen oder XML Dateien zu laden. Hier bietet die Schnittstelle zu Automap einen Vorteil gegenüber anderen Tools. Dadurch, dass durch Automap Netzwerke zugeschnitten auf ORA generiert werden können, ist der Import in wenigen Schritten erledigt.

Eine besondere Stärke ist hier, dass die Thesaurusdateien, welche Automap nutzt, in ORA ebenfalls geöffnet bzw. bearbeitet werden können. Daher können zu einem sehr frühen Stadium der Netzanalyse inhaltliche Konzeptgruppen erkannt und miteinander verglichen werden. Einträge einzelner Gruppen lassen sich problemlos in andere Gruppen verschieben, sollte dies nötig sein.

Lediglich die standardisierten Prozessabläufe können ORA gegenüber anderen Tools "schlechter" erscheinen lassen. Dadurch, dass der Fokus von ORA auf Organisationsdaten liegt, welche in unterschiedliche Gruppen aufgeteilt werden können, sind einige Reports darauf ausgelegt, diese Unter-netzwerke zu analysieren. Es waren nur wenige Methoden ersichtlich, welche das Meta-Netzwerk als gesamtes analysieren konnten. Wie beispielsweise in Kapitel 2.2.2.4.2 geschildert, konnten lediglich die Strukturen der quadratischen Netzwerke betrachtet werden.

ORA bietet hier viele Methoden, welche eine Netzwerkanalyse ermöglichen, so wie Clusterbildung, Cliquenanalysen, Egonetzwerke und "einfache" Maßzahlenberechnung.

3.2.2 Forschungsfrage 3

Wie in 2.2.1.2.4 geschildert, lassen sich vor allem die in ORA vertretenen Klassen "Agent", "Organization", "Knowledge", "Resource", "Belief", "Event", "Task", "Location", "Role", "Action" und "Unknown" betrachten. Alle darüber hinausgehenden Konzepte müssen auf eine dieser Klassen heruntergebrochen werden können. Man ist in der Lage, neue Klassen zu erstellen, welche allerdings immer einen der oben genannten Typen als Oberklasse besitzen müssen.

Weiter lassen sich beliebige Netzwerke erstellen. Dies kann entweder manuell im Visualizer geschehen oder über die Berechnung von Maßzahlen. Zusätzlich zu den eigentlichen Textbausteinen, welche vor allem für die Textanalyse relevant sind, lassen sich aber Relationen zwischen diesen Konzepten betrachten und verändern.

Im Visualizer von ORA lassen sich existierende Netze anhand von bestimmten Werten auch verfeinern. In dieser Arbeit wurde häufig der Wert der anzuzeigenden Links auf über "10" gesetzt. Das bedeutet, dass der Visualizer nur Relationen anzeigt, die eine berechnete Bedeutung von 10 oder höher besitzen. So lassen sich schrittweise Subnetzwerke auch manuell erstellen.

Wie im Kapitel zur Analyse der Egonetzwerke zu erkennen, können ebenfalls mehrere Egonetzwerke miteinander verbunden werden, was eine Art Cliquenansicht bietet. Der Nutzer kann die Reichweite des Egonetzes erweitern oder vermindern. Ebenfalls kann der Nutzer sagen, er möchte nur die Schnittmenge von zwei Egonetzwerken sehen. Dies bietet den Vorteil, dass nur Konzepte angezeigt werden, welche Verbindungen zu allen Egokonzepten besitzen.

Durch die Anonymisierung der Daten geht in der Visualisierung der Ergebnisse einiges an Inhalt verloren. So im Beispiel der Zusammenfassungen einzelner Konzepte. Für einen, mit dem Text nicht vertrauten, Leser ist nicht ersichtlich ob die Konzepte Organization 31 - 38 alle zur selben Unternehmensgruppe gehören oder nicht. Allerdings ist diese Tatsache für den Aspekt der Geldwäsche und damit für die Polizeiarbeit durchaus relevant, denn die einzelnen Ausprägungen eines Unternehmens (LTD, NV, BV, SA oder AG) können, innerhalb der genutzten Daten, ein Indikator für die Geldwäsche sein.

3.2.3 Forschungsfrage 4

Das Verhältnis der verschiedenen Konzepte zueinander liegt vor allem in der Textnähe, welche in Automap für die Erstellung der Netze genutzt wird. Allerdings lassen sich durch die Verbindungen interessante Konstellationen innerhalb des Netzwerkes identifizieren. Die Wichtigkeit von einzelnen Konzepten wird in Kapitel 2.2.2.4.4 dargestellt und dient als Grundlage für die im nächsten Kapitel erstellte Grundlagenontologie. Es lassen sich Knoten mit vielen Verbindungen und Knoten mit wenigen Ver-

bindungen identifizieren, ebenfalls werden ausgehende und eingehende Verbindungen voneinander unterschieden, wodurch sich eventuelle Prozessabläufe besser bestimmen lassen können.

3.3 Ergebnisse der Ontologiebildung

Für die Ontologiebildung sind folgende Forschungsfragen relevant:

- FF3: Welche Einheiten lassen sich aus dem Text extrahieren?
- FF4: Wie stehen diese Einheiten miteinander im Verhältnis?

An dieser Stelle werden die beiden Forschungsfragen mit Hilfe einer Grafik beantwortet. Diese enthält die wichtigsten Konzepte, die innerhalb dieser Arbeit bezüglich der Geldwäsche identifiziert wurden. Dabei liegt der Fokus auf dem Verschleierungsaspekt der Geldwäsche, also wie die Gelder über verschiedene Einheiten "legalisiert" werden. Dabei werden unter anderem direkte Konzepte aus dem Netzwerk genutzt oder von ihnen abstrahiert. So werden beispielsweise Personen in Sender und Empfänger unterschieden, je nachdem, ob sie eine bestimmte Aufgabe ausführen oder das Ziel einer bestimmten Aufgabe sind. Wie eingangs erwähnt, handelt es sich an dieser Stelle lediglich um eine Grundlagenontologie, die die Zusammenhänge von verschiedenen potentiell wichtigen Konzepten im Bereich der Geldwäsche grob darstellen soll. Weitere Arbeiten auf diesem Feld sollten daher die hier geschilderte Ontologie eher als Einstiegspunkt für eine noch einzuführende formal korrekte Ontologie betrachten.

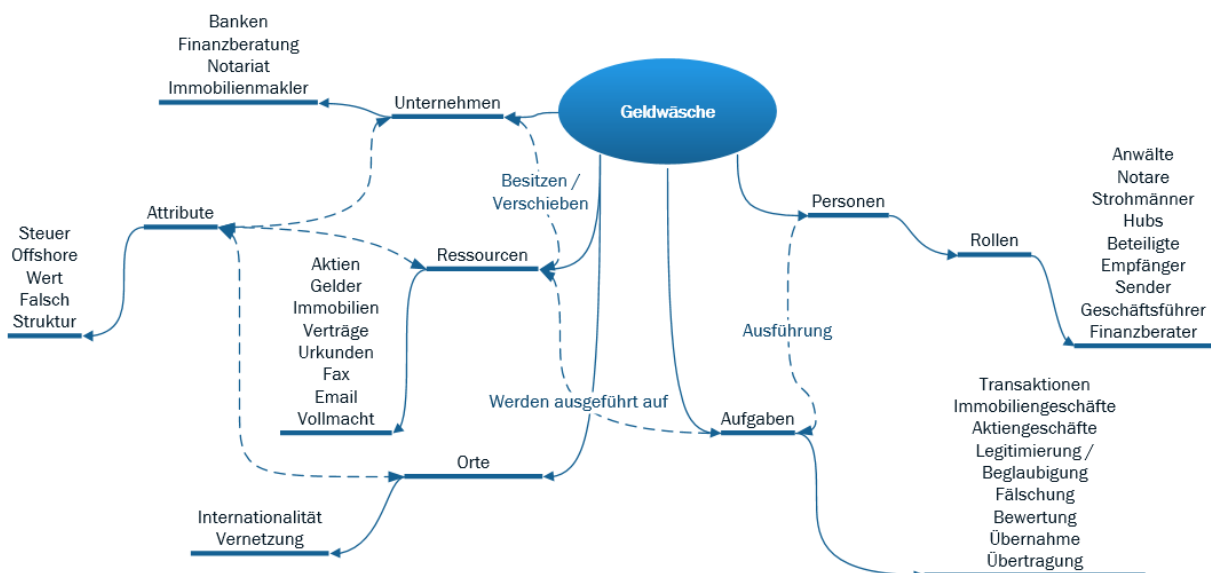


Abbildung 3.2: Grundlagenklassifizierung im Bereich Geldwäsche

4 Fazit und Ausblick

Innerhalb dieses Kapitel werden kurz die Ergebnisse des vorherigen Kapitels zusammengefasst und eine Wertung der vorgenommenen Schritte, der Restriktionen dieser Arbeit, der möglichen weiteren Forschungsarbeiten und eine Wertung der Tools gegeben.

4.1 Inhaltliches Fazit

Die Erkenntnisse der Geldwäsche, welche durch die Polizei im Bericht festgehalten wurden, konnten durch die inhaltliche Nähe der Konzepte zueinander zumindest untermauert werden. Weiter ließen sich Beziehungen zwischen Unternehmen, Personen und Geldbeträgen aufzeigen, welche der Polizei entweder nicht bekannt oder zumindest nicht in einer kompakten Darstellung präsent sein könnten.

Für die Geldwäsche scheinen besonders international agierende Personen und Unternehmen wichtig zu sein. Häufig wurden Steuern, Rechnungen oder Finanzierungen in den Netzen gesehen. Die Existenz der Konzepte "Fehler", "Falsch", "Urkunde" und "Urkundenfälschung", untermauert zusammen mit den Beziehungen dieser Konzepte zum Konzept „Rechnung“ den Schluss, dass immer eine Form von Nachweis erbracht wird, der die Herkunft von Geldern oder Ressourcen legitimieren soll, aber nicht immer rechtens ist.

Es ließen sich somit einige Eigenheiten von Geldwäschekonzepten erkennen. Eine Prüfung dieser Zusammenhänge durch eventuell andere Data Mining Tools oder andere Textgrundlagen könnte diese Ergebnisse eventuell validieren.

Eine konkrete Existenz von Geldwäsche bzw. die Existenz der Beteiligung bestimmter Personen oder Organisationen kann allerdings nicht gewährleistet werden, da diese Arbeit einigen Restriktionen unterliegt, welche im nächsten Abschnitt erläutert werden. Alle inhaltlichen Aussagen dieser Arbeit sind daher hypothetischer Natur.

4.2 Restriktionen der Arbeit

Wie an einigen Stellen der Arbeit erwähnt, unterliegt diese Arbeit bestimmten Restriktionen. Einerseits sind die Daten, die für die Analyse genutzt wurden, in sehr unstrukturierter Form geliefert worden, sodass viel Zeit in die manuelle Aufbereitung der Textdatei aufgewendet wurde. Andererseits unterliegen die Daten einem bestimmten inhaltlichen Fokus, welcher sich aus der Polizeiarbeit ergibt. Durch diesen Fokus kann sich die Wichtigkeit einiger Konzepte innerhalb der Analyse gegenüber der wirklichen Wichtigkeit verschieben. Ebenfalls basieren die Analysen auf der Nähe von Textpassagen untereinander, sodass die Erstellung der Dokumente und die Verschriftlichung der Interviews ebenfalls eine Rolle spielt.

Eine letzte Restriktion dieser Arbeit ist, dass lediglich deutsche Daten innerhalb der Programme analysiert wurden, was im folgenden Unterkapitel weiter beleuchtet wird.

4.3 Wertung der genutzten Tools

Sobald man etwas mit Automap und ORA und deren Menüführung vertraut ist, bieten die Programme eine einfache Möglichkeit für Nutzer, eine Text- und Netzwerkanalyse durchzuführen. Je nach Datenqualität und Ursprung der Daten kann sich der Nutzen dieser Programme aber erheblich ändern. Für Forschungen auf deutschen Textdaten, die nicht vorverarbeitet wurden und zu denen keine Thesauri bestehen, ist der Einsatz von Automap gut zu überlegen.

Innerhalb der Arbeit wurden an einigen Stellen Fehlernachrichten oder Unzulänglichkeiten bei der Dokumentation der Programme aufgezeigt. Bewertet man die Zeit, die man in Fehlerbehebung bzw. -identifikation investiert, steigt der manuelle Aufwand gegenüber der erwünschten Automatisierung stark an. Innerhalb dieser Forschungsdatei wurden die Thesaurusdateien komplett manuell erstellt und lediglich durch Automap auf Fehler geprüft und von Automap sortiert.

Für englische Texte wurden die Prozesse dieser Analysen allerdings bis zu einem bestimmten Punkt nachvollzogen und die Ergebnisse, welche Automap automatisch erstellen kann, sehen gegenüber den Ergebnissen aus der deutschen Analyse sehr vielversprechend aus. Der Vorteil dieser Tools wird sich daher wohl lediglich in englischen Analysen wirklich auswirken.

Weiter lassen sich die Reports, welche mit ORA erstellt werden können, nicht beliebig anpassen. Die standardisierten Reports müssen daher je nach Forschungsfrage manuell weiterverarbeitet werden. So lässt sich beispielsweise für die Analyse der Schlüsselkonzepte und ihrer Egonetzwerke keine Aggregation der Inhalte in neue Form generieren. Alle Tabellen, die die Aussagen der Reports zusammengefasst haben, sind manuell entstanden.

Ebenfalls lassen sich in einigen Reports Unstimmigkeiten entdecken, welche beispielsweise die Gruppierung betreffen. So war es nicht möglich, stichhaltige Gruppen von ORA automatisch erstellen zu lassen. Die Tatsache, dass trotz unterschiedlicher Einstellungen gleiche Reports erstellt wurden bzw. die Reports sich je nach Dateiformat widersprachen, führte dazu, dass eine Cliquenanalyse in dieser Arbeit nicht möglich war.

4.4 Ausblick

Die inhaltlichen Ergebnisse, die in dieser Arbeit erzielt wurden, sind zum Teil sehr aussagekräftig und eignen sich für weitere Forschungsarbeiten. Mögliche Aufgaben für diese Anschlussarbeiten könnten entweder eine Validierung bzw. Gegenprüfung der Ergebnisse mit anderen bekannten Data Mining Tools oder Analysen auf englischsprachigen Textkörpern sein.

Ein weiterer Punkt ist die scheinbar fehlerhafte Funktion der Cliquenanalyse bzw. der Clustererstellung. Für diesen Punkt lohnt es sich, mit bewährten, eventuell auch kommerziellen, Data Mining Tools eine Analyse der hier betrachteten Daten einzubringen.

In dieser Arbeit wurden die relevanten Schritte aufgezeigt, die durch die Textanalyse zur Netzwerkanalyse führen. Eine Anwendung von englischen Daten würde daher zeigen ob die Anzahl der Fehler abnimmt und somit auf eventuelle Sprachprobleme der Tools zurückzuführen ist oder ähnliches.

Die Ergebnisse dieser Arbeit werden in Klartext den Experten des GLODERS Projektes übergeben, die die Ergebnisse mit den Stakeholdern der niederländischen Polizei besprechen werden.

Ein weiterer Punkt, der in folgenden Forschungsschritten durchgeführt werden könnte, ist die Fokussierung auf den Teilbereich der Ontologie. Hierfür könnten die Konzepte, die in der Grundlagenontologie enthalten sind, einen ersten Einstieg geben.

5 Literaturverzeichnis

Gesetze im Internet - Strafgesetzbuch. §261. Online verfügbar unter http://www.gesetze-im-internet.de/stgb/___261.html, zuletzt geprüft am 26.06.2014.

Gesetze im Internet - Strafgesetzbuch. §255. Online verfügbar unter http://www.gesetze-im-internet.de/stgb/___255.html, zuletzt geprüft am 28.09.2014.

Gesetze im Internet - Strafgesetzbuch. §253. Online verfügbar unter http://www.gesetze-im-internet.de/stgb/___253.html, zuletzt geprüft am 28.09.2014.

Aggarwal, Charu C.; Zhai, ChengXiang (Hg.) (2012): Mining Text Data. 2012. Aufl. Boston, MA: Springer US. Online verfügbar unter <http://dx.doi.org/10.1007/978-1-4614-3223-4>.

Altenkirch, Lars (2002): Techniken der Geldwäsche und ihre Bekämpfung. 1. Aufl. Frankfurt am Main: Bankakademie-Verlag (Banking & finance aktuell, Bd. 10).

Bennett, Brandon; Fellbaum, Christiane (2006): Formal ontology in information systems. Proceedings of the fourth international conference (FOIS 2006). Amsterdam, Washington, DC: IOS Press (Frontiers in artificial intelligence and applications / FAIA, v. 150).

Bundeskriminalamt (2013): Organisierte Kriminalität. Bundeslagebericht 2013.

Cambridge Dictionaries Online: Extortion Racket. Online verfügbar unter <http://dictionary.cambridge.org/dictionary/business-english/extortion-racket>, zuletzt geprüft am 06.10.2014.

Carley, Kathleen M.; Columbus, Dave; Landwehr, Peter (2013a): Automap User's Guide 2013. CMU-ISR-13-105. Institute for Software Research Carnegie Mellon University. Online verfügbar unter www.casos.cs.cmu.edu/publications/papers/CMU-ISR-13-105.pdf, zuletzt aktualisiert am 07.06.2013, zuletzt geprüft am 17.11.2014.

Carley, Kathleen M.; Pfeffer, Jürgen; Reminga, Jeff; Storrack, Jon; Columbus, Dave (2013b): ORA User's Guide 2013. CMU-ISR-13-108. Institute for Software Research Carnegie Mellon University. Online verfügbar unter www.casos.cs.cmu.edu/publications/papers/CMU-ISR-13-108.pdf, zuletzt geprüft am 20.11.2014.

CASOS (Hg.): DyNetML. Center for Computational Analysis of Social and Organizational Systems; Carnegie Mellon University. Online verfügbar unter <http://www.casos.cs.cmu.edu/projects/dynetml/>, zuletzt geprüft am 27.11.2014.

Chapman, Pete; Clinton, Julian; Kerber, Randy; Khabaza, Thomas; Reinartz, Thomas; Shearer, Colin; Wirth, Rüdiger (2000): CRISP-DM 1.0. Step-by-step data mining guide. Tech. rep. SPSS. Online verfügbar unter <http://www.the-modeling-agency.com/crisp-dm.pdf>, zuletzt geprüft am 19.11.2014.

Cocchiarella, Nino B. (2007): Formal ontology and conceptual realism. Dordrecht: Springer (Synthese library, v. 339).

- Diesner, Jana (2012): Uncovering and Managing the Impact of Methodological Choices for the Computational Construction of Socio-Technical Networks from Texts. PHD Thesis. Carnegie Mellon University, Pittsburgh. Institute for Software Research.
- Diesner, Jana; Carley, Kathleen M. (2008): Revealing Social Structure from Texts. Meta-Matrix Text Analysis as a novel method for network text analysis.
- Diesner, Jana; Carley, Kathleen M.; Tambayong, Laurent (2012): Extracting socio-cultural networks of the Sudan from open-source, large-scale text data. In: *Comput Math Organ Theory* 18 (3), S. 328–339. DOI: 10.1007/s10588-012-9126-x.
- EUROPOL (2013): Threat Assessment. Italian Organised Crime.
- GLODERS.eu (Hg.) (2012): Concept and Objectives. Online verfügbar unter <http://www.gloders.eu/project/concept-and-objectives>, zuletzt geprüft am 28.08.14.
- Gruber, Thomas R. (1993): Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In: *International Journal Human-Computer Studies* (43), S. 907–928.
- Jansen, Dorothea (2003): Einführung in die Netzwerkanalyse. Grundlagen, Methoden, Forschungsbeispiele. 2. erweiterte Auflage. Wiesbaden, s.l.: VS Verlag für Sozialwissenschaften. Online verfügbar unter <http://dx.doi.org/10.1007/978-3-663-09875-1>.
- Jiang, Jing (2012): Information Extraction from text. In: Charu C. Aggarwal und ChengXiang Zhai (Hg.): Mining Text Data. 2012. Aufl. Boston, MA: Springer US, S. 11–35.
- Krukow, Oliver (2013): Concept Network Extraction from Text. Method and Tools for the Research into Extortion Racket Systems.
- Kun, Jeremy (2013): The Erdős-Rényi Random Graph. Online verfügbar unter <http://jeremykun.com/2013/08/22/the-erdos-renyi-random-graph/>, zuletzt geprüft am 14.12.2014.
- Lietz, Haiko (2007): Mit neuen Methoden zu neuen Aussagen. Semantische Netzwerkanalyse am Beispiel der Europäischen Verfassung. Working Paper. Universität Duisburg-Essen. Online verfügbar unter www.haikolietz.de/docs/verfassung.pdf, zuletzt aktualisiert am 05.09.07, zuletzt geprüft am 04.12.14.
- Manning, Christopher D.; Raghavan, Prabhakar; Schütze, Hinrich (2008): Introduction to information retrieval. New York: Cambridge University Press.
- Mcauley, Julian; Leskovec, Jure (2014): Discovering social circles in ego networks. In: *ACM Trans. Knowl. Discov. Data* 8 (1), S. 1–28. DOI: 10.1145/2556612.
- Mori, Shunji; Nishida, Hirobumi; Yamada, Hiromitsu (1999): Optical character recognition. New York: J. Wiley (Wiley series in microwave and optical engineering).
- Nenkova, Ani; McKeown, Kathleen (2012): A survey of text summarization techniques. In: Charu C. Aggarwal und ChengXiang Zhai (Hg.): Mining Text Data. 2012. Aufl. Boston, MA: Springer US, S. 43–66.

Nooy, Wouter de; Mrvar, Andrej; Batagelj, Vladimir (2005): Exploratory social network analysis with Pajek. New York: Cambridge University Press (Structural analysis in the social sciences).

Reichenberger, Klaus (2010): Grundlagen semantischer Netze. In: Klaus Reichenberger (Hg.): Kompendium semantische Netze. Berlin, Heidelberg: Springer Berlin Heidelberg (X.media.press), S. 3–19.

Sage, Andrew P.; Rouse, William B. (2009): Handbook of systems engineering and management. 2nd ed. Hoboken, N.J.: John Wiley & Sons (Wiley series in systems engineering and management).

Schneider, Friedrich; Dreer, Elisabeth; Riegler, Wolfgang (2006): Geldwäsche. Formen, Akteure, Größenordnung - und warum die Politik machtlos ist. 1. Aufl. Wiesbaden: Gabler.

Schweizer Bundesrat (2002): Lage- und Gefährdungsanalyse Schweiz nach den Terroranschlägen vom 11. September 2001. Bericht des Bundesrates an das Parlament.

TRANSCRIME (2009): Study on Extortion Racketeering. the Need for an Instrument to Combat Activities of Organised Crime.

Tsvetovat, Maksim; Reminga, Jeff; Carley, Kathleen M. (2004): DyNetML: Interchange Format for Rich Social Network Data. CMU-ISRI-04-105. Pittsburgh, Februar 2004. Online verfügbar unter www.casos.cs.cmu.edu/.../CMU-ISRI-04-105.pdf, zuletzt geprüft am 24.11.2014.

United Nations Office on Drugs and Crime (Hg.): The Money-Laundering Cycle. Online verfügbar unter <https://www.unodc.org/unodc/en/money-laundering/laundrycycle.html>, zuletzt geprüft am 27.11.2014.

United Nations Office on Drugs and Crime (2011): Criminal Intelligence. Manual for Analysts. New York.

Wasserman, Stanley; Faust, Katherine (1994): Social network analysis. Methods and applications. Cambridge, New York: Cambridge University Press (Structural analysis in the social sciences, 8).

Anhang

	A	B	C	D	E	F	G	H	I	J	K	L
1				Agents						organization		
2												
3	Total Degree Centrality	1	2	3	4	5		1	2	3	4	5
4	Hub Centrality	Agent-344	Agent-346	Agent-174	Agent-2	Agent-188		Organization-31	Organization-447	bank	Organization-256	Organization-215
5	Betweenness	Agent-174	Agent-346	Agent-2	Agent-344	Agent-188		Organization-215	Organization-46	Organization-447	Organization-31	Organization-99
6	Eigenvektor	Agent-344	Agent-346	Agent-2	Agent-174	Agent-188		Organization-31	Organization-215	Organization-447	Organization-46	Organization-99
7	Closeness	Agent-344	Agent-181	Agent-346	Agent-119	Agent-231		Organization-31	Organization-36	Organization-31	Organization-46	bank
8				Attributes						knowledge		
9												
10	Total Degree Centrality	1	2	3	4	5		1	2	3	4	5
11	Hub Centrality	unternehmen	betrag	signatur	struktur	wert		steuerrecht	oft_shore_beratung	befugnisse	immobilien_welt	hypothekpfandrecht
12	Betweenness	unternehmen	struktur	auslandisch	bedingung	wert		befugnisse	steuerrecht			
13	Eigenvektor	unternehmen	bedingung	signatur	betrag	finanzuell		befugnisse	steuerrecht	finanzpolitik	finanzwelt	hypothekpfandrecht
14	Closeness	unternehmen	bedingung	signatur	betrag	finanzuell		befugnisse	steuerrecht			
15				task						resource		
16												
17		1	2	3	4	5		1	2	3	4	5
18	Total Degree Centrality	zahlung	verkauf	vereinbarung	fax	erpressung		aktien	rechnung	daten_kabel	geld	konto
19	Hub Centrality	zahlung	vereinbarung	verkauf	erpressung	kauf		aktien	rechnung			
20	Betweenness	zahlung	bezahlung	beteiligung	verkauf	vereinbarung		geld	aktien	eigentum	bankkonten	rechnung
21	Eigenvektor	zahlung	verkauf	vereinbarung	erpressung	tax		rechnung	aktien	aktien	daten_kabel	gewinn
22	Closeness	zahlung	gegrundet	kontakt	bezahlung	verkaufserfung		gebäude	kosten	geld	grundstück	credit
23												
24												
25				geld						location		
26		1	2	3	4	5		1	2	3	4	5
27	Total Degree Centrality	währung_855000	währung_10000000	währung_87500	währung_5000000	€_607_70		amsterdam	Location-58	Location-59	Location-126	Location-57
28	Hub Centrality	-	-	-	-	-		amsterdam	amsterdam	Location-59	Location-126	Location-126
29	Betweenness	währung_750000	währung_1500000	währung_1000000	währung_2000000	währung_12000000		amsterdam	linnuden	Location-57	Location-58	Location-126
30	Eigenvektor	währung_15000	währung_6000	währung_60000	währung_750000	nlg_827957		amsterdam	Location-59	Location-57	amsterdam	Location-55
31	Closeness	währung_750000	währung_1500000	währung_1000000	währung_220000	währung_29130000		amsterdam	Gemeinde	Location-222	Location-223	Location-224

Anhang 1: Top 5 Key Entities der relevanten Knotenklassen

Anhang 2: Context Report (Zusammengefasst)

Resource x Resource	968	Betweenness	0.222	0.015	6.334	no	0.002	6.756	no	0.000	6.798	no	0	6.805	no	0	6.805	no	0	6.805	no
		Closeness	0.021	-0.064	208.425	no	-0.001	56.396	no	0.001	50.868	no	0	54.362	no	0	54.362	no	0	54.362	no
		Inverse Closeness	14.654	0.099	30.506	no	0.263	30.163	no	0.003	30.709	no	0	30.714	no	0	30.714	no	0	30.714	no
Unknown x Unknown	142	Betweenness	0	0.000	Means test could not be computed	0.000	Means test could not be computed	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Closeness	1.389	0.001	3.639.796	no	0.001	3.639.723	no	0	3.641.891	no	0	3.641.891	no	0	3.641.891	no	0	3.641.891	no
		Inverse Closeness	0.039	0.001	1.406	yes	0.001	1.389	yes	0	1.424	yes	0	1.424	yes	0	1.424	yes	0	1.424	yes
Group x Group	30	Betweenness	0	0	0	yes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Closeness	6.302	0.083	174.777	no	0.084	174.758	no	0	177.126	no	0	177.126	no	0	177.126	no	0	177.126	no
		Inverse Closeness	0.833	0	1.464	yes	0.009	1.448	yes	0	1.464	yes	0	1.464	yes	0	1.464	yes	0	1.464	yes

SL	Zentralitäts Maß	Wert	Erdos Renyi			Core Periphery			Cellular (meanCellSI- ze=4)			Cellular (meanCellSI- ze=10)			Cellular (meanCellSI- ze=18)			Cellular (meanCellSI- ze=25)			
			Rand.	T-Value	Sa-me	Rand.	T-Value	Sa-me	Rand.	T-Value	Sa-me	Rand.	T-Value	Sa-me	Rand.	T-Value	Sa-me	Rand.	T-Value	Sa-me	
Geld x Geld	1432	Betweenness	0.002	0.001	1.702	yes	0.000	1.913	yes	0	2.349	no	0	2.349	no	0	2.349	no	0	2.349	no
		Closeness	0.012	0.003	706.529	no	0.002	821.996	no	0	990.572	no	0	990.572	no	0	990.572	no	0	990.572	no
		Inverse Closeness	0.476	0.006	10.489	no	0.028	10.001	no	0	10.626	no	0	10.626	no	0	10.626	no	0	10.626	no
Attribute x Attribut	344	Betweenness	0.603	0.026	4.238	no	0.003	4.408	no	0	4.428	no	0	4.428	no	0	4.428	no	0	4.428	no
		Closeness	0.035	0.015	17.084	no	0.012	21.634	no	0	32.470	no	0	32.470	no	0	32.470	no	0	32.470	no
		Inverse Closeness	15.16	0.131	17.999	no	0.181	17.939	no	0	18.155	no	0	18.155	no	0	18.155	no	0	18.155	no
Knowledge x Knowledge	30	Betweenness	0	0	0	yes	0	0	yes	0	0	yes	0	0	yes	0	0	yes	0	0	yes
		Closeness	6.302	0.083	174.777	no	0.084	174.758	no	0	177.126	no	0	177.126	no	0	177.126	no	0	177.126	no
		Inverse Closeness	0.833	0	1.464	yes	0.009	1.448	yes	0	1.464	yes	0	1.464	yes	0	1.464	yes	0	1.464	yes
Agent x Agent	712	Betweenness	0.376	0.024	4.796	no	0.002	5.087	no	0.009	5.000	no	0	5.118	no	0	5.118	no	0	5.118	no
		Closeness	0.003	-0.028	107.877	no	0.006	-43.211	no	0.053	-35.832	no	0	46.949	no	0	46.949	no	0	46.949	no
		Inverse Closeness	15.90	0.130	29.015	no	0.261	28.773	no	0.108	29.056	no	0	29.254	no	0	29.254	no	0	29.254	no
Event x Event	1782	Betweenness	0.007	0.000	4.950	no	0.000	5.003	no	0	5.111	no	0	5.111	no	0	5.111	no	0	5.111	no
		Closeness	0.008	0.001	488.326	no	0.001	523.341	no	0	584.270	no	0	584.270	no	0	584.270	no	0	584.270	no
		Inverse Closeness	0.546	0.003	13.270	no	0.013	13.031	no	0	13.337	no	0	13.337	no	0	13.337	no	0	13.337	no
Role x Role	378	Betweenness	0.392	0.002	5.035	no	0.001	5.043	no	0	5.058	no	0	5.058	no	0	5.058	no	0	5.058	no
		Closeness	0.041	0.010	31.492	no	0.008	33.754	no	0	41.639	no	0	41.639	no	0	41.639	no	0	41.639	no
		Inverse Closeness	6.559	0.021	13.112	no	0.071	13.012	no	0	13.155	no	0	13.155	no	0	13.155	no	0	13.155	no
Organization x Organization	942	Betweenness	0.368	0.015	6.753	no	0.002	7.011	no	0.000	7.035	no	0	7.041	no	0	7.041	no	0	7.041	no
		Closeness	0.013	-0.061	277.424	no	-0.000	56.140	no	0.002	45.127	no	0	54.525	no	0	54.525	no	0	54.525	no
		Inverse Closeness	14.74	0.098	35.211	no	0.257	34.829	no	0.004	35.436	no	0	35.446	no	0	35.446	no	0	35.446	no
Task x Task	758	Betweenness	0.310	0.036	5.759	no	0.003	6.445	no	0.016	6.168	no	0	6.507	no	0	6.507	no	0	6.507	no
		Closeness	0.007	-0.023	65.802	no	0.008	-10.110	no	0.098	-36.433	no	0	49.216	no	0	49.216	no	0	49.216	no
		Inverse Closeness	10.33	0.184	32.318	no	0.335	31.836	no	0.201	32.264	no	0	32.903	no	0	32.903	no	0	32.903	no
Location x Location	704	Betweenness	0.221	0.158	0.776	yes	0.003	2.685	no	0.072	1.842	yes	0.005	2.662	no	0	2.727	no	0	2.727	no
		Closeness	0.010	0.201	-60.921	no	0.109	-46.837	no	0.571	-50.518	no	0.106	-84.741	no	0	49.494	no	0	49.494	no
		Inverse Closeness	11.88	0.773	22.880	no	0.540	23.360	no	1.020	22.372	no	0.126	24.212	no	0	24.472	no	0	24.472	no

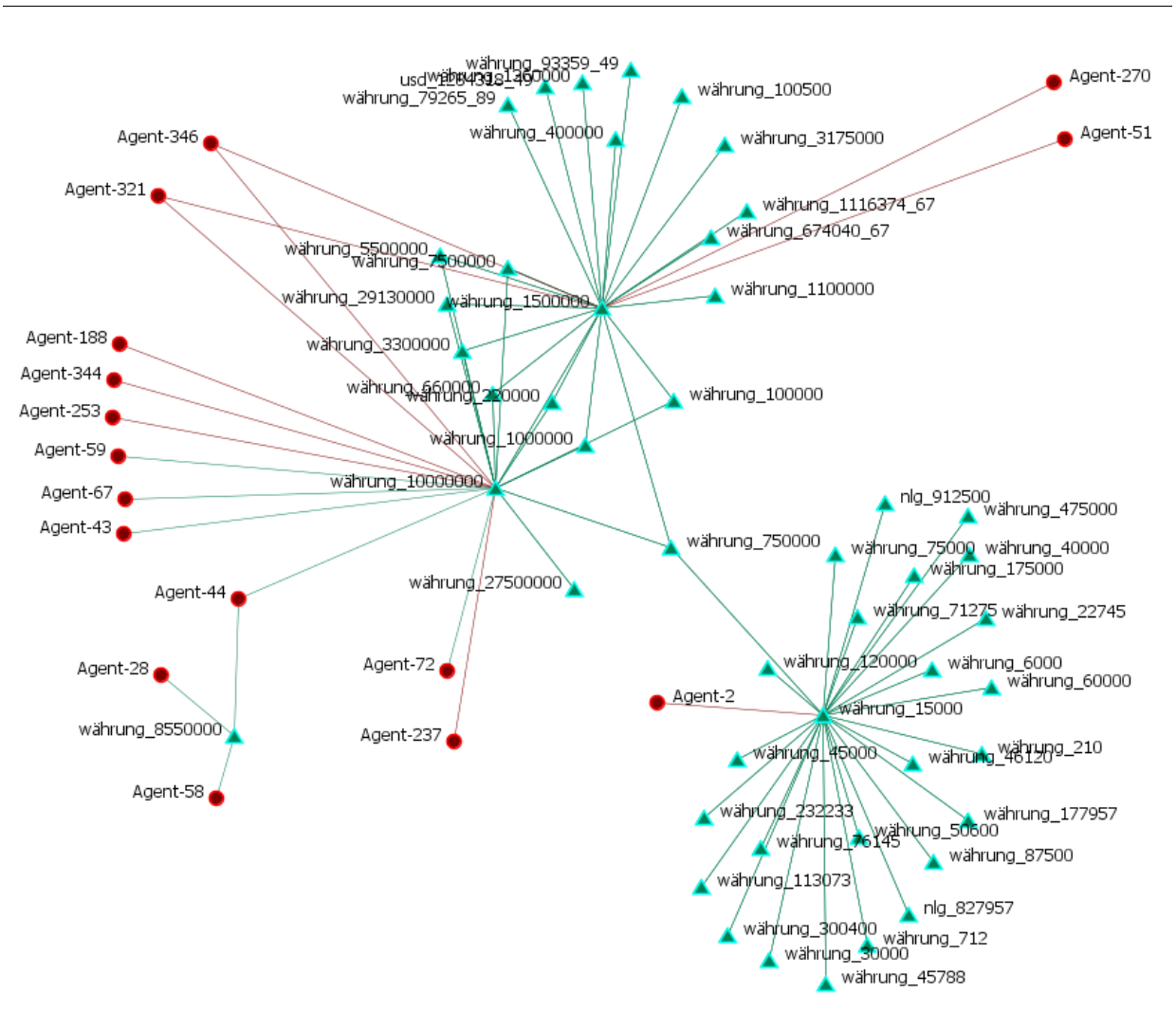
Group	Agent	Attribute	Event	Geld	Group	Knowledge	Location	Organization	Resource	Role	Task	Unknown	Density	Internal links	External links	Percent inner links	E/I index	Cheeger value
1	115	67	297	196	2	4	73	81	97	68	230	33	0.008	12487	12811	49.36%	0.013	10.143
2	87	39	305	241	11	4	121	148	153	32	14	12	0.014	19564	18370	51.57%	-0.031	15.741
3	54	23	98	120	1	5	19	89	93	10	14	3	0.015	4221	7937	34.72%	0.306	15.004
4	25	13	96	41	0	0	82	86	75	30	36	6	0.007	1747	5815	23.10%	0.538	11.867
5	26	13	56	85	0	0	37	40	41	21	15	3	0.028	3159	7027	31.01%	0.380	20.852
6	22	7	14	10	2	1	2	7	4	6	2	0	0.046	268	701	27.66%	0.447	9.104
7	5	0	1	3	0	0	7	2	3	3	21	14	0.253	860	1055	44.91%	0.102	17.881
8	9	1	12	7	0	2	6	6	9	3	2	0	0.085	270	986	21.50%	0.570	17.298
9	3	4	6	1	0	0	5	1	1	0	3	0	0.234	129	301	30%	0.400	12.542
10	3	3	0	0	0	0	0	0	0	0	0	0	1	30	10	75%	-0.500	1.667
11	3	0	0	1	0	0	0	0	0	0	0	0	0.250	3	41	6.82%	0.864	10.250
12	0	0	2	0	0	0	0	1	1	0	0	0	0.250	3	23	11.54%	0.769	5.750
13	0	0	1	0	0	0	0	2	0	0	0	0	0.667	4	17	19.05%	0.619	5.667
14	1	0	0	0	0	0	0	0	0	0	1	0	0.500	1	0	100%	-1	0

Anhang 3: Newman Cluster

	A	B	C	D
1	Nodeset Name	Node Name	Node Title	newman
2	Agent		Agent-79	14
3	Attribute		international	14
4	Event		Event-76	14
5	Geld		nlg_550000	14
6	Location		bosnien	14
7	Organization		Organization-76	14
8	Resource		baukredit	14
9	Role		hauptclient	14
10	Task		client_registrierung	14
11	Agent		Agent-21	11
12	Agent		Agent-31	11
13	Agent		Agent-170	11
14	Attribute		beglaubigte	11
15	Attribute		business	11
16	Attribute		abhängig	11
17	Event		Event-20	11
18	Event		Event-29	11
19	Event		Event-167	11
20	Geld		nlg_53693701_98	11
21	Geld		nlg_100000000	11
22	Geld		usd_15010	11
23	Location		alkmaar	11
24	Location		Location-29	11

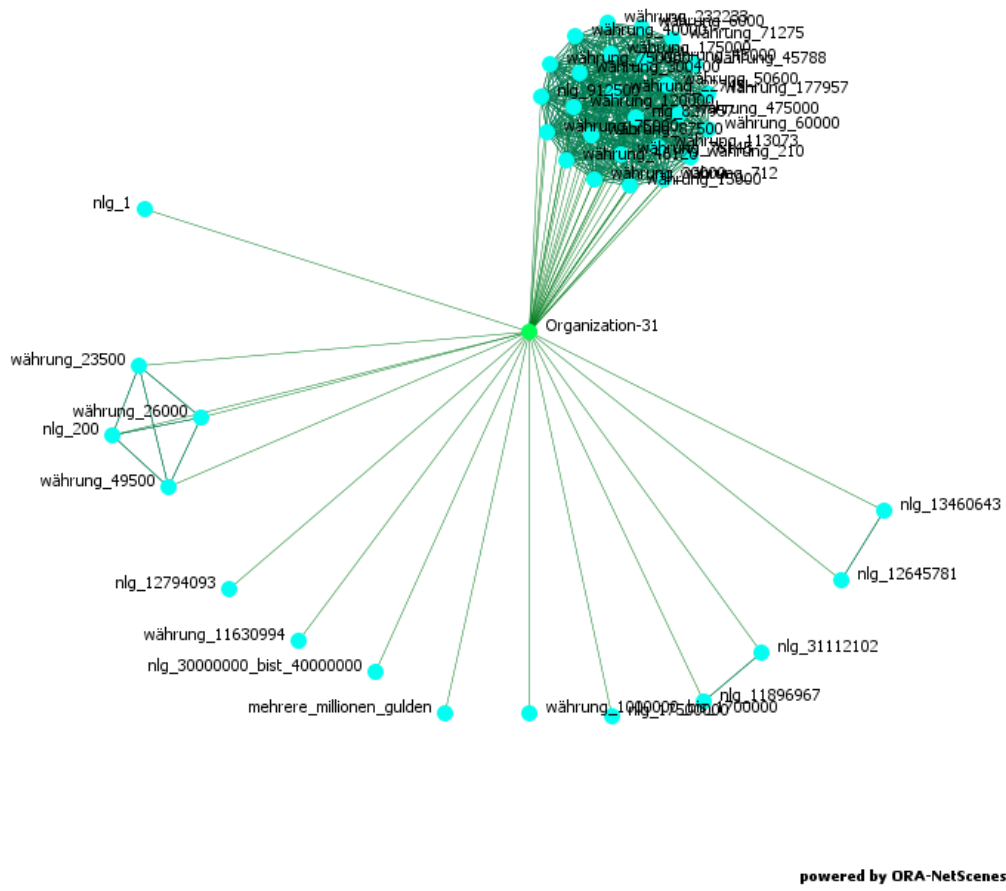
Anhang 4: Newman Cluster CSV

(Sortiert nach newman absteigend, Nodeset Name aufsteigend, Spalte Node Name aus Datenschutzgründen geleert)



Anhang 5: Verbindungen von Geldern und Agenten

2014_12_05_anonymisiert



Anhang 6: Verbindungen von Geldern und Organisation 31