

# Analyse der Netzwerkstruktur des österreichischen E-Tourismus- Marktes und Bestimmung der zentralen Knoten

DIPLOMARBEIT

zur Erlangung des akademischen Grades

**Diplom-Ingenieur**

im Rahmen des Studiums

**Wirtschaftsinformatik**

eingereicht von

**Roland Piazza**

Matrikelnummer 0026785

an der  
Fakultät für Informatik der Technischen Universität Wien

Betreuung

Betreuer: Univ.Prof. Dipl.-Ing. Dr.techn. Hannes Werthner

Mitwirkung: Univ.Ass. Mag.rer.nat. Julia Neidhardt

Wien, 08.02.2011

\_\_\_\_\_  
(Unterschrift Verfasser)

\_\_\_\_\_  
(Unterschrift Betreuer)

## **Erklärung zur Verfassung der Arbeit**

Roland Piazzi  
Lagergasse 29/13  
8020 Graz

„Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich jene Stellen – einschließlich der Tabellen, Karten und Abbildungen – der Arbeit, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.“

Wien, am 8. Februar 2011

## Abstract

In dieser Arbeit wird eine Analyse der Netzwerkstruktur des österreichischen E-Tourismus-Marktes durchgeführt und die zentralen Knoten dieses Netzwerks werden mit Hilfe der wichtigsten Zentralitätskennzahlen (Grad-basierte Zentralität, Nähe-basierte Zentralität, Zwischenzentralität) der sozialen Netzwerkanalyse bestimmt.

Das Netzwerk besteht aus den Websites der Tourismusanbieter Österreichs, die durch Links miteinander verbunden sind. In den Kapiteln 2 und 3 wird eine theoretische Einführung in die Themengebiete E-Tourismus und Netzwerkanalyse gegeben. Danach folgt in Kapitel 4 die Beschreibung des Crawlers, mit dem die Daten für die Analyse gesammelt werden. In Kapitel 5 wird die methodische Vorgehensweise beim Sammeln der Daten und der Analyse der Ergebnisse dargelegt.

Schließlich werden in Kapitel 6 die Ergebnisse präsentiert, wobei das Netzwerk zuerst graphisch dargestellt wird und die berechneten Zentralitätskennzahlen für die Knoten des Netzwerks angeführt werden. Auch wird das Netzwerk mit einem ähnlichen Netzwerk aus dem Tourismusbereich und mit dem Netzwerk des World Wide Webs anhand wichtiger Eigenschaften von Netzwerkmodellen verglichen. In der Folge wird die Grad-basierte Zentralität allgemein und die Zentralitätskennzahl des Eingangsgrads speziell, als relevanteste Kennzahl für die Ermittlung der zentralen Knoten eines Netzwerks des World Wide Webs ermittelt, und die zentralen Knoten werden anhand dieser Kennzahl bestimmt und beschrieben.

Insgesamt ergab diese Untersuchung, dass die zentralsten Websites im Netzwerk die sozialen Medien (Facebook, Youtube und Twitter), das Tourismusportal Tiscover, das Portal der ÖBB sowie ein Destinationsmanagement-Portal aus Oberösterreich sind.

This work analyses the networks structure of the Austrian eTourism market and identifies the central points within this network with the application of the most important centrality measures (Degree, Closeness, Betweenness) of the Social Network Analysis. The network consists of the websites of the Austrian tourism players, which are connected with each other through links.

Chapter 2 and 3 give an introduction into eTourism and Network Analysis. Chapter 4 contains a description of the Crawler, which was developed to gather the data for the analysis. Chapter 5 continues with the description of the method while crawling the web and analysing the gained results.

Finally chapter 6 presents the results. At first the network will be shown graphically and the calculated centrality measures will be pointed out. In addition the network will be compared with a similar network of a tourism destination and the network of the World Wide Web, regarding to important properties of network models. Then the centrality measure of the Degree, and especially the In-Degree, will be determined as the most relevant kind of measure when asking for the central points of a network of the World Wide Web and after that the central points of the network will be described regarding to this kind of measure.

This study shows that the central points are social media sites (Facebook, Youtube and Twitter), the tourism platform Tiscover, the site of the Austrian Rail Company, as well as one destination management site from Upper Austria.

# Inhaltsverzeichnis

1	Einleitung .....	1
1.1	Problembeschreibung .....	1
1.2	Methodische Vorgehensweise .....	2
2	Der E-Tourismus .....	4
2.1	Begriffsdefinition .....	4
2.2	Zahlen und Fakten über den Tourismus .....	4
2.2.1	Der Tourismus aus internationaler Perspektive .....	4
2.2.2	Der Tourismus aus nationaler (österreichischer) Perspektive .....	6
2.3	Der Tourismusmarkt .....	7
2.3.1	Das Tourismusprodukt .....	8
2.3.2	Die Distribution der Tourismusprodukte .....	9
2.3.3	Die Akteure des Tourismusmarktes .....	9
2.4	Informationstechnologien im Tourismus .....	15
2.4.1	Tourismus als ein Informationsprodukt .....	15
2.4.2	Informationstechnologische Entwicklungen im Tourismus .....	15
2.4.3	Änderungen in der Tourismusnachfrage und der Zusammenhang mit Informationstechnologien .....	18
2.4.4	Auswirkungen der Informationstechnologien auf Produktion und Distribution im Tourismus .....	19
3	Die Netzwerkanalyse .....	21
3.1	Graphentheorie .....	21
3.1.1	Relevanz der Graphentheorie für die soziale Netzwerkanalyse .....	21
3.1.2	Graphentheorie – Eine Einführung .....	21
3.1.3	Weiterführende graphentheoretische Konzepte .....	24
3.2	Soziale Netzwerkanalyse .....	27
3.2.1	Geschichte der sozialen Netzwerkanalyse .....	27
3.2.2	Bestimmung und Arten von Netzwerkdaten .....	28
3.2.3	Darstellung von Netzwerkdaten in Form von Matrizen .....	29
3.2.4	Wichtige Kennzahlen der sozialen Netzwerkanalyse .....	30
3.3	Netzwerkmodelle .....	39
3.3.1	Reguläre Netzwerke .....	39
3.3.2	Zufallsnetzwerke .....	39
3.3.3	Kleine-Welt-Netzwerke .....	40
3.3.4	Skalenfreie Netzwerke .....	42
3.3.5	Zusammenfassung zu den Netzwerkmodellen .....	44
3.4	Das World Wide Web als Graph .....	45
3.4.1	Die Wahrscheinlichkeitsverteilung der Knotengrade im World Wide Web .....	45
3.4.2	Verbundenheit und durchschnittliche Weglänge im World Wide Web .....	46
3.4.3	Der Clusterkoeffizient des World Wide Webs .....	46
3.4.4	Die Fliegen-Form des World Wide Webs (Bow-Tie) .....	47
4	Der Crawler .....	50
4.1	Allgemeine Beschreibung .....	51
4.2	Die Klassen .....	52
4.2.1	Klasse Hauptklasse .....	52
4.2.2	Klasse Hauptfenster .....	52
4.2.3	Klasse CrawlerFenster .....	52
4.2.4	Klasse Crawler .....	54
4.2.5	Klasse CrawlThread .....	54
4.2.6	Klasse Database .....	55

4.3	Das Datenmodell .....	56
4.4	Pseudocode zu den wichtigsten Funktionen des Crawlers .....	57
4.5	Technische Details .....	60
4.5.1	Properties-Datei .....	60
4.5.2	Threads .....	60
4.6	Problemereiche eines Crawlers .....	62
5	Die Analysemethode .....	64
5.1	Das Sammeln der Daten mit dem Crawler .....	64
5.1.1	Wahl der Startadressen .....	64
5.1.2	Art und Weise des Durchsuchens .....	67
5.1.3	Die Programmabfolge des Crawlers als Sequenzdiagramm .....	70
5.2	Transformation der Daten .....	71
5.2.1	Transformation der Daten in eine Textdatei .....	72
5.2.2	Konvertierung der Textdatei in eine Pajek-Datei .....	74
5.3	Import der Daten .....	75
5.4	Darstellung der Daten .....	76
5.5	Berechnung der Kennzahlen .....	78
5.5.1	Berechnung Grad-basierter Zentralität .....	79
5.5.2	Berechnung der Nähe-basierter Zentralität .....	80
5.5.3	Berechnung der Zwischenzentralität .....	81
5.5.4	Berechnung der Dichte .....	83
5.5.5	Berechnung des Clusterkoeffizienten .....	83
5.5.6	Berechnung der durchschnittlichen Weglänge .....	85
6	Die Ergebnisse .....	87
6.1	Vergleiche zur Beurteilung der Crawler-Qualität und den Suchdurchläufen .....	87
6.1.1	Beurteilung der Crawler-Qualität .....	88
6.1.2	Vergleich der Ergebnisse mit einem externen Crawler .....	90
6.1.3	Vergleich der vier Suchdurchläufe .....	93
6.2	Grafische Darstellung des Netzwerks .....	96
6.2.1	Netzwerke nach eingehenden Links .....	97
6.2.2	Netzwerk nach ausgehenden Links .....	100
6.2.3	Gesamtes Netzwerk .....	103
6.3	Ergebnisse für die Zentralitätskennzahlen und Vergleich zusätzlicher Ergebnisse mit ähnlichen Netzwerken .....	105
6.3.1	Ergebnisse für die Grad-basierte Zentralität .....	105
6.3.2	Ergebnisse für die Nähe-basierte Zentralität .....	107
6.3.3	Ergebnisse für die Zwischenzentralität .....	110
6.3.4	Berechnung und Vergleich zusätzlicher Ergebnisse .....	111
6.4	Bestimmung einer Zentralitätskennzahl zur Ermittlung der zentralen Knoten .....	119
6.4.1	Betrachtungen zur Relevanz der Zwischenzentralität .....	119
6.4.2	Vergleich mit zwei weiteren Maßzahlen aus der Webometrie und der Netzwerkanalyse .....	120
6.4.3	Schlussbetrachtungen und logische Folgerungen zur Ermittlung der relevantesten Zentralitätskennzahl .....	122
6.5	Beschreibung der zentralen Knoten .....	123
6.5.1	Facebook .....	123
6.5.2	Tiscover .....	125
6.5.3	Youtube .....	126
6.5.4	ÖBB .....	126
6.5.5	Genussland.at .....	127
6.5.6	Twitter .....	129

6.6 Zusätzliche Ergebnisse und Beobachtungen .....	130
7 Zusammenfassung und Ausblick .....	132
Literaturverzeichnis.....	135
Anhang 1: Startliste für Suchdurchlauf 3 .....	143
Anhang 2: Startliste für Suchdurchlauf 4.....	146

# 1 Einleitung

In diesem einleitenden Kapitel wird zuerst kurz das Problem dieser Arbeit definiert und die Ziele festgelegt. Danach wird näher auf die methodische Vorgehensweise bei der Durchführung der Arbeit eingegangen.

## 1.1 Problembeschreibung

Ziel dieser Arbeit ist die Analyse der Netzwerkstruktur des österreichischen E-Tourismus-Marktes, wobei die Ermittlung der zentralen Knoten in diesem Netzwerk eine wesentliche Rolle spielt. [1] definiert den Begriff E-Tourismus als die Anwendung von Informations- und Kommunikationstechnologien im Tourismus. Wichtig für diese Arbeit ist vor allem der Bereich des Internets bzw. speziell des World Wide Webs innerhalb des E-Tourismus. Das Netzwerk des österreichischen E-Tourismus-Marktes besteht in diesem Sinne aus den Websites der Tourismusanbieter, welche im österreichischen Tourismus tätig sind. Diese Websites bilden zusammen mit den Links, durch welche sie untereinander verbunden sind, ein Netzwerk. Jedem Netzwerk liegt ein Graph zugrunde, weshalb einige grundlegende Begriffe der Graphentheorie in dieser Arbeit erläutert werden müssen. Dabei entsprechen die Websites des Netzwerks den Knoten und die Links zwischen den Websites den Kanten des Graphen.

Die Analyse dieses Netzwerks bzw. Graphen erfolgt mit Hilfe der wissenschaftlichen Methoden der Netzwerkanalyse. Hierbei spielt die soziale Netzwerkanalyse, welche zur Bestimmung der zentralen Knoten des Netzwerks herangezogen wird und ein Untergebiet der Netzwerkanalyse ist, eine besondere Rolle. Die soziale Netzwerkanalyse eignet sich nicht nur zur Untersuchung von Personennetzwerken, sondern auch zur Untersuchung von Organisations- oder Webnetzwerken. Sie bezieht sich unter anderem auf Methoden, mit denen sich die Zentralität eines Knotens innerhalb eines Netzwerks bestimmen lässt. Diese Methoden spiegeln sich in unterschiedlichen Zentralitätskennzahlen wider. Die wichtigsten dieser Kennzahlen, die auch zur Bestimmung der zentralen Knoten des Netzwerks des österreichischen E-Tourismus-Marktes herangezogen werden, sind die Grad-basierte Zentralität (Degree centrality), die Nähe-basierte Zentralität (Closeness centrality) und die Zwischenzentralität (Betweenness centrality).

Da diese Zentralitätskennzahlen Zentralität in unterschiedlicher Art und Weise definieren und deren Ergebnisse nicht miteinander kombiniert werden können, wird schließlich nur die für die Ermittlung der zentralen Knoten eines Netzwerks des World Wide Webs relevanteste Zentralitätskennzahl (allgemein die Grad-basierte Zentralität, speziell der Eingangsgrad) herangezogen, um die zentralen Knoten des Netzwerks zu bestimmen. Die Knoten, welche für diese Kennzahl am zentralsten liegen, sind dann die zentralen Knoten, die im Rahmen dieser Arbeit bestimmt werden sollen. Die Wahl dieser Zentralitätskennzahl wird unter anderem mit der unterschiedlichen Natur des World Wide Webs im Vergleich zu anderen Netzwerken (zum Beispiel Transport- oder Personennetzwerke) oder auch mit den Grundlagen anderer Kennzahlen und Methoden (Web Impact Factor und HITS) zur Bewertung von Knoten im Netzwerk des World Wide Webs begründet. Für die Zentralität der Knoten in Bezug auf den Eingangsgrad gilt es dann, Erklärungen zu suchen. Schließlich soll auch die Netzwerkstruktur des österreichischen E-Tourismus-Marktes allgemein untersucht und mit ähnlichen Ergebnissen verglichen werden.

Insgesamt ist es das Ziel dieser Arbeit, die zentralen Knoten des Netzwerks des österreichischen E-Tourismus-Marktes, welches die Websites der österreichischen Teilnehmer am Tourismusmarkt sind, zu bestimmen, die Möglichkeit einer Anwendung der Methoden der sozialen Netzwerkanalyse auf ein solches Netzwerk zu untersuchen, aber auch deren Grenzen zu erkennen.

## **1.2 Methodische Vorgehensweise**

Zu Beginn der Arbeit wurde eine ausführliche Literaturrecherche betrieben. Dabei wurden hauptsächlich die beiden Teilbereiche E-Tourismus und Netzwerkanalyse, welche in dieser Arbeit auch eigene Kapitel (2 und 3) bilden, studiert. Die hier verwendeten wichtigsten Literaturquellen für den Bereich E-Tourismus sind [4], [5], [8], [12], [34] und [35] sowie für den Bereich der Netzwerkanalyse [44], [45], [46], [47], [54], [55], [86] und [88]. Zusätzlich wurden [64], [80], [81], [82] und [83] studiert, welche von anderen Netzwerkanalysen im Tourismusbereich handeln. Um die Funktionsweise eines Crawlers zu verstehen, wurden auch Dokumente zu diesem Thema studiert, diese werden hier aber nicht explizit genannt.

Während und vor allem nach der Literaturrecherche wurde ein Crawler für die Datensammlung gesucht. Da nach langer Suche kein passender Crawler gefunden wurde, weil der Crawler entweder Ergebnisse in unbrauchbarer Form lieferte oder überhaupt nicht zum Laufen gebracht werden konnte, wurde beschlossen, selbst einen Crawler zu entwickeln. Dazu wurde nur kurz studiert wie ein Crawler grundsätzlich funktioniert und der Crawler wurde daraufhin entwickelt. Die Funktionsweise, Architektur und Art der Datenspeicherung lehnt sich dabei nur unwesentlich an andere, bereits existierende Crawler an. Als Programmierumgebung wurde Java gewählt, weil Webseiten mit dieser Sprache einfach durchsucht werden können. Für die Speicherung der Daten wurde die Datenbank MySQL verwendet, weil diese Datenbank allgemein sehr funktionstüchtig und auch frei verfügbar ist. Der Crawler wird genauer in Kapitel 4 beschrieben.

Nach der Entwicklung des Crawlers wurden erste Suchdurchläufe ausgeführt. Währenddessen wurde das Analysetool UCINET [85] mit integriertem Darstellungstool zur Analyse und Darstellung von (sozialen) Netzwerken studiert und ermittelt, in welcher Form UCINET die Daten benötigt. Um die Daten in die richtige Form zu bringen, wurde in Java ein Transformationsprogramm geschrieben, welches in Kapitel 5 beschrieben wird. Sobald der erste Suchdurchlauf beendet war, wurden die Daten in UCINET importiert und es wurde versucht, die Zentralitätskennzahlen zu berechnen sowie das Netzwerk mit dem im Analysetool integrierten Darstellungstool NetDraw darzustellen. Wie die Zentralitätskennzahlen berechnet und das Netzwerk dargestellt wird, wird auch in Kapitel 5 beschrieben. In der Folge wurden weitere Suchdurchläufe mit verschiedenen Startadressen durchgeführt, um das Ergebnis zu verbessern.

Während dessen wurde mit dem Verfassen der Arbeit begonnen. Zuerst wurden die Kapitel 2 und 3, über den E-Tourismus und die Netzwerkanalyse, geschrieben. Gleich danach wurde auch Kapitel 4 über den Crawler und Kapitel 5 über die Methode beim Sammeln und Analysieren der Daten fertig gestellt. Nachdem dann Suchdurchlauf Nummer 4 beendet worden war, wurde mit der Beschreibung der Ergebnisse in Kapitel 6 begonnen.

Zuerst wurden in diesem Kapitel die Ergebnisse aller Suchdurchläufe miteinander verglichen. Da es grundsätzlich nicht möglich ist, ein vollständiges Netzwerk des österreichischen E-Tourismus-Marktes zu erhalten (siehe dazu die Begründung am Anfang des Abschnitts 5.1.1), musste das repräsentativste Netzwerk ermittelt werden, anhand dessen die notwendigen Analysen durchgeführt werden konnten. Um die Qualität des Crawlers zu unterstreichen wurde versucht dessen Ergebnisse mit den Ergebnissen eines externen Crawlers zu vergleichen, welcher am ehesten geeignete Ergebnisse für eine derartige Analyse lieferte. Trotzdem wurde dabei festgestellt, dass ein Vergleich aufgrund der Verschiedenartigkeit der Ergebnisse und der unzureichenden Dokumentation über den externen Crawler nicht möglich ist.

In der Folge wurde das repräsentativste Netzwerk graphisch dargestellt und die Zentralitätskennzahlen dafür berechnet. Auch wurden Vergleiche zusätzlicher Ergebnisse (Netzwerkdichte und Durchschnittsgrad, durchschnittliche Weglänge, Clusterkoeffizient,



Potenzgesetzverteilung der Knotengrade, Fliegen-Form im Netzwerk und die Anzahl nicht verbundener Knoten) mit einer ähnlichen Untersuchung über die Netzwerkstruktur der Websites der Tourismusanbieter der Insel Elba und der Netzwerkstruktur des World Wide Webs durchgeführt. Anschließend wurde durch den Vergleich mit anderen Kennzahlen und Methoden zur Bewertung von Knoten im World Wide Web (Web Impact Factor und HITS), sowie durch logische Schlussfolgerung die relevanteste Zentralitätskennzahl (Eingangsgrad) für die Bestimmung der zentralen Knoten des Netzwerks ermittelt. Zum Schluss wurden diese zentralen Knoten anhand der Zentralitätskennzahl des Eingangsgrads bestimmt und kurz beschrieben, welche Funktionalität ihre Websites bieten, sowie Gründe dargelegt, weshalb gerade diese Knoten im Zentrum des Netzwerks des österreichischen E-Tourismus-Marktes stehen.

## 2 Der E-Tourismus

In diesem Kapitel werden die wichtigsten Informationen zum Tourismus im Allgemeinen und speziell zum E-Tourismus festgehalten. Begonnen wird mit einer Definition des Begriffs E-Tourismus. Danach werden einige Zahlen und Fakten über den Tourismus vorgestellt, um seine wichtige Rolle in der Wirtschaft und seine Entwicklungen aufzuzeigen. In der Folge wird der Tourismusmarkt beschrieben. Hierbei geht es um das Produkt des Tourismus, die Distributionsarten im Tourismus und die Kategorisierung der Akteure des Tourismusmarktes. Am Schluss dieses Kapitels wird speziell auf den E-Tourismus und somit auch auf die Informationstechnologien im Tourismus eingegangen.

### 2.1 Begriffsdefinition

Wörtlich übersetzt bedeutet E-Tourismus soviel wie „elektronischer Tourismus“. Der Begriff bezeichnet aber nicht eine Art von Tourismus, der in elektronischer Weise geschieht, denn das würde natürlich keinen Sinn machen.

[1] bezeichnet als E-Tourismus die Anwendung von Informations- und Kommunikationstechnologien bzw. die technologischen Innovationen im touristischen Umfeld. Damit sind hauptsächlich alle mit dem Internet verknüpften Anwendungsfelder wie E-Business (Tourismusmarketing, Web 2.0 im Tourismus) und E-Government gemeint; man verwendet diesen Begriff aber auch, wenn generell von der Einführung neuer Hard- und Softwareentwicklungen in den Tourismusbereich die Rede ist. Um dies zu ermöglichen, analysiert der E-Tourismus alle damit zusammenhängenden betriebs- und volkswirtschaftlichen Prozesse und Strukturen. Das Ziel des E-Tourismus ist die Digitalisierung der gesamten touristischen Wertschöpfungskette sowie die Einbindung der unterschiedlichen touristischen Leistungsträger in diesen Prozess. Die damit verbundene Optimierung der Prozesse in der touristischen Wertschöpfungskette führt zu einer verbesserten Wettbewerbsfähigkeit der Tourismusindustrie.

Nach [2] bezeichnet E-Tourismus alle digitalisierten und über eine webbasierte Technologie realisierten Transaktionen, die von der Tourismuswirtschaft zur Optimierung der Wertschöpfungsprozesse eingesetzt werden, um dadurch die Wettbewerbsfähigkeit zu steigern.

Die korrekte Schreibweise des Begriffs E-Tourismus (engl. *eTourism*) ist nicht eindeutig, denn neben der in dieser Arbeit bevorzugten Variante findet man in der Literatur auch die Form „eTourismus“.

### 2.2 Zahlen und Fakten über den Tourismus

In diesem Abschnitt werden Zahlen zum Tourismus präsentiert, zunächst zum internationalen und dann zum nationalen, österreichischen Tourismus.

#### 2.2.1 Der Tourismus aus internationaler Perspektive

Die Einnahmen des weltweiten Tourismus steigen schon seit vielen Jahren, besonders seitdem es Flugangebote zu günstigeren Preisen gibt (ungefähr ab den 1960er Jahren). Zwischen 1960 und 1980 stiegen die weltweiten Einnahmen im Tourismus um über 1400%. Die nachfolgende Tabelle 1 zeigt die Veränderung bei den Tourismuseinnahmen in Milliarden US-Dollar an. [3]

Die Bevölkerung der Länder USA, Deutschland und Großbritannien, welche den Großteil der Nachfrage im Tourismus ausmachen, sieht das Reisen als einen wichtigen Konsumbestandteil, und es wird deshalb oft eher auf andere Konsumformen verzichtet, um Reisen zu können. [3]

<b>Jahr</b>	<b>Nordamerika</b>	<b>Welt</b>
1950	0,7	2,1
1960	1,4	6,9
1970	3,5	17,9
1980	12,3	101,4
1990	54,8	263,4
2000	104,2	475,8

**Tabelle 1: Veränderung der Einnahmen in Milliarden US-Dollar im Tourismus in Nordamerika und der Welt zwischen 1950 und 2000. [3]**

Wie in der Tabelle 1 zu sehen ist, sank die Wachstumsrate ab 1980 in der Welt und ab 1990 in Nordamerika. Zwischen 1980 und 1990 wuchsen die Einnahmen in der Welt vergleichsweise nur mehr um das 2,5-fache. Trotzdem gibt es noch Platz für Wachstum, denn nur 4% der Weltbevölkerung reisen international und 80% der Einnahmen im Tourismus werden von weniger als 20 Nationen erzeugt. 1980 prophezeiten Experten, dass sich die Tourismusausgaben in der Welt bis 1995 auf 367 Milliarden US-Dollar vervierfachen werden. Auch wurde vorausgesagt, dass das verfügbare Einkommen und die Einwohnerzahlen der Städte steigen werden und dass die Arbeitszeiten der Bevölkerung zurückgehen werden. Nicht zuletzt sollte die Umwelt aufgrund der sinkenden Anzahl natürlich verbliebener Ressourcen auch im Tourismus zu einem wichtigen Thema werden. [3]

Diese Voraussagen waren größtenteils richtig. Die Urbanisierung, das Interesse an Umweltfragen und an Reisen von kürzerer Dauer sind gestiegen. Das vorausgesagte Wachstum wurde sogar übertroffen. Die weltweiten Einnahmen des Tourismus stiegen im Jahr 2000 auf 475,8 Milliarden US-Dollar. Die durchschnittlichen Einnahmen pro Ankunft waren 680 US-Dollar und die internationalen Tourismusankünfte stiegen insgesamt auf 699 Millionen an. Das schnellste Wachstum wurde in Ostasien und der Pazifik-Region mit 14% zwischen 1999 und 2000 verbucht. In letzter Zeit wuchs die Tourismusindustrie weniger stark. 2001 schrumpfte sie um 0,6%, nach einem durchschnittlichen jährlichen Wachstum von 4,5% zwischen 1990 und 2000. Das war der erste Rückgang der Industrie seit 1982. [3]

[4] sagt, dass die Tourismusbranche mit etwa 260 Millionen Menschen, die direkt oder indirekt darin tätig sind, der weltweit größte Jobanbieter ist. Die Tourismusindustrie erwirtschaftet etwa 10% des Welt-Bruttosozialprodukts und etwa 13,4% in der EU15.

Im letzten Jahr wurde die Tourismusindustrie hart von der Weltwirtschaftskrise getroffen. Nach einem Rückgang des weltweiten Tourismus-Bruttosozialprodukts von 4,8% im Jahr 2009, wurde für 2010 wieder ein leichtes Wachstum von 0,5% prognostiziert. Diese Prognose verbessert sich weiter für das Jahr 2011 mit 3,2%. [5]

Ein paar Prognosen für die Veränderungen in den verschiedenen ökonomischen Kategorien des Tourismus werden in der folgenden Tabelle 2 gegeben.

<b>Bruttosozialprodukt</b>
Der Anteil des Tourismus am Weltbruttosozialprodukt wird 2010 9,2% und im Jahr 2020 9,6% betragen
<b>Wachstum</b>
Das reale Wachstum stieg nach einem Minus von 4,8% im Jahr 2009, im Jahr 2010 auf 0,5%. Für die nächsten 10 Jahre wird ein durchschnittliches Wachstum von 4,4% pro Jahr prognostiziert.
<b>Arbeit</b>
Die Anzahl der Arbeitskräfte die im Tourismus tätig sind, wird von 8,1% und 235.785.000 Jobs im Jahr 2010 auf 9,2% und 303.019.000 Jobs im Jahr 2020 anwachsen, d.h. im Jahr 2020 wird bereits fast jeder 10. Job ein touristischer Job sein.
<b>Exportanteile der Besucher</b>
Der Exportanteil der durch internationale Besucher für den Tourismus erwirtschaftet wird, sinkt von 6,1% im Jahr 2010 auf 5,2% im Jahr 2020.
<b>Investitionen</b>
Die Investitionen wachsen von 9,2% im Jahr 2010 auf 9,4% der gesamten weltweit getätigten wirtschaftlichen Investitionen im Jahr 2020 an.

**Tabelle 2: Prognosen zu den Veränderungen der ökonomischen Kategorien im weltweiten Tourismus zwischen 2010 und 2020 [5]**

## 2.2.2 Der Tourismus aus nationaler (österreichischer) Perspektive

[6] sagt, dass Österreich sowohl bei den Einnahmen als auch bei den Ankünften unter den Top 10 der Tourismusdestinationen liegt. Deshalb nimmt der Tourismus eine sehr wichtige Rolle in der österreichischen Volkswirtschaft ein.

Doch kamen 2009 laut [7] 32,3 Millionen und damit 0,9% weniger Gäste als 2008 nach Österreich. Die Österreicher bildeten mit 11 Millionen Gästen die größte Kundengruppe im eigenen Land, gefolgt von den Gästen aus Deutschland. Insgesamt gab es bei den ausländischen Gästen in Österreich gegenüber 2008 einen Rückgang von 2,6% auf 21,3 Millionen. Die Zahl der Nächtigungen erreichte im Jahr 2009 124,3 Millionen, was einem Rückgang von 1,9% gegenüber dem Jahr 2008 entspricht. Damit erreichte man in einem wirtschaftlich schwierigen Jahr das fünftbeste jemals erzielte Ergebnis. Die Nächtigungen aus dem Inland erhöhten sich gegenüber 2008 um 1,6% auf 34,4 Millionen, was einem neuen Rekordwert entspricht. Bei den Nächtigungen aus dem Ausland gab es einen Rückgang von 3,2% auf 89,9 Millionen, was aber angesichts eines sehr guten Vergleichsjahres 2008 noch immer den zweitbesten Wert der letzten 13 Jahre bedeutete. [7]

Tabelle 3 zeigt wiederum die Prognosen für die einzelnen ökonomischen Kategorien bis zum Jahr 2020, diesmal für den österreichischen Tourismus.

Wie in dieser Tabelle zu sehen ist, arbeiten derzeit etwa 559.000 Menschen in Österreich in der Tourismusbranche. Diese Menschen arbeiten in ca. 90.000 Betrieben. Daraus kann geschlossen werden, dass der österreichische Tourismussektor vor allem aus Klein- und Kleinstbetrieben besteht. Diese Betriebe zeichnet vor allem ihre Größe, Unabhängigkeit, Flexibilität, Saisonalität, die Art der Führung in Form eines Familienbetriebs, sowie vielfach die geographische Abgeschiedenheit aus. Daraus folgt, dass meist traditionelle Managementpraktiken angewandt werden. Diese sind gekennzeichnet durch ein Management,

das geprägt ist von Kurzsichtigkeit, geringer Kooperationsbereitschaft und einer Abneigung gegen Informations- und Kommunikationstechnologien. [6]

<p><b>Bruttosozialprodukt</b></p> <p>Der Anteil des Tourismus am österreichischen Bruttosozialprodukt wird von 12,5% im Jahr 2010 auf 15,4% im Jahr 2020 anwachsen.</p>
<p><b>Wachstum</b></p> <p>2010 gibt es wahrscheinlich einen Rückgang beim Wachstum von 3,9%. Jedoch wird mit einem durchschnittlichen Wachstum von 4,1% pro Jahr in den nächsten 10 Jahren gerechnet.</p>
<p><b>Arbeit</b></p> <p>Die Anzahl der Arbeitskräfte die im Tourismus tätig sind wird von 13,4% und 559.000 Jobs im Jahr 2010 auf 17,5% und 753.000 Jobs im Jahr 2020 anwachsen, d.h. im Jahr 2020 könnte bereits fast jeder 7,5te Job ein touristischer Job sein.</p>
<p><b>Exportanteile der Besucher</b></p> <p>Der Exportanteil der durch internationale Besucher im Tourismus erwirtschaftet wird, sinkt von 10,1% im Jahr 2010 auf 9,6% im Jahr 2020.</p>
<p><b>Investitionen</b></p> <p>Der Anteil der Investitionen in den Tourismus im Verhältnis zu den Investitionen in der gesamten Wirtschaft sinkt minimal von 12,6% im Jahr 2010 auf 12,5% im Jahr 2020.</p>

**Tabelle 3: Prognosen zu den Veränderungen der ökonomischen Kategorien im österreichischen Tourismus zwischen 2010 und 2020 [8]**

Seit kurzer Zeit gibt es in Österreich einen Trend zum Qualitätstourismus, um dem stärkeren Wettbewerb, der durch bestimmte Entwicklungen im Tourismus aufgekommen ist, besser begegnen zu können. Dieser stärkere Wettbewerb hat mehrere Gründe: die Etablierung von Billigfluglinien, was zu einer größeren Auswahl von Destinationen, billigeren und mehr kurzfristiger beschlossenen Reisen führt; zusätzlich die Öffnung der Ostmärkte, der stärkere Trend zu Kurzurlauben sowie der immer geringer werdende Stammkundenanteil unter den Gästen. Da der Tourismus, wie in Abschnitt 2.4 noch gezeigt werden wird, ein informationsintensives Geschäft ist, hängt eine Service- und Qualitätsoffensive hauptsächlich von der Qualität der zur Verfügung gestellten Informationen ab. Dies erfordert für einen Tourismussektor wie den österreichischen, der vor allem aus kleineren Unternehmen besteht, in Zukunft eine stärkere Vernetzung der Teilnehmer, um sich mit differenzierteren Angeboten am immer härter werdenden internationalen Markt behaupten zu können. [6]

### **2.3 Der Tourismusmarkt**

Nach den Zahlen und Fakten zum Tourismus wird in diesem Abschnitt ein Überblick über den Tourismusmarkt gegeben. Zuerst wird kurz ein Blick auf das Tourismusprodukt geworfen, danach werden die Distributionskanäle im Tourismus beschrieben. Weiters werden die Hauptakteure des Tourismus genannt – dies sind die Hauptanbieter (Beherbergungsbetriebe, Restaurants und Anbieter von Attraktionen), die Fluglinien, die Reiseveranstalter, die Reisebüros, die Reservierungssysteme und die Destinationsmanagement-Organisationen.

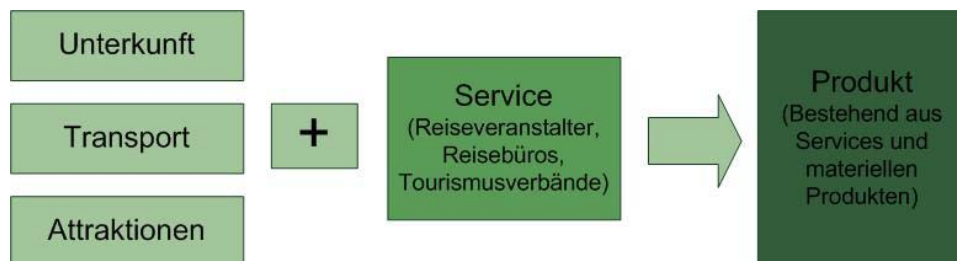
### 2.3.1 Das Tourismusprodukt

Im Zentrum des Tourismusmarktes steht das Produkt des Tourismus. Das Tourismusprodukt besteht aus einer Vielzahl von einzelnen Dienstleistungen und Services, die von verschiedenen Anbietern erbracht werden. Da ein Konsument meist mehrere verschiedene Dienstleistungen und Services beim Konsum des Tourismusproduktes in Anspruch nimmt, müssen diese Teildienstleistungen zu einem Gesamtprodukt gebündelt werden. Auch müssen die Anbieter dieser Teildienstleistungen selbst gebündelt werden, damit der Konsument aus mehreren Anbietern leichter eine Auswahl treffen kann. Aus dieser Logik folgt eine notwendige Zweiteilung des Tourismusprodukts, die sich laut [4] folgendermaßen darstellt:

- Zum ersten Teil gehören die Bereiche Transport, Unterkunft und Attraktionen. Die Produzenten dieser Services sind Luftfahrt-, Schifffahrt- und Bahntransporteure, Hotels und Beherbergungsbetriebe sowie Anbieter von anderen Services und Attraktionen wie Schigebieten, Naturparks, Museen und ähnliche. Diese Anbieter werden Serviceanbieter genannt.
- Der zweite Teil umfasst jene Gruppe von Akteuren, welche die Services zum Kunden transportieren. Dieser Part ist sehr wichtig, denn Tourismusprodukte sind Produkte, die beim Kauf nicht inspiziert oder ausprobiert werden können. Es geht also darum, wie das Produkt von dieser Gruppe beim Kunden präsentiert wird. Zu dieser Gruppe gehören Reiseveranstalter, Reisebüros und Tourismusverbände.

Diese Teilung des Tourismusproduktes kann wie ein Geschäft mit Produkten oder Waren verstanden werden: Der erste Teil des Tourismusprodukts sind die Waren oder Produkte, die in den Geschäften angeboten werden und den zweiten Teil bilden die Geschäfte, welche dem Kunden zueinander passende Produkte von unterschiedlichen Herstellern direkt anbieten.

In Abbildung 1 wird diese Unterteilung des Gesamtproduktes Tourismus noch einmal grafisch dargestellt.



**Abbildung 1: Bestandteile des Tourismusprodukts, adaptiert aus [4]**

Nachdem nun das Tourismusprodukt beschrieben wurde, ist es auch wichtig zu verstehen, wie das Produkt zum Kunden transportiert wird. Die dafür zur Verfügung stehenden Distributionskanäle werden im folgenden Abschnitt beschrieben. Im einfachsten Fall, wenn der Konsument das Produkt direkt beim Anbieter (z.B. Hotel) kauft, fällt der zweite Teil des Produktes, wie er in diesem Abschnitt beschrieben wurde, weg. In diesem Fall braucht es keinen Serviceanbieter, der das Produkt zum Kunden transportiert. Der Kunde muss sich in diesem Fall die einzelnen Teilprodukte seines Urlaubs aber auch selbst zusammenstellen. Obwohl diese Praxis vor allem unter jüngeren Personen immer beliebter wird, ist es nicht die gängige Vorgehensweise. Deshalb kann an dieser Produktunterteilung, wie sie in diesem Abschnitt vorgenommen wurde, durchaus festgehalten werden.

### **2.3.2 Die Distribution der Tourismusprodukte**

[9] nennt mehrere Autoren, die definieren was Distribution ist. Wahab S., Crampon L.J. und Rothfield L.M. sagten 1976, dass Distribution ist was Produkte zur Verfügung stellt. Alcazar Martinez nannte 2002 als Ziel der Distribution, das Produkt dem Konsumenten zu einem gewünschten Zeitpunkt, in der gewünschten Menge und am gewünschten Ort zur Verfügung zu stellen.

Speziell im Tourismussektor sind die Aufgaben der Distribution folgende: Den Touristen Informationen über ihren Urlaub zur Verfügung zu stellen, einzelne touristische Produkte zu einem Gesamtprodukt zusammen zu fügen und dem Touristen Möglichkeiten zu eröffnen, um Reservierungen zu machen, zu bestätigen und zu bezahlen. [10]

Bei der Produktdistribution im Tourismus wird laut [11] zwischen zwei Arten unterschieden.

- Direkte Distribution
- Indirekte Distribution

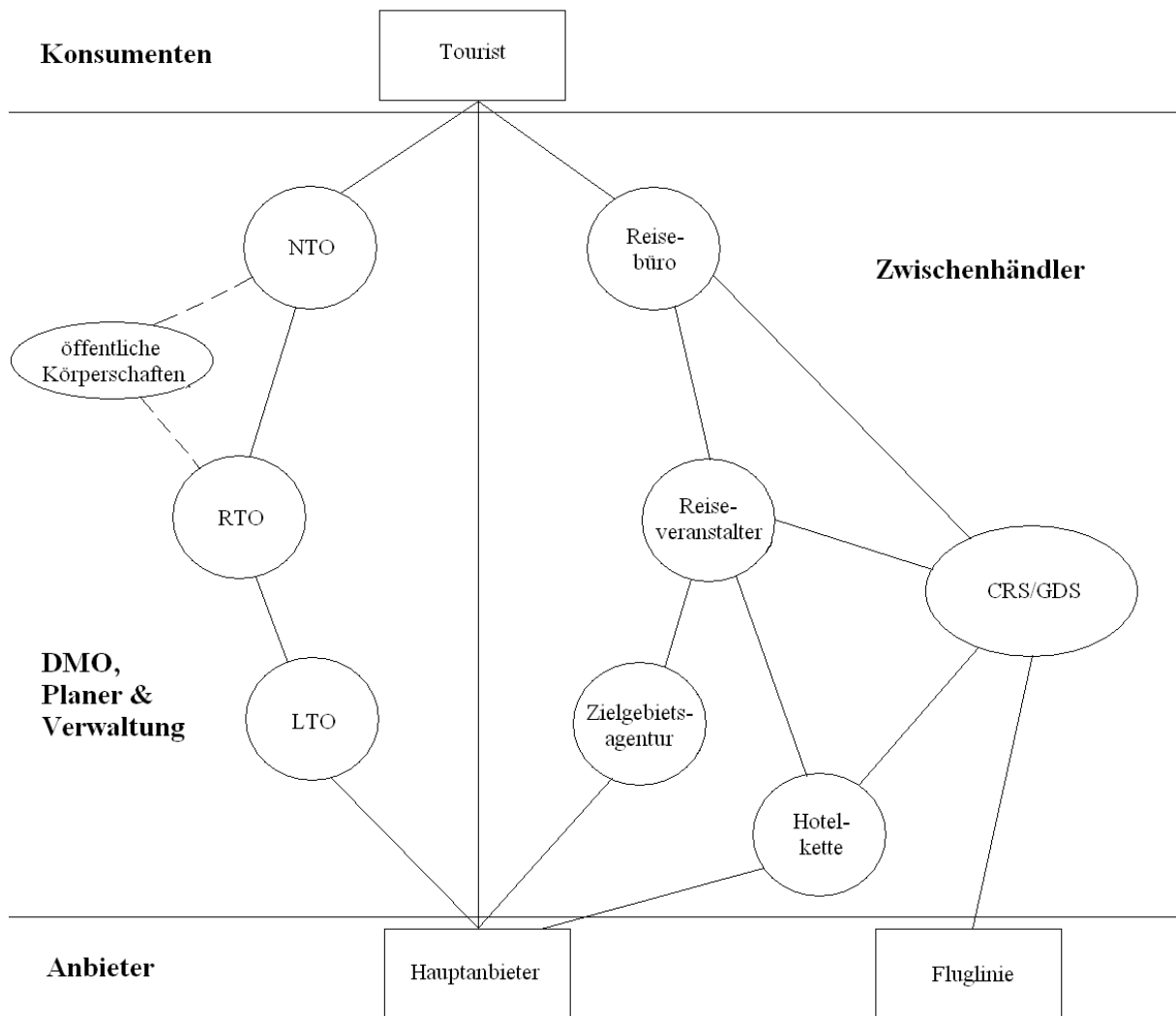
Bei der direkten Distribution tritt der Anbieter des touristischen Produkts (z.B. Hotel, Museumsbetreiber) direkt mit dem Touristen in Verbindung. Die Abwicklung des Geschäfts erfolgt dabei so, dass der Tourist entweder eine eigene Buchungsstelle des Anbieters aufsucht oder den Anbieter über ein elektronisches Medium wie Telefon, Fax oder das Internet kontaktiert.

Bei der indirekten Distribution stehen zwischen dem Anbieter des touristischen Produktes und dem Touristen ein oder mehrere Vermittler. Hier gibt es mehrere mögliche Wege, je nachdem welcher Vermittler gewählt wird. Zu den wichtigsten Vermittlern zählen Reiseveranstalter, Reisebüros oder Destinationsmanagement-Organisationen. Dazu kommen in jüngster Zeit neue Arten von Vermittlern im Internet wie z.B. Flugsuchmaschinen, Online-Reisebüros, Auktionsportale oder ähnliche auf. Gleich wie bei der direkten Distribution erreicht der Tourist den Vermittler entweder persönlich im Geschäft des Vermittlers oder über ein elektronisches Medium.

In beiden Fällen, sowohl bei der direkten als auch bei der indirekten Distribution, spielt das Internet in jüngster Zeit eine wichtige Rolle. Vor allem erleichtert das Internet die direkte Distribution, gleichzeitig eröffnet es aber auch Möglichkeiten für neue Akteure im Markt, die als Zwischenhändler auftreten. Bevor das Internet aufkam, gab es fast ausschließlich die indirekte Distribution. Zu den Auswirkungen des Internets bzw. speziell des World Wide Webs auf den Tourismusmarkt wird in Abschnitt 2.4.2.3 mehr gesagt.

### **2.3.3 Die Akteure des Tourismusmarktes**

Zu den Akteuren des Tourismusmarktes gehören die Hauptanbieter wie Hotels, Restaurants, Anbieter von Unterhaltungsmöglichkeiten und ähnliche, die Fluglinien, die Reiseveranstalter, die Zielgebietsagenturen, die Reisebüros, die Reservierungssysteme und zuletzt die Destinationsmanagement-Organisationen. Hier muss darauf hingewiesen werden, dass neben den Fluglinien auch andere Unternehmen aus dem Transportsektor zum Tourismusmarkt gehören, diese aber in dieser von [12] vorgenommenen Einteilung des Tourismusmarktes nicht direkt genannt werden. Diese Einteilung wird zuerst in graphischer Form (siehe Abbildung 2) dargestellt und daraufhin werden die Akteure etwas detaillierter beschrieben. Die Kästchen und Kreise in Abbildung 2 sind dabei die Akteure im Tourismusmarkt und die Linien dazwischen stellen die wichtigsten Beziehungen zwischen den Akteuren dar. Der Akteur Zielgebietsagentur wird in der Folge nicht in einem eigenen Abschnitt, sondern im Zuge der Beschreibung des Reiseveranstalters beschrieben.



**Abbildung 2: Akteure im Tourismusmarkt.** Für die Abkürzungen CRS/GDS siehe Abschnitt 2.3.3.6 und für die Abkürzungen NTO, RTO und LTO siehe Abschnitt 2.3.3.7. Die Grafik wurde adaptiert aus [12]

### 2.3.3.1 Hauptanbieter

Die Hauptanbieter „produzieren“ das eigentliche Tourismusprodukt indem sie etwa Unterkünfte und Gaststätten zur Verfügung stellen oder Unterhaltungsmöglichkeiten anbieten. Zu dieser Gruppe von Akteuren gehören vor allem kleine und mittelgroße Betriebe.

Laut [13] waren im Jahr 2003 95% der europäischen Unternehmen im Unterkunft- und Restaurantsektor kleine Unternehmen mit neun oder weniger Angestellten. Im selben Jahr beschäftigten kleine bis mittelgroße Unternehmen im europäischen Gastgewerbe 83% der gesamten Angestellten dieses Sektors.

Es ist laut [14] unumstritten, dass neue kleine bis mittelgroße Unternehmen ein sehr wichtiges Potential für die Schaffung von neuen Arbeitsplätzen, für unternehmerischen Wandel und für Innovationsdenken darstellen.

Ein Vorteil gegenüber größeren Unternehmen liegt auch in der Flexibilität der Klein- und Mittelbetriebe. Wegen der sich schnell ändernden Nachfrage im Tourismus ist Flexibilität im Angebot, zum Beispiel durch speziell maßgeschneiderte Produkte und Services, eine wichtige Funktion die solche Unternehmen gegenüber großen Unternehmen haben. [15]



Trotz allen hier genannten positiven Eigenschaften von kleinen bis mittelgroßen Unternehmen, nennt [14] zusätzlich zu den in Abschnitt 2.2.2 genannten negativen Eigenschaften dieser Unternehmen, vor allem die geringe Stabilität am Markt, geringe Investitionsbereitschaft, ein schwaches Management und Widerstände gegenüber Ratschlägen und Veränderungen als negative Ausprägungen von kleinen und mittelgroßen Unternehmen.

Zudem sind diese Unternehmen aufgrund ihrer geringen finanziellen Mittel gegenüber großen Unternehmen benachteiligt und müssen deshalb von der Politik durch Zusammenschluss unter dem Dach einer Destinationsmanagement-Organisation geschützt und gefördert werden.

### **2.3.3.2 Fluglinien**

Die Fluglinien sind ein sehr wichtiger und technologisch fortschrittlichster Akteur auf dem Tourismusmarkt [12]. Laut [16] hatte der Flugliniensektor weit reichende Einflüsse auf den gesamten Tourismus. Positiven Einfluss hatten technische Entwicklungen im Flugzeugbau, die Einführung und Verbesserung von Kommunikations- und Informationstechnologien vor allem durch die Fluglinien sowie komplexere Management- und Marketingmethoden. Diese Veränderungen im Flugliniensektor führten zu einer höheren Qualität und niedrigeren Preisen im Flugverkehr. Doch trotz eines bestehenden Trends zu mehr Liberalisierung, welche auch schon zu niedrigeren Preisen geführt hat, gibt es im Flugverkehr noch immer zu viel staatliche Regulierung, die einen negativen Einfluss auf den Tourismus ausübt. [16]

Die oben angeführten Änderungen im Flugliniensektor führten laut [16] und [12] auch zu einem ständigen Zuwachs im Flugverkehr in den letzten Jahrzehnten. Zum Beispiel wuchs der Flugverkehr zwischen 1980 und 1992 um 92% und die Kapazität um 94%.

Zu betonen ist, dass die Deregulierung im Flugverkehr zu Markteintritten von Billigfluglinien führte. Letztere haben zu großen Veränderungen in der Fluglinienindustrie und im Tourismus geführt, weil die Verfügbarkeit von billigen Flugtickets - wie bereits in Abschnitt 2.2.1 herausgestellt wurde - einer der wichtigsten Faktoren für Wachstum im Tourismus ist. [17]

### **2.3.3.3 Hotelketten**

Laut [18] durchdringen Hotelketten immer mehr den Markt der Unterkünfte im Tourismus. In den USA gehören bereits 70 % der Hotels zu einer Hotelkette. Und auch in Europa gewinnen internationale Hotelketten zunehmend an Einfluss, auch wenn diese Branche dort noch hauptsächlich von kleinen und mittelgroßen Betrieben dominiert wird.

300-400 Hotelketten weltweit besitzen etwa 5,6 Millionen Zimmer. 86 der größten internationalen Hotelketten kontrollieren 3.987.595 Zimmer, was einem Anteil von 24 % der weltweiten Zimmeranzahl entspricht. [19]

In Europa haben laut [20] die sieben größten internationalen Hotelketten mehr als die Hälfte des Marktes der internationalen Hotelketten inne. Zu diesen sieben größten Hotelketten gehören die Firmen: Accor, Carlson Hospitality, Choice Hotels International, HHC, Intercontinental Hotel Group (IHG), Marriott International und Starwood Hotels and Resorts. Diese internationalen Hotelketten operieren in Europa mit 39 Markennamen, welche 557.000 Zimmer zur Verfügung stellen.

Dominiert wird der Markt der Hotelketten laut [12] von multinationalen US-Firmen, welche 1992 13 der Top 20 Ketten besaßen. Diese Ketten richten ihren Fokus auf das Marktsegment mit höheren Preisen und besitzen gut funktionierende Reservierungssysteme.

### **2.3.3.4 Reiseveranstalter**

Die Aufgabe der Reiseveranstalter ist laut [21] die Sammlung und das Angebot von Informationen über Destinationen sowie das Bewerten von Destinationen, beruhend auf

Eigen- und auf Nutzerbewertung, und schließlich die Zusammensetzung von Angeboten aus dem Hotel-, Transport- und Unterhaltungsbereich zu einem Produkt, welches dann von ihnen verkauft wird.

Die Einnahmen erzielen die Reiseveranstalter aus dem Verkauf der zusammengesetzten Produkte, welche sie einzeln von den Hauptanbietern bzw. Transportanbietern kaufen. Wichtige Gründe für Hauptanbieter mit Reiseveranstaltern zusammenzuarbeiten sind folgende: die Reiseveranstalter haben einen guten Zugang zum Markt und zum Kunden, der auch durch ihren bekannten Markennamen gegeben ist, und das finanzielle Risiko kann teilweise vom Hauptanbieter auf den Reiseveranstalter übertragen werden. Vorteil für die Konsumenten ist, dass der starke Wettbewerb unter den Reiseveranstaltern zu niedrigen Preisen führt. [12]

Reiseveranstalter haben einen großen Einfluss im Tourismusmarkt. Laut [22] halten sie 25% des Tourismusmarktes und organisierten im Jahr 2000 175 Millionen internationale Reisen. Demzufolge haben die Reiseveranstalter nicht nur einen großen Einfluss auf den Umsatz der Hauptanbieter, sondern auch auf Destinationen insgesamt. Laut [23] ist der Markt der Reiseveranstalter in Europa hart umkämpft und Reiseveranstalter sind der am öftesten gewählte Partner, den ein Tourist wählt, um seine Reise zu planen und durchzuführen. 55% der Touristen aus Großbritannien und 43% der Touristen aus Deutschland wählten im Jahr 2000 eine Paketreise, die von einem Reiseveranstalter angeboten wurde.

In Summe gibt es in diesem Sektor mehr kleine und mittelgroße, als große Betriebe. Dominiert wird der Sektor aber von etwa 10 großen Reiseveranstaltern, die im Jahr 2000 etwa 2/3 des Marktanteils bei den Reiseveranstaltern besaßen. Die vier größten Unternehmen waren laut [24] folgende: TUI (Deutschland), Airtours Travel Group (Großbritannien), Condor & Neckermann (Deutschland) und Thomson Travel Group (Großbritannien). Die beiden britischen Reiseveranstalter verkauften im selben Jahr 42% der in ihrem Land angebotenen Reisepakete (Thomson Travel Group 24% und Airtours 18%), während ihre Partner in Deutschland 44% aller im Land angebotenen Reisepakete verkauften (TUI 23% und Condor & Neckermann 21%).

Eine Zielgebietsagentur agiert innerhalb einer Destination neben einem oder mehreren Reiseveranstaltern, kümmert sich um die lokalen Operationen und pflegt den Kontakt mit den Hauptanbietern. Sie arbeitet entweder im Namen des Reiseveranstalters oder eigenständig; in manchen Fällen führt der Reiseveranstalter diese Aktivitäten auch selbst durch. Zielgebietsagenturen sehen sich als Konkurrenten von DMOs. [12]

### **2.3.3.5 Reisebüros**

Reisebüros agieren hauptsächlich als Vermittler zwischen Reiseveranstaltern und Kunden, aber auch zwischen Fluglinien und Hotels und dem Kunden. Ihr Einkommen besteht aus Kommissionen, die aus einem Prozentsatz des Produktpreises zusammengesetzt sind. Die meisten traditionellen Reisebüros sind Klein- oder Mittelbetriebe. [12]

Laut [25] sind die Reisebüros durch die neuesten Entwicklungen in den Informationstechnologien einem zunehmenden Wettbewerb ausgesetzt, da die Hauptanbieter, Reiseveranstalter und Fluglinien durch das World Wide Web versuchen, ihre Dienste und Produkte direkt an den Kunden zu verkaufen. Durch die vermehrte Durchdringung des World Wide Webs drängen zusätzlich Online-Reisebüros wie Expedia oder Travelocity auf den Markt, traditionelle Reisebüros gehen auch Online und so genannte Broker bieten in Auktionen Flugtickets und andere Reiseprodukte an. Dies führt zu zusätzlichen Konkurrenten und verstärkter Konkurrenz zwischen den Reisebüros.

Reisebüros benutzen in hohem Maße Global Distribution Systems (GDS) (siehe Abschnitt 2.3.3.6 und 2.4.2.2), mit denen sie zum Beispiel die Verfügbarkeit der touristischen Produkte prüfen können. Laut [26] entwickelten die Eigentümer von Amadeus, einem der führenden Anbieter eines GDS, für Reisebüros in Deutschland das so genannte START-System, welches Interfaces für Reservierungsvorgänge zur Verfügung stellt. Durch diese Interfaces können die Reisebüros Reservierungsmöglichkeiten für ihre Kunden auf ihren eigenen Portalen im Internet zur Verfügung stellen. Laut [12] konnte durch dieses System der Umsatz um 17% gesteigert und die Kommunikationskosten der Reisebüros um 22% gesenkt werden.

Laut [28] trat 1995 das erste Online-Reisebüro im World Wide Web auf. Seitdem sind viele weitere in das Geschäft eingestiegen. [27] nennt als die zehn wichtigsten Online-Reisebüros: Expedia, Travelocity, Orbitz.com, Southwest Airlines, CheapTickets.com, American Airlines, Yahoo! Travel, Delta, Priceline und AOL Travel.

### **2.3.3.6 Computergestützte Reservierungssysteme (CRS/GDS)**

Die Abkürzung CRS steht für Computerized Reservation System und GDS für Global Distribution Systems. Der Unterschied zwischen beiden Begriffen wird in den Abschnitten 2.4.2.1 und 2.4.2.2 genauer beschrieben. Beide Systeme sind im Grunde computergestützte Reservierungssysteme.

Laut [29] wurden in den 1950er Jahren die ersten CRS von den Fluglinien entwickelt und in Betrieb genommen. In den 1980er Jahren wurden diese Systeme dann in GDS umgewandelt. Diese Systeme verbinden die Zwischenhändler mit den Haupt- und Transportanbietern. Zum Beispiel sind Fluglinien und Hotels sowie Reiseveranstalter und Reisebüros an diese Systeme angeschlossen und dadurch können Verfügbarkeiten überprüft und eventuell Reservierungen durchgeführt werden. Laut [12] haben sich vier Systeme am Markt durchgesetzt. Diese sind Amadeus, Galileo, Sabre und Worldspan. Die Eigentümer dieser Systeme sind noch immer hauptsächlich Fluglinien. In der Zwischenzeit wurden auch andere tourismusrelevante Produkte wie alternative Transportmittel, Unterkünfte und Produkte von Reiseveranstaltern in die Datenbanken dieser Systeme aufgenommen.

Laut [30] dominiert Sabre in Nordamerika, wenn die daran angeschlossenen Reisebüros als Bewertungskriterium für die Dominanz herangezogen werden. In Europa ist Amadeus der Spitzenreiter. Und Galileo ist am gleichmäßigsten auf der ganzen Welt verteilt. Alle vier Systemanbieter arbeiten weltweit. Neben diesen vier global operierenden GDS-Systemen, gibt es auch noch eine größere Anzahl von regionalen CRS, welche im Jahr 1996 18% der weltweiten Destinationen in ihren Datenbanken führten. Im Jahr 1988 waren 95% der amerikanischen Reisebüros an ein computergestütztes Reservierungssystem angeschlossen. In Europa hinkte diese Entwicklung ein bisschen hinterher. Trotzdem waren im Laufe der 1990er Jahre in den meisten europäischen Ländern 80-90% der Reisebüros an computergestützte Reservierungssysteme angeschlossen waren.

### **2.3.3.7 Destinationsmanagement-Organisationen (DMO)**

Die Aufgaben von Destinationsmanagement-Organisationen, kurz DMO, sind vielfältig. [31] nennt folgende Aufgaben:

1. Die Koordination der verschiedenen Interessensgemeinschaften, welche mit dem Tourismus in Verbindung stehen (politische, bürgerliche, wirtschaftliche und andere).
2. Sodann hat eine DMO eine Führungs- und Vertretungsrolle für den Tourismus in einer bestimmten Destination inne, sodass die Bürger dieser Destination die Bedeutung des Tourismus für die Region zu schätzen wissen.

3. Als weitere Aufgabe wird die Mithilfe zur Entwicklung von touristischen Angeboten und eines Images der Destination genannt.
4. Eine DMO ist unter anderem auch zur Unterstützung der Touristen vor Ort durch Informationen da.
5. Letzte und nicht zu vergessende Aufgabe ist die Funktion als Kontaktorganisation zu externen Organisationen, wie Reiseveranstalter, Reisebüros und Organisatoren von Geschäftstreffen, welche Besucher zur Destination bringen.

DMOs werden laut [12] in NTO, RTO und LTO unterteilt (siehe Abbildung 2). NTO steht für „National Tourism Organisation“ und vertritt eine ganze Nation als Destination. RTO steht für „Regional Tourism Organisation“ und ist die Vertretung einer Destination auf regionaler Ebene. Schließlich steht LTO für „Local Tourism Organisation“ und ist die Vertretung der Destination auf lokaler Ebene. Diese Organisation vertritt also zum Beispiel eine Gemeinde oder einen Bezirk.

Da viele der Hauptanbieter im Tourismus kleine und mittelgroße Betriebe sind und daher nicht soviel Wissen, Möglichkeiten und Macht haben, ist es wichtig, dass diese Betriebe durch DMOs unterstützt und von ihnen vertreten werden.

Die technologischen Entwicklungen durch das Internet und die damit einhergehende Möglichkeit einer schnelleren Abwicklung der touristischen Geschäfte, veranlasste die DMOs über ihre rein marketingtechnische und informative Aufgabe hinaus zu gehen. So genannte Destinationsmanagement-Systeme (DMS) wurden entwickelt. [32] Diese ermöglichen dem Konsumenten auch direkte Buchungen, da der Kunde nicht nur Informationen über die Destination, sondern auch das Produkt kaufen will. [12] Die DMS enthalten laut [32] drei Komponenten: eine Produktdatenbank mit Informationen über Sehenswürdigkeiten, Unterkünfte und Reiseinformationen, eine Kundendatenbank und ein Buchungs- und Reservierungssystem. Letzteres macht das gesamte System viel mächtiger und stärkt somit die Position der DMOs gegenüber ihrer Konkurrenz.

Laut [12] ist ein neu auftretendes Problem der DMOs in Österreich, dass sie zum Beispiel rechtlich nicht in der Lage sind, dem Konsumenten Empfehlungen zu bestimmten Anbietern oder Produkten zu geben. Wegen ihrer Finanzierung durch die Politik, müssen sie alle Anbieter in gleichem Maße vertreten und dürfen nicht einzelne bevorzugen. Der Konsument fragt aber zunehmend nach solchen Empfehlungen. Um dieses rechtliche Problem zu umgehen, gründen DMOs private Unternehmen, welche diese Aufgaben erfüllen und mit denen sie Partnerschaften eingehen. Dies führt wiederum zu Wettbewerbsverzerrungen, da öffentlich finanzierte Unternehmen indirekt in der Privatwirtschaft mitmischen. Dieser Trend der Partnerschaften von DMOs mit privaten Unternehmen wird sich aber in Zukunft weiter verstärken.

Auch gibt es laut [33] zusätzliche Veränderungen, welche durch die Veränderungen im globalen Tourismus in Zukunft auf die DMOs zukommen werden. Dazu gehört zum Beispiel, dass DMOs nachhaltigen Tourismus anbieten sollten, in dem Qualität vor Quantität gilt. Auch sollte der Klimawandel zunehmend berücksichtigt werden, indem der Tourismusbetrieb die Notwendigkeit der Reduktion von Treibhausgasen nicht außer Acht lässt. Das Marketing der DMOs muss sich verändern und sich an „den drei E-Wörtern“ Entertainment, Excitement und Education orientieren. Auch das richtige Risikomanagement ist ein wichtiger Aufgabenbereich der DMOs für die Zukunft, da Sicherheit und Bequemlichkeit in einer Destination entscheidende Wettbewerbsvorteile sind. Zuletzt spielt noch die Ausbildung eine wichtige Rolle bei den zukünftigen Aufgaben der DMOs. Um wettbewerbsfähig zu bleiben, müssen sich die Mitarbeiter ständig weiterbilden; ebenso sollte auch die Gesellschaft als

Ganzes in Aus- und Weiterbildungsprogramme, welche den Tourismus betreffen, mit einbezogen werden.

In diesem Abschnitt wurden das Tourismusprodukt, die Distribution im Tourismus und die Akteure des Tourismus beschrieben. Dabei wurde der Tourismus insgesamt behandelt. Im nächsten Abschnitt wird speziell auf den E-Tourismus eingegangen, indem die Anwendung von Informationstechnologien im Tourismus eingehender erörtert wird.

## **2.4 Informationstechnologien im Tourismus**

In diesem Abschnitt wird besprochen, welcher Zusammenhang zwischen Informationstechnologien und dem Tourismus besteht. Zuerst wird versucht, zu erklären, wieso der Tourismus ein Informationsprodukt ist. Danach werden die drei wichtigsten informationstechnologischen Entwicklungen im Tourismus vorgestellt.

Anschließend werden Zusammenhänge zwischen den Veränderungen in der Tourismusnachfrage, die in den letzten Jahrzehnten auftraten, und den Informationstechnologien aufgezeigt. Letztlich werden die Auswirkungen der Informationstechnologien auf Produktion und Distribution im Tourismus behandelt.

### **2.4.1 Tourismus als ein Informationsprodukt**

[34] sagt: „Information is the lifeblood of tourism“. Dies ist ein vielfach zitierter Satz. Informationen sind deshalb ein unverzichtbarer Bestandteil des Tourismus. Laut [35] sind Kauf und Konsum des Tourismusprodukts zeitlich getrennt, weshalb das Produkt zum Zeitpunkt des Kaufs nicht getestet werden kann. Aus diesem Grund sind Informationen sehr wichtig. Die Beschreibung des Produktes in Form von Information ist das Einzige, was der Konsument hat, um sich zu entscheiden und einen Kauf zu tätigen. Deshalb sind Informationstechnologien im Tourismus überaus wichtig. [35]

Im folgenden Abschnitt werden die drei wichtigsten informationstechnologischen Entwicklungen für den Tourismus beschrieben.

### **2.4.2 Informationstechnologische Entwicklungen im Tourismus**

#### **2.4.2.1 Computer Reservation Systems (CRS)**

Die ersten Unternehmen welche Informations- und Kommunikationstechnologien im Tourismus anwandten waren die Fluglinien in den 1950er Jahren. Im Jahr 1962 führten die American Airlines zum Beispiel das bekannte CRS mit dem Markennamen Sabre ein. Zu jener Zeit war dies noch ein technisches Wunder, welches sogar mehr Codierungsaufwand erforderte als das Mercury-Projekt der NASA. [36] Zu Beginn war ein CRS eine Datenbank in welcher das Inventar eines Betriebes gespeichert wurde [34]. Doch schon in den 1970er Jahren haben sich die Aufgaben eines CRS über die Inventarfunktion hinaus entwickelt. In der Folge wurden CRS auch benützt um Flugpläne zu erstellen, die Mitarbeiter einzuteilen und Entscheidungen bezüglich des Managements zu unterstützen. [36]

Die Deregulierung des US-Fluglinienmarktes in den 1970er Jahren führte zu einem erhöhten Flugverkehr. Um dieses zusätzliche Angebot und die Nachfrage besser zu managen waren CRS unabdingbar. [36] Schließlich waren alle Fluglinien gezwungen CRS zu benützen, um wettbewerbsfähig zu bleiben, da der Einsatz von CRS die Kosten senkte und damit niedrigere Preise ermöglichte. Die Unternehmen entwickelten entweder ihr eigenes CRS oder benützten das CRS einer anderen Fluglinie gegen Entrichtung einer Gebühr. In den 1970er Jahren, begannen die ersten Fluglinien mit Reisebüros zusammenzuarbeiten und ihnen ihre Reservierungssysteme zur Verfügung zu stellen. Die Fluglinien lernten in dieser Zeit, wie wichtig die Art und Weise ist, in welcher die Informationen den Reisebüros zur Verfügung

gestellt wurden, und dass dies die Auswahl der Fluglinie durch das Reisebüro bei der Planung der Reise beeinflusste. [37]

Nach den Fluglinien entwickelten auch Hotelketten und Reiseveranstalter CRS. Durch CRS können Produkte global organisiert und verkauft werden. Und wie bereits festgestellt wurde, unterstützen CRS zusätzliche Geschäftsfunktionen und tragen damit allgemein zu Kostenreduktion und höheren Profiten bei. Unter anderem ist mit Hilfe von CRS nicht nur eine flexible Preisgestaltung, sondern auch die Reduktion der Kommunikationskosten möglich. CRS sind also von fundamentaler Bedeutung für die gesamte Tourismusindustrie. [34]

#### 2.4.2.2 Global Distribution Systems (GDS)

Durch die Globalisierungstendenzen im Tourismus, mussten auch die CRS global operieren. Aus diesem Grund entwickelten sich die GDS aus den CRS. Unterschiedliche Fluglinien integrierten ihre CRS in ein GDS. [34] Auch wurden Anfang der 1990er Jahre zusätzliche alternative Services, wie zum Beispiel Unterkünfte und Autoverleih, in die Datenbanken der GDS mit aufgenommen [30].

Die Integration von lokalen oder unternehmensinternen CRS in die globalen GDS erfolgte durch neu auftretende Verbindungsunternehmen, so genannte „switch companies“. Dazu gehören zum Beispiel THISCO oder WizCom. Diese Unternehmen stellen die technologische Verbindung zwischen CRS und GDS über Interfaces zur Verfügung. Diese Integration der CRS in die GDS gibt den GDS und den Verbindungsunternehmen ziemlich viel Macht über die Hauptanbieter. Deshalb sind die GDS und die Verbindungsunternehmen auch in der Lage, hohe Gebühren für ihren Dienst zu verlangen. [38]

Die vier führenden GDS waren im Jahr 1995: Sabre, Galileo, Amadeus und Worldspan. Tabelle 4 enthält ein paar Zahlen zu diesen vier wichtigsten GDS.

GDS	Anzahl der Orte in denen Reisebüros die GDS benutzen	Terminals (1996)	Teilnehmende Fluglinien (1995)
Sabre	33.453	124.800	377
Galileo	36.614	125.000	525 <sup>a</sup>
Amadeus	42.328	111.400	Nicht verfügbar
Worldspan	17.325	49.500	420 <sup>a</sup>

**Tabelle 4: Zahlen zu den vier führenden GDS in den Jahren 1995 und 1996. <sup>a</sup> gilt für das Jahr 1997. [30]**

In letzter Zeit boten GDS immer mehr Produkte an, sowohl für Freizeitreisende als auch für Geschäftsreisende. Dazu gehörten zum Beispiel Theatertickets oder ganze Urlaubspakete. Die Entwicklung von Destinationsmanagement-Systemen und deren Integration in ein GDS ermöglicht es auch den Klein- und Mittelbetrieben der Tourismusbranche im System präsent zu sein. [34]

#### 2.4.2.3 Das World Wide Web

Anfang der 1990er Jahre wurde das World Wide Web entwickelt und eingeführt. Diese Entwicklung hat seitdem zu großen Veränderungen in vielen Bereichen des menschlichen Lebens geführt. Durch die mögliche weltweite Vernetzung aller Rechner und durch das Angebot von Texten, Graphiken, Videos und Musik konnten verteilte Informationssysteme entwickelt werden, welche auf jegliche Art von Daten des gesamten World Wide Webs

zugreifen können. Dadurch wurde es unter anderem möglich, Bankdienstleistungen zu Hause am eigenen Computer zu erledigen, Käufe von zu Hause aus zu tätigen oder auch von zu Hause aus zu arbeiten.

Natürlich hatte die Entwicklung des World Wide Webs auch enorme Auswirkungen auf den Tourismus, in dem Informationstechnologien ja schon immer eine wichtige Rolle gespielt haben. Durch das World Wide Web wird der direkte Kontakt zwischen Anbieter und Kunde erleichtert, wodurch ein Nachteil für die Zwischenhändler entstehen kann. Denn der Anbieter kann auf einer eigenen Website seine Dienste anbieten und durch die weltweite Vernetzung des Systems fast jeden Kunden erreichen. Trotzdem pflegen nicht alle Kunden den direkten Kontakt zum Anbieter und benützen auch weiterhin Zwischenhändler wie Reisebüros oder Reiseveranstalter für die Planung und Abwicklung ihres Urlaubs. [39]

Der Tourist hat durch das World Wide Web auch viel mehr Auswahlmöglichkeiten. Durch das Netz hat er praktisch Zugang zu jedem Anbieter auf der ganzen Welt. Dies erhöht die Auswahlmöglichkeiten enorm. Die vielen Auswahlmöglichkeiten können aber auch zu Verwirrung führen und dem Kunden einen hohen Suchaufwand aufbürden. Hier können so genannte Empfehlungssysteme (Recommender Systems) helfen, indem sie dem Benutzer bei der Suche nach bestimmten touristischen Angeboten Empfehlungen auf Basis seines Nutzerprofils, Bewertungen von anderen Benutzern oder der preislichen Gestaltung geben. Durch die Bewertungen oder den Kommentar von touristischen Dienstleistungen kann der Tourist seine Meinung auf der ganzen Welt verbreiten und so die Kaufentscheidung von vielen Touristen bzw. das Image der Tourismusanbieter beeinflussen. Gute und qualitativ hochwertige Leistungsbereitstellung wird deshalb noch wichtiger für den Anbieter. Auch kann der Konsument der touristischen Dienstleistung durch die Empfehlung anderer leichter qualitativ schlechte Anbieter umgehen. [40]

Das World Wide Web eröffnet aber auch den Tourismusanbietern große Chancen. Hierzu zählt nicht nur der direkte Kontakt zum Kunden. Zum Beispiel können auch Preise, Informationen oder Verfügbarkeiten sehr schnell angepasst und diese Anpassung kann sofort an den Kunden weitergegeben werden. Dadurch kann das Angebot flexibler gestaltet werden und zusätzliche Einnahmen generieren oder Verluste einschränken. Auch die Bewerbung eines Produktes wird vereinfacht. Durch das World Wide Web sind Bewerbung und Kauf nicht mehr notwendigerweise zeitlich und örtlich voneinander getrennt. Wird ein Kunde auf ein Produkt aufmerksam, kann er sofort auf der gleichen Seite oder auf der durch einen Link verbundenen Seite einen Kauf tätigen. Durch E-Mails werden die Möglichkeiten des Eins-zu-eins-Marketings besser ausgeschöpft. Nicht zu letzt hat ein Anbieter durch massives Vorhandensein von persönlichen Daten, aus welchen Nutzerprofile erstellt werden können, die Möglichkeit zu persönlicherem Marketing und Angebot. [41]

Schließlich treten durch das World Wide Web neue Zwischenhändler in den Tourismusmarkt ein. Inhaber von GDS entwickelten zum Beispiel ihre eigenen Portale, um direkt mit den Kunden in Kontakt zu treten (*Travelocity.com* von Sabre und *TRIP.com* von Galileo). Einige dieser neuen Zwischenhändler sind aber Neueinsteiger in der Tourismusindustrie. Durch die Erfahrung mit Informationstechnologien, durch die zur Verfügung stehenden finanziellen Mittel sowie der Bekanntheit des Unternehmens konnten zum Beispiel Firmen wie Microsoft (Expedia) oder Yahoo (Yahoo! Travel) durch die Entwicklung ihrer eigenen Online-Reisebüros in den Markt eintreten. Zusätzlich gab es Neueinsteiger, welche die Angebote verschiedener Anbieter miteinander vergleichen und damit Nutzen aus der Verwirrung der Benutzer durch das Überangebot ziehen. Dazu gehören zum Beispiel *checkfelix.com* für mehrere Anbieter aus unterschiedlichen Bereichen oder das Portal *orbitz.com*, das speziell für Fluglinien eingerichtet wurde. Die erhöhte Flexibilität des Angebots durch das World Wide Web ermöglichte auch den Einstieg von so genannten Last-Minute-Portalen wie *lastminute.com*, um freie Plätze in letzter Minute noch an den Mann zu bringen. Zuletzt seien

noch Auktionsseiten genannt, wie zum Beispiel *Priceline.com*. Auf diesem Portal können Touristen ihre Nachfrage und den Preis spezifizieren und Anbieter können darauf eingehen oder nicht. [42]

Laut [34] traten mit dem Aufkommen des World Wide Webs auch einige neue Problem-bereiche, welche im Folgenden angeführt sind, auf:

- Sicherheit bei der Datenübertragung
- Glaubwürdigkeit bzw. Vertraulichkeit von Daten
- Geistiges Eigentum und Copyrights
- Bandbreite- und Geschwindigkeitsbegrenzungen
- Verwirrung der User durch Überangebot und daraus folgende Unzufriedenheit
- Das Fehlen von fähigen Spezialisten
- Ungleicher Zugang und Preis für die Benützung des World Wide Webs

Diese Probleme stellen in Zukunft eine nicht zu unterschätzende Herausforderung dar, und das nicht nur für den Tourismus.

### **2.4.3 Änderungen in der Tourismusnachfrage und der Zusammenhang mit Informationstechnologien**

Die Interessen der Touristen und ihre Nachfrage haben sich in den letzten Jahrzehnten sehr verändert. Früher – hauptsächlich in den 1950er, 60er und 70er Jahren – war der Tourismus ein Massenphänomen, geprägt von vorgefertigten Urlaubspaketen in denen alles strikt durchgeplant war, wobei die Touristen in ihrem Urlaub vor allem „Sonne tanken“ wollten. Heute sucht der Tourist nach authentischeren Urlaubserlebnissen und will eher flexibel bleiben. [43] Auch sind Touristen heutzutage technologisch und sprachlich gebildeter und reisen viel mehr. Sie versuchen in der Freizeit und im Urlaub hauptsächlich ihren Interessen nachzugehen und sich dabei auch weiterzubilden. Dies wird im Allgemeinen als „edutainment“ bezeichnet. [40] Nicht zuletzt steigt die Anzahl der Reisenden weltweit an.

Diese Änderungen können durch Informationstechnologien leichter bewältigt werden. Informationstechnologien helfen dabei, die steigende Nachfrage schneller und kostengünstiger zu befriedigen. Der selbständigere Reisende kann durch Informationstechnologien und speziell durch das World Wide Web leichter direkt mit dem Anbieter in Kontakt treten und verbraucht dadurch nicht wertvolle Zeit durch das Einschalten eines Zwischenhändlers. Auch erleichtern die Informationstechnologien das individuelle Zusammenstellen von Produkten, um den zunehmend nachgefragten authentischeren Urlauben gerecht zu werden. Nicht zuletzt kann die für den Tourismus so wichtige Information zum Beispiel durch das Internet einfach und bequem vom Touristen zu Hause oder im Büro abgefragt werden.

Um weiterhin wettbewerbsfähig zu bleiben müssen alle Tourismusanbieter diese neuen Trends bei den Informationstechnologien umsetzen. Die Kundenzufriedenheit hängt vor allem vom Umfang und von der Genauigkeit der zur Verfügung stehenden Informationen über eine Destination bzw. die Verfügbarkeit von Unterkünften, Attraktionen und Aktivitäten ab. Das ist deswegen so, weil der Unterschied zwischen den Erwartungen des Kunden und der tatsächlichen Erfahrung durch die Informationen kleiner und damit die Wahrscheinlichkeit einer nicht gewollten Überraschung gering wird. [34]



Laut [34] gibt es noch zusätzliche Entwicklungen und Services, die durch die Informationstechnologien ermöglicht werden und zu höherer Kundenzufriedenheit führen können:

- Das Mehr an Information erhöht die Auswahlmöglichkeiten für die Kunden und somit die Kundenzufriedenheit.
- Schnellere Abwicklung durch das Wegfallen von Bürokratie und Papierwirtschaft lässt den Kunden Zeit sparen und ermöglicht es ihm, sich in seiner Freizeit mit „wichtigeren“ Dingen zu beschäftigen, was wiederum die Kundenzufriedenheit erhöht.
- Individuelle Zusammensetzung von Produkten und Eins-zu-eins-Marketing durch Empfehlungen werden durch die Informationstechnologien erst ermöglicht oder zumindest erleichtert und erhöhen damit die Kundenzufriedenheit.
- Zusätzliche Services können durch Informationstechnologien angeboten werden (zum Beispiel TV-Unterhaltung während des Fluges, Benützung von Bürogeräten vor Ort, neue Informationskanäle).
- Vereinfachung von touristischen Leistungen durch Informationstechnologien erhöht die Kundenzufriedenheit (zum Beispiel Video-Vorschau eines Trips im TV des Hotels oder über das Internet, Navigationssysteme).
- Personalisierte Services durch Informationstechnologien erhöhen die Kundenzufriedenheit (zum Beispiel kennt der Telefonoperator den Gast beim Namen, durch Einloggen im Web können dem Gast personalisierte Services angeboten werden).
- Besseres Funktionieren der Organisationen und bessere Integration der Abteilungen eines Unternehmens durch Informationstechnologien verbessert nicht zuletzt die Services und damit die Kundenzufriedenheit.

#### **2.4.4 Auswirkungen der Informationstechnologien auf Produktion und Distribution im Tourismus**

In diesem Abschnitt werden die weit reichenden Auswirkungen der Informationstechnologien auf die Produktion und die Distribution im Tourismus behandelt.

Im Produktionsprozess wirken mehrere Teilnehmer – vom Hauptanbieter bis hin zu den Zwischenhändlern – mit, indem sie Informationen untereinander austauschen, zum Produkt eigene zusätzliche Informationen angeben und dadurch zusätzliche Wertschöpfung erzielen. Darüber hinaus braucht es auch beim Zusammensetzen der Produkte zu Produktbündeln und schließlich zum Endprodukt Informationen. Jedes einzelne Produkt muss gut beschrieben werden und bei der Zusammensetzung muss auf die Kundenwünsche, die unterschiedlichen Preise und verschiedene Distributionskanäle Rücksicht genommen werden. Dies alles benötigt Information und kann durch Informationstechnologien wesentlich erleichtert, automatisiert und beschleunigt werden. [35]

Auch in der Distribution der Tourismusprodukte spielen die Informationstechnologien eine wichtige Rolle. Die Informationstechnologien veränderten hier vor allem die Art und Weise, wie Tourismusprodukte an den Kunden verteilt wurden. Traditionellerweise waren hier die CRS und GDS besonders wichtig. [42] Durch diese Systeme konnten Reiseveranstalter oder Reisebüros die Verfügbarkeit von touristischen Produkten überprüfen und eventuell auch direkt Buchungen durchführen. In jüngster Zeit trat das Internet zunehmend als Distributionskanal in Erscheinung. Dieses veränderte wiederum die Art und Weise, wie die Tourismusprodukte an den Kunden verteilt wurden. [12] Durch das Internet können Benutzer direkt auf die Reservierungssysteme von Fluglinien, Hotelketten und Autoverleihfirmen

zugreifen [42]. Aber auch andere Anbieter wie Destinationen und Hotels nutzen das Internet, indem sie die Produkte über Reservierungssysteme direkt an den Kunden verkaufen (zum Beispiel über Tiscover). Zusätzlich gibt es, wie bereits erwähnt, neue Firmen, die erst über das Internet in das Reise- und Tourismusgeschäft eingestiegen sind. Dies sind zum Beispiel Online-Reisebüros wie Expedia von Microsoft oder Yahoo! Travel oder Vergleichsportale wie *Checkfelix.com* oder *Ski.com*. Durch das Internet haben die Anbieter Zugang zu einem größeren Kundenkreis und können direkter mit dem Kunden in Kontakt treten, und zwar das ganze Jahr über rund um die Uhr. [42] Dies ist vor allem ein Vorteil des Internets gegenüber den traditionellen Informationssystemen wie CRS/GDS, für die es immer ein Reisebüro oder einen Reiseveranstalter als Vermittler brauchte.

## **3 Die Netzwerkanalyse**

Nachdem verschiedene theoretische Aspekte des E-Tourismus behandelt wurden, wird nun näher auf die Netzwerkanalyse, speziell die soziale Netzwerkanalyse, eingegangen. Da jedes Netzwerk durch einen Graphen dargestellt werden kann, können graphentheoretische Methoden auch in der Netzwerkanalyse verwendet werden. Die Netzwerkanalyse wird in unterschiedlichen wissenschaftlichen Gebieten angewandt. Die soziale Netzwerkanalyse ist ein Untergebiet der Netzwerkanalyse und beschäftigt sich mit Netzwerken bestehend aus Individuen, Organisationen, Staaten oder auch Webseiten. Die soziale Netzwerkanalyse eignet sich deshalb unter anderem zur Untersuchung der Netzwerkstruktur und zur Bestimmung der zentralen Knoten des österreichischen E-Tourismus Marktes, welcher sich als Netzwerk im World Wide Web, bestehend aus den Websites der im Tourismus tätigen Unternehmen, darstellen lässt.

In diesem Kapitel werden zuerst einige grundlegende Konzepte der Graphentheorie vorgestellt. In der Folge wird in Abschnitt 3.2 die soziale Netzwerkanalyse erklärt, vor allem die für diese Arbeit wichtigen Zentralitätskennzahlen zur Bestimmung der zentralen Knoten des Netzwerks. Abschnitt 3.3 beschäftigt sich mit den gängigsten Arten von Netzwerkmodellen, die in der Wissenschaft der Netzwerkanalyse bekannt sind. Abschnitt 3.4 gibt schließlich noch einen Aufschluss darüber, welchem Netzwerkmodell der Graph bzw. das Netzwerk des World Wide Webs entspricht und welche speziellen Kennzeichen bzw. welche Form dieses Netzwerk hat.

Um die Konzepte dieses Kapitels verständlicher zu machen werden auch Graphen dargestellt. Diese wurden immer mit der Anwendung aus [84] gezeichnet.

### **3.1 Graphentheorie**

Zu Beginn dieses Abschnitts wird kurz erklärt wieso die Graphentheorie für die soziale Netzwerkanalyse so wichtig ist.

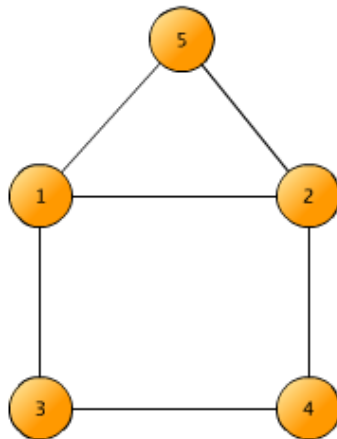
#### **3.1.1 Relevanz der Graphentheorie für die soziale Netzwerkanalyse**

Die Graphentheorie ist aus verschiedenen Gründen wichtig für die soziale Netzwerkanalyse. In der Literatur werden hierfür drei wichtige Gründe genannt: Erstens stellt die Graphentheorie das Vokabular für die Benennung der verschiedenen Bestandteile, Strukturen und Eigenschaften der sozialen Netzwerkanalyse zur Verfügung. Zweitens bietet die Graphentheorie mathematische Operationen und Ideen, um diese Strukturen und Eigenschaften zu messen und zu quantifizieren. Als dritter Grund wird genannt, dass auf Basis der Strukturen und der dazugehörigen mathematischen Operationen verschiedene Theoreme formuliert, aus denen später Schlüsse gezogen werden können. [44c]

Die Graphentheorie stellt also die Konzepte und Elemente zur Verfügung, mit denen das Modell eines sozialen Systems dargestellt und analysiert werden kann. Das Modell ist dabei eine Abstraktion der Wirklichkeit. Obwohl nicht alle real vorhandenen Eigenschaften des Systems in diesem Modell dargestellt werden können, wird doch ein Großteil davon erfasst. [44c] Die Punkte der Graphentheorie, auch Knoten genannt, sind dabei die Akteure bzw. Websites des Modells und die Linien, auch Kanten genannt, sind die Verbindungen bzw. Links zwischen den Akteuren oder den Websites.

#### **3.1.2 Graphentheorie – Eine Einführung**

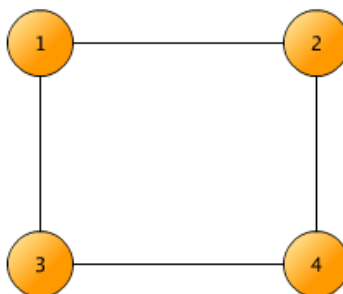
Ein Graph besteht aus Knoten und Kanten, die Verbindungen zwischen den Knoten darstellen. Der bekannteste Graph ist wohl das „Haus vom Nikolaus“, welches in der folgenden Abbildung 3 dargestellt ist. [45a]



**Abbildung 3: Das Haus vom Nikolaus, aus [45a]**

Dieser in Abbildung 3 angeführte Graph hat 5 Knoten und 8 Kanten. Die Knoten und Kanten eines Graphen repräsentierend besteht jeder Graph aus zwei Mengen von Informationen: aus einer Menge von Knoten  $N=\{n_1, n_2, \dots, n_g\}$  und einer Menge von Kanten  $K=\{k_1, k_2, \dots, k_i\}$ . [44b]

Ein Teilgraph eines Graphen besteht immer aus Teilmengen der Knoten- und Kantenmengen. Das heißt, ein Graph T ist ein Teilgraph des Graphen G, wenn die Knoten des Teilgraphs T eine Teilmenge der Knoten des Graphs G und die Kanten des Teilgraphs T eine Teilmenge der Kanten des Graphs G sind. [45c]



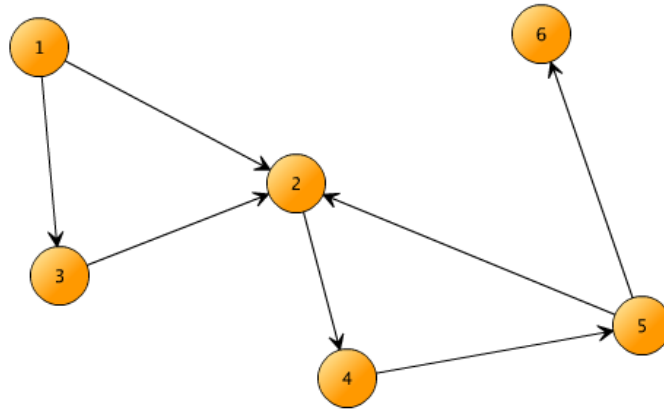
**Abbildung 4: Teilgraph des Graphen in Abbildung 3**

In der obigen Abbildung 4 ist ein Teilgraph des Graphen aus Abbildung 3 dargestellt.

In den vorherigen Beispielen wurden speziell ungerichtete Graphen behandelt. Da in dieser Arbeit aber gerichtete Graphen, auch Digraphen genannt, vorkommen, wird in den folgenden Ausführungen speziell darauf eingegangen.

Ungerichtete Graphen bestehen aus ungerichteten Kanten, gerichtete Graphen hingegen aus gerichteten Kanten. In Abbildung 5 ist ein beispielhafter gerichteter Graph dargestellt. [45d]

Um zwei wichtige Konzepte der Graphentheorie erklären zu können, muss zuerst der Begriff des Kantenzuges erklärt werden. Ein Kantenzug ist eine zusammenhängende Folge von Kanten. Diese Folge kann sozusagen gezeichnet werden, ohne den Stift abzusetzen. Das heißt, ein Kantenzug ist zum Beispiel eine Folge von zwei Kanten, bei der die erste Kante im selben Knoten endet, von welchem aus die zweite Kante weitergeht. [45b] Natürlich ist bei einem Kantenzug in einem gerichteten Graphen die Richtung der Kanten zu beachten. So ist zum Beispiel die Verbindung der Knoten 1, 2 und 4 in Abbildung 5 ein gerichteter Kantenzug. Hingegen ist die Verbindung der Knoten 1, 2 und 5 kein gerichteter Kantenzug.



**Abbildung 5: Ein gerichteter Graph**

Eines der beiden wichtigen Konzepte der Graphentheorie ist der Weg. Ein Weg ist ein Kantenzug, in dem jeder Knoten nur einmal enthalten ist. Zum Beispiel besteht in Abbildung 5 ein Weg durch die Kantenverbindungen der Knoten 1, 2, 4 oder der Knoten 1, 3, 2, 4. Das zweite wichtige Konzept, der Kreis, ist ein Weg, der einem so genannten geschlossenen Kantenzug entspricht. Ein Kantenzug wird dann als geschlossen bezeichnet, wenn der Ausgangsknoten des Kantenzuges gleich dem Zielknoten ist. Ein Kreis wäre in Abbildung 5 der Kantenzug zwischen den Knoten 2, 4, 5 und 2. Ein Weg bzw. ein Kreis ist ein gerichteter Weg bzw. Kreis, wenn der dazugehörige Kantenzug gerichtet ist. Siehe dazu [45c] für ungerichtete Wege bzw. Kreise und [45d] für die gerichtete Version der beiden Konzepte. In der Folge ist bei Verwendung des Begriffs Weg immer ein gerichteter Weg gemeint.

Die Dichte  $d$  eines gerichteten Graphen ist laut [44d] das Verhältnis der im Graph vorhandenen Kanten und der Anzahl der Kanten, die theoretisch möglich wären, wenn es zwischen jedem Knotenpaar zwei Kanten, eine in jede Richtung, geben würde. Die Formel zur Berechnung der Dichte lautet:

$$d = \frac{L}{g(g-1)}$$

$L$  ist dabei die Anzahl der vorhandenen Kanten und  $g$  die Anzahl der Knoten im Graph. Würde jeder Knoten mit jedem anderen verbunden werden, so würde das in einem gerichteten Graphen  $g(g-1)$  Verbindungen (und  $\frac{g(g-1)}{2}$  Verbindungen in einem ungerichteten Graphen) ergeben, eine weniger als die Anzahl der Knoten, weil Verbindungen des Knotens zu sich selbst nicht mitgezählt werden. In Abbildung 5 ist  $L = 7$  und  $g = 6$ . Das heißt, die Dichte ist:

$$d = \frac{7}{6 \cdot 5} = \frac{7}{30} = 0,233$$

Die geodätische Distanz eines Knotens zu einem anderen Knoten ist die Anzahl der Kanten auf dem kürzesten Weg zwischen diesen beiden Knoten. Es gilt zu beachten, dass bei gerichteten Graphen die (geodätische) Distanz von einem Knoten  $X$  zu einem Knoten  $Y$  nicht gleich der (geodätischen) Distanz von Knoten  $Y$  zu Knoten  $X$  sein muss. [44e] Zum Beispiel ist die geodätische Distanz von Knoten 1 zu Knoten 4 im Graphen aus Abbildung 5 2. In diesem Fall gibt es zwei Wege von Knoten 1 zu Knoten 4 – den Weg über die Knoten 1, 2, 4 und jenen über die Knoten 1, 3, 2 und 4. Der kürzere Weg ist mit einer Länge von 2 Kanten der über die Knoten 1, 2 und 4.

Der Durchmesser eines gerichteten Graphen ist die Länge des längsten aller kürzesten Wege aller Knotenpaare. Das heißt, es werden die kürzesten Wege aller Knotenpaare herangezogen und davon wird das Maximum bestimmt. Dieses Maximum entspricht dem Durchmesser des Graphen. Die Berechnung des Durchmessers eines Graphen ist eigentlich nur für stark verbundene Graphen möglich. Ein gerichteter Graph ist dann stark verbunden, wenn es von jedem Knoten zu jedem anderen Knoten einen Weg gibt. Ist dies nicht der Fall, so ist die Berechnung des Durchmessers nicht definiert. [44e] In Abbildung 5 ist der Durchmesser des Graphen undefiniert, da es nicht von jedem Knoten zu allen anderen einen Weg gibt. Würden zur Berechnung des Durchmessers jedoch nur jene Knoten herangezogen, zwischen denen es auch einen Weg gibt, so wäre der Durchmesser 4, weil der längste aller kürzesten Wege, der eine Länge von 4 hat, jener von Knoten 1 zu Knoten 6 ist.

Nun wird kurz der in der Folge dieser Arbeit so wichtige Begriff des Knotengrades erklärt. Der Grad eines Knotens entspricht bei ungerichteten Graphen der Anzahl der Kanten mit welchen dieser Knoten verbunden ist. [107a] Bei gerichteten Graphen wird zwischen Eingangsgrad und Ausgangsgrad unterschieden. Der Eingangsgrad entspricht der Anzahl der Kanten die zu einem bestimmten Knoten hinführen, wobei die Pfeilspitze der Kante in Richtung dieses Knotens zeigt. Der Ausgangsgrad hingegen entspricht der Anzahl der Kanten die von einem bestimmten Knoten weg führen, wobei die Pfeilspitze der Kanten vom Knoten weg zeigt. [107c] In Abbildung 5 hat der Knoten 1 einen Ausgangsgrad von 2 und einen Eingangsgrad von 0 weil zwei Kanten vom Knoten weg führen (zu Knoten 2 und 3) und keine Kante zu Knoten 1 hinführt. Knoten 2 hingegen hat einen Eingangsgrad von 3 und einen Ausgangsgrad von 1 weil drei Kanten zu diesem Knoten hin führen (von den Knoten 1,3 und 5 ausgehend) und eine Kante vom Knoten weg führt (zu Knoten 4).

Am Schluss dieses Abschnitts wird noch der Begriff des isolierten Knotens definiert. Ein Knoten in einem ungerichteten Graphen heißt isoliert, wenn dessen Knotengrad null ist. In einem gerichteten Graphen heißt ein Knoten isoliert, wenn sowohl sein Eingangs- als auch sein Ausgangsgrad gleich 0 ist. [107a] Der Begriff isoliert kann sowohl auf einen einzelnen Knoten, als auch auf einen Teilgraphen eines Graphen angewandt werden. Ein Teilgraph wird als isoliert bezeichnet, wenn keiner seiner Knoten durch eine Kante mit dem Rest des Graphen verbunden ist.

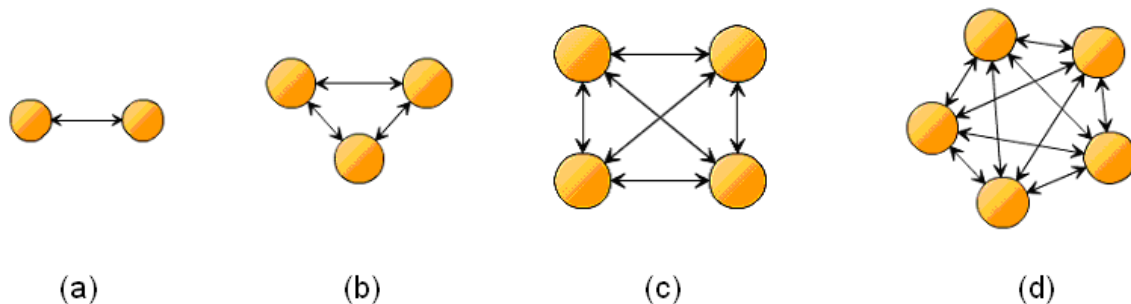
### **3.1.3 Weiterführende graphentheoretische Konzepte**

Im Folgenden werden einige zusätzliche graphentheoretische Konzepte vorgestellt. Abschnitt 3.1.3.1 behandelt den Begriff des vollständigen Graphen und Abschnitt 3.1.3.4 beschreibt was ein regulärer Graph ist; beide Begriffe werden zum Beispiel im Abschnitt 3.3.1 über die regulären Netzwerke gebraucht. Darüber hinaus werden in Abschnitt 3.1.3.2 und Abschnitt 3.1.3.3 zwei für die soziale Netzwerkanalyse wichtige graphentheoretische Konzepte, jene der Clique und der Komponente, vorgestellt.

In der Folge werden oft Graphen dargestellt deren Kanten manchmal zwei Pfeile in jeweils eine Richtung haben. Dies ist eine vereinfachte Darstellung von zwei Kantenverbindungen, jede in eine Richtung, zwischen zwei Knoten. Dies wird in der Folge immer so dargestellt.

#### **3.1.3.1 Vollständige Graphen**

Ein vollständiger ungerichteter Graph ist ein ungerichteter Graph bei dem jeder Knoten mit jedem anderen Knoten direkt verbunden ist. [107b] Ein vollständiger gerichteter Graph ist ein gerichteter Graph in dem jeder Knoten mit jedem anderen direkt und in beide Richtungen verbunden ist. [108] Abbildung 6 zeigt mehrere vollständige gerichtete Graphen.



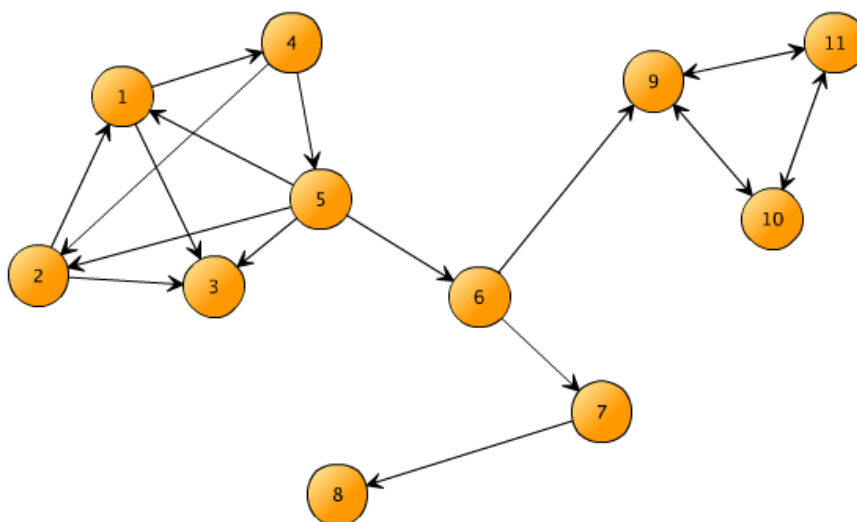
**Abbildung 6: Beispiele für gerichtete vollständige Graphen. (a) bestehend aus 2 Knoten, (b) bestehend aus 3 Knoten, (c) bestehend aus 4 Knoten und (d) bestehend aus 5 Knoten.**

### 3.1.3.2 Cliques

Das Konzept der Clique ist in der sozialen Netzwerkanalyse deswegen so wichtig, weil damit bestimmte Teile einer sozialen Gesellschaft gefunden werden können, in welchen die Akteure enger miteinander vernetzt sind als in der gesamten Gesellschaft. Bestimmte Berechnungen können dann zum Beispiel nicht nur allgemein für das gesamte Netzwerk, sondern auch speziell nur für die Cliques durchgeführt werden.

Zuerst die Beschreibung einer Clique für ungerichtete Graphen: hier ist eine Clique ein Teilgraph eines Graphen, in dem jeder Knoten mit jedem anderen direkt verbunden ist. Dieser Teilgraph kann nun in der sozialen Netzwerkanalyse zum Beispiel auf die Zentralitätsmaße untersucht werden, welche im Abschnitt 3.2.4 beschrieben werden. [46c]

Diese Definition einer Clique mag in manchen Fällen zu restriktiv sein, denn von jedem Knoten muss eine direkte Verbindung zu jedem anderen Knoten bestehen. Eine weniger restriktive Alternative dazu, in welcher die Knoten nicht notwendigerweise direkt miteinander verbunden sein müssen, ist das Konzept der n-Clique. Das „n“ steht hier für die maximal mögliche Weglänge um zwei Knoten in einer Clique miteinander zu verbinden. Für  $n=2$  würde das bedeuten, dass die Knoten der Clique mindestens über zwei Kanten miteinander verbunden sein müssen. Sind zwei Knoten „nur“ über drei Kanten miteinander verbunden so gehören sie nicht zur 2-Clique. Dieses Konzept gilt analog für alle  $n>2$ . [46c]



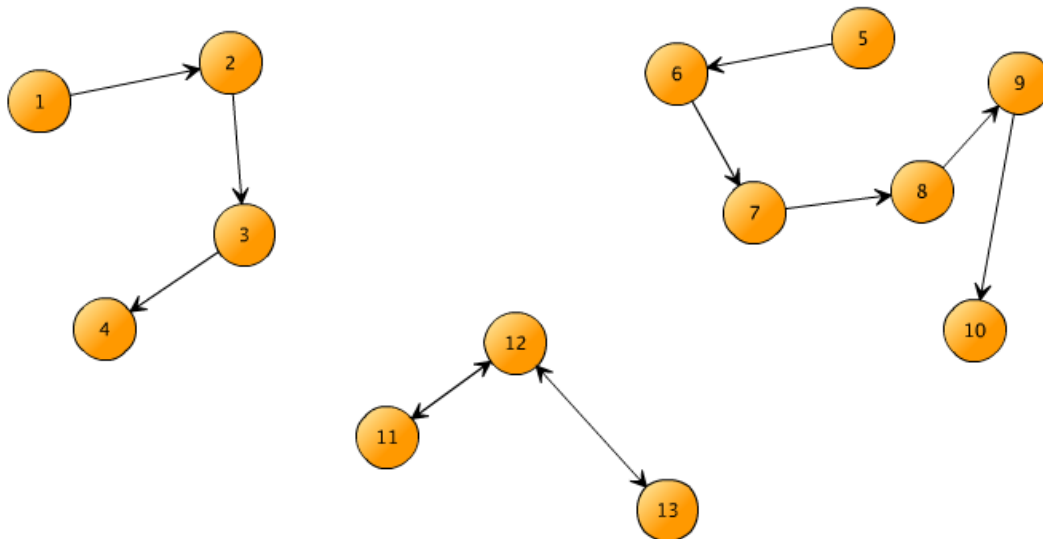
**Abbildung 7: Zeigt eine schwache und eine starke Clique. Die Knoten 1, 2, 3, 4 und 5 bilden die schwache Clique. Die starke Clique besteht aus den Knoten 1, 2 und 3.**

Bei gerichteten Graphen wird zwischen starken und schwachen Cliques unterschieden. Bei starken Cliques muss jeder Knoten mit jedem anderen direkt und in beiden Richtungen verbunden sein. Bei schwachen Cliques reicht eine „generelle“ Verbundenheit, das heißt jeder Knoten muss mit jedem anderen in mindestens eine Richtung direkt verbunden. Demzufolge existiert genau dann eine schwache Clique in einem gerichteten Graphen, wenn nach dem Weglassen der Pfeile an den Kanten jeder Knoten mit jedem anderen direkt verbunden ist. [47c]

In Abbildung 7 sind eine starke und eine schwache Clique innerhalb eines gerichteten Graphen dargestellt. Die schwache Clique wird durch den Teilgraphen, bestehend aus den Knoten 1, 2, 3, 4 und 5 gebildet. Die starke Clique besteht aus den Knoten 9, 10 und 11.

### 3.1.3.3 Komponenten

Ein weiteres wichtiges Konzept der sozialen Netzwerkanalyse ist jenes der Komponente, welches wiederum zuerst für ungerichtete Graphen erklärt wird. Eine Komponente ist ein Teilgraph eines ungerichteten Graphen, in dem jeder Knoten mit jedem anderen über einen Weg verbunden ist, der aber mit dem Rest des Graphen nicht verbunden ist. Dieser Teilgraph ist sozusagen ein isolierter Teilgraph (siehe am Ende des Abschnitts 3.1.2). [46c]



**Abbildung 8:** Zeigt zwei schwache und eine starke Komponente. Die Knoten 1, 2, 3 und 4, sowie die Knoten 5, 6, 7, 8, 9 und 10 bilden jeweils eine schwache Komponente. Die starke Komponente besteht aus den Knoten 11, 12 und 13.

Bei den gerichteten Graphen ist das Konzept der Komponenten auszuweiten. Hier wird in ähnlicher Weise wie bei den Cliques zwischen schwachen und starken Komponenten unterschieden. Eine schwache Komponente ist ein Teilgraph, der isoliert ist und in dem jeder Knoten mit jedem anderen in mindestens eine Richtung durch einen Weg miteinander verbunden ist. Eine starke Komponente hingegen ist ein Teilgraph, der auch isoliert ist, wobei aber jeder Knoten dieses Teilgraphs mit jedem anderen Knoten des Teilgraphs in beide Richtungen durch einen Weg verbunden sein muss. [46c]

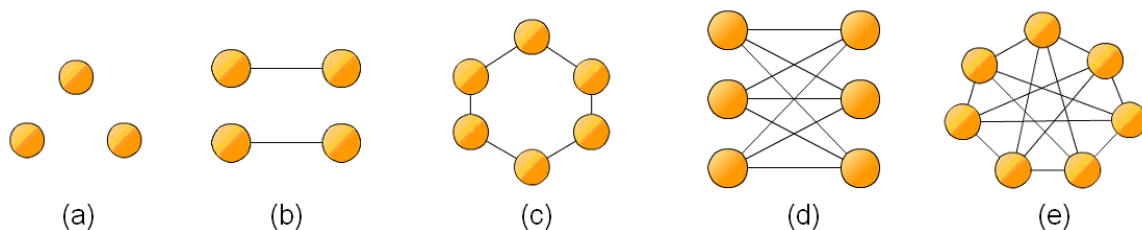
Abbildung 8 zeigt einen gerichteten Graphen mit einer starken und zwei schwachen Komponenten. Die Knoten 1, 2, 3 und 4, sowie die Knoten 5, 6, 7, 8, 9 und 10 bilden jeweils eine schwache Komponente in der es jeweils zu jedem Knoten einen Weg in eine Richtung gibt, die aber selbst isoliert ist. Die starke Komponente besteht aus den Knoten 11, 12 und 13. In dieser gibt es von jedem Knoten zu jedem anderen einen Weg in beiden Richtungen und sie ist auch isoliert.



### 3.1.3.4 Reguläre Graphen

Der Begriff des regulären Graphen ist wichtig für das Verständnis der regulären Netzwerke im Abschnitt 3.3.1. Hier werden nur ungerichtete Graphen behandelt, da auch die regulären Netzwerke, sowie die anderen Netzwerkmodelle, meist nur in dieser Form dargestellt und behandelt werden. Diese Netzwerkmodelle und deren Charakteristiken gelten jedoch in gleichem Maße auch für gerichtete Graphen, welche dem Netzwerk des österreichischen E-Tourismus-Marktes zugrunde liegen.

Ein Graph ist dann regulär, wenn jeder Knoten des Graphs den gleichen Knotengrad hat. Zusätzlich ist ein Graph  $k$ -regulär wenn jeder Knoten den Knotengrad  $k$  hat, wobei  $k \geq 0$  ist. Abbildung 9 zeigt Beispiele für  $k$ -reguläre Graphen für alle  $0 \leq k \leq 4$ . Abbildung 9a zeigt einen 0-regulären Graphen, Abbildung 9b einen 1-regulären Graphen, Abbildung 9c einen 2-regulären Graphen, Abbildung 9d einen 3-regulären Graphen und Abbildung 9e einen 4-regulären Graphen. [107b]



**Abbildung 9: Beispiele von  $k$ -regulären Graphen. (a) 0-regulärer Graph, (b) 1-regulärer Graph, (c) 2-regulärer Graph, (d) 3-regulärer Graph, (e) 4-regulärer Graph. Aus [107b].**

## 3.2 Soziale Netzwerkanalyse

In diesem Abschnitt wird zuerst ein Überblick über die Entwicklung der sozialen Netzwerkanalyse gegeben. Danach folgt eine Beschreibung der Daten und Matrizen, welche in der sozialen Netzwerkanalyse verwendet werden. Zum Schluss werden die wichtigsten Zentralitätsmaße für die Analyse von sozialen Netzwerken vorgestellt.

### 3.2.1 Geschichte der sozialen Netzwerkanalyse

Die Theorien der sozialen Netzwerkanalyse lassen sich bereits in den Anfängen der Sozialwissenschaften finden. Der im Jahre 1798 geborene Franzose Auguste Comte war einer der ersten, welcher den Begriff der Sozialwissenschaften prägte und versuchte, Gesetzmäßigkeiten hinter den Interaktionen von Individuen in einem sozialen System (zum Beispiel der Gesellschaft) zu erkennen. Parallel dazu gab es im 19. Jahrhundert auch die ersten Datensammlungen über und die ersten graphischen Repräsentationen von sozialen Systemen. [48]

In den 1930er Jahren entwickelte Jacob L. Moreno den Begriff der Soziometrie. Das war der nächste wichtige Schritt zur Entwicklung der sozialen Netzwerkanalyse. Soziometrie ist der Name für eine Technik, die angewandt wird, um die gegenseitigen Einflüsse von Individuen in sozialen Netzwerken aufzuzeigen und graphisch zu repräsentieren. Dabei sind die Links zwischen den Individuen Kanäle für die sozialen Einflüsse und Ideen. [49]

Diese Ideen zur die Soziometrie wurden in der Folge von einer Gruppe in die USA emigrierter Deutscher, welche im Bereich der soziale und kognitiven Psychologie arbeiteten, weiter entwickelt. Gleichzeitig arbeiteten Soziologen und Anthropologen der Harvard Universität an der Weiterentwicklung einiger Ideen des britischen Sozialanthropologen Radcliffe-Brown. Ihre Arbeit produzierte wichtige Studien über Gemeinschaften und Gesellschaften, welche ihre Betonung auf informative und interpersonale Verbindungen in

sozialen Systemen legte. In Großbritannien gab es einige Parallelen dazu. Auch auf den Sozialanthropologen Radcliffe-Brown aufbauend entwickelten einige Forscher, vor allem von der Universität Manchester, Ideen welche den Konfliktbereich von sozialen Akteuren in einem sozialen System betrafen. Sie erzielten sehr große Fortschritte durch die Anwendung der Mathematik in der sozialen Theorie. In den 1960er Jahren kam dann der endgültige Durchbruch der sozialen Netzwerkanalyse. Zusammen mit seinen Studenten erweiterte der Harvard-Professor Harrison White seine Untersuchungen zur mathematischen Basis sozialer Strukturen und verband sie mit Erkenntnissen seiner Vorgänger zu einer einzigartigen Synthese. In der Folge wurden die Argumente von White und die Arbeit der britischen Forscher zu einem kohärenten Framework der sozialen Netzwerkanalyse vereint. [47a]

### **3.2.2 Bestimmung und Arten von Netzwerkdaten**

Bevor eine Untersuchung des Netzwerks gemacht werden kann, muss zuerst die zu untersuchende Population bestimmt werden. Das heißt, es muss eine Grenze gezogen werden zwischen jenen Personen oder Akteuren, die in der Studie eine Rolle spielen und jenen, die hierfür nicht relevant sind. In der wissenschaftlichen Literatur gibt es zwei mögliche Ansätze zur Bestimmung der Grenze der Population: den realen und der nominalen Ansatz. Beim realen Ansatz bestimmen die Personen selbst, ob sie sich zu einer Population zugehörig fühlen oder nicht. Der nominale Ansatz hingegen beruht auf der Einschätzung des Wissenschaftlers. Er bestimmt, welche Eigenschaften einer Person die Zugehörigkeit zu einer Population bestimmen. [44a]

In dieser Arbeit wurde bei der Bestimmung des Netzwerks des österreichischen E-Tourismus-Marktes der nominale Ansatz gewählt. Welche Akteure in der wissenschaftlichen Literatur zum Tourismus und damit auch zum E-Tourismus gehören, wurde in Abschnitt 2.3.3 beschrieben. Die Population sollte also relativ einfach zu bestimmen sein – dies war in dieser Untersuchung aber nicht gänzlich der Fall.

Denn zum einen sollten auch Akteure im Netzwerk berücksichtigt werden die nicht zu einer der in Abschnitt 2.3.3 genannten Akteure des Tourismusmarktes gehören, welche aber indirekt mit dem Tourismus zu tun haben. Dazu zählen zum Beispiel die sozialen Medien (Facebook, Twitter usw.) die deswegen berücksichtigt werden müssen, weil Akteure des Tourismusmarktes in diesen sozialen Medien für sich werben. Und zum anderen ist der im Rahmen dieser Arbeit entwickelte Crawler nicht in der Lage zwischen touristischen und nicht touristischen Akteuren zu unterscheiden. Deshalb musste der Crawler öfters gestartet werden und am Ende musste dann händisch jenes Netzwerk ausgewählt werden, welches die wenigsten für diese Analyse irrelevante Knoten enthält, damit das gefundene Netzwerk das Netzwerk des österreichischen E-Tourismus-Marktes am besten repräsentiert.

Bei der Bestimmung der Population für eine Untersuchung ist es leider nicht immer möglich, alle Akteure einer zu untersuchenden Population in der Untersuchung zu berücksichtigen, weil es u.a. zuviel Aufwand bedeuten würde, die gesamte Population aufgrund ihrer Größe zu untersuchen. In diesem Fall kann nur ein Teil der Population für die Untersuchung berücksichtigt werden. Dieser Teil der Population muss mit einer hohen Wahrscheinlichkeit alle Akteure der Population repräsentieren, ansonsten wäre die Untersuchung nicht aufschlussreich. [44a] In der hier vorliegenden Untersuchung mit Hilfe eines Crawlers war es aufgrund von Zeit- und Strukturgründen nicht möglich, alle zum E-Tourismus-Markt gehörenden Websites zu erfassen. Einerseits war es wegen der Geschwindigkeit des Crawlers und der Größe des Netzwerks nicht möglich, in der zur Verfügung stehenden Zeit alle relevanten Websites zu erfassen. Andererseits ist dies aufgrund der Struktur des Netzwerks und der Abhängigkeit von den Startknoten sowie dem Vorgehen des Crawlers nicht möglich, denn nicht alle tourismusbezogenen Websites sind mit anderen touristischen Websites verbunden und können deswegen nicht gefunden werden. Auch startet der Crawler von einer

Website und wechselt zur nächsten über die Links dieser Website, weshalb die Wahl der Startwebsite ausschlaggebend dafür ist, welches Netzwerk gefunden wird. Leider gibt es keine touristische Website, die durch Links mit allen anderen Websites des österreichischen E-Tourismus-Marktes verbunden ist, weshalb nie alle zum österreichischen E-Tourismus-Markt gehörenden Websites gefunden werden können.

Es muss also ein Teilnetzwerk als Repräsentation des gesamten Netzwerks dienen. [44a] nennt als Möglichkeit zur Findung eines Teilnetzwerks, welches das gesamte Netzwerk repräsentiert, das so genannte „snowball sampling“. Dabei werden von einem Startknoten aus alle mit ihm verbundenen Knoten ermittelt und ins Netzwerk aufgenommen. Von den neu hinzugekommenen Knoten ausgehend werden wiederum alle Knoten ermittelt, mit denen sie verbunden sind und die so gefundenen Knoten werden wiederum ins Netzwerk aufgenommen. Dieser Vorgang setzt sich so lange fort, bis der Wissenschaftler meint, genug repräsentative Knoten für das Netzwerk zu haben. Genau auf diese Art und Weise funktioniert auch ein Crawler. Von einem Startknoten aus findet dieser über die Links die Knoten in seinem Umkreis und durchsucht dann diese Knoten nach Links zu anderen Knoten, usw.

Wichtig im Zusammenhang mit den Netzwerkdaten ist auch der Unterschied zwischen so genannten One-mode- und Two-mode-Netzwerkdaten. Bei One-mode-Netzwerkdaten werden nur die Daten von einer Gruppe von Akteuren gesammelt. Diese Akteure können sowohl Empfänger als auch Sender von Verbindungen sein. Bei Two-mode-Netzwerken werden die Daten zu zwei verschiedenen Gruppen von Akteuren gesammelt. Zu einer Gruppe gehören dann zum Beispiel nur Sender einer Verbindung und zur anderen Gruppe nur Empfänger von Verbindungen. [50] Im Falle dieser Arbeit sind die Daten One-mode-Netzwerkdaten, da es sich um eine Gruppe von Akteuren, die vorwiegend dem Tourismussektor zugehörige Websites sind, handelt und alle im Netzwerk präsenten Websites sowohl Links „senden“ als auch Links „empfangen“ können.

### 3.2.3 Darstellung von Netzwerkdaten in Form von Matrizen

Jeder Graph und jedes Netzwerk kann auch in Form einer Matrix dargestellt werden. Für die Beispielsmatrix, die im Folgenden angeführt wird, wird das in Abbildung 10 dargestellte Netzwerk verwendet.

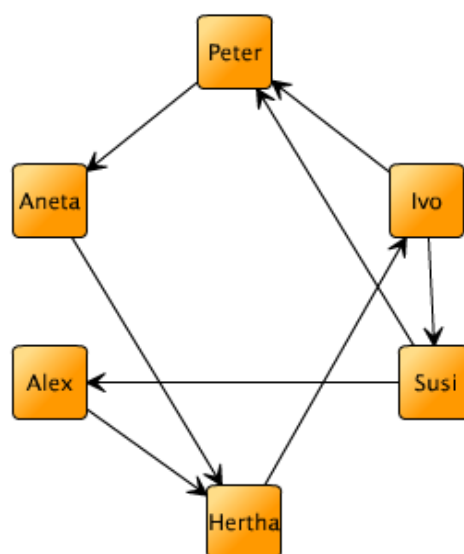


Abbildung 10: Ein Beispiel-Netzwerk bestehend aus sechs Personen

[44f] nennt zwei Arten von Matrizen, die zur Darstellung eines Netzwerks verwendet werden können: die Adjazenzmatrix und die so genannte Inzidenzmatrix. In dieser Arbeit wird immer die Adjazenzmatrix verwendet, da auch die meisten Programme zur Darstellung von sozialen Netzwerken mit Adjazenzmatrizen arbeiten. Im Folgenden wird deshalb auch nur die Adjazenzmatrix beschrieben.

### 3.2.3.1 Adjazenzmatrix

In Tabelle 5 ist eine Adjazenzmatrix – in der sozialen Netzwerkanalyse auch Soziomatrix genannt – dargestellt, welche den Graphen aus Abbildung 10 repräsentiert.

	<b>Peter</b>	<b>Aneta</b>	<b>Alex</b>	<b>Ivo</b>	<b>Hertha</b>	<b>Susi</b>
<b>Peter</b>	-	1	0	0	0	0
<b>Aneta</b>	0	-	0	0	1	0
<b>Alex</b>	0	0	-	0	1	0
<b>Ivo</b>	1	0	0	-	0	1
<b>Hertha</b>	0	0	0	1	-	0
<b>Susi</b>	1	0	1	0	0	-

**Tabelle 5: Die Matrixdarstellung des Netzwerks aus Abbildung 10**

In der Graphentheorie wird diese Matrix meist Adjazenzmatrix genannt, weil die Einträge in der Matrix zeigen, ob zwei Knoten adjazent zueinander (d.h. benachbart, weil sie durch eine Kante miteinander verbunden sind) sind oder nicht. Jedoch wird diese Art Matrix in der sozialen Netzwerkanalyse immer als Soziomatrix bezeichnet. [44f]

Die Größe der Soziomatrix ist  $g \times g$  ( $g$  Zeilen und  $g$  Spalten) für One-mode-Netzwerke. Es gibt eine Zeile und eine Spalte für jeden Knoten. Die Knoten sind als  $1, 2, \dots, g$  gekennzeichnet. Die Einträge in der Soziomatrix ( $x_{ij}$ ) zeigen, welche Knoten zueinander adjazent sind oder nicht. Es gibt eine 1 in Zelle  $(i, j)$ , wenn es eine Verbindung vom Knoten  $n_i$  zum Knoten  $n_j$  gibt. Sonst gibt es eine 0 in Zelle  $(i, j)$ . Dabei muss der Knoten  $n_i$  der Ausgangsknoten und der Knoten  $n_j$  der Zielknoten sein. [44f]

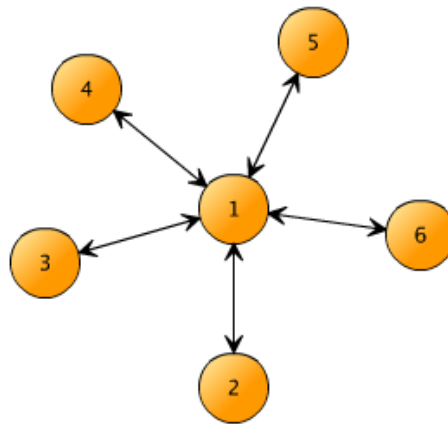
Die Elemente der Hauptdiagonale sind die Verbindungen jedes Knotens zu sich selbst. Die Soziomatrix eines Graphen in dem jeder Knoten mit jedem andern direkt verbunden ist enthält, abgesehen von den Elementen in der Hauptdiagonale, alles 1en. Die Soziomatrix eines Graphen in dem es nur Knoten und keine Kanten gibt enthält nur 0en. [44f]

### 3.2.4 Wichtige Kennzahlen der sozialen Netzwerkanalyse

In diesem Abschnitt werden vor allem die drei wichtigsten Zentralitätskennzahlen (Grad-basierte Zentralität, Nähe-basierte Zentralität und Zwischenzentralität) der sozialen Netzwerkanalyse vorgestellt. Am Ende des Abschnitts wird noch kurz auf eine weitere wichtige, im Rahmen dieser Arbeit verwendete Kennzahl (Clusterkoeffizient) eingegangen.

Das Konzept der Zentralität ist eines der am meisten untersuchten und wichtigsten Konzepte der sozialen Netzwerkanalyse [51]. Im Laufe der Zeit wurden verschiedene Messmethoden zur Bestimmung der Zentralität eines Knotens innerhalb eines sozialen Netzwerks entwickelt. Diese Messmethoden wurden für die Untersuchung von verschiedensten Netzwerken eingesetzt und ihre Ergebnisse wurden auf verschiedenste Art und Weise interpretiert. [52] Obwohl es so viele wichtige Messmethoden für Zentralität gibt, sind sich die Wissenschaftler nicht ganz darüber klar, was unter dem Begriff Zentralität zu verstehen ist. Den gängigsten Definitionen nach bedeutet Zentralität die strukturelle Wichtigkeit oder Prominenz eines Knotens.

Den in der sozialen Netzwerkanalyse wahrscheinlich am meisten anerkannten Beitrag zum Zentralitätsbegriff lieferte Linton C. Freeman im Jahre 1979. Er reduzierte dabei die vielen verschiedenen zentralitätsbezogenen Messmethoden auf drei Zentralitätskennzahlen, welche seitdem zu den wichtigsten in der sozialen Netzwerkanalyse gehören und in der Folge vorgestellt werden. Ein Knoten, der zum Beispiel im Zentrum des in der Theorie so wichtigen Stern-Netzwerks steht, erhält bei jeder dieser drei Kennzahlen den maximalen Wert. [52] und [53] Dieses Stern-Netzwerk ist in Abbildung 11 dargestellt. Knoten 1 ist dabei der zentrale Knoten, welcher bei allen drei Zentralitätskennzahlen die maximalen Werte erreicht.



**Abbildung 11: Das Stern-Netzwerk**

Die hier in der Folge beschriebenen drei wichtigen Zentralitätskennzahlen (Grad-basierte Zentralität, Nähe-basierte Zentralität und Zwischenzentralität) werden Punktzentralitäten genannt, weil die Zentralität eines Punktes berechnet wird. Neben der Punktzentralität gibt es auch das Konzept der Graphzentralisierung. Ein gesamter Graph kann zentral oder nicht zentral ausgerichtet sein. Ein Graph ist zum Beispiel dann zentral ausgerichtet, wenn alle Knoten gleichmäßig um einen zentralen Knoten oder eine Menge von zentralen Knoten aufgebaut sind. [47b] Das beste Beispiel für einen zentralen Graphen ist der in Abbildung 11 dargestellte Graph in Sternform bzw. das Stern-Netzwerk.

Zur Erklärung der drei Zentralitätskennzahlen dient das Netzwerk in Abbildung 12 und die dazugehörige Tabelle 6 mit den Werten für die Zentralitätskennzahlen. Dieses Netzwerk, welches gut zur Erklärung der Zentralitätsmaße verwendet werden kann, wurde in ungerichteter Form aus [47b] entnommen und in ein gerichtetes Netzwerk umgewandelt, wobei jede ungerichtete Kante einfach in eine in beide Richtungen verlaufende, gerichtete Kante umgewandelt wurde, um die gute beispielhafte Darstellung des Netzwerks zur Illustration der Unterschiede bei den drei Zentralitätskennzahlen beizubehalten. Auch die Informationen zu den Abschnitten 3.2.4.1 wurden komplett aus [47b] entnommen, wobei die darin angeführten Formeln in [47b] nur in wörtlicher Form vorhanden waren und mit eigenen Parameterbezeichnungen angeführt. Für die Abschnitte 3.2.4.2 und 3.2.4.3 waren die Formeln zwar teilweise in den angeführten Quellen vorhanden, wurden jedoch auch mit eigenen Parameterbezeichnungen versehen.

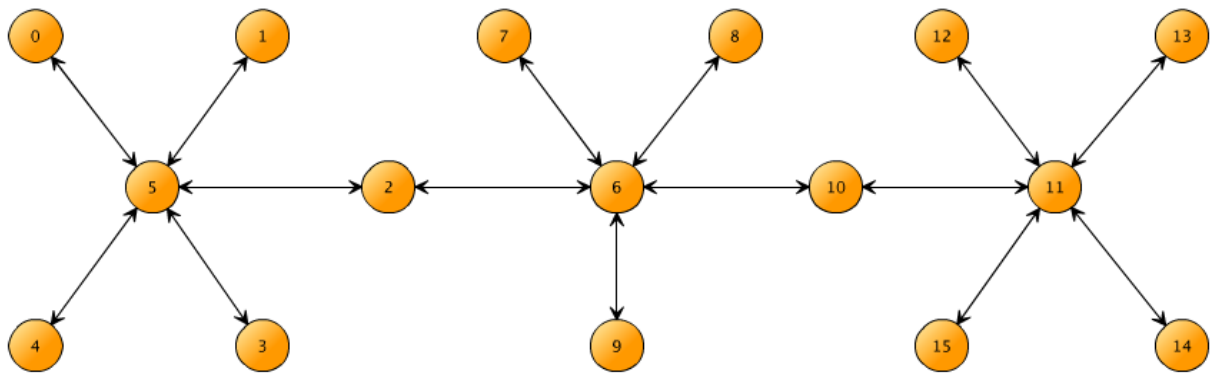


Abbildung 12: Ein beispielhafter Graph anhand dessen die drei Zentralitätskennzahlen gut erklärt werden können. Adaptiert aus Quelle [47b]

Knoten			6	2,10	5,11	7,8,9	Restliche Knoten
<b>Grad-basierte Zentralität</b>	Eingehend	Absolut	5	2	5	1	1
		Relativ	0,33	0,133	0,33	0,0667	0,0667
	Ausgehend	Absolut	5	2	5	1	1
		Relativ	0,33	0,133	0,33	0,0667	0,0667
<b>Nähe-basierte Zentralität</b>	Eingehend	Absolut	33	37	43	47	57
		Relativ	0,45455	0,40541	0,34884	0,31915	0,26316
	Ausgehend	Absolut	33	37	43	47	57
		Relativ	0,45455	0,40541	0,34884	0,31915	0,26316
<b>Zwischenzentralität</b>	Absolut	75	50	50	0	0	
	Relativ	0,35714	0,2381	0,2381	0	0	

Tabelle 6: Zentralitätswerte des Graphen in Abbildung 12

### 3.2.4.1 Grad-basierte Zentralität (Degree)

Die Grad-basierte Zentralität ist, im Gegensatz zur Nähe-basierten Zentralität, ein lokales Zentralitätsmaß, weil bei dessen Berechnung nur die direkten Nachbarn herangezogen werden. Es gibt zwar auch die Möglichkeit, den Begriff so zu definieren, dass auch die indirekten Nachbarn zur Berechnung der Grad-basierten Zentralität hinzugezogen werden. Das ist jedoch eher unüblich und wird auch bei der Berechnung der Grad-basierten Zentralität des Netzwerks des österreichischen E-Tourismus-Marktes nicht angewandt.

Ein Knoten hat bei ungerichteten Graphen einen hohen Grad-basierten Zentralitätswert, wenn er einen hohen Knotengrad hat und kann deswegen im lokalen Umfeld als zentraler Knoten angesehen werden. Bei gerichteten Graphen wird zwischen Eingangsgrad-basierter Zentralität und Ausgangsgrad-basierter Zentralität unterschieden. Das heißt, ein Knoten mit hohem Ausgangsgrad hat einen hohen Ausgangsgrad-basierten Zentralitätswert, hingegen hat ein Knoten mit hohem Eingangsgrad einen hohen Eingangsgrad-basierten Zentralitätswert. Hier wird auch zwischen absoluter und relativer Eingangs- bzw. Ausgangsgrad-basierter Zentralität unterschieden. Zu den Begriffserklärungen des Ein- und Ausgangsgrads siehe Abschnitt 3.1.2.

Formal gesehen ist die absolute Eingangsgrad-basierte Zentralität eines Knotens:

$$EGZ_i^{abs} := EG_i$$

$EGZ_i^{abs}$  ist dabei die absolute Eingangsgrad-basierte Zentralität des Knotens  $i$  und  $EG_i$  der Eingangsgrad des Knotens  $i$ . Die relative Eingangsgrad-basierte Zentralität wird wie folgt berechnet:

$$EGZ_i^{rel} = \frac{EG_i}{EGZ_{max}^{abs}}$$

$EGZ_i^{rel}$  entspricht hier der relativen Eingangsgrad-basierten Zentralität,  $EG_i$  dem Eingangsgrad von Knoten  $i$  und  $EGZ_{max}^{abs}$  dem maximalen absoluten Eingangsgrad den ein Knoten im Graph haben kann. Dieser entspricht immer der Anzahl der Knoten im Graph minus 1.

Dies gilt analog für die Ausgangsgrad-basierte Zentralität. Auch hier wird zwischen relativem und absolutem Wert unterschieden. Die absolute Ausgangsgrad-basierte Zentralität ist deswegen:

$$AGZ_i^{abs} = AG_i$$

$AGZ_i^{abs}$  entspricht der absoluten Ausgangsgrad-basierten Zentralität und  $AG_i$  dem Ausgangsgrad des Knotens  $i$ . Die relative Ausgangsgrad-basierte Zentralität lässt sich mit der folgenden Formel errechnen:

$$AGZ_i^{rel} = \frac{AG_i}{AGZ_{max}^{abs}}$$

$AGZ_i^{rel}$  entspricht hier der relativen Ausgangsgrad-basierten Zentralität,  $AG_i$  dem Ausgangsgrad von Knoten  $i$  und  $AGZ_{max}^{abs}$  dem maximalen absoluten Ausgangsgrad den ein Knoten im Graph haben kann. Wiederum ist das die Anzahl der Knoten minus 1.

Nun ein Beispiel für die Berechnung der Grad-basierten Zentralität. In Abbildung 12 haben zum Beispiel die Knoten 5,6 und 11 die höchste Ein- und Ausgangsgrad-basierte Zentralität haben, weil sie die meisten ein- und ausgehenden Verbindungen bzw. den höchsten Ein- und Ausgangsgrad besitzen. Tabelle 6 bestätigt dies. Alle drei Knoten haben einen Aus- und Eingangsgrad von 5. Dabei wird, wie bereits erwähnt, zwischen relativem und absolutem Ein- bzw. Ausgangsgrad unterschieden. In Abbildung 12 gibt es insgesamt 16 Knoten, das heißt der maximale Knotengrad eines Knotens ist 15. Wenn nun ein Knoten mit 5 anderen Knoten verbunden ist, dann ist sein relativer Ein- bzw. Ausgangsgrad 5 dividiert durch 15, was den Wert 0,33 ergibt. Die relative Grad-basierte Zentralität eines Knotens hat den Vorteil, dass sie, im Gegensatz zur absoluten Grad-basierten Zentralität, von der Größe des Graphen (Anzahl der Knoten im Graph) unabhängig ist und sich somit die relativen Grad-basierten Zentralitätswerte von Knoten aus verschiedensten Graphen miteinander vergleichen lassen.

### 3.2.4.2 Nähe-basierte Zentralität (Closeness)

Die Nähe-basierte Zentralität ist im Gegensatz zur Grad-basierten Zentralität ein globales Zentralitätsmaß, weil für dessen Berechnung alle Knoten des Netzwerks, und nicht nur die direkten Nachbarn eines Knotens, herangezogen werden. Auch die Nähe-basierte Zentralität kann auf gerichtete und ungerichtete Graphen angewandt werden, wobei hier wiederum das Augenmerk nur auf die gerichteten Graphen gelegt wird. Bei der Nähe-basierten Zentralität sind die kürzesten Wege von einem Knoten zum anderen von Bedeutung. Dabei werden für jeden Knoten alle kürzesten Wege zu allen anderen Knoten berechnet. [47b]

Auch hier muss zwischen absoluter und relativer Nähe-basierter Zentralität unterschieden werden. [113] Die Summe aller kürzesten Wege von einem Knoten zu jeden anderen ergibt die absolute Nähe-basierte Zentralität, wobei beachtet werden muss dass ein Knoten zentraler liegt, je geringer sein Nähe-basierter Zentralitätswert ist. Je nach dem, in welche Richtung die

kürzesten Wege berechnet werden, wird zwischen Ein- und Ausgangsnähe-basierter Zentralität unterschieden. [47b] Die relative Nähe-basierte Zentralität eines Knotens ergibt sich aus dem Quotienten seiner absoluten Nähe-basierten Zentralität und der minimalen Nähe-basierten Zentralität. [113]

Die absolute Eingangsnähe-basierte Zentralität eines Knotens entspricht der Summe der geodätischen Distanzen zu allen anderen Knoten des Graphen [47b]:

$$ENZ_i^{abs} = \sum_{j=1}^n GDK_{ij}, \text{ für alle } j \neq i$$

$ENZ_i^{abs}$  bezeichnet den absoluten Eingangsnähe-basierten Zentralitätswert;  $GDK_{ij}$  steht für die geodätische Distanz (kürzester Weg, siehe Abschnitt 3.1.2) zwischen Knoten  $i$  und Knoten  $j$ .  $n$  entspricht der Anzahl der Knoten im Graphen. Analog berechnet sich die absolute Ausgangsnähe-basierte Zentralität [47b]:

$$ANZ_i^{abs} = \sum_{j=1}^n GDK_{ji}, \text{ für alle } j \neq i$$

$ANZ_i^{abs}$  ist der absolute Ausgangsnähe-basierte Zentralitätswert;  $GDK_{ji}$  entspricht der geodätischen Distanz von Knoten  $j$  zu Knoten  $i$ .  $n$  steht wiederum für die Anzahl der Knoten im Graphen. Insgesamt gilt, je geringer der absolute Ein- bzw. Ausgangsnähe-basierte Zentralitätswert, desto zentraler liegt dieser Knoten im Netzwerk bezüglich der Nähe-basierten Zentralitätskennzahl.

Nun zur relativen Nähe-basierten Zentralität. Für die eingehende Richtung berechnet sie sich wie folgt [113]:

$$ENZ_i^{rel} = \frac{n-1}{\sum_{j=1}^n GDK_{ij}}, \text{ für alle } j \neq i$$

$ENZ_i^{rel}$  ist der relative Eingangsnähe-basierte Zentralitätswert; die anderen Parameter wurden bereits oben erklärt. Der relative Eingangsnähe-basierte Zentralitätswert ergibt sich also aus dem Verhältnis der minimalen Eingangsnähe-basierten Zentralität, wenn jeder kürzeste Weg zu jedem Knoten eine Länge von 1 hat, und der absoluten Eingangsnähe-basierten Zentralität. In diesem Fall gilt im Gegensatz zur absoluten Nähe-basierten Zentralität, dass ein Knoten um so zentraler liegt, je höher sein relative Nähe-basierter Zentralitätswert ist. Analog dazu berechnet sich die relative Ausgangsnähe-basierte Zentralität [113]:

$$ANZ_i^{rel} = \frac{n-1}{\sum_{j=1}^n GDK_{ji}}, \text{ für alle } j \neq i$$

Hierfür gelten die Erklärungen wie bei der relativen Eingangsnähe-basierten Zentralität. Der Parameter  $ANZ_i^{rel}$  ist in diesem Fall die relative Ausgangsnähe-basierte Zentralität.

Im Netzwerk aus Abbildung 12 erhält der Knoten 6 den geringsten absoluten bzw. den höchsten relativen Nähe-basierten Zentralitätswert, wie in Tabelle 6 zu sehen ist. Beim lokalen Zentralitätsmaß in Form der Grad-basierten Zentralität hatten die Knoten 6, 5 und 11 die gleichen Werte und waren demzufolge gleich zentral. Wenn aber das gesamte Netzwerk in Abbildung 12 betrachtet wird, ist zu sehen, dass die drei Knoten optisch gesehen nicht gleich zentral liegen. Am zentralsten liegt hier eindeutig Knoten 6. Sein relativer Ein- bzw. Ausgangsnähe-basierter Zentralitätswert ergibt sich aus der Division von 15 durch 33. Das Maß der Nähe-basierten Zentralität ist deswegen so wichtig, weil es im Unterschied zur Grad-



basierten Zentralität nicht nur die lokale Umgebung eines Knotens, sondern das gesamte Netzwerk betrachtet.

### 3.2.4.3 Zwischenzentralität (Betweenness)

Das dritte hier vorgestellte Zentralitätsmaß ist die Zwischenzentralität. Dieses Zentralitätsmaß misst, inwieweit ein Knoten eine Vermittlerrolle bei der Verbindung der anderen Knoten spielt. Knoten mit einem hohen Zwischenzentralitätswert haben manchmal einen niedrigen Ein- oder Ausgangsgrad und liegen dabei an einem wichtigen Verbindungspunkt zwischen zwei Teilen eines Netzwerks, sodass viele Verbindungen zwischen zwei Knoten aus diesen beiden Netzwerkteilen notwendigerweise über diese Knoten laufen müssen. Das Konzept der Zwischenzentralität hat also viel mit der gegenseitigen Abhängigkeit von Knoten in einem Netzwerk zu tun. Ein Knoten ist dann in hohem Maße abhängig von einem anderen, wenn die meisten Verbindungen zu den übrigen Knoten über diesen anderen Knoten erfolgen. [47b]

Die Berechnung der Zwischenzentralität basiert, ähnlich wie die Nähe-basierte Zentralität auf dem Konzept der geodätischen Distanzen zwischen zwei Knoten, ist aber etwas komplizierter. Sie wird zwar vom Konzept her für ungerichtete und gerichtete Graphen ähnlich durchgeführt, zuerst muss aber das Verfahren für ungerichtete Graphen erklärt werden, da dieses die Basis für die Anwendung auf gerichtete Graphen bildet.

Auch wird bei der Zwischenzentralität, gleich wie bei den beiden vorher beschriebenen Zentralitätskennzahlen zwischen absoluter und relativer Zwischenzentralität unterschieden. Die absolute Zwischenzentralität ist relativ einfach berechenbar, wenn es nur eine mögliche geodätische Distanz zwischen zwei Knoten gibt. In diesem Fall wäre der absolute Zwischenzentralitätswert eines Knotens die Anzahl der geodätischen Distanzen (oder kürzesten Wege) aller Knotenpaare, welche über diesen Knoten führen. Der Fall, dass es nur eine mögliche geodätische Distanz zwischen zwei Knoten gibt ist aber in der Praxis relativ selten. Im Falle von mehreren möglichen geodätischen Distanzen muss bei der Erklärung etwas ausgeholt werden. Es sei das Knotenpaar  $i$  und  $j$  und es existieren mehrere geodätische Distanzen (d.h. mehrere kürzeste Wege) zwischen diesem Knotenpaar. [113]

Zuerst muss nun der absolute Zwischenzentralitätswert eines Knotens nur für ein bestimmtes Knotenpaar wie folgt berechnet werden [113]:

$$b_{ij}(p_k) = \frac{1}{g_{ij}} * g_{ij}(p_k)$$

$g_{ij}$  entspricht der Anzahl der geodätischen Distanzen zwischen Knoten  $i$  und  $j$ ; das heißt der Quotient von 1 durch  $g_{ij}$  entspricht der Wahrscheinlichkeit einer geodätischen Distanz, von allen zwischen diesen Knoten existierenden geodätischen Distanzen, wobei die Wahrscheinlichkeit für alle Distanzen als gleich angenommen wird.  $g_{ij}(p_k)$  ist die Anzahl der geodätischen Distanzen zwischen Knoten  $i$  und  $j$ , welche über Knoten  $k$  führen.

Um nun den absoluten Zwischenzentralitätswert eines Knotens zu berechnen werden die  $b_{ij}(p_k)$  für alle Knotenpaare bestimmt und summiert [113]:

$$ZZ_k^{abs} = \sum_i^n \sum_j^n b_{ij}(p_k), \text{ für alle } i \neq j \neq k$$

$ZZ_k^{abs}$  ist der absolute Zwischenzentralitätswert des Knotens  $k$ ,  $n$  bezeichnet die Anzahl der Knoten im Graph.

Da absolute Zwischenzentralitätswerte verschiedener Graphen von unterschiedlicher Knotenanzahl nicht miteinander vergleichbar sind, wurde der relative Zwischenzentralitäts-

wert eingeführt. Dieser berechnet sich aus dem absoluten Zwischenzentralitätswert wie folgt [113]:

$$ZZ_k^{rel} = \frac{2 * ZZ_k^{abs}}{n^2 - 3 * n + 2}$$

Die Ergebnisse für den relativen Zwischenzentralitätswert liegen zwischen 0 und 1. In dieser Formel wird der absolute Zwischenzentralitätswert eines Knotens mit dem maximalen Zwischenzentralitätswert eines Knotens in einem ungerichteten Graphen verglichen, indem ersterer durch letzteren dividiert wird. Denn der maximale Zwischenzentralitätswert eines Knotens beträgt [113]:

$$\frac{n^2 - 3 * n + 2}{2}$$

Diesen maximalen Zwischenzentralitätswert erhält der zentrale Knoten (Knoten 1) im Stern-Netzwerk das in Abbildung 11 dargestellt ist. Sein relativer Zwischenzentralitätswert ist damit 1 und alle anderen Knoten haben den Wert 0.

Bei gerichteten Graphen wird die absolute Zwischenzentralität gleich wie bei ungerichteten Graphen berechnet. Die Berechnung der relativen Zwischenzentralität verläuft ähnlich analog. Auch hier wird die absolute Zwischenzentralität eines Knotens durch die maximale Zwischenzentralität dividiert. Jedoch ist die maximale Zwischenzentralität eines Knotens in einem gerichteten Graphen nicht eindeutig und hängt von der Anzahl der Knoten mit ausgehenden Kanten  $n_a$ , der Anzahl der Knoten mit eingehenden Kanten  $n_e$  und der Anzahl der Knoten mit reziproken Kanten  $n_r$  (d.h. in beide Richtungen führende Kanten) ab. [114]

Unter diesen Voraussetzungen berechnet sich der relative Zwischenzentralitätswert eines Knotens k wie folgt [114]:

$$ZZ_k^{rel} = \frac{ZZ_k^{abs}}{(n_a - 1) * (n_e - 1) - (n_r - 1)}$$

Wobei  $ZZ_k^{abs}$  wieder den absoluten Zwischenzentralitätswert von Knoten k darstellt, diesmal für einen gerichteten Graphen und der Nenner in diesem Bruch entspricht der maximalen Zwischenzentralität eines Knotens mit den vorgegebenen Anzahlen der Knoten mit eingehenden, ausgehenden und reziproken Kanten.

Längere Zeit war die Vergleichbarkeit der relativen Zwischenzentralitätswerte von Knoten verschiedener Graphen umstritten. [114] Jedoch behauptet [114], dass sich die relativen Zwischenzentralitätswerte von Knoten aus verschiedenen Graphen, wenn sie mit obiger Formel berechnet werden, nicht nur bei unterschiedlicher Knotenanzahl, sondern auch bei unterschiedlichem Grad an Reziprozität miteinander vergleichen lassen. Dies scheint dem Autor unverständlich, hängt doch obige Formel von der Anzahl der Knoten mit eingehenden, ausgehenden und reziproken Kanten ab. Demnach wäre das Ergebnis auch unterschiedlich wenn sich der Grad der Reziprozität ändert und könnte dann nur mit guter Kenntnis der unterschiedlichen Verhaltensweisen der Zwischenzentralitäten bei verschieden-reziprok ausgeprägten Graphen interpretiert werden.

Nichtsdestotrotz liegt ein Knoten gemäß der Zwischenzentralität umso zentraler je höher sein Zwischenzentralitätswert, sowohl im relativen, als auch im absoluten Sinn ist. Auch ist es für diese Arbeit nicht wichtig Zwischenzentralitätswerte von Knoten aus unterschiedlichen Graphen miteinander zu vergleichen, denn es müssen nur die zentralen Knoten des Netzwerks des österreichischen E-Tourismus-Marktes anhand dieser Kennzahl ermittelt und nicht mit anderen Netzwerken verglichen werden.

Tabelle 6 zeigt die Zwischenzentralitätswerte für die Knoten aus dem Netzwerk von Abbildung 12. Da das Netzwerk gleichmäßig um ein Zentrum, nämlich Knoten 6, verteilt ist und deshalb die meisten kürzesten Wege über diesen Knoten führen, hat dieser Knoten auch den höchsten Zwischenzentralitätswert von 75 (absolut) und 0,35714 (relativ). Die Knoten 2,5,10 und 11 folgen mit einem Wert 50 (absolut) und 0,2381 (relativ) darauf (siehe Tabelle 6). Alle anderen Knoten, die im Netzwerk weiter außen liegen, sind in keinem Fall Vermittlerknoten zweier Knotenpaare und haben deshalb einen Zwischenzentralitätswert von 0. Es gilt zu beachten dass die relative Zwischenzentralität in UCINET [85], dem Programm zur Analyse von sozialen Netzwerken das in dieser Arbeit verwendet wird, immer nach der ungerichteten Variante berechnet wird. In diesem Fall erhält Knoten 6 aus Abbildung 12 zwar denselben absoluten Zwischenzentralitätswert von 75, jedoch den doppelten relativen Zwischenzentralitätswert von 0,71429, als er nach der Berechnung mit der Formel für gerichtete Graphen erhalten würde. Da das Netzwerk des österreichischen E-Tourismus-Marktes jedoch in Bezug auf die Zwischenzentralität nicht mit einem anderen Netzwerk verglichen wird, ist dies unwichtig, da in diesem Fall die Knoten mit den höchsten absoluten bzw. relativen Zwischenzentralitätswerten schon als die Zentralsten identifiziert werden können.

### 3.2.4.4 Der Clusterkoeffizient

Der Clusterkoeffizient misst, inwieweit ein Netzwerk aus Clustern besteht, d.h. inwieweit sich ein Netzwerk in lokale Nachbarschaften aufteilen lässt. Diese lokalen Nachbarschaften wären dann von einer eher hohen Dichte gekennzeichnet, während es zwischen den Nachbarschaften nur wenige Verbindungen gibt. Dabei kann sowohl der Clusterkoeffizient eines gesamten Netzwerks, als auch jener eines einzelnen Knotens berechnet werden. Häufigere Verwendung findet jedoch der Clusterkoeffizient des gesamten Netzwerks. Um diesen zu berechnen, werden zuerst die Dichten in den Nachbarschaften jedes Knotens berechnet. Diese bilden sozusagen die Clusterkoeffizienten der einzelnen Knoten. Um nun den Clusterkoeffizient des gesamten Netzwerks zu erhalten wird das arithmetische Mittel aus allen Dichten in allen Nachbarschaften eines jeden Knotens ermittelt. [46a]

Formal gesehen lässt sich der Clusterkoeffizient (nach Watts und Strogatz - die auch in Abschnitt 3.3.3 in Zusammenhang mit Kleine-Welt-Netzwerken genannt werden) eines Knotens in einem gerichteten Graphen folgendermaßen berechnen:

$$C_i = \frac{n}{k_i(k_i - 1)}$$

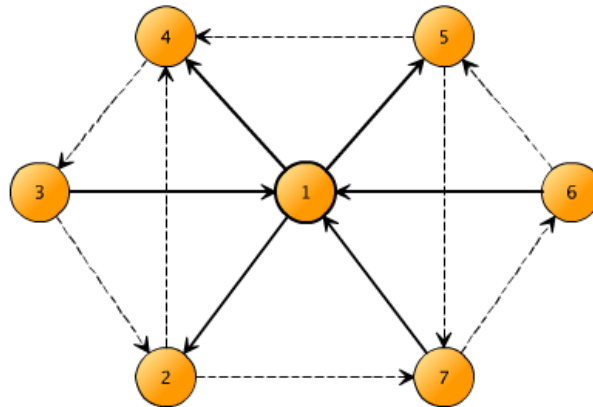
Wobei  $C_i$  der Clusterkoeffizient des Knotens  $i$ ,  $n$  die Anzahl der verbundenen Knoten in der Nachbarschaft von Knoten  $i$  und  $k_i$  der Knotengrad (in diesem Fall die Summe aus Ein- und Ausgangsgrad) des Knotens  $i$  ist. [110]

Der Clusterkoeffizient für den gesamten Graphen ist definiert als:

$$C = \frac{1}{n} \sum_{i=1}^n C_i$$

Dabei ist  $C_i$  der Clusterkoeffizient des Knotens  $i$  und  $n$  die Anzahl der Knoten im Graph. [110] Der Wert des Clusterkoeffizienten eines vollständigen Graphen ist 1. Die obige Definition des Clusterkoeffizienten von Watts und Strogatz wurde vor allem für die Charakterisierung von Kleine-Welt-Netzwerken formuliert. Sie hat unter anderem einige Nachteile und es gibt deshalb auch andere Definitionen von Clusterkoeffizienten. [111] Jedoch berechnet das Programm zur Analyse von Netzwerken UCINET, welches in dieser Arbeit verwendet wird, den Clusterkoeffizienten nach der hier beschriebenen Methode.

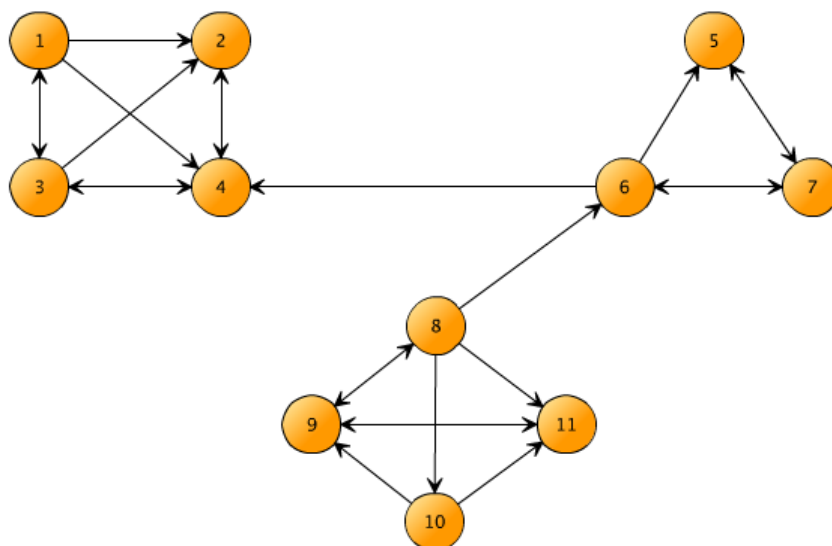
Als Beispiel für die Berechnung des Clusterkoeffizienten eines Knotens (Knoten 1) in einem gerichteten Graphen soll der Graph in Abbildung 13 dienen. Knoten 1 hat sechs Nachbarn (Knoten 2, 3, 4, 5, 6 und 7), mit denen er über eine Kante verbunden ist. Innerhalb dieser sechs Nachbarn gibt es 8 (dargestellt als gestrichelte Linien in Abbildung 13) von insgesamt  $6 \cdot 5$  möglichen Verbindungen (Kanten). Das heißt der Clusterkoeffizient von Knoten 1 beträgt 8 dividiert durch 30 was einem Wert von 0,267 entspricht.



**Abbildung 13: Beispielgraph für die Berechnung des Clusterkoeffizienten von Knoten 1**

Der Clusterkoeffizient ist auch ein Maß dafür, inwieweit sich cliquenähnliche Gebilde innerhalb eines Graphen befinden. Das heißt, damit es Cluster in einem Graphen gibt, müssen diese nicht wie Cliques einen vollständigen Teilgraphen bilden, aber sie müssen relativ stark miteinander verbunden sein.

In Abbildung 14 ist ein Graph mit relativ hohem Clusterkoeffizienten dargestellt. Der Graph hat einen durchschnittlichen Clusterkoeffizienten von 0,606. In der Abbildung ist auch zu erkennen, dass sich drei Cluster gebildet haben, deren Knoten enger miteinander vernetzt sind als die Cluster untereinander.



**Abbildung 14: Graph mit relativ hohem Clusterkoeffizienten**

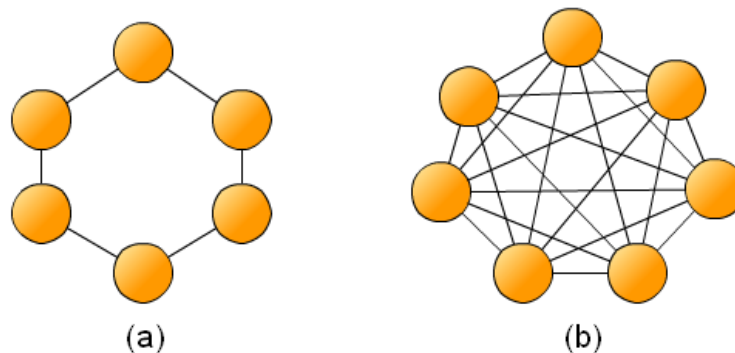
Nachdem nun die wichtigsten Zentralitätskennzahlen der sozialen Netzwerkanalyse und der Clusterkoeffizient erläutert wurden, werden im nächsten Abschnitt kurz ein paar Netzwerkmodelle vorgestellt.

### 3.3 Netzwerkmodelle

In der Netzwerktheorie gibt es unterschiedliche Modelle von Netzwerken mit unterschiedlichen Formen, Aussehen und Eigenschaften. Die verschiedenen Netzwerkmodelle werden in diesem Abschnitt vorgestellt. Dabei wird ein kurzer Überblick über die gängigsten Netzwerkmodelle und deren wichtigste Eigenschaften gegeben. Meistens werden die hier beschriebenen Netzwerkmodelle nur in Zusammenhang mit ungerichteten Graphen erwähnt, weshalb die in diesem Abschnitt dargestellten beispielhaften Graphen ungerichtet sind. Trotzdem gelten die Charakteristiken dieser Netzwerke in gleichem Maße auch für gerichtete Graphen und damit auch zum Beispiel für das World Wide Web oder für das Netzwerk des österreichischen E-Tourismus-Marktes.

#### 3.3.1 Reguläre Netzwerke

Begonnen hat in der Netzwerktheorie alles mit Untersuchungen von so genannten regulären Netzwerken [86a]. Ein reguläres Netzwerk ist ein Netzwerk, dem ein regulärer Graph zugrunde liegt. Ein regulärer Graph ist, wie bereits in Abschnitt 3.1.3.4 definiert wurde, ein Graph bei dem jeder Knoten denselben Grad hat [87]. In Abbildung 15 sind zwei bekannte Beispiele für reguläre Netzwerke dargestellt. Abbildung 15a zeigt ein so genanntes Ring Netzwerk und Abbildung 15b zeigt ein so genanntes vollständiges Netzwerk, welchem ein vollständiger Graph zugrunde liegt. Laut [88] und [89] gehören aber auch Netzwerke wie ein Baum, ein Gitternetz oder das in Abbildung 11 dargestellte Sternnetzwerk zu den regulären Netzwerken, obwohl bei diesen Netzwerken nicht alle Knoten exakt den gleichen Grad haben. Es geht also bei dieser Art von Netzwerken vor allem um die geordnete Struktur. Vielfach existieren diese Netzwerke auch nur in der Theorie. Sie sind relativ einfach, weshalb die Wissenschaft ihren Fokus bald auf komplexere Netzwerke gerichtet hat, welche real existierende Systeme besser beschreiben können. Diese werden in der Folge beschrieben.



**Abbildung 15: Zwei Beispiele für reguläre Netzwerke. (a) zeigt einen so genannten Ring-Netzwerk und (b) ein so genanntes vollständiges Netzwerk in dem jeder Knoten mit jedem anderen direkt verbunden ist**

#### 3.3.2 Zufallsnetzwerke

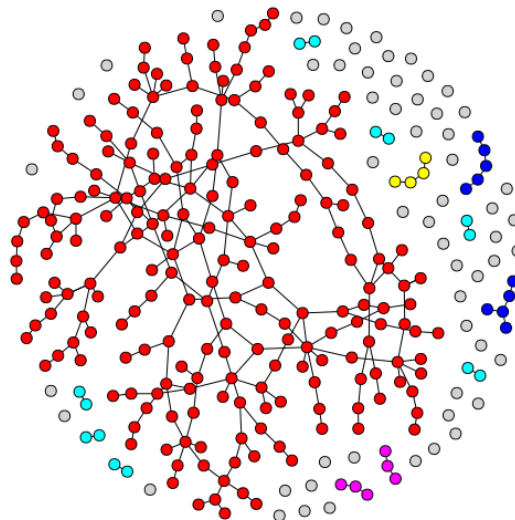
Die Idee die hinter den Zufallsnetzwerken steht ist, dass diese Netzwerke rein „zufällig“ entstehen und sich verändern. Obwohl Netzwerke in der Praxis eher selten „zufällig“ entstehen und sich „zufällig“ verändern, da meist spezifische Faktoren die Struktur des Netzwerks bestimmen, wurde versucht, künstliche Netzwerke durch einen Zufallsprozess zu erzeugen. Dabei gibt es zu Beginn des zufälligen Erzeugungsprozesses eine bestimmte Anzahl isolierter Knoten, welche dann zufällig und mit gleicher Wahrscheinlichkeit  $p$  miteinander verbunden werden.

Erdős und Renyi studierten, wie sich die Struktur eines Zufallsnetzwerks mit Zugabe oder Wegnahme von Knoten ändert und entwickelten eines der wichtigsten Modelle für die

Konstruktion von Zufallsgraphen. Dabei sei  $n$  die Anzahl der Knoten und  $l$  die Anzahl der Kanten. Solange die Anzahl der Kanten klein ist, bleibt der Graph in einzelne isolierte Knoten oder Komponenten zerteilt. Mit zunehmender Anzahl von Kanten werden zuerst andere isolierte Knoten mit den bestehenden kleinen Komponenten verbunden und dann verbinden sich auch Komponenten untereinander. Ein wichtiger Strukturübergang erfolgt wenn  $l = \frac{n}{2}$  ist.

Ab diesem Zeitpunkt beginnen sich die großen Komponenten untereinander zu verbinden und formen so hauptsächlich eine einzige Komponente. Die Anzahl der Knoten in dieser größten Komponente wächst linear und beträgt  $n$  mit  $n \rightarrow \infty$ , während die Anzahl der Knoten in den verbleibenden kleinen und isolierten Komponenten maximal  $\log n$  beträgt und somit fast konstant bleibt. [88] und [90]

Die Wahrscheinlichkeitsverteilung der Grade entspricht bei Zufallsnetzwerken einer Binomialverteilung, falls die Anzahl  $n$  der Knoten klein ist, und einer Poisson-Verteilung, falls die Anzahl  $n$  der Knoten groß ist [91] und [86b]. Der Durchmesser des Graphen bzw. die durchschnittliche Distanz zwischen zwei Knoten sind, wenn die Verbindungswahrscheinlichkeit  $p$  nicht zu klein ist, eher gering und hängen logarithmisch von der Anzahl der Knoten  $n$  ab. Da die Wahrscheinlichkeit, dass zwei Knoten miteinander verbunden sind, für alle Knoten gleich ist, entspricht diese Wahrscheinlichkeit auch genau dem Clusterkoeffizienten. Dies lässt sich auch durch den Quotienten  $C = p = \frac{\bar{k}}{n}$  ausdrücken, wobei  $p$  der Wahrscheinlichkeit, dass zwei Knoten miteinander verbunden sind, entspricht, und  $\bar{k}$  der durchschnittliche Knotengrad ist. [86b] Daraus kann geschlossen werden, dass der Clusterkoeffizient relativ niedrig ist. Denn der Clusterkoeffizient ist genau dann hoch, wenn die Wahrscheinlichkeit für die Verbindung zweier Knoten unter benachbarten Knoten höher ist als unter anderen. Da diese Wahrscheinlichkeit bei Zufallsnetzwerken aber für alle Knoten gleich ist, bleibt der Clusterkoeffizient relativ niedrig und es werden keine Cluster im Netzwerk gebildet.



**Abbildung 16: Zufallsnetzwerk nach Erdős und Renyi mit 300 Knoten und 250 Kanten, aus [92]**

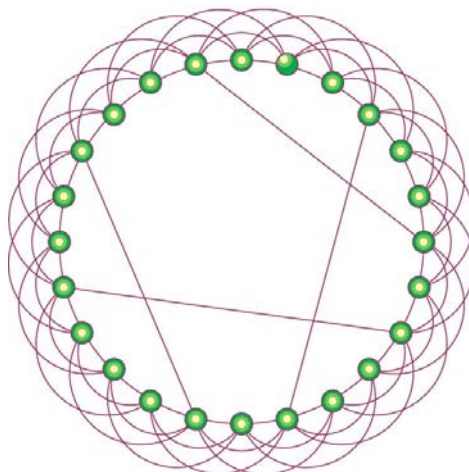
Abbildung 16 zeigt ein Zufallsnetzwerk nach Erdős und Renyi mit 300 Knoten und 250 Kanten, welche sich mit gleicher Wahrscheinlichkeit verbunden haben.

### 3.3.3 Kleine-Welt-Netzwerke

Netzwerke der realen Welt sind weder vollständig geordnet wie reguläre Netzwerke noch sind sie total zufällig wie Zufallsnetzwerke. Viele Netzwerke der realen Welt liegen irgendwo

dazwischen. [88] Der Begriff Kleine-Welt-Netzwerk stammt von der geringen durchschnittlichen Weglänge zwischen zwei Knoten und bezog sich in seinen Ursprüngen vor allem auf soziale Netzwerke. Stanley Milgram untersuchte die Anzahl der notwendigen Verbindungen um einen Brief von einer Seite der USA zur anderen zu senden und kam dabei auf durchschnittlich 6 Verbindungen, was sehr wenig ist. Neben der geringen Distanz bei vielen Netzwerken der realen Welt haben diese oft auch einen hohen Clusterkoeffizienten. Zum Beispiel ist die Wahrscheinlichkeit in einem sozialen Netzwerk hoch, dass zwei Personen, welche denselben Freund haben, auch untereinander befreundet sind. Dies führt zur Bildung von Clustern. [64a]

Wie auch bei Zufallsnetzwerken, wurde auch für Kleine-Welt-Netzwerke versucht, künstliche Netzwerke zu erzeugen, welche Charakteristiken von Kleine-Welt-Netzwerken haben, damit das Modell des Kleine-Welt-Netzwerkes anhand dieser künstlichen Netzwerke weiter erforscht werden konnte. Watts und Strogatz entwickelten im Jahr 1998 [86c] das wichtigste Verfahren zur künstlichen Erzeugung von Kleine-Welt-Netzwerken. Ausgangspunkt dieses Verfahrens ist ein reguläres Ring Netzwerk, in welchem jeder Knoten mit  $K$  Nachbarn verbunden ist. In der Folge werden die bestehenden Kanten zufällig und mit einer bestimmten Wahrscheinlichkeit mit neuen Knoten verbunden, oder es können laut [64a] auch zusätzliche Kanten hinzugefügt werden. Somit ergeben sich nach genügender Anzahl an Neuverbindungen Netzwerke mit den typischen Kleine-Welt-Netzwerkeigenschaften wie einer kurzen durchschnittlichen Weglänge zwischen zwei Knoten und, anders als bei Zufallsnetzwerken, einem hohen Clusterkoeffizienten, zumindest für hohe Verbindungswahrscheinlichkeiten. Die durchschnittliche Weglänge verhält sich aber bei Kleine-Welt-Netzwerken für relativ hohe Verbindungs-wahrscheinlichkeit ähnlich wie jene bei Zufallsnetzwerken, welche bei hoher Verbindungswahrscheinlichkeit auch logarithmisch von der Anzahl der Knoten abhängt. Bei geringer Verbindungswahrscheinlichkeit steigt die durchschnittliche Weglänge mit zunehmender Knotenanzahl linear an. [86c]

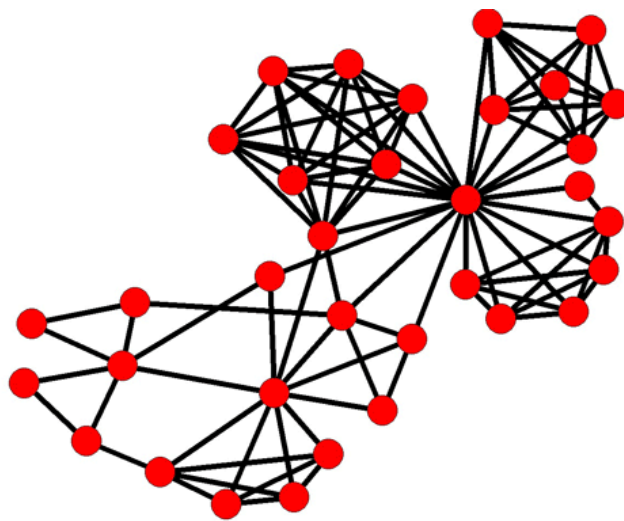


**Abbildung 17: Künstlich erzeugtes Kleine-Welt-Netzwerk nach dem Modell von Watts und Strogatz, nach dem Hinzufügen von ein paar Kanten. Aus [88]**

Abbildung 17 zeigt ein künstlich erzeugtes Kleine-Welt-Netzwerk nach dem Modell von Watts und Strogatz. Dabei wurde jene Variante gewählt, bei der neue Kanten zufällig hinzugefügt werden, anstatt bestehende Kanten neu zu positionieren. Dieses Modell ist somit eine Mischung aus einem komplett geordneten regulären Netzwerk und einem vollständig zufälligen Netzwerk. Nach dem Hinzufügen von nur wenigen Kanten sinkt die durchschnittliche Weglänge zwischen zwei Knoten bereits beträchtlich. Gleichzeitig sichert die ursprüngliche Form einen hohen Clusterkoeffizienten.

Die Wahrscheinlichkeitsverteilung der Knotengrade ist nach dem Modell von Watts und Strogatz für Kleine-Welt-Netzwerke jener eines Zufalls-Netzwerks sehr ähnlich. Die Spitze der Verteilung liegt beim Durchschnittsgrad und für große Durchschnittsgrade fällt die Verteilung exponentiell ab. Insgesamt haben alle Knoten ungefähr die gleiche Anzahl an Kanten und damit denselben Grad. [86c] In Abbildung 17 ist zu erkennen, dass der Grad der Knoten sehr gleichmäßig verteilt ist.

Abbildung 18 zeigt ein typisches, in der Realität existierendes Kleine-Welt-Netzwerk. Das Netzwerk zeigt die Vernetzungen von Wörtern im menschlichen Gehirn. Die Knoten entsprechen den Wörtern. Auch hier ist zu erkennen, dass die durchschnittlichen Weglängen zwischen zwei Knoten relativ niedrig sind, gleichzeitig aber die Tendenz zur Bildung von Clustern besteht und somit zu einem hohen Clusterkoeffizienten führt. Auch die Anzahl der Kanten pro Knoten und somit der Knotengrad ist bis auf ein paar Ausnahmen über die meisten Knoten ungefähr gleich groß.



**Abbildung 18:** Ein typisches in der Realität existierendes Kleine-Welt-Netzwerk, welches die Vernetzung von Wörtern im menschlichen Gehirn zeigt, aus [93]

### 3.3.4 Skalenfreie Netzwerke

Viele reale Netzwerke genügen zwar in manchen Eigenschaften den vorher beschriebenen Netzwerkmodellen, doch die Wahrscheinlichkeitsverteilung von deren Knotengrade entspricht nicht einer Binomialverteilung oder einer Poisson-Verteilung, sondern fällt viel steiler und schneller ab, ist aber gleichzeitig viel breiter als letztere. Hingegen entsprechen die Wahrscheinlichkeitsverteilungen dieser Netzwerke sehr oft und zumindest annähernd einer Potenzgesetzverteilung, sodass  $p_k = k^{-x}$  ist. [86c] und [88] Der Begriff skalenfrei kommt daher, dass in der Gleichung, welche die Potenzgesetzverteilung beschreibt, kein typischer Skalierungsparameter wie etwa in der Poisson-Verteilung oder anderen existiert [64a]. Das heißt die Form und Positionierung der Verteilungskurve im Koordinatensystem bleibt weitestgehend unabhängig von der Knotenanzahl oder dem durchschnittlichen Knotengrad.

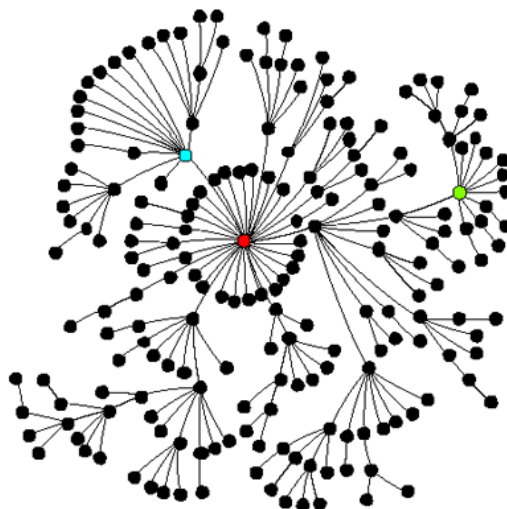
Ein großer Unterschied besteht auch in der theoretischen (künstlichen) Konstruktion dieser skalenfreien Netzwerkmodelle. Während bei Zufalls- und Kleine-Welt-Netzwerken von einer fixen Anzahl an Knoten ausgegangen wird und die Verbindungswahrscheinlichkeit für alle Knoten gleich ist, wird bei der Erzeugung der skalenfreien Netzwerke mehr Augenmerk auf die Evolution des Netzwerks und die damit zusammenhängende Entstehung einer bestimmten Topologie durch die Art der Evolution wertgelegt. Die Netzwerkmodellierung startet dazu bei ein paar wenigen Knoten und es werden ständig neue Knoten hinzugefügt, welche mit



größerer Wahrscheinlichkeit mit jenen Knoten verbunden werden, die bereits viele Verbindungen haben. Dies führt dazu, dass einige wenige Knoten, welche viele Verbindungen haben, immer mehr Verbindungen erhalten, während der Großteil der Knoten sehr wenige Verbindungen hat. Die Eigenschaft, dass sich das Netzwerk im Laufe der Zeit entwickelt und ständig neue Knoten dazukommen, trifft auch eher auf in der Realität existierende Netzwerke zu. [86c], [88] und [94].

Albert und Barabasi entwickelten zur Erzeugung von skalenfreien Netzwerkmodellen das Barabasi-Albert-Modell. Dieses startet, wie oben erwähnt, bei einigen wenige Knoten im Netzwerk und es werden ständig neue Knoten hinzugefügt, welche mit den bestehenden Knoten verlinkt werden, wobei die Wahrscheinlichkeit, dass ein Knoten mit dem neuen Knoten verbunden wird, von dessen Grad abhängt. Theoretische Messungen zeigten, dass diese Herangehensweise zu einem Netzwerkmodell führt, dessen Grade nach einem Potenzgesetz verteilt sind und das skalenfrei ist, weil sich das Verhältnis der Anzahl der Knoten mit sehr hohem Grad zu jenen mit niedrigem Grad bei Vergrößerung des Netzwerks kaum ändert. [86c]

Die durchschnittliche Weglänge eines mit dem Barabasi-Albert-Modell erzeugten skalenfreien Netzwerkes ist geringer als jene bei einem Zufallsnetzwerk mit selber Netzwerkgröße und demselben durchschnittlichen Knotengrad. Der Clusterkoeffizient hingegen ist beim Barabasi-Albert-Modell etwa fünf Mal so hoch wie der eines Zufallsgraphen. Auch ist der Clusterkoeffizient des Barabasi-Albert-Modells gleich wie der eines Zufallsnetzwerks und anders als der des Watts-Strogatz-Modells für Kleine-Welt-Netzwerke, abhängig von der Größe des Netzwerks. Insgesamt gesehen und für eine hohe Anzahl an Knoten ist der Clusterkoeffizient für skalenfreie Netzwerke, die nach Barabasi und Albert erzeugt wurden, um einiges geringer als jener von Kleine-Welt-Netzwerken nach dem Watts-Strogatz-Modell, weil der Clusterkoeffizient bei letzterem Modell unabhängig von der Knotenanzahl immer gleich hoch ist. Hingegen sinkt beim Barabasi-Albert-Modell der Clusterkoeffizient mit steigender Knotenanzahl. [86c]



**Abbildung 19: Ein typisches skalenfreies Netzwerk, aus [95]**

Abbildung 19 zeigt ein typisches skalenfreies Netzwerk. Es ist zu erkennen, dass einige wenige Knoten einen sehr hohen Grad haben während die meisten Knoten einen geringen Grad haben. Dies ist die typische Eigenschaft eines skalenfreien Netzwerkes. Gleichzeitig ist zu erkennen, dass die durchschnittlichen Weglängen zwischen zwei Knoten relativ gering sind und auch der Clusterkoeffizient, wie besprochen, relativ niedrig ist.

### 3.3.5 Zusammenfassung zu den Netzwerkmodellen

In diesem Abschnitt werden die Eigenschaften der hier besprochenen Netzwerkmodelle, sowie die gemessenen Kennzahlen in einer Tabelle zusammengefasst. Tabelle 7 zeigt diese Zusammenfassung. Teilweise wurden hier Formeln angeführt, welche bei den oben beschriebenen Netzwerkmodellen nicht in dieser Detailliertheit vorkommen. In diesem Fall wurden die Quellenangaben dazu in der Beschriftung der Tabelle 7 angegeben.

	<b>Reguläre Netzwerke</b>	<b>Zufallsnetzwerke</b>	<b>Kleine-Welt-Netzwerke</b>	<b>Skalenfreie Netzwerke</b>
<b>Typische Eigenschaft</b>	Jeder Knoten hat denselben Grad	„Zufällige“; künstliche Erzeugung, wobei jeder Knoten mit gleicher Wahrscheinlichkeit verbunden wird	Sehr kurze durchschnittliche Weglänge und relativ hoher Clusterkoeffizient	Potenzgesetzverteilung der Knotengrade; Skalenfreiheit;
<b>Wahrscheinlichkeitsverteilung der Knotengrade</b>	Delta-Verteilung mit Spitze $k^a$	Für kleine N: Binomialverteilung Für große N: Poisson-Verteilung	Ähnlich wie bei Zufallsnetzwerken mit Spitze $\bar{k}$ und exponentiell abfallend für große k	Potenzgesetzverteilung
<b>Durchschnittliche Weglänge</b>	Abhängig von Netzwerkart; üblicherweise sehr hoch; Zb: Ringnetzwerk = $n/2$ ; Ausnahme: Vollständiges Netzwerk = 1 oder Sternnetzwerk = 2;	Für nicht zu kleine Verbindungswahrscheinlichkeit p: $l \sim \frac{\ln N}{\ln k}^c$	Für kleine p und N: $l \sim N^d$ Für größere p und N: $l \sim \ln(N)^d$	Geringer als bei Zufallsnetzwerken; $l \sim \frac{\ln N^d}{\ln \ln N}$
<b>Clusterkoeffizient</b>	Üblicherweise sehr gering; Ausnahme: Vollständiges Netzwerk; $C = \frac{k}{b}$	$C = p = \frac{\bar{k}}{N}$	Relativ hoch; jedoch beim künstlich erzeugten Modell von Watts und Strogatz abhängig von p und nur für große p relativ hoch; $C''(p) = \frac{3K(K-1)}{2K(K-1) + 8pK^2 + 4p^2K^2}^d$	Sinkt mit steigender Knotenanzahl; 5 Mal so hoch wie der eines Zufallsnetzwerks aber insgesamt kleiner als der eines Kleine-Welt-Netzwerks; entspricht ungefähr einer Potenzgesetzverteilung mit $C \sim N^{-0,75}^d$

**Tabelle 7: Zusammenfassung der Eigenschaften der in diesem Abschnitt vorgestellten Netzwerkmodelle.** <sup>a</sup> aus [86e], <sup>b</sup> laut [86d] abhängig von der Form, <sup>c</sup> aus [86b], <sup>d</sup> aus [86c], N = Knotenanzahl, k = durchschnittlicher Knotengrad, l = durchschnittliche Weglänge, C = Clusterkoeffizient, p = Verbindungswahrscheinlichkeit, K entspricht der Anzahl der verbundenen benachbarten Knoten (siehe Abschnitt 3.3.3), C' entspricht einer etwas anderen Version des Clusterkoeffizienten als die in Abschnitt 3.2.4.4 genannte (siehe dazu [86c])

Am Schluss bleibt noch zu sagen, dass viele in der Realität existierenden Netzwerke einerseits typische Charakteristiken von mehreren verschiedenen Netzwerkmodellen aufweisen, andererseits aber einige typische und vermutete Charakteristiken von diesen Netzwerkmodellen nicht haben. Auch ist es oft schwierig, durch die Entwicklung von Modellen die Netzwerke der Realität genau zu repräsentieren. Bei der Modellentwicklung wird meist von bestimmten Eigenschaften ausgegangen, die das zu entwickelnde Modell haben sollte, dabei werden aber andere Eigenschaften außer Acht gelassen, die vielleicht auch wichtig wären. Dies führt notwendigerweise, wie die Bedeutung des Begriffs Modells impliziert, zu einer Abstraktion der Wirklichkeit, welche der Wirklichkeit nicht immer gerecht wird.

Im folgenden Abschnitt wird versucht festzustellen, welche Eigenschaften das Netzwerk bzw. der Graph des World Wide Webs hat.

### 3.4 Das World Wide Web als Graph

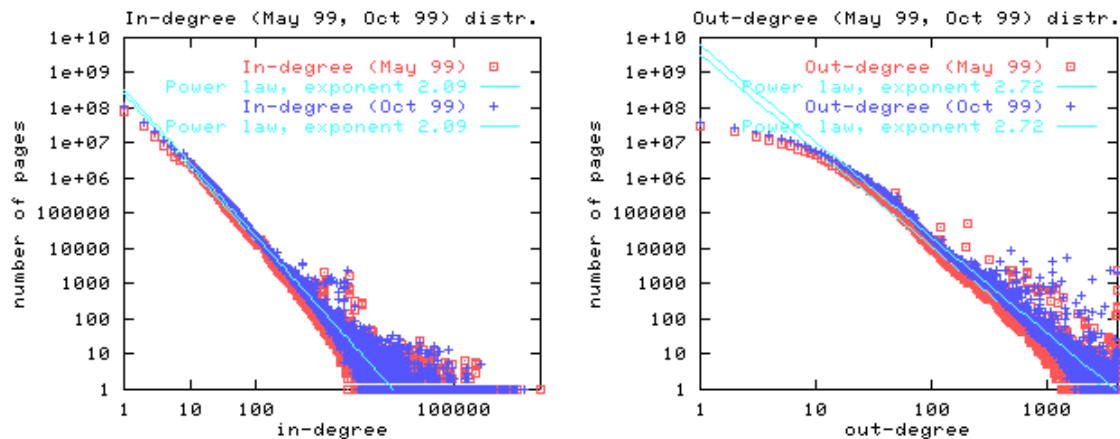
Wie unschwer zu erkennen ist, kann auch das World Wide Web als Graph dargestellt werden. Dabei ist jede Webseite bzw. Website ein Knoten und die Links zwischen den Webseiten bzw. Websites sind die Kanten. In diesem Fall handelt es sich um einen gerichteten Graphen, da eine Webseite oder Website ausgehende und eingehende Links haben kann. In den meisten Untersuchungen über das World Wide Web wurden die Webseiten und nicht die Websites als Knoten gezählt.

#### 3.4.1 Die Wahrscheinlichkeitsverteilung der Knotengrade im World Wide Web

Eine der wichtigsten Charakteristiken des Graphen des World Wide Webs, welche in der Vergangenheit festgestellt wurden, ist die Existenz von Potenzgesetz-Wahrscheinlichkeitsverteilungen der Ein- und Ausgangsgrade der Knoten im Netzwerk des World Wide Webs, so wie dies auch für die im vorigen Kapitel festgestellten skalenfreien Netzwerke gilt. Eine Potenzgesetzverteilung der Grade besagt, dass die Wahrscheinlichkeit, dass ein Knoten einen Grad von  $k$  hat, dem Wert aus  $k^{-x}$  entspricht. Im Allgemeinen kann im World Wide Web von einem Parameterwert der Potenzgesetzverteilung für die Eingangsgrade von  $x_{in}=2,1$  und für die Ausgangsgrade von  $x_{out}=2,7$  ausgegangen werden. Dies wurde auch von einem der größten, jemals durchgeführten Untersuchungen des World Wide Webs mit einem Crawler bestätigt. Im Mai und Oktober 1999 wurden von Altavista zwei Untersuchungen mit einem Crawler durchgeführt, welche dieses Ergebnis bestätigten. Dabei wurden 200 Millionen Webseiten mit 1,5 Milliarden Links untersucht. [54] und [55]

Die Existenz der Potenzgesetzverteilung der Knotengrade wurde, zumindest für die eingehenden Knotengrade, auch durch vier später durchgeführte Untersuchungen mit Hilfe von Crawlern bestätigt. Der Unterschied bei diesen Untersuchungen war, dass der Verteilungsparameter  $x_{in}$  insgesamt etwas niedriger war. Für die ausgehenden Knotengrade wurde jedoch eine Wahrscheinlichkeitsverteilung festgestellt, die der Potenzgesetzverteilung zwar ähnlich ist, im Gegensatz dazu aber schneller abnimmt und bald abbricht. Dies wird in der Untersuchung dadurch erklärt, dass die Anzahl der ausgehenden Links pro Webseite begrenzt ist, was aus Übersichtlichkeitsgründen ziemlich einleuchtend scheint. [96] [98] nennt diese Art der Verteilung der ausgehenden Knotengrade eine so genannte „unperfekte“ oder nur annähernd zutreffende Potenzgesetzverteilung. Insgesamt und Allgemein gesehen kann also von einer Potenzgesetzverteilung der Knotengrade im World Wide Web ausgegangen werden. Siehe dazu zum Beispiel auch [64c] und [97].

Abbildung 20 zeigt die logarithmierten Gradverteilungen im World Wide Web nach einem Potenzgesetz. Typisch für die Potenzgesetzverteilung der Wahrscheinlichkeiten ist eine gerade Linie wenn die  $x$ - und  $y$ -Werte vorher logarithmiert wurden. Im rechten Diagramm der Abbildung 20 ist zu erkennen, dass die Ausgangsgrade nicht exakt auf der geraden Linie liegen und deswegen, wie oben bereits erwähnt, nur annähernd eine Potenzgesetzverteilung darstellen. Aber auch die Verteilung der Eingangsgrade, dargestellt im linken Diagramm der Abbildung 20, weist für besonders niedrige Eingangsgrade eine kleine Abweichung von der geraden Linie auf. Auch beim Netzwerk des österreichischen E-Tourismus-Marktes werden in der Folge solche Abweichungen identifiziert und damit bestimmt, dass es sich auch hier nur annähernd um eine Potenzgesetzverteilung handelt. In diesem Zusammenhang ist [103] erwähnenswert, in der aufgezeigt wird, dass die meisten realen Netzwerke, welche eine Potenzgesetzverteilung der Knotengrade aufweisen, dies nur annäherungsweise tun. Das heißt, dass die strikte Annahme einer Potenzgesetzverteilung der Grade nur oberhalb eines bestimmten Knotengrades vorliegt. Im linken Diagramm der Abbildung 20 ist dies zum Beispiel ab einem Knotengrad von etwas über 10 der Fall.



**Abbildung 20: Ein- und Ausgangsgradverteilung des World Wide Webs nach einem Potenzgesetz, aus [102]**

Aus diesen Erkenntnissen folgt, dass das World Wide Web – zumindest bei der Verteilung der Grade – einem skalenfreien Netzwerk entspricht. Die Skalierungsfreiheit ist darin erkennbar, dass das Verhältnis der Anzahl der stark verbundenen Knoten zu jenen, die nur einen geringen Grad besitzen, mit zunehmender Knotenanzahl ungefähr gleich bleibt. Das heißt, es gibt bestimmte Knoten, die immer wieder von neu entstehenden Knoten verbunden werden und damit einen immer höheren Knotengrad erreichen, während die Masse der Knoten einen geringen Grad beibehält. Dies entspricht auch der Evolution eines skalenfreien Netzwerkmodells von Barabasi und Albert, wie in Abschnitt 3.3.4 beschrieben wurde.

### 3.4.2 Verbundenheit und durchschnittliche Weglänge im World Wide Web

Die erwartete hohe Verbundenheit der Knoten im World Wide Web konnte jedoch durch die von Altavista im Jahre 1999 durchgeführte Untersuchung nicht bestätigt werden. Insgesamt gibt es nur für etwa  $\frac{1}{4}$  der Knotenpaare einen möglichen Weg. [55] und [64c] Für die verbundenen Knotenpaare weist das World Wide Web jedoch die Kleine-Welt-Netzwerkcharakteristik der kurzen Wege auf. Die durchschnittliche Weglänge zwischen zwei verbundenen Knoten beträgt 16. [55] Diese Werte wurden laut [64c] von vielen weiteren Untersuchungen bestätigt. Siehe dazu zum Beispiel auch [100] und [101]. Die durchschnittlichen Weglängen sind dabei fast so gering wie jene eines Zufallsnetzwerks von gleicher Größe und mit hoher Verbindungswahrscheinlichkeit  $p$ , und entsprechen somit einer Kleine-Welt-Netzwerkcharakteristik. Denn die durchschnittlichen Weglängen beider Netzwerkmodelle verhalten sich – wie in den Abschnitten 3.3.2 und 3.3.3 festgestellt wurde – bei hoher Verbindungswahrscheinlichkeit  $p$  ähnlich logarithmisch zur Knotenanzahl.

### 3.4.3 Der Clusterkoeffizient des World Wide Webs

Laut einigen Untersuchungen weist das World Wide Web im Unterschied zu einem Zufallsgraphen auch einen relativ hohen Clusterkoeffizienten, wie es bei Kleine-Welt-Netzwerken üblich ist, auf. [100] nennt eine Untersuchung, welche 1998 durchgeführt wurde und in der die Knoten des ungerichteten Netzwerks den Websites entsprechen. Nach Wegfall der Blattknoten (jene Knoten mit einem Knotengrad von 1) verblieben 153.127 Knoten im Netzwerk. Der Clusterkoeffizient betrug hierfür 0,1078, im Vergleich zu 0,00023 in einem Zufallsnetzwerk bei gleicher Anzahl von Knoten und Kanten. Als nächstes wurde der Clusterkoeffizient für das dazugehörige gerichtete Netzwerk berechnet. Hier war die Anzahl der Knoten 64.826, wobei jeder Knoten von jeden anderen aus erreichbar sein musste, und der Clusterkoeffizient betrug immer noch 0,081 im Vergleich zu 0,00105 bei einem Zufallsnetzwerk mit gleicher Knoten- und Kantenanzahl. Zuletzt wurde das Netzwerk der durch einen Weg verbundenen Websites der .edu-Domain, bestehend aus 3.456 Websites,

untersucht. Das Ergebnis brachte einen etwas höheren Clusterkoeffizienten von 0,156 im Vergleich zu 0,0012 beim Zufallsgraphen mit derselben Knoten- und Kantenanzahl.

[101] spricht von einer anderen Untersuchung der Websites der Domains *stanford.edu* und *mit.edu*. Diese Domains enthalten Homepages von Studenten, Fakultäten und Mitarbeitern der Universität. Zuerst wurden die nicht verbundenen Komponenten, also wie oben jene Knoten, welche nicht durch einen Weg miteinander verbunden sind, aus dem Netzwerk herausgelöst. Übrig blieben 58% oder 1265 Homepages für die Domain *stanford.edu* und 85,6% oder 1281 Homepages für die Domain *mit.edu*. Der Clusterkoeffizient war für die Domain *stanford.edu* 0,22 und für die Domain *mit.edu* 0,21, was insgesamt etwas höher ist als in der vorherigen Untersuchung, bei welcher zumindest ein größerer Teil des gesamten World Wide Webs untersucht wurde.

Auch [64b] nennt relativ hohe Clusterkoeffizienten von 0,29 bei 269.504 Webseiten der *nd.edu* Domain. Diese Untersuchungen lassen klar einen im Vergleich zu einem Zufallsnetzwerk viel höheren Clusterkoeffizient im World Wide Web erkennen. Damit kann das World Wide Web als Kleine-Welt-Netzwerk mit Potenzgesetzverteilungscharakteristik der Knotengrade und damit auch als skalenfreies Netzwerk angesehen werden. Wie in Abschnitt 3.3.5 festgestellt wurde, weisen eben viele Netzwerke der realen Welt Charakteristiken von mehreren Netzwerkmodellen auf.

### 3.4.4 Die Fliegen-Form des World Wide Webs (Bow-Tie)

In der weiter oben in Abschnitt 3.4.1 genannten, von Altavista im Jahr 1999 durchgeführten Untersuchung des World Wide Webs wurde auch festgestellt, dass der Graph des World Wide Webs eine so genannte Fliegen-Form aufweist [55]. Diese Form ist in Abbildung 21 dargestellt.

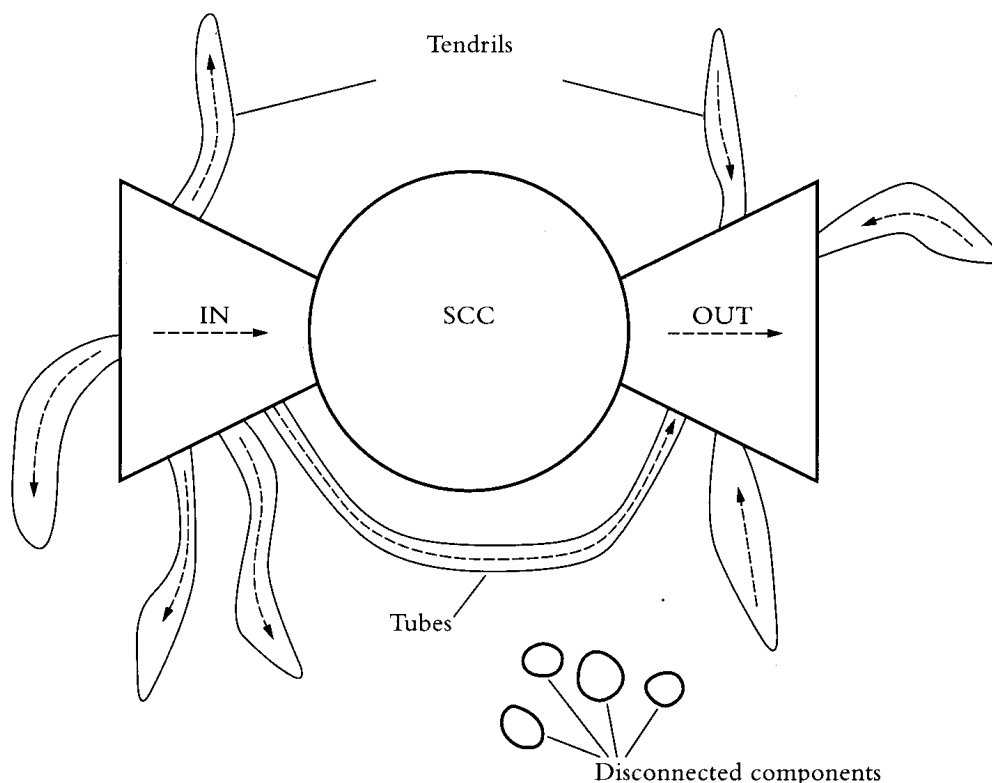


Abbildung 21: Die Fliegen-Form des World Wide Webs, adaptiert aus [56]

Diese Fliegen-Form besteht aus sechs verschiedenen Komponenten, welche in der Folge kurz beschrieben werden:

- Hauptkomponente (SCC für Strongly Connected Component): Dieser Teil besteht aus Knoten, wobei jeder dieser Knoten mit jedem anderen durch einen Weg verbunden ist.
- Eingangskomponente (IN): Diese besteht aus Knoten, von denen ein Weg zu Knoten der Hauptkomponente führt, die aber nicht von Knoten aus der Hauptkomponente erreicht werden können.
- Ausgangskomponente (OUT): Diese besteht aus allen Knoten, die durch einen Weg von der Hauptkomponente aus erreicht werden können, aber nicht umgekehrt.
- Ranken (Tendrils): Entweder bestehen sie aus Knoten, die nur von Knoten der Eingangskomponente erreicht werden können; sie sind weder von Knoten der Hauptkomponente noch von jenen der Ausgangskomponente erreichbar. Oder es handelt sich um Knoten, die zwar von der Ausgangskomponente her zugänglich sind, die aber weder von den Knoten der Eingangs- noch von jenen der Hauptkomponente erreicht werden können.
- Tubes: Zu den Tubes gehören alle Knoten, die von Knoten der Eingangskomponente erreicht werden können und deren Links zu Knoten der Ausgangskomponente führen. Die Tubes stellen aber laut [56] eine sehr kleine Komponente dar und werden deshalb in Tabelle 8 auch nicht berücksichtigt.
- Nicht verbundene Komponenten (Disconnected components): Sie bestehen aus Knoten, die zu keiner der oben genannten Komponenten gehören.

Nachfolgende Tabelle 8 zeigt die Verteilung der Websites auf die unterschiedlichen Komponenten der Fliegen-Form des Webgraphs.

	<b>Hauptkomponente</b>	<b>Eingangskomponente</b>	<b>Ausgangskomponente</b>	<b>Ranken</b>	<b>Nicht verbundene Komponente</b>
<b>Crawl von Altavista (1999)</b>	28%	21%	21%	22%	9%

**Tabelle 8: Größe der verschiedenen Komponenten der Fliegen-Form nach der Durchsichtung des World Wide Webs mit einem Crawler, durchgeführt von Altavista 1999 [54]**

Diese Fliegen-Form des World Wide Webs ist weitestgehend anerkannt. Auch [96] und [98] bestätigten mit anderen Untersuchungen diese Form, jedoch mit unterschiedlichen Mengenaufteilungen zwischen den Komponenten. Diese unterschiedliche Aufteilung wird sich auch beim Netzwerk des österreichischen E-Tourismus-Marktes zeigen. Die Autoren von [96] erklärten sich die unterschiedlichen Aufteilungen durch unterschiedliche Strategien der Crawler und unterschiedliche Startpunkte der Untersuchungen. Laut [96] ist es zudem aufgrund der Größe und der Dynamik des World Wide Webs auch mit dem größten Crawler nicht möglich, ein vollständiges und reales Abbild des Webgraphs zu erhalten, weswegen die Untersuchungsergebnisse wiederum verschieden ausfallen können. Laut [96] gibt es auch Untersuchungen zu Teilgraphen des World Wide Webs, um zu zeigen, dass der Teilgraph gleiche Charakteristiken aufweist wie der gesamte Graph. [98] bestätigt die Fliegen-Form und die Potenzgesetzverteilung der Knotengrade für Teilgraphen, bestehend aus Webseiten zu gleichen Themengebieten. [97] stellte hingegen bei der Untersuchung der Teilkomponenten der Fliegen-Form fest, dass die Knoten dieser Teilkomponenten unterschiedlich stark

miteinander verbunden sind und unterschiedliche Strukturen aufweisen. Die Teilgraphen weisen also ähnliche Charakteristiken auf, wenn sie aufgrund eines bestimmten Themengebietes gewählt wurden. Dies ist jedoch nicht der Fall, wenn die Komponenten der Fliegen-Form die Teilgraphen bilden.

## 4 Der Crawler

In diesem Kapitel wird der Webcrawler beschrieben, mit welchem die Daten über das Netzwerk des österreichischen E-Tourismus-Marktes gesammelt werden. In der Folge wird der Webcrawler nur Crawler genannt.

Ein Crawler ist ein Programm, welches das World Wide Web durchsucht und gegebenenfalls analysiert. Vor allem Suchmaschinen verwenden Crawler. Die grundlegende Technik eines Crawlers ist, dass von einer Startseite ausgehend die Links dieser Seite gesucht werden. Diese Links werden gespeichert und die Seiten, auf welche die Links verweisen, werden als Nächstes durchsucht. Einfach gesagt durchsucht ein Crawler das World Wide Web, indem er von einer Seite über die Links auf die nächste Seite gelangt.

Die dabei gefundenen Seiten können auch nach bestimmten Inhalten durchsucht werden, damit zum Beispiel nur Seiten eines bestimmten Themenbereichs gefunden werden. Solche Crawler werden auch fokussierende Crawler genannt. Der für diese Arbeit entwickelte Crawler ist jedoch nicht auf ein bestimmtes Themengebiet beschränkt und nimmt alle Seiten in das gefundene Netzwerk auf. Man könnte meinen, dass es in diesem Fall wichtig wäre, nur Seiten von Tourismusangebietern in die Analyse mit einzubeziehen und deshalb einen fokussierenden Crawler zu verwenden. Doch es gibt durchaus auch Seiten, die nicht zum Tourismussektor gehören, aber trotzdem eine wichtige Rolle in diesem Netzwerk spielen können, wie auch in der Analyse der Ergebnisse gezeigt werden wird.

Der Crawler wurde neu entwickelt weil die Suche nach einem frei verfügbaren und bereits vorhandenen Crawler nicht erfolgreich war. Die meisten Crawler konnten aufgrund diverser Probleme nicht lauffähig gemacht werden. Etwa an die fünf Crawler konnten gestartet werden, lieferten jedoch nicht brauchbare Ergebnisse oder die falschen Daten. Aufgrund dessen und auch weil ein Crawler im Grunde ein mit relativ wenig Aufwand zu erstellendes Programm, welches Webseiten durchsucht und die Ergebnisse in eine Datenbank schreibt, ist, wurde beschlossen den Crawler selbst zu entwickeln.

In Kapitel 6, Abschnitt 1 wird dann versucht die Qualität dieses selbst entwickelten Crawlers zu beurteilen. Einerseits werden dabei Ergebnisse des eigenen Crawlers miteinander verglichen, andererseits wird auch versucht einen Vergleich mit den Ergebnissen des am ehesten geeigneten und lauffähigen externen Crawler anzustellen. Es wird sich jedoch zeigen dass ein Vergleich mit dem externen Crawler, sowie eine mögliche Beurteilung des eigenen Crawlers dadurch nicht möglich ist. Trotzdem wird beim Vergleich der Ergebnisse des eigenen Crawlers festgestellt werden, dass die Qualität des Crawlers gut ist.

In diesem Kapitel wird zuerst eine allgemeine Beschreibung des Crawlers gegeben. Danach werden die Klassen des Programms anhand des Klassendiagramms und einer kurzen Beschreibung näher betrachtet. In der Folge wird auch das Datenmodell, in welchem die Daten in der Datenbank gespeichert werden, erläutert und die Funktionen der Datenbanktabellen werden beschrieben. Um einen genaueren Einblick in die Funktionsweise des Crawlers zu erhalten, werden im vierten Abschnitt die zwei Klassen mit den wichtigsten Routinen in Pseudocode-Form wiedergegeben und erklärt. Dann gibt es noch eine Beschreibung von technischen Details bezüglich der Properties-Datei sowie der Anwendung der Thread-Technologie im Programm. Am Schluss folgt ein kurzer Abschnitt mit verschiedenen Problembereichen und Einschränkungen eines Crawlers in Bezug auf die Ermittlung eines Netzwerks des World Wide Webs bzw. des Webgraphen.



## 4.1 Allgemeine Beschreibung

Der Crawler wurde in Java in Verbindung mit einer MySQL Datenbank geschrieben. Der Code besteht aus ca. 1000 Zeilen. Nachfolgende Abbildung 22 zeigt eine Darstellung des Startfensters.

Im Textbereich mit dem Label Start URL werden die Startpunkte (gedacht als Websites) durch ein Komma getrennt angegeben. Im Textfeld darunter kann die Durchsuchung einer Website auf eine bestimmte Anzahl von Seiten begrenzt werden. Das heißt, es wird nur eine bestimmte Anzahl (etwa 5000) von Webseiten innerhalb einer Website durchsucht. Dies ist bei Websites welche aus vielen Webseiten bestehen, wichtig, um die Dauer des Suchvorgangs nicht zu lange werden zu lassen. Darunter kann noch angeklickt werden, ob ein Suchvorgang, mit der in der Properties-Datei (siehe Abschnitt 4.5.1) definierten Datenbank von neuem gestartet, d.h. die Datenbank wenn sie schon vorhanden ist gelöscht wird, oder, falls die Checkbox nicht angeklickt wird, ein zu einem früheren Zeitpunkt gestarteter Suchvorgang mit der eingestellten Datenbank fortgesetzt werden soll. Wenn der Start eines neuen Suchvorgangs gewählt wurde, kann als zweite Möglichkeit, ausgewählt werden, ob mit einer neuen Datenbank gestartet oder eine bereits existierende benützt werden soll. Die Datenbank wird dann gegebenenfalls gelöscht. Welche Datenbank beim jeweiligen Crawl benützt werden soll, wird in einer Properties-Datei festgelegt. Auf die Inhalte der Properties-Datei wird in einem der folgenden Abschnitte noch näher eingegangen.

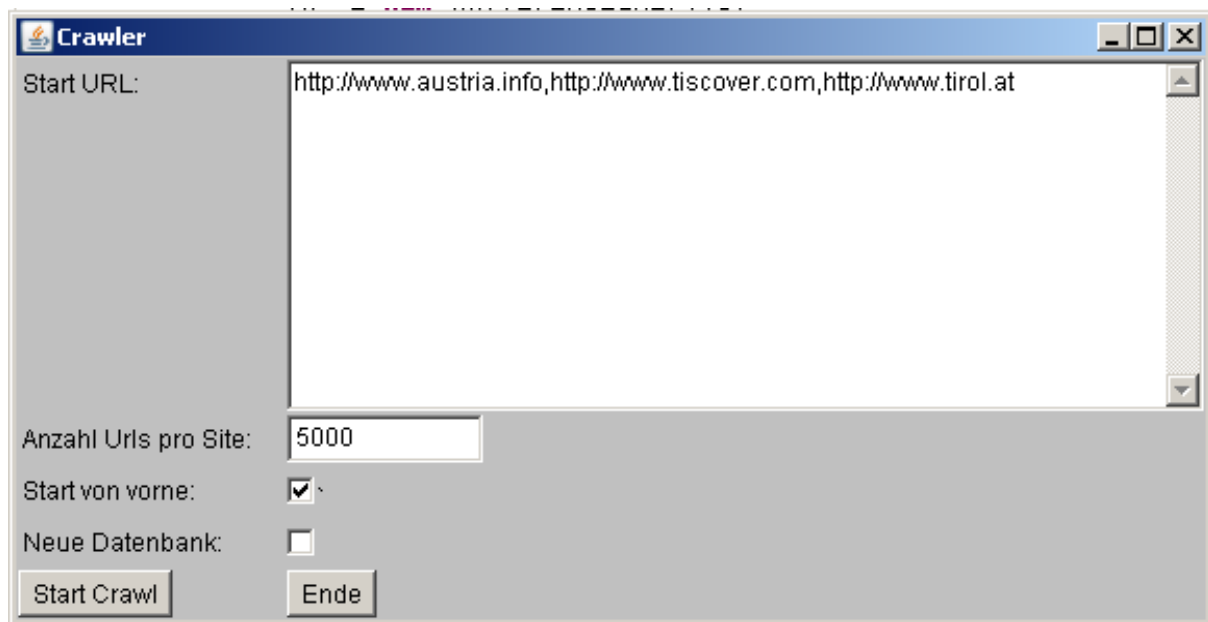


Abbildung 22: Das Startfenster des Crawlers

Die nachfolgende Abbildung 23 zeigt den Crawler bei der Arbeit. Darin ist dargestellt, welche Webseite (Aktuelle Url) gerade durchsucht wird, die wievielte Webseite der Website diese ist (Nr.), welche Website (Domain) gerade durchsucht wird und wie viele Websites insgesamt bereits durchsucht wurden (Soviel Domains gecrawlt). Rechts unten ist noch der Knopf „Stop Crawl“ angebracht, mit welchem der Suchvorgang abgebrochen werden kann. In diesem Fall erscheint wieder das Startfenster aus Abbildung 22. Die nicht editierbaren Textfelder in Abbildung 23 werden während jedem Durchsuchen einer neuen Webseite von außen neu beschrieben.

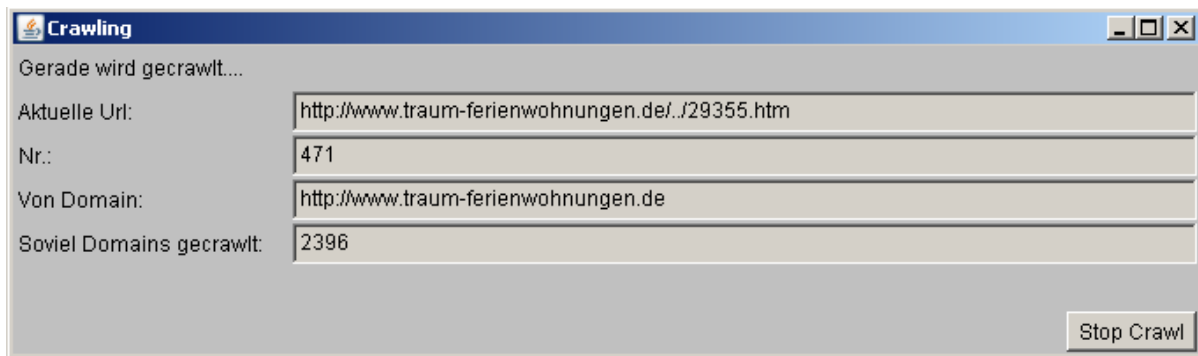


Abbildung 23: Der Crawler während des Suchvorgangs

## 4.2 Die Klassen

In diesem Abschnitt wird eine kurze Beschreibung der Klassen des Crawlers gegeben. Abbildung 24 zeigt hierzu das Klassendiagramm.

### 4.2.1 Klasse Hauptklasse

Diese Klasse dient nur zum Starten des Programms.

### 4.2.2 Klasse Hauptfenster

Die Klasse *Hauptfenster* wird für die Darstellung des Fensters, welches in Abbildung 22 zu sehen ist, benötigt. Im Folgenden werden die Methoden dieser Klasse kurz beschrieben.

Der Konstruktor ist dazu da, um das Fenster sichtbar zu machen.

Die Methode *getCrawlerStand* wird aufgerufen, wenn ein zu einem früherem Zeitpunkt gestarteter Suchvorgang fortgesetzt werden soll. In diesem Fall schaut diese Methode in der Tabelle *Webseite* nach und sucht den Datensatz bei dem das Attribut *GeradeDurchsucht* auf 1 gesetzt ist (d.h. das während des vorherigen Abbruchs des Suchvorgangs gerade durchsucht wurde). Bei diesem Datensatz wird dann mit dem Suchvorgang fortgesetzt. Rückgabewert dieser Methode ist ein Integer-Array, welches die beiden Zähler für die Websites und die Webseiten enthält (siehe Listing 2 in Abschnitt 4.4 die Variablen *zaehlerWebsites* und *zaehlerWebseiten*).

Die Methode *makegbc* ist dazu da, ein benutzerdefiniertes Layout des Fensters zu erzeugen.

Die Methode *actionPerformed* ist nötig, um die Ereignisse, die von einem Klick ausgelöst, werden abzufangen. Hier wird dann bei Klick auf den Button „Start Crawl“ der Suchvorgang gestartet und bei Klick auf den Button „Ende“ das Programm beendet.

### 4.2.3 Klasse CrawlerFenster

Die Klasse *CrawlerFenster* wird für die Darstellung des in Abbildung 23 dargestellten Fensters benötigt.

Der Konstruktor übernimmt ein Objekt der Klasse *Crawler*. Dieses wird in der Folge dazu gebraucht, den Thread der Klasse *Crawler* zu stoppen und damit den Suchvorgang zu beenden, wenn der Benutzer den Button „Stop Crawl“ drückt.

Die Methode *updateFenster* wird benötigt, um im Fenster die Textfelder mit den Informationen zum gegenwärtigen Suchvorgang zu verändern. Im Fenster wird dabei, wie in Abbildung 23 zu sehen ist, die aktuell durchsuchte Website und Webseite sowie die Anzahl der bereits durchsuchten Webseiten einer Website und die der Websites insgesamt angezeigt. Um diese Textfelder von außen ändern zu können, bedarf es dieser Methode.

Die Methode *makegbc* dient wieder, wie oben beschrieben, der benutzerdefinierten Darstellung des Fensterlayouts.

Die Methode *actionPerformed* ist dazu da, um das Ereignis eines Klicks auf einen Button aufzufangen und entsprechend darauf zu reagieren. In diesem Fenster gibt es nur das Ereignis, wodurch der Suchvorgang gestoppt wird.

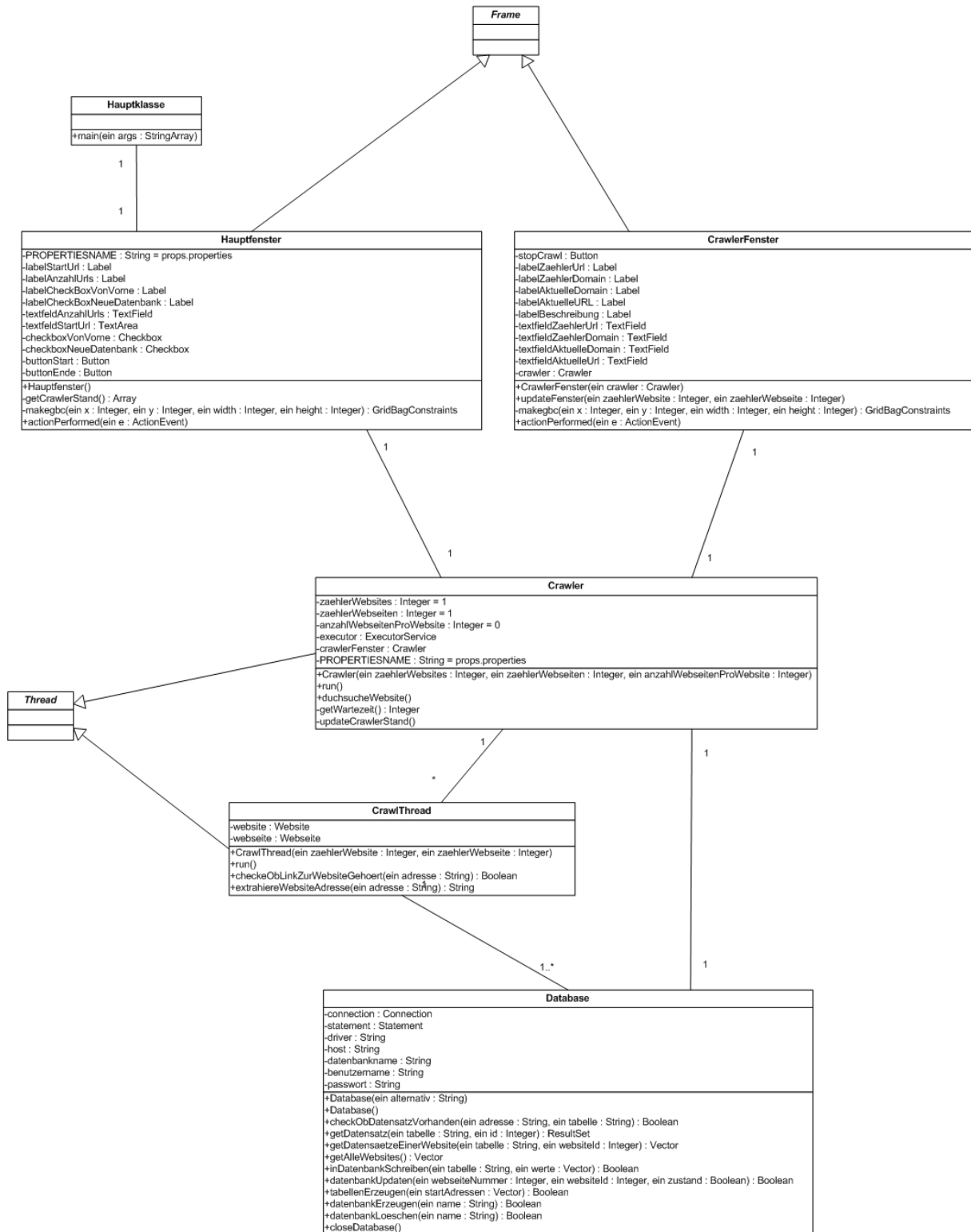


Abbildung 24: Das Klassendiagramm des Crawlers

#### 4.2.4 Klasse Crawler

Die Klasse *Crawler* ist für die Programmlogik zuständig. In ihr wird in einer Schleife nacheinander eine Liste von Websites, und innerhalb dieser eine Liste von Webseiten, die zu dieser Website gehören, durchsucht. Die zwei wichtigsten Methoden dieser Klasse werden in Abschnitt 4.4 in Pseudocode-Form genauer beschrieben. Hier folgt nur eine kurze Beschreibung der Methoden dieser Klasse.

Der Konstruktor erhält von der Klasse Hauptfenster die Parameter *zaehlerWebsites*, *zaehlerWebseiten* und *anzahlWebseitenProWebsite*. Die ersten beiden dienen zur Information für das Programm die wievielte Website (*zaehlerWebsites* enthält die *ID* einer Website in der gleichnamigen Tabelle) und die wievielte Webseite innerhalb der Website (*zaehlerWebseite* enthält den Inhalt des Attributes *WebseiteNR* in der Tabelle *Webseite*) gerade durchsucht wird. Diese Zählvariablen werden in der Schleife abgearbeitet. Der Parameter *anzahlWebseitenProWebsite* enthält den Inhalt des Textfeldes mit dem Label „Anzahl URLs pro Site“ aus dem Fenster in Abbildung 22.

Die Methode *run* durchsucht eine Liste von Websites (siehe Tabelle *Websites* im Datenmodell in Abschnitt 4.3). Innerhalb dieser Methode gibt es eine While-Schleife, welche – solange es noch zu durchsuchende Websites gibt – eine nach der anderen durchsucht.

Die Methode *durchsucheWebsite* durchsucht alle Webseiten einer Website, die in der Tabelle *Webseiten* gespeichert sind. Dabei werden, ausgehend von einer Start-Webseite, neue relative Links zu Webseiten innerhalb der Website gefunden, diese in die Datenbanktabelle *Webseiten* gespeichert und dann nacheinander durchsucht. In dieser Methode gibt es auch eine While-Schleife die durchgeführt wird, solange es zusätzliche Webseiten innerhalb der Website gibt und solange nicht bereits mehr Webseiten als angegeben durchsucht wurden (siehe Eingabe im Fenster in Abbildung 22). Für jede Adresse einer Webseite führt diese Methode dann die Klasse *CrawlThread* aus, mit welcher der HTML-Code der Webseite nach Links durchsucht wird.

Die Methode *getWartezeit* wird benötigt, um die Wartezeit für das Programm zu berechnen, welche von der Anzahl der bereits durchsuchten Webseiten pro Website abhängt. Der Rückgabewert ist vom Typ Integer und in Millisekunden angegeben und die Wartezeiten für das Programm werden aus einer Properties-Datei gelesen (siehe Abschnitt 4.5.1). Mehr dazu warum das Programm eine Wartezeit hat in Abschnitt 4.5.2 zu den Threads.

Die Methode *updateCrawlerStand* ruft die Methode *updateFenster* der Klasse *CrawlerFenster* auf, welche dafür sorgt, dass die Daten zu Website und Webseite die gerade durchsucht werden in diesem Fenster stehen. Diese Methode übergibt als Parameter die Zählvariablen für die Website und die Webseite.

#### 4.2.5 Klasse CrawlThread

Die Klasse *CrawlThread* erledigt das eigentliche Durchsuchen des HTML-Codes der Webseiten nach Links. In diesem Abschnitt werden die Methoden dieser Klasse nur kurz beschrieben. Eine genauere Beschreibung, vor allem die des Suchalgorithmus, folgt in Abschnitt 4.4 in Pseudocode-Form.

Der Konstruktor der Klasse dient zum initialisieren der Variablen *zaehlerWebsite* und *zaehlerWebseite* welche von der Klasse *Crawler* übergeben wurden und dazu dienen, damit der Crawler weiß welche Webseite er gerade durchsucht und zu welcher Website diese Webseite gehört.

Die Methode *run* ist die wichtigste Methode dieser Klasse. In ihr werden die Links einer Webseite herausgesucht und in der Datenbank gespeichert.

Die Methode *checkeObLinkZurWebsiteGehoert* prüft, ob ein Link relativ zur aktuell durchsuchten Website ist oder nicht, d.h. ob die Adresse des Links der gleichen Website entstammt wie die aktuell durchsuchte Website. Die Adresse des Links wird in Form eines String-Parameters an die Methode übergeben. Wenn der Link relativ ist, ist der Rückgabewert der Methode „true“.

Die Methode *extrahiereWebsiteAdresse* steht für das Herausfiltern der Adresse einer Website aus einer URL. Ein Beispiel dafür: Der Inhalt des an die Methode übergebenen String-Parameters sei die zu extrahierende URL *http://www.austria.info/hotels/steiermark/index.php*, so lautet das zurückgegebene Ergebnis nach Ausführen dieser Methode *http://www.austria.info*. Allerdings gilt es zu beachten, dass diese Methode auch Subsites als eigenständige Websites erkennt. Zum Beispiel wird *http://at.austria.info/hotels/index.php* zu *http://at.austria.info* und stellt somit eine „neue“ Website neben der eigentlichen Website *http://www.austria.info* dar. Es gibt aber generell nur wenige Subsites, weshalb sich dieser Fehler auf das Ergebnis des Suchvorgangs nur minimal auswirkt. Da die Behebung dieses Fehlers ziemlich aufwendig wäre, die Auswirkungen aber wie gesagt nur gering sind, wurde auf eine Korrektur verzichtet.

#### 4.2.6 Klasse Database

Die Klasse *Database* ist für die Kommunikation mit der Datenbank zuständig. Sie hat zuerst einmal zwei verschiedene Konstruktoren. Die Konstruktoren sind dazu da, die Benutzerdaten aus der Properties-Datei einzulesen und die Verbindung zur Datenbank herzustellen. Es gibt deshalb zwei Konstruktoren, weil es vorkommen kann, dass eine Datenbank zuerst gelöscht werden muss, bevor sie wieder erzeugt wird. Um das Löschen zu ermöglichen, muss das Datenbankobjekt zuerst mit einem anderen Datenbanknamen initialisiert werden, um dann von dieser Datenbank aus die andere Datenbank löschen zu können.

Die Methode *checkeObDatensatzVorhanden* schaut ob ein Datensatz in einer der Tabellen *Websites*, *Webseiten*, *InLinks* oder *OutLinks* vorhanden ist, indem sie als Parameter die Adresse und den Tabellennamen dieses Datensatzes (alle genannten Tabellen haben das Attribut Adresse, welches auch eindeutig ist) übergeben erhält und falls der Datensatz vorhanden ist „true“ zurückgibt.

Die Methode *getDatensatz* liefert als Rückgabewert einen bestimmten Datensatz aus einer der vier Tabellen *Websites*, *Webseiten*, *InLinks* und *OutLinks*. Ihre Parameter sind der Tabellename und die *ID* des Datensatzes.

Die Methode *getDatensaetzeEinerWebsite* beschafft alle zu einer Website mit einer bestimmten *ID* (Parameter *websiteId*) gehörenden Webseiten, ein- oder ausgehenden Links. Diese Methode wird vor allem für die Transformation der Daten für die Analyse (siehe Abschnitt 5.2) benötigt.

Die Methode *getAlleWebsites* holt alle Websites aus der Tabelle *Websites*. Auch diese Methode wird hauptsächlich bei der Transformation der Daten für die Netzwerkanalyse in Abschnitt 5.2 gebraucht.

Die Methode *inDatenbankSchreiben* dient dazu, Daten in die Datenbank zu schreiben. Sie erhält als Parameter den Tabellennamen und einen Vektor mit den Daten. Je Tabelle weiß die Methode dann welche Parameter im Vektor enthalten sind und wie sie das SQL-Statement definieren soll.

Die Methode *datenbankUpdaten* dient zum Verändern des Datenbankinhaltes der Tabelle *Webseiten*, speziell des Attributes *GeradeDurchsucht*, das speichert welche Webseite gerade durchsucht wird. Wenn dieses Attribut den Wert 1 hat, wird die dazugehörige Webseite gerade durchsucht. In dieser Methode wird sowohl der Wert des Attributes der Webseite die

gerade durchsucht wurde auf 0, als auch jener des Attributes der Webseite bei der das Durchsuchen in der Folge begonnen wird auf 1 gesetzt.

Die Methode *tabellenErzeugen* dient dem Erzeugen der Datenbanktabellen wenn der Crawler gestartet wird. Der übergebene Parameter enthält die Startadressen die in die erzeugten Tabellen geschrieben werden müssen.

Die Methode *datenbankErzeugen* ist dazu da, eine neue Datenbank zu erzeugen. Wie oben gilt auch hier, dass der Parameter der übergeben wird, den Namen der Datenbank enthält.

Die Methode *datenbankLoeschen* dient dem Löschen einer Datenbank. Der übergebene Parameter enthält den Namen der zu löschenden Datenbank.

Außerdem gibt es noch die Methode *closeDatabase*. Diese Methode schließt die Verbindung zur Datenbank.

### 4.3 Das Datenmodell

In diesem Abschnitt wird das Datenmodell zur Speicherung der Daten in einer MySQL-Datenbank beschrieben. Abbildung 25 zeigt das gesamte Datenmodell, welches hier beschrieben wird.

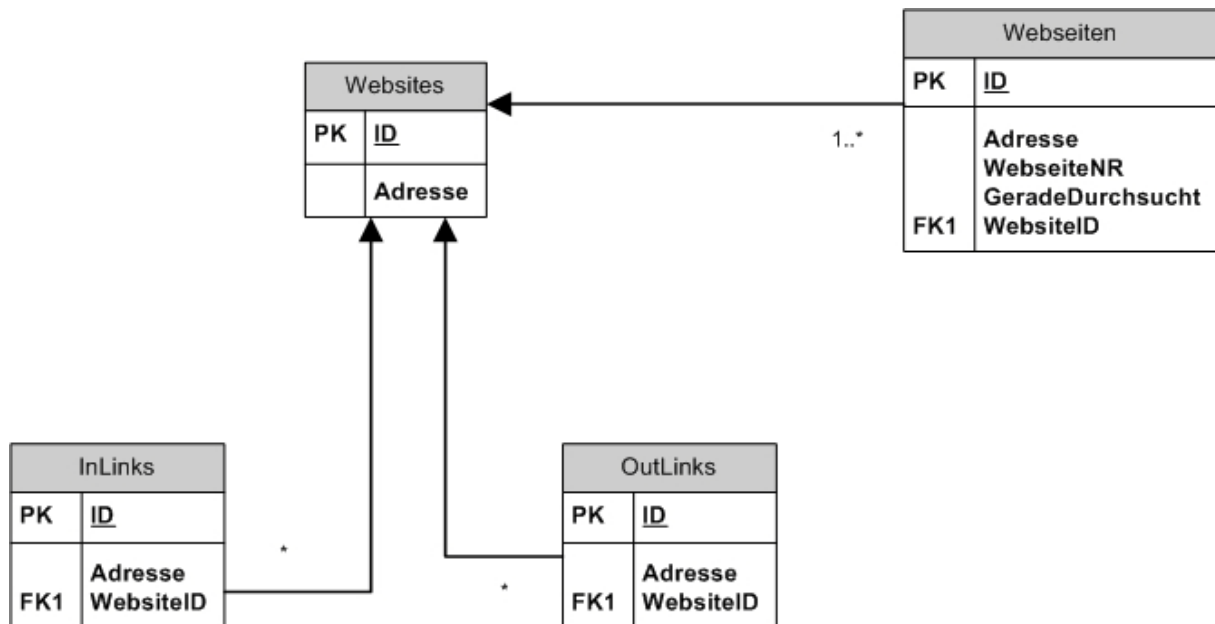


Abbildung 25: Das Datenmodell des Crawlers

Die Namen der Tabellen im Datenmodell aus Abbildung 25 bezeichnen auch den Inhalt der Tabellen. Für jeden Datensatz in der Tabelle *Websites* (enthält alle zu durchsuchenden Websites) gibt es entweder keinen, einen oder mehrere Einträge in den Tabellen *InLinks* und *OutLinks* (d.h. jede Website kann keinen oder viele ein- oder ausgehende Links haben). Gleichzeitig gibt es für jede Website mehrere, aber zumindest eine zugehörige Webseite. Alle Tabellen bestehen jeweils aus einer eindeutigen *ID* (vom Typ *Integer*) mit der Eigenschaft *auto\_increment* (für automatisches Erhöhen der *ID*) und einer Adresse (vom Typ *VarChar*, zum Beispiel in der Tabelle *OutLinks* die Adresse *http://www.tiscover.at* oder in der Tabelle *Webseiten* die Adresse *http://www.austria.info/it/destinazioni/tirol/index.php*). Zusätzlich dazu enthalten die Tabellen *Webseiten*, *InLinks* und *OutLinks* noch ein Attribut *WebsiteID* (vom Typ *Integer*) welches die *ID* der zugehörigen Website darstellt und somit ein Fremdschlüssel ist, damit für jede Website die dazugehörigen ein- und ausgehenden Links, sowie die Webseiten identifiziert werden können. Die Tabelle *Webseiten* enthält des Weiteren

noch zwei Attribute die kurz beschrieben werden müssen. Das Attribut *WebseiteNR* (vom Typ *Integer*) benennt die wievielte Webseite innerhalb der dazugehörigen Website dieser Eintrag ist. Dieses Attribut ist notwendig um die Webseiten einer Website eindeutig zu identifizieren. Zusätzlich entspricht dieses Attribut der Variablen *zaehlerWebseiten* in der Klasse *Crawler*. Das Attribut *GeradeDurchsucht* enthält entweder den Wert 1 oder 0 und informiert über die gerade durchsuchte Webseite. Hat das Attribut den Wert 1 so heißt das, dass die zum Datensatz gehörende Webseite gerade durchsucht wird, ansonsten hat das Attribut den Wert 0. So weiß der Crawler, im Falle einer Wiederaufnahme des Suchvorgangs wo er gestoppt wurde.

#### 4.4 Pseudocode zu den wichtigsten Funktionen des Crawlers

In diesem Abschnitt wird der Pseudocode für einige wichtige Routinen (speziell das eigentliche Durchsuchen) des Programms dargestellt. Es gilt zu beachten, dass nicht alle Klassen des Programms darin involviert sind. Zum Beispiel werden die Klassen der GUI hier nicht in Pseudocode dargestellt.

Die Darstellung in Pseudocode betrifft die zwei Klassen *Crawler* und *CrawlThread*. Beide sind, wie im Klassendiagramm in Abbildung 24 zu sehen ist, von der Klasse *Thread* abgeleitete Klassen.

```
public class CrawlThread extends Thread {

    public void run() {
        while (esGibtEineZeileInWebseite) {
            while (esGibtEinenLinkInDerZeile) {
                if (schaueObLinkZuWebsiteGehoert) {
                    if (webseiteVonLinkNochNichtVorhanden) {
                        speichereWebseiteInDatenbank;
                    }
                }
                else {
                    if (websiteNochNichtVorhanden) {
                        speichereWebsiteInDatenbank;
                        speichereOutLinkInDatenbank;
                        speichereInLinkInDatenbank;
                    }
                    else {
                        if (geradeGefundeneWebsiteID > durchsuchteWebsiteID) {
                            speichereWebseiteInDatenbank;
                        }
                        if (outLinkNichtVorhanden) {
                            speichereOutLinkInDatenbank;
                        }
                        if (inLinkNichtVorhanden) {
                            speichereInLinkInDatenbank;
                        }
                    }
                }
            }
        }
    }
}
```

#### Listing 1: Pseudocode der Klasse *CrawlThread*

Zuerst wird die Klasse *CrawlThread* beschrieben. Diese Klasse durchsucht eine Seite nach Links und speichert sie in der Datenbank. (Allgemeine Informationen zu dieser Klasse finden

sich in Abschnitt 4.2.5) In Listing 1 ist der zugehörige Pseudocode (Parameter- und Methodenbezeichnungen stimmen nicht immer mit dem Code überein) dargestellt.

Zuerst wird in Listing 1 die Webseite Zeile für Zeile durchsucht. Anschließend wird die Zeile solange durchsucht, solange es einen Link in der Zeile gibt. Mit dem Aufruf *schauObLinkZuWebsiteGehört* wird kontrolliert ob es sich beim gefundenen Link um einen relativen Link (d.h. die Adresse des Links gehört zur gerade durchsuchenden Website) oder um einen externen Link (d.h. der Link zeigt auf eine andere als die gerade durchsuchende Website) handelt. Falls es ein relativer Link ist muss kontrolliert werden (Aufruf *webseiteVonLinkNochNichtVorhanden*) ob diese Webseite bereits in der Tabelle *Webseiten* vorhanden ist. Falls nicht, wurde ein neuer Link (bzw. eine neue Webseite) gefunden, ansonsten ist das Programm in dieser Klasse bereits fertig.

Handelt es sich bei dem gefundenen Link um einen externen Link so muss zuerst geschaut werden ob die Adresse des Links bereits als Website in der gleichnamigen Tabelle vorhanden ist. Falls es diese Website noch nicht gibt wird sie eingetragen und der gefundene Link wird auch in den Tabellen *OutLinks* (als ausgehender Link der gerade durchsuchenden Website) und *InLinks* (die gerade durchsuchende Website als eingehender Link der jetzt gefundenen Website) eingetragen. Existiert die Website dieses Links bereits so ist es zunächst wichtig zu wissen, ob diese Website bereits durchsucht wurde oder nicht. Falls sie noch nicht durchsucht wurde (d.h. deren *ID* ist größer als die *ID* der gerade durchsuchenden Website) muss der gefundene Link in der Tabelle *Webseiten* gespeichert werden. Die Website dieses Links ist zwar schon vorhanden, aber um die unterschiedlichen, aber auf die gleiche Website verweisenden Links zu berücksichtigen, werden sie als Webseiten gespeichert und somit werden gleich am Anfang des Durchsuchens dieser Website genau jene Webseiten durchsucht die verlinkt wurden. In der Folge wird dann kontrolliert ob die Website des gefundenen Links schon als ausgehender Link der gerade durchsuchenden Website in der Tabelle *OutLinks* eingetragen ist und wenn nicht wird dies gemacht. Das gleiche wird für die Tabelle *InLinks* durchgeführt.

Nun noch kurz zwei Dinge zum Durchsuchen der Webseite, welche im Pseudocode in Listing 1 nicht genannt werden. Das Finden eines Links in einer Zeile der Webseite wird mit dem Befehl *String.indexOf(„<a href=“)* durchgeführt. Falls ein Link gefunden wurde muss überprüft werden um welche Art von Link es sich handelt, denn es sollen nur Links die auf Webseiten verweisen berücksichtigt werden:

```
if ((String.indexOf(„.jpg“)==-1)    && (String.indexOf(„.png“)==-1)    &&
    (String.indexOf(„.pdf“)==-1 && (String.indexOf(„.xml“)==-1) usw.
```

Mit diesem Stück Code werden nur die „unnützen“ Links wie etwa Bilder, pdf-Dateien oder xml-Dateien herausgefiltert. Natürlich könnte das auch umgekehrt gelöst werden, indem geschaut wird, ob der Link mit *.html*, *.php*, *.shtml*, *.jsp* oder anderem endet. Aufgrund der Schwierigkeit, alle möglichen Endungen für „richtige“ Links zu finden, hat sich der Autor für die erste Variante entschieden. Somit war es dann von Anfang an leichter, jene Links, die nicht gebraucht werden oder keine „richtigen“ Links bzw. Links zu anderen Webseiten sind, auszuschließen.

Nun folgt die Beschreibung des Pseudocodes der Klasse *Crawler* in Listing 2 (Parameter- und Methodenbezeichnungen müssen nicht den im Code verwendeten entsprechen). Es gilt zu beachten, dass die Klasse *Crawler* für jede zu durchsuchende Webseite, welche in der Datenbank gefunden wurde, einen Thread der Klasse *CrawlThread* wie oben beschrieben startet. Wie dabei vorgegangen wird, wird in Listing 2 beschrieben und dann genauer erklärt.

Diese Klasse hat, außer zusätzlichen Variablen, welche hier nicht näher erklärt werden, zwei wichtige Klassenvariablen: *zaehlerWebsites* und *zaehlerWebseiten*. Diese Variablen zählen



die wievielte Webseite einer Website oder die wievielte Website insgesamt gerade durchsucht wird und entsprechen der *ID* in der Tabelle *Websites* und dem Attribut *WebseiteNR* in der Tabelle *Webseiten* (siehe Abschnitt 3.3).

Die erste Methode der Klasse ist wieder die Methode *run*, weil die Klasse aus der Klasse *Thread* abgeleitet ist. Von dieser Klasse wird zwar nur ein einzelner Thread zeitgleich ausgeführt, doch es musste alles in einen Thread gepackt werden, um den Zugriff zur Laufzeit auf den Thread bzw. dessen Ausführung zu behalten. Ansonsten wäre es für den Benutzer nicht möglich, die Ausführung des Programms über die GUI zu stoppen.

```
public class Crawler extends Thread {

    int zaehlerWebsites;
    int zaehlerWebseiten;
    int anzahlWebseitenProWebsite;

    public void run() {
        while (websiteVorhanden) {
            durchsucheWebsite();
            zaehlerWebsites++;
            zaehlerWebseiten = 1;
        }
    }

    public void durchsucheWebsite() {
        new ThreadPool(5);
        while (webseiteVorhanden) &&
            (zaehlerWebseiten <= anzahlWebseitenProWebsite) {
            CrawlThread ct = new CrawlThread(...);
            ct.start();
            Thread.sleep(getWartezeit(zaehlerWebseiten));
            zaehlerWebseiten++;
        }
    }
}
```

## Listing 2: Pseudocode der Klasse *Crawler*

In der Methode *run* werden alle Websites durchsucht, beginnend bei der Startwebsite (siehe Abschnitt 5.1.1) und gefolgt von den danach gefundenen Websites. Dabei gilt es zu beachten, dass jene Websites, deren Links zuerst gefunden wurden, auch zuerst durchsucht werden. Es gibt hierzu wieder eine While-Schleife. Diese Schleife wird solange durchgeführt, solange es noch zu durchsuchende Websites gibt, was zum Beispiel für das gesamte World Wide Web praktisch immer der Fall ist. Zu Beginn wird in dieser Schleife gleich die Methode *durchsucheWebsite* aufgerufen, welche im nächsten Absatz beschrieben wird. Danach muss nur die Zählvariable für die Websites *zaehlerWebsites* um eins erhöht und die Zählvariable für die Webseiten *zaehlerWebseiten* auf 1 gesetzt werden, weil nun eine neue Website durchsucht wird.

In der Methode *durchsucheWebsite* werden die Webseiten einer Website nacheinander durchsucht. Hier gibt es dann einen Thread-Pool mit fünf Threads, welche alle für die Klasse *CrawlThread* (siehe Pseudocode in Listing 1) erzeugt werden. Zuerst wird wieder eine While-Schleife solange ausgeführt, solange es noch zu durchsuchende Webseiten einer Website gibt und solange nicht die Begrenzung der Anzahl der zu durchsuchenden Webseiten pro Website (Variable *anzahlWebseitenProWebsite* in Listing 2) erreicht ist. Darin wird jeweils ein Thread gestartet (und ihm gleichzeitig die Zählvariablen übergeben, was aber in Listing 2 nicht angeführt ist). Dieser Thread durchsucht dann, wie vorher in diesem Abschnitt beschrieben eine Webseite. Danach muss das Programm kurz mit dem Aufruf *Thread.sleep(Zeit)* unterbrochen werden, wobei die Länge der Unterbrechung von der Anzahl der bereits

durchsuchten Webseiten innerhalb einer Website abhängt (näheres dazu in Abschnitt 4.5.2). Schließlich muss noch die Zählvariable für die Webseiten *zaehlerWebseiten* um eins erhöht werden.

## 4.5 Technische Details

In diesem Abschnitt werden ein paar technische Details zum Programm beschrieben. Zuerst wird der Inhalt der Properties-Datei beschrieben und anschließend wird genauer auf die Vorgehensweise bei den Threads eingegangen.

### 4.5.1 Properties-Datei

Die Daten für die Datenbank und die Wartezeiten, sowie die dazugehörigen Grenzen für die Threads (siehe dazu Abschnitt 4.5.2) werden aus einer Properties-Datei ausgelesen. Die Datei heißt *props.properties*. Die Endung der Datei muss *.properties* sein. Die Struktur einer Properties-Datei muss immer so sein, dass in einer Zeile immer eine Eigenschaft, dann ein Gleichheitszeichen (=) und danach der Wert für diese Eigenschaft steht.

```
datenbanknameAlternativ = mysql

host = localhost
datenbankname = crawler
username = maxmustermann
password = maxmustermannpassword

#Grenzen Geschwindigkeit
geschwindigkeit1 = 2500
geschwindigkeit2 = 1500
geschwindigkeit3 = 1000
geschwindigkeit4 = 750
geschwindigkeit5 = 350

#Grenzen Anzahl
anzahl1 = 500
anzahl2 = 750
anzahl3 = 1000
anzahl4 = 2500
anzahl5 = 5000
```

#### Listing 3: Inhalt der Properties-Datei

Für die Bearbeitung von Properties-Dateien gibt es in Java das vorgefertigte Paket *java.util.properties*, welches nur am Anfang der Klasse importiert werden muss. Mit den Klassen und Methoden dieses Pakets kann zum Beispiel die Datei sequentiell durchlaufen oder auch sofort auf den Wert einer Eigenschaft durch Angabe der Eigenschaft zugegriffen werden.

Die Properties-Datei, die in diesem Programm verwendet wurde, wird im obigen Listing 3 dargestellt. Kommentare in der Datei werden mit einem vorangehenden #-Zeichen markiert.

Auf die Datenbankdetails muss hier nicht näher eingegangen werden. Jedoch werden die Grenzen für die Anzahl und die Geschwindigkeiten, die in der Properties-Datei weiter unten angeführt sind, im folgenden Kapitel näher erklärt.

### 4.5.2 Threads

Dieser Crawler wurde mit Hilfe der in Java zur Verfügung stehenden Thread-Technologie geschrieben. [57] definiert Threads folgendermaßen: „A thread is a thread of execution in a

program. The Java Virtual Machine allows an application to have multiple threads of execution running concurrently“.

Das heißt, mit Hilfe von Threads können mehrere Aufgaben gleichzeitig gelöst werden. Ohne den Einsatz von Threads war der Crawler überaus langsam, weshalb diese Technologie angewandt wurde.

Das Paket *java.util* enthält die Klasse *Thread*. Die Implementierung von Threads in Java ist ziemlich einfach. Eine Klasse die eine Teilaufgabe parallel ausführen soll, muss von der Klasse *Thread* abgeleitet und die Methode *run* muss in dieser Klasse implementiert werden. In diese Methode muss dann der Code, der parallel ausgeführt werden soll, enthalten sein. Gestartet wird der Thread indem zuerst ein Objekt der Klasse, welche den Thread implementiert, erzeugt wird. Danach muss der Thread nur noch über das erzeugte Objekt durch den Aufruf *Objekt.start()* gestartet werden. Der Thread kann ganz einfach mit *Objekt.stop()* oder *Objekt.interrupt()* beendet werden. Der Thread wird dann parallel ausgeführt, während das Programm nach dem Aufruf des Threads an dieser Stelle fortfährt.

Um die Geschwindigkeit des Crawlers zu erhöhen, sollten mehrere solche Threads gleichzeitig ausgeführt werden. Die Verwaltung der Threads – das heißt, wann welcher Thread an die Reihe kommt – gestaltet sich im Gegensatz zum Aufruf des Threads schon ein bisschen schwieriger. Jedoch gibt es seit kurzer Zeit in Java im Paket *java.util.concurrent* das Interface *Executor* [58], welches das Handhaben von Threads erleichtert. Um dieses Interface benutzen zu können, muss zuerst ein Objekt der Klasse *Executor* erzeugt und dabei die Anzahl der gewünschten Threads übergeben werden: *Executor executor = Executors.newFixedThreadPool(5)*. Danach wird ein Objekt der Klasse erzeugt, welche den Thread implementiert; dieses Objekt wird beim Aufruf von *executor.execute(crawler)* übergeben. Damit werden die Threads gestartet. In diesem Fall wird eine Anzahl von 5 Threads angenommen. Mehr als 5 Threads zu verwenden kostet mit der Zeit zuviel Speicher und es kann passieren, dass eine *OutOfMemoryException* passiert.

Wenn die Klasse *CrawlThread* als Thread ausgeführt wird, kommt es jedoch zu Problemen. Da das Auffinden der Links in einer Webseite und das Speichern der Links in die Datenbank relativ viel Zeit kostet, in der Zwischenzeit aber bereits alle fünf Threads gestartet sind und neue Threads benötigt werden, kommt es zu Abbrüchen von noch nicht abgeschlossenen Threads, um neue Threads beginnen zu können. In diesem Fall wird nur ein Teil der Webseiten durchsucht und auch nur ein Teil der Links gespeichert.

Um dieses Problem zu lösen, wurde ein Algorithmus implementiert, der durch das Aufrufen der Methode *Thread.sleep(Zeit in Millisekunden)* den Programmablauf für eine gewisse Zeit stoppt. Auf diese Weise können die Threads ihre Arbeit beenden, ohne vorher unterbrochen zu werden.

Der Algorithmus verwendet für unterschiedliche Bereiche bei der Anzahl von durchsuchten Webseiten unterschiedliche Wartezeiten. Diese sind in der folgenden Tabelle 9 dargestellt. Die Erfahrung des Autors hat gezeigt, dass die meisten ausgehenden, externen Links am Anfang des Durchsuchens einer Website gefunden werden. Das ist der Grund dafür, dass am Anfang einer Domain lange gewartet wird. Je größer die Anzahl der durchsuchten Webseiten einer Website wird, desto weniger ausgehende Links werden gefunden und desto größer ist auch die Wahrscheinlichkeit, dass das weitere Durchsuchen dieser Website eigentlich schon hinfällig wäre, weil es innerhalb dieser Website nicht so viele zu durchsuchende Webseiten gibt. Oftmaliges und Probieren und Beobachten hat dem Autor gezeigt, dass die meisten Websites aus weniger als 1000 Webseiten bestehen. Deswegen wird beim Durchsuchen der ersten Webseiten einer Website länger gewartet und mit zunehmender Anzahl von durchsuchten Webseiten je Domain wird das Programm schneller. Dadurch kann Zeit gespart werden. Würde nach jedem Durchsuchen einer Webseite sowie innerhalb der ersten Periode

2500 Millisekunden gewartet werden, so würde für das Durchsuchen einer Website, welche aus 5000 Webseiten besteht, 12500 Sekunden, also fast vier Stunden, dauern. Um dann 1000 Websites von gleicher Größe zu durchsuchen, würden 4000 Stunden benötigt, was ungefähr 166 Tagen entspricht. Es würde als ungefähr fünfeneinhalb Monate dauern, um ein Netzwerk, bestehend aus nur 1000 Websites, zu erhalten. Doch nicht jede Website besteht, wie schon vorher gesagt wurde, aus 5000 Webseiten und die Beschleunigung des Programmablaufs mit größerer Anzahl von durchsuchten Webseiten je Website verbessert die Geschwindigkeit um einiges, auch wenn der Crawler noch immer zu langsam arbeitet.

Anzahl durchsuchter Webseiten je Website	Wartezeit in Millisekunden
0-499	2500
500-749	1500
750-999	1000
1000-1499	500
1500-1999	250
>=2000	50

**Tabelle 9:** Zeigt die verschiedenen Perioden je nach Anzahl der bereits durchsuchten Webseiten pro Website und die dazugehörige Zeitspanne in Millisekunden, während welcher der Thread jeweils nach dem Durchsuchen einer Webseite pausiert.

Die obigen Geschwindigkeitsangaben sind allerdings an den Computer angepasst, auf dem der Suchvorgang durchgeführt wurde. Je leistungsfähiger ein Computer ist, desto kürzer muss das Programm angehalten werden und desto mehr Threads können gleichzeitig laufen. Der PC auf dem der Crawler lief, ist ein Pentium 2, mit einem 1.77 GHz Prozessor und etwa 1300 MB RAM. Ein oder mehrere schnellere PC's würden die Geschwindigkeit des Crawlers deutlich verbessern.

#### **4.6 Problembereiche eines Crawlers**

Ein Crawler ist vor allem bei größeren Netzwerken ein gutes und ab einer bestimmten Netzwerkgröße sogar ein unverzichtbares Werkzeug um ein Netzwerk bestehend aus Webseiten oder Websites zu bestimmen. Zum Beispiel ist eine Suchmaschine für das World Wide Web auf einen Crawler angewiesen. Nichtsdestotrotz hat ein Crawler auch seine Schwachstellen und Problembereiche, wodurch seine Ergebnisse zum Teil unvollständig sein können. Diese Problembereiche eines Crawlers im Allgemeinen, sowie des im Rahmen dieser Arbeit entwickelten Crawlers im Speziellen, sollen in diesem Abschnitt kurz angesprochen werden.

Erstens ist es mit einem Crawler schwierig alle Links innerhalb einer Webseite zu finden. Zum Beispiel können Webseiten in denen die Links in Java-Script Code eingebettet sind oder solche wo die Links im Rahmen eines Formulars gesetzt sind mit dem hier entwickelten Crawler nicht gefunden werden. Das Auffinden dieser Links wäre durch eine andere Suchstrategie, welche mit zusätzlichem Code umgesetzt werden könnte noch möglich. Auch die Handhabung von HTML-Frames erschwert die Arbeit eines Crawlers, da auch hier eine andere Suchstrategie gewählt werden muss, welche bei dem hier entwickelten Crawler nur teilweise umgesetzt wurde; Webseiten die aus HTML-Frames bestehen, können aber prinzipiell mit einem Crawler auf Links durchsucht werden. Jedoch können zum Beispiel Links, welche in zugangsbeschränkten Webseiten oder etwa in Flash generierten Seiten enthalten sind, aufgrund von technologischen oder sicherheitstechnischen Gründen mit keinem Crawler gefunden werden.

Zweitens erschwert die dynamische Natur des World Wide Webs die Arbeit eines Crawlers. „Zur Laufzeit“ erzeugte Webseiten können zum Beispiel mit einem Crawler nicht gefunden und somit nach Links durchsucht werden. Dies ist ein relativ großes Problem in Zusammenhang mit der Ermittlung des gesamten Webgraphen, da nicht wenige Webseiten erst „zur Laufzeit“ als Reaktion auf Benutzereingaben erzeugt werden. Auch ist das World Wide Web kein statisches Gebilde das sich nicht verändert. Websites verändern sich durch Hinzufügen, Umgestalten oder Entfernen von Webseiten und neue Websites werden nicht nur monatlich, wöchentlich oder täglich, sondern in jeder Minute und Sekunde hinzugefügt. Durch dieses rasante Wachstum des World Wide Webs ist das Ergebnis eines Crawlers, nachdem er seine Arbeit beendet hat, bereits „veraltet“ und spiegelt deshalb nur in unzureichendem Maße das gewünschte Ergebnis wieder. Dieser Umstand kommt zwar beim Netzwerk des österreichischen E-Tourismus-Marktes nicht so sehr zum Tragen wie beim World Wide Web insgesamt, darf aber nicht vergessen werden.

Dies sind nur einige Problembereiche von Crawlern, die deshalb auch bei der Erforschung des Webgraphen auftreten. Weitere Schwachstellen könnten genannt werden, dies würde aber den Rahmen dieser Arbeit sprengen. Im nächsten Kapitel wird beschrieben, wie bei der Analyse des Netzwerks des österreichischen E-Tourismus-Marktes vorgegangen wird.

## 5 Die Analysemethode

In diesem Kapitel wird beschrieben nach welcher Methode die Analyse der Netzwerksstruktur des österreichischen E-Tourismus-Marktes erfolgt. Begonnen wird mit der Beschreibung der Strategie beim Sammeln der Daten durch den Crawler. Danach müssen die Daten aus der Datenbank für den Import in das Analysetool transformiert werden. Dies wird im Abschnitt 5.2 beschrieben. Abschnitt 5.3 zeigt dann, wie die Daten in das Analysetool importiert werden. Anschließend folgt in Abschnitt 5.4 die Beschreibung der graphischen Darstellung des Netzwerks. Abgeschlossen wird mit Abschnitt 5.5, in welchem erläutert wird, wie die für die Analyse notwendigen Kennzahlen mit Hilfe des Analysetools berechnet werden können.

### 5.1 Das Sammeln der Daten mit dem Crawler

Hier wird beschrieben, wie die Daten mit dem Crawler gesammelt werden. Wie der Crawler funktioniert, wurde ja bereits in Kapitel 4 beschrieben. In diesem Abschnitt geht es um bestimmte Einstellungen, die vorgenommen werden müssen.

#### 5.1.1 Wahl der Startadressen

Ein wichtiger Aspekt beim Betreiben des Crawlers ist die Wahl der Startadressen. Es geht vor allem darum, jene Startadressen zu finden, die möglichst viele Tourismusanbieter verlinken, um so ein aufschlussreiches Netzwerk des österreichischen E-Tourismus-Marktes zu erhalten. Das aufschlussreichste Netzwerk wird dann für die Analyse herangezogen. Dies muss deshalb so sein, weil ein vollständiges Netzwerk des österreichischen E-Tourismus-Marktes nicht ermittelt werden kann.

Denn erstens können aufgrund von technischen Problemen in Zusammenhang mit einem Crawler (siehe Abschnitt 4.6) zum Beispiel nie alle Webseiten bzw. Websites des Netzwerks gefunden werden. Zweitens ist die Leistung des PC's auf dem der Crawler lief und auch die zur Verfügung stehende Zeit für die Arbeit begrenzt, weshalb das Netzwerk nicht vollständig ermittelt werden kann. Drittens entfernt sich der Suchvorgang des Crawlers mit zunehmender Dauer immer mehr von der Tourismusdomäne und der österreichischen Domäne, weil der Crawler kein fokussierter Crawler ist und somit alle Webseiten die er findet speichert und weiter durchsucht. Aufgrund dessen werden anstatt Knoten, die dem österreichischen E-Tourismusnetzwerk angehören, mit zunehmender Dauer immer mehr „artfremde“ Knoten gefunden und die Websites der Akteure des österreichischen E-Tourismus-Marktes geraten in der Anzahl in die Minderheit, weshalb man nicht mehr von einem Netzwerk des österreichischen E-Tourismus-Marktes sprechen kann. Wegen dieser Gegebenheiten ist die Wahl der Startadressen sehr wichtig. Diese sollen in hoher Anzahl andere zum E-Tourismusnetzwerk zugehörige Knoten verlinken, damit das ermittelte Netzwerk möglichst repräsentativ für den österreichischen E-Tourismus-Markt ist.

Insgesamt wurde der Crawler vier Mal mit vier verschiedenen Startadressen gestartet. Bei Suchdurchlauf 1 wurde als Startadresse <http://www.austria.info> verwendet. In Suchdurchlauf 2 wurde <http://www.tiscover.com> als Startadresse angegeben. Bei den Suchdurchläufen 3 und 4 wurden mehrere Startadressen, die in Anhang 1 und 2 aufgelistet sind, angegeben.

Jedoch wurde bei den Suchdurchläufen 3 und 4 jeweils als erste Startadresse <http://www.austria.info> und bei Suchdurchlauf 3 als zweite Adresse <http://www.tiscover.com> verwendet. *Austria.info* ist das Destinationsmanagementportal für Österreich und ist deswegen eine gute Startadresse, weil es viele Tourismusbetriebe und regionale sowie lokale Destinationsmanagementportale verlinkt. Siehe dazu eine kurze zusammenfassende Darstellung der Suchdurchläufe 1-4 in Tabelle 10.

	Suchdurchlauf 1	Suchdurchlauf 2	Suchdurchlauf 3	Suchdurchlauf 4
<b>Start-adressen</b>	<i>http://www.austria.info</i>	<i>http://www.tiscover.com</i>	Siehe Liste in Anhang 1	Siehe Liste in Anhang 2
<b>Anzahl durchsuchter Websites</b>	2116	1385	8276	2395

**Tabelle 10: Startadressen und Anzahl durchsuchter Websites bei den Suchdurchläufen 1-4**

Nach Suchdurchlauf 1 und 2 wurde festgestellt, dass es notwendig ist, weitere Websites in die Liste der Startadressen aufzunehmen, um so ein dichteres Netz von Tourismusanbietern zu erhalten. Für Suchdurchlauf 3 wurden deswegen mit Hilfe der Suchmaschine Google Tourismusportale gesucht, von denen angenommen wurde, dass sie möglichst viele Tourismusanbieter verlinken. Dabei wurden für den Tourismus spezifische Begriffe wie Urlaub, Österreich, Ferien, Ski, Campen und andere in die Suchmaschine eingegeben. Aus den gefundenen Portalen wurde jene Liste ausgewählt, die in Anhang 1 angeführt ist. Dies sind hauptsächlich Portale, die für bestimmte Urlaubsarten werben (Familienurlaub, Urlaub für Verliebte, Urlaub am Bauernhof, Bergurlaub, Schiurlaub, Badeurlaub, Campingurlaub usw.) und auf denen sich Empfehlungen zu Anbietern und Urlaubsmöglichkeiten finden; zum Teil werden dort auch Buchungsmöglichkeiten geboten.

Im Laufe der Durchführung von Suchdurchlauf 3 wurde festgestellt, dass viele der als Startadressen angegebenen Portale gar nicht viele Links zu Portalen anderer Tourismusanbieter aufweisen. Deswegen wurde für Suchdurchlauf 4 eine Liste von Portalen aus dem Tourismusbereich, welche die meisten ausgehenden Links zu verzeichnen hatten, als Startadressen zusammengestellt. Diese Liste ist in Anhang 2 angeführt. Auf diese Art und Weise kann sichergestellt werden, dass ein möglichst umfangreiches Netzwerk von Portalen verschiedenster Tourismusanbieter erhalten wird.

Je größer die Anzahl der Startadressen, die auch zum Tourismusbereich gehören, desto besser sind die Chancen, ein möglichst vollständiges Netzwerk des österreichischen E-Tourismus-Marktes zu bekommen. Die Chancen, ein möglichst vollständiges Netzwerk des Tourismusmarktes zu bekommen, sind deswegen besser, weil die Länge eines Suchdurchlaufs begrenzt ist und so am Anfang sofort wichtige, tourismusrelevante Websites durchsucht werden, die auch Links zu anderen tourismusrelevanten Websites setzen. Der Suchdurchlauf muss deswegen begrenzt werden, weil nur eine begrenzte Zeit zur Verfügung steht und auch weil das Netzwerk hauptsächlich die Tourismusdomäne repräsentieren soll. Denn je länger ein Suchdurchlauf dauert, desto mehr Links werden gefunden, die von Portalen sind, welche nicht aus der Tourismusdomäne stammen. Angenommen, der Startpunkt ist z.B. ein touristisches Portal, welches natürlicherweise auch Links zu Websites enthält, die nicht aus dem Tourismusbereich stammen. Wenn diese Websites später durchsucht werden, dann werden mit großer Wahrscheinlichkeit nur mehr Links zu Websites gefunden, die nicht zum Tourismus gehören. Es ist zwar notwendig auch Websites in die Analyse aufzunehmen, welche nicht direkt mit dem Tourismus zu tun haben, es sollte aber eine Grenze gezogen werden, denn das Ziel dieser Arbeit ist es schließlich, die Struktur des Netzwerks des österreichischen E-Tourismus-Marktes zu analysieren.

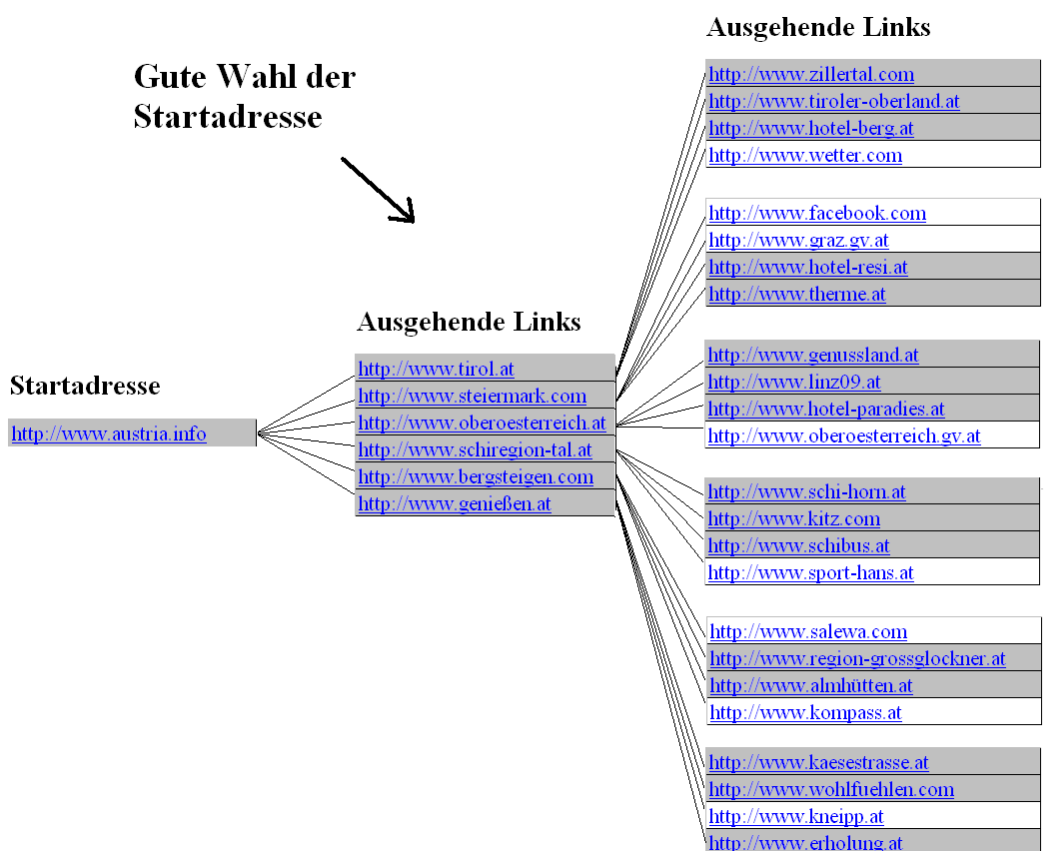
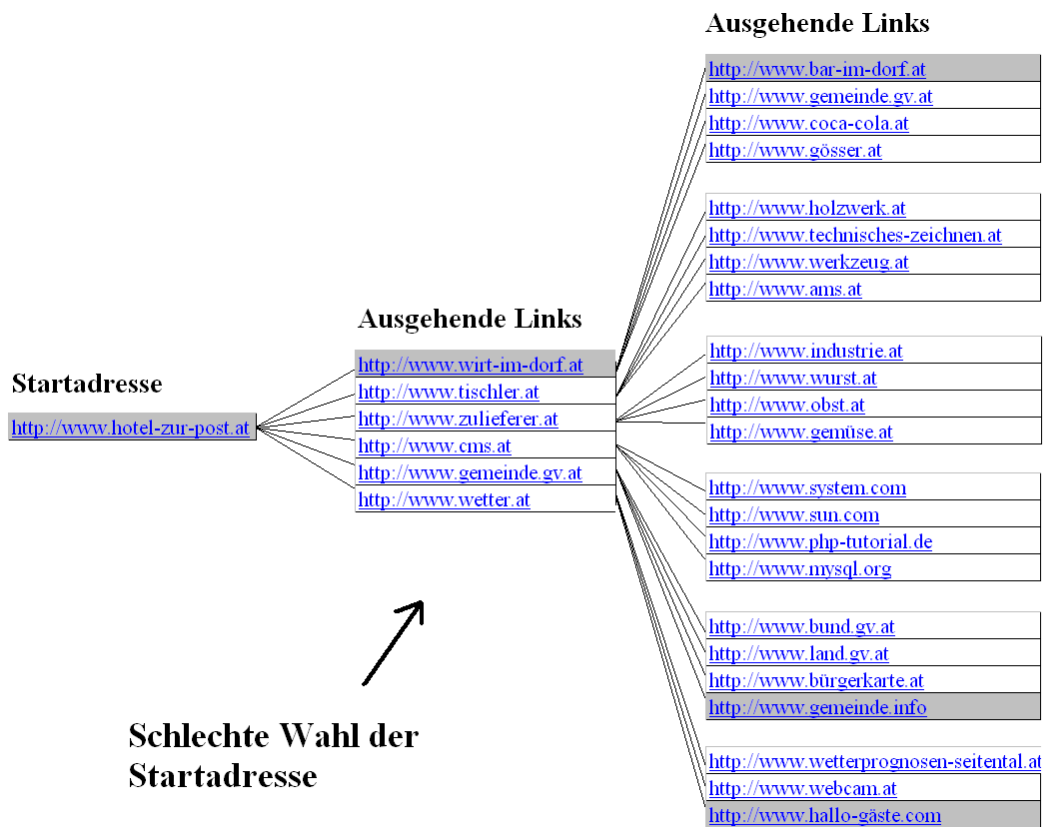


Abbildung 26: Darstellung der Unterschiede bei den gefundenen Netzwerken abhängig von der Wahl der Startadresse. Oben ein gefundenes Netzwerk mit einer schlecht gewählten Startadresse, unten ein Netzwerk mit einer gut gewählten Startadresse.



Abbildung 26 zeigt beispielhaft, wie wichtig es ist, die richtige Startadresse(n) zu wählen. Die grau markierten Adressen sind Teil des E-Tourismus-Marktes (zum Beispiel DMOs, Hotels, Gaststätten oder Transportunternehmen – siehe dazu Abschnitt 2.3.3). Die anderen Adressen gehören nicht zum E-Tourismus-Markt. Ganz links sind jeweils die Startadressen angegeben. Die Linien zwischen den Adressen markieren die gefundenen ausgehenden Links einer Adresse (Website). Im oberen Teil der Abbildung 26 ist ein Netzwerk mit einer schlecht gewählten Startadresse zu sehen. Dieses beinhaltet insgesamt nur wenige Websites von Akteuren des Tourismus-Marktes (5). Darunter ist ein Netzwerk mit einer gut gewählten Startadresse zu sehen. Hier sind nahezu alle gefundenen Adressen (23) aus dem Tourismusbereich. Die Adressen in dieser Abbildung wurden zum Teil erfunden.

### **5.1.2 Art und Weise des Durchsuchens**

In dieser Arbeit wurden verschiedene Möglichkeiten ausprobiert, wie das World Wide Web durchsucht und die Links gewertet werden können.

#### **5.1.2.1 Websites oder Webseiten als Knoten**

Wichtig ist es, zu bestimmen ob die Webseiten oder die Websites als Knoten des Netzwerks gerechnet werden sollen. In der Netzwerkanalyse des World Wide Webs werden meistens Webseiten als Knoten gerechnet. Da es in dieser Arbeit aber um die Bestimmung der zentralsten bzw. wichtigsten Organisationen und Unternehmen im Netzwerk geht, macht es eher Sinn, die gesamten Websites dieser Unternehmen als Knoten zu zählen, was auch so gehandhabt wurde. Es wäre zwar auch möglich, die Zentralität bzw. Wichtigkeit einer Organisation oder eines Unternehmens ungefähr zu ermitteln, wenn die Knoten des Netzwerks Webseiten wären, jedoch können bestimmte Zentralitätswerte nicht einfach zusammengezählt werden, um dadurch den Zentralitätswert für die gesamte Website und somit für die gesamte Organisation oder Unternehmen zu ermitteln.

#### **5.1.2.2 Die mehrfache Wertung von mehreren gleichen Links zwischen Websites**

Nachdem nun bestimmt wurde, dass die Knoten des Netzwerks den Websites entsprechen sollen, gilt es zu spezifizieren, wie die Links von einer Website zur anderen gewertet werden. Einige Websites enthalten mehr als nur einen Link auf ein und dieselbe Website, so wird z.B. auf verschiedenen Webseiten einer Website öfters durch Links auf die gleiche Website verwiesen. In diesem Fall gilt es zu klären, ob die Verbindung zwischen zwei Websites gleich gewertet werden soll, egal wie oft beide Websites miteinander verlinkt sind oder nicht. Die einfachste Variante ist, die Verbindung zwischen zwei Websites gleich zu gewichten, egal wie viele Links zwischen beiden bestehen, denn zwei Knoten eines Netzwerks werden pro Richtung üblicherweise nur durch eine Kante verbunden. Diese Variante wurde dann auch bei allen vier Suchdurchläufen angewandt.

Die zweite Variante, bei der die Anzahl von Links zwischen zwei Websites mitberücksichtigt wird, kann alternativ zur Anbringung mehrerer Kanten durch die Gewichtung der Kanten, welche zwei Knoten verbinden, umgesetzt werden. Die Höhe der Gewichtung richtet sich nach der Anzahl der Links, die zwischen zwei Websites gesetzt werden. Diese Variante wurde nur bei Suchdurchlauf 3 durchgeführt. Für diese Variante spricht, dass die Gewichtung der Kanten, also die Anzahl gleicher Links zwischen zwei Websites, auch etwas über die Wichtigkeit eines Knotens aussagen könnte. Das heißt, je mehr Links von einer Website auf eine andere gesetzt würden, desto wichtiger wäre die verbundene Website. Ob sich dieses Argument jedoch in der Praxis als haltbar erweist, ist fraglich. Dies wird versucht im folgenden Absatz zu erklären.

Denn die meisten Verbindungen zwischen zwei Websites bestehen durch nur einen Link. Daneben sind die viele Fälle von mehrfach eingebauten gleichen Links zwischen zwei Websites oft bloß auf die Struktur oder den Aufbau der Website zurückzuführen. Es ist dann ein und derselbe Link auf mehreren Webseiten der Website zu finden, nur weil dieser Link zum Beispiel im Hauptmenü oder einem anderen, auf jeder Seite angezeigten Menü untergebracht ist. Dies heißt aber nicht automatisch, dass die verlinkte Website deswegen wichtiger wäre. Dadurch würde die Website, die in diesem Fall öfters verlinkt wird, auch gegenüber den anderen Websites, die nur einmal verlinkt werden, überbewertet. Auch Links zu denselben Webseiten der sozialen Netzwerksites sind zum Beispiel auf mehreren Webseiten einer Website zu finden. Wenn diese mehrfach gezählt werden würden, so erhielten zum Beispiel die sozialen Netzwerke vergleichsweise zu viele eingehende Links, obwohl immer wieder dieselbe Webseite innerhalb des sozialen Netzwerks angelinkt würde.

Dazu ein Beispiel: ein Hotel hat auf jeder Webseite seiner Website den „Like“-Knopf von Facebook und einen Link auf das Portal *austria.info* auf nur einer Webseite. Einer realistischen Annahme zufolge könnte die Anzahl der Webseiten innerhalb einer Website eines mittelgroßen Hotels auf etwa 50 Webseiten geschätzt werden. In diesem Fall würde das soziale Netzwerk Facebook 50 eingehende Links nur von diesem Hotel erhalten. Die Berücksichtigung dieser 50 Links in der Gewichtung der Kanten zwischen diesen beiden Knoten unter der Annahme, dass die Gewichtung etwas über die Wichtigkeit der Beziehung aussagt, würde dazu verleiten zu meinen, dass nun diese Beziehung 50 mal so wichtig wäre als zum Beispiel die Beziehung zum Knoten *austria.info*. Dies ist aber in diesem Verhältnis sicherlich falsch. Auch würde in diesem Fall Facebook 49 Links mehr erhalten als *austria.info* und so wiederum zur Ansicht verleiten, Facebook wäre um das 50-fache wichtiger als das Portal *austria.info*, unter der Annahme, dass die Anzahl der eingehenden Links etwas über die Wichtigkeit eines Knotens aussagt. Deshalb würde auch diese Interpretation einen Knoten im Vergleich mit einem Anderen unverhältnismäßig hoch bewerten.

Bei der Auswertung dieser Variante mit der Gewichtung der Kanten hat sich aber gezeigt, dass diese Variante keine sinnvollen Ergebnisse liefert, weswegen diese Herangehensweise auch nicht weiter verfolgt wurde. Die Darstellung dieses Netzwerks zeigte, zumindest im Netzwerk mit den Knoten mit den 50 höchsten Eingangsgraden, nicht wenige isolierte Komponenten. Sie waren isoliert, weil sie zum Beispiel insgesamt viele eingehende Links von nur wenigen Knoten erhielten. Aufgrund dessen waren diese Knoten zwar im Netzwerk mit den 50 höchsten Eingangsgraden vertreten, dadurch aber, dass sie ihre Links hauptsächlich von einigen wenigen anderen Seiten erhielten und diese selbst nicht im Netzwerk der 50 Knoten mit den höchsten Eingangsgraden vertreten waren, hatten sie keine Verbindungen zu den übrigen Knoten im Netzwerk und waren deswegen meist isoliert oder das Netzwerk war in mehrere Komponenten aufgeteilt.

Abbildung 27 zeigt dieses Netzwerk. Die Knotengröße und die Größe der Knotenbezeichnung entsprechen der Anzahl der eingehenden Links. Die Linienstärke der Kanten entspricht der Gewichtung der Kanten nach der Anzahl der Links die es zwischen zwei Knoten gibt. Es ist zu erkennen, dass es mehrere Komponenten im Netzwerk gibt, sowie nicht wenige isolierte Knoten (links oben in der Abbildung 27). Auch sind zum Beispiel die Knoten der Subdomain „Bergfex“ vom Rest des Netzwerks isoliert und ihre Knotengröße ist unverhältnismäßig hoch da sie viele eingehende Links von ihren eigenen Subdomains erhalten. Unverhältnismäßig dick erscheint in Abbildung 27 auch die Verbindung zwischen *addthis.com* und *ausflugstipps.at*. *Addthis.com* ist sicher auch für andere Knoten wichtig, in der Abbildung scheint es aber so als wäre die Verbindung mit *ausflugstipps.at* von besonderer Bedeutung. Andererseits scheint die Dicke der Verbindung zwischen den Knoten *austria.info* und *austriatourism.com* gerechtfertigt. Dies sind beides Websites derselben DMO und sollten deswegen stärker verlinkt sein um sich so gegenseitig zu empfehlen.

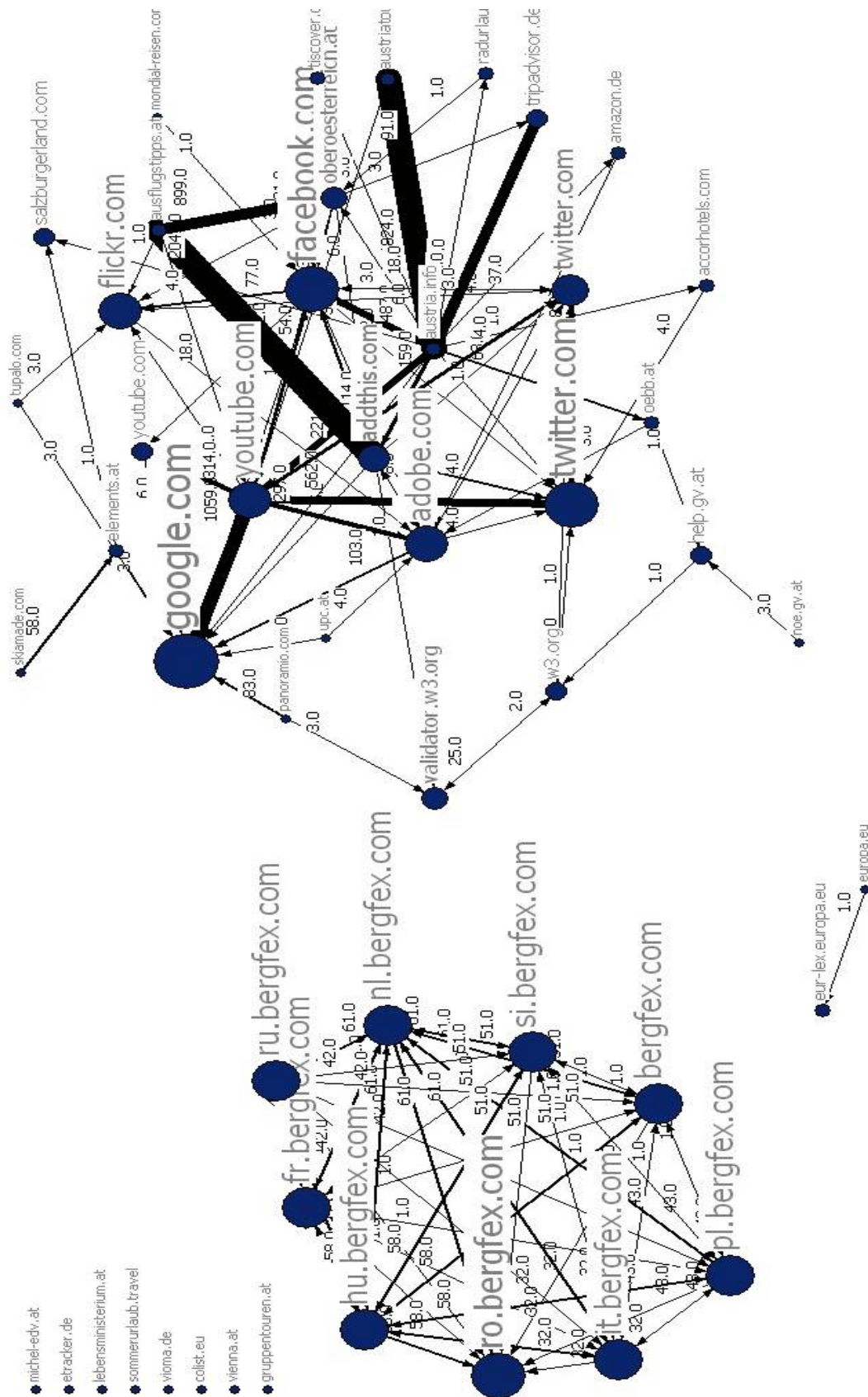


Abbildung 27: Netzwerk nach Suchdurchlauf 3 mit den 50 Knoten mit den meisten eingehenden Links und Gewichtung der Kanten durch die Anzahl eingehender Links zwischen zwei Knoten

In Summe kann also gesagt werden, dass die einfache Zählung von mehreren Links zwischen zwei Websites, so wie sie schließlich auch durchgeführt wurde, angebrachter ist.

### 5.1.2.3 Nur statische oder auch dynamische Webseiten

Eine zweite wichtige Frage bezüglich der Einstellungen des Crawlers ist, ob beim Durchsuchen der Seiten nur statische oder auch dynamische Links gefunden bzw. deren Webseiten durchsucht werden sollen. Dynamische Links sind URLs, die auch Parameter und Werte für diese Parameter enthalten. Bei den Suchdurchläufen 1 und 2 wurde nur nach statischen Links gesucht, bei den Suchdurchläufen 3 und 4 hingegen wurde auch nach dynamischen Links in den Webseiten gesucht. Die zusätzliche Suche nach dynamischen Links erhöht sicher die Anzahl der gefundenen Webseiten, die mit der durchsuchten Webseite verlinkt sind, und es kann deswegen davon ausgegangen werden, dass der gesamte Suchdurchlauf ein genaueres Netzwerk hervorbringt. Zwar sind diese dynamischen Links meist relative Links innerhalb einer Website, jedoch werden so auch dynamische Webseiten durchsucht, welche externe Links enthalten könnten, die sonst nicht gefunden werden würden. Trotzdem ist diese Möglichkeit eher begrenzt, weil die dynamisch erzeugten Teile einer Webseite meist nur verschiedene dynamisch erzeugte Inhalte liefern, jedoch eher keine neuen externen Links die nicht schon auf einer anderen dynamisch erzeugten oder sogar statischen Webseite zu finden sind. Gleichzeitig erhöht die Suche nach dynamischen Links, wie die Praxis gezeigt hat, den Suchaufwand für bestimmte Websites drastisch.

### 5.1.3 Die Programmabfolge des Crawlers als Sequenzdiagramm

In diesem Abschnitt wird die Programmabfolge des Crawlers beim Suchvorgang als Sequenzdiagramm dargestellt und kurz beschrieben. In Abbildung 28 ist das Sequenzdiagramm abgebildet, welches die Programmabfolge und die Interaktionen zwischen den Klassen widerspiegelt.

Zuerst wird ein Objekt der Klasse *Hauptfenster* erstellt. Im Konstruktor der Klasse *Hauptfenster* wird dann die Methode *setVisible*, welche das Fenster am Bildschirm darstellt, aufgerufen. Danach wird ein Objekt der Klasse *Crawler* erzeugt und durch den Aufruf *Crawler.start()* wird der Thread dieser Klasse gestartet. Nun wird in der Klasse *Crawler* ein Objekt der Klasse *CrawlerFenster* erstellt. Durch das Erstellen dieses Objektes wird der Konstruktor ausgeführt, in dem wie auch oben die Methode *setVisible*, die für die Darstellung des Fensters sorgt, ausgeführt wird.

Danach wird eine While-Schleife ausgeführt. Innerhalb dieser Schleife wird jedes Mal zuerst die Methode *holeWebsiteAusDatenbank(id)* ausgeführt. Dieser Befehl holt immer die nächste zu durchsuchende Website aus der Datenbank. Im Anschluss daran wird die Methode *durchsucheWebsite* aufgerufen. Diese Methode durchsucht dann Schritt für Schritt jede interne Webseite dieser Website. In letzterer Methode wird wiederum eine While-Schleife ausgeführt. Innerhalb dieser While-Schleife wird jedes Mal zuerst die als Nächste zu durchsuchenden Webseite mit dem Methodenaufruf *holeWebseiteAusDatenbank(id)* aus der Datenbank geholt. Mit dem Methodenaufruf *CrawlThread.start()* wird dann ein Thread der Klasse *CrawlThread* gestartet, welcher die Seite nach Links durchsucht und diese in der Datenbank speichert.

Es gilt zu beachten, dass die Methodennamen in Abbildung 28 im Vergleich zu ihren Bezeichnungen im Programm, etwas abgeändert sind, um verständlicher zu sein.

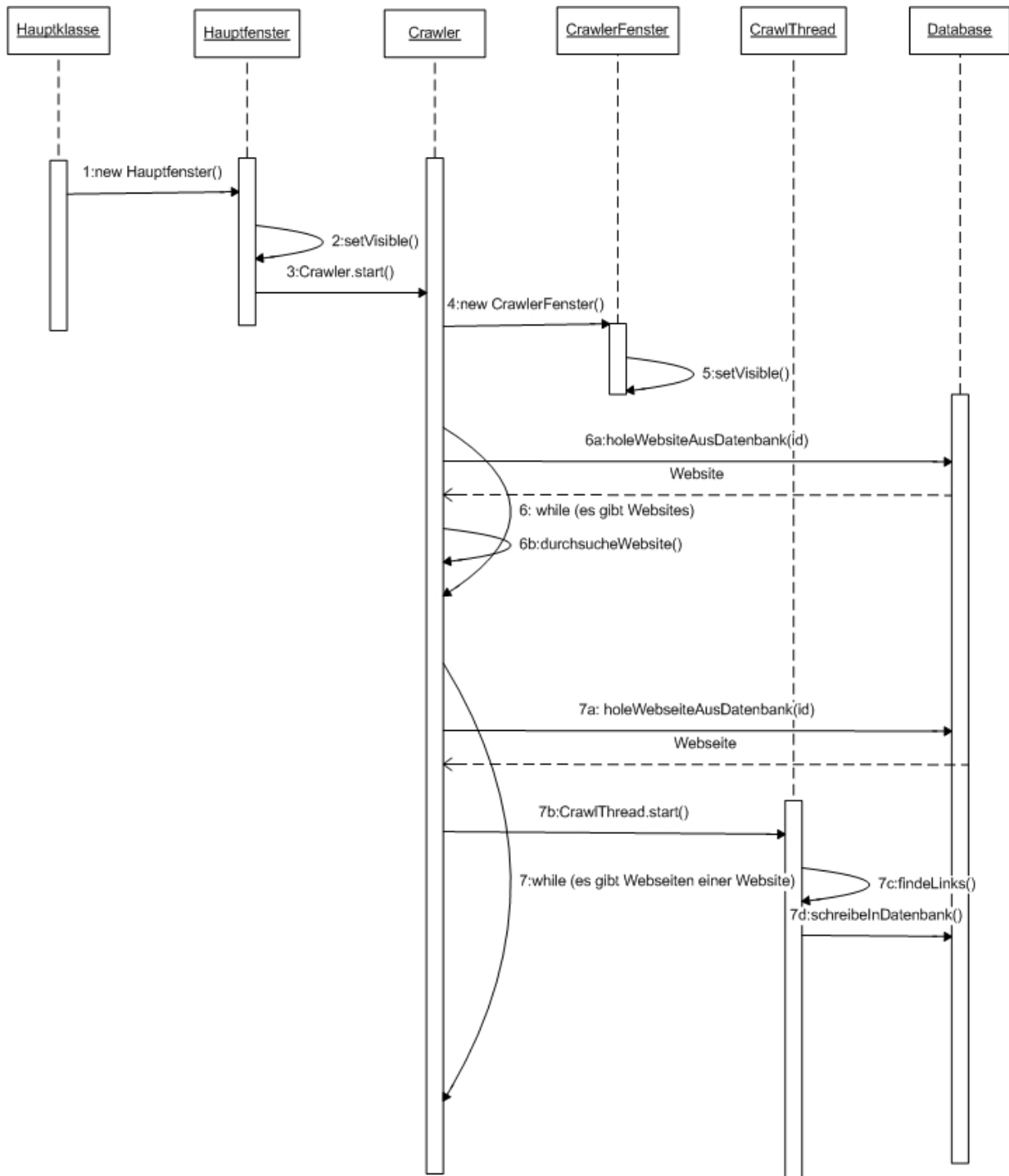


Abbildung 28: Sequenzdiagramm des Crawlers

## 5.2 Transformation der Daten

Nachdem die Daten mit dem Crawler gesammelt wurden und – wie in Abschnitt 4.4 beschrieben – in die Datenbanktabellen geschrieben wurden, müssen sie für das Analysetool umgewandelt werden. Zur Verfügung standen die bekannten Tools für die Analyse von Netzwerken UCINET [85] und Pajek [106]. Die Auswahl fiel auf UCINET, da dessen Bedienung benutzerfreundlicher ist als jene von Pajek. Allerdings hat UCINET die Restriktion, dass die Matrix in einem Datensatz maximal 256 Zeilen und Spalten haben darf, wenn er zum Beispiel aus einer Excel-Datei importiert wird. Das heißt, dass das Netzwerk maximal 256 Knoten haben darf. Um diese Restriktion zu umgehen, wird aus den Daten mit

Hilfe des frei erhältlichen Tools txt2pajek [59] eine *.net* Datei für Pajek erzeugt und diese in UCINET importiert. In diesem Fall gibt es keine Begrenzung der Anzahl der Knoten.

Zuerst müssen die Daten in bestimmter Form in eine Textdatei geschrieben werden. Das Tool txt2pajek kann diese Textdatei dann in eine Pajek-Datei umwandeln.

### 5.2.1 Transformation der Daten in eine Textdatei

Zuerst müssen die Daten in einen Vector geschrieben werden. Für jede Website wird ein Vektor erzeugt, der die Anzahl der eingehenden und ausgehenden Links sowie die Anzahl der durchsuchten Webseiten dieser Website enthält. Die Vektoren aller Websites werden wiederum in einen Vektor gespeichert. Dieser Vorgang ist als Pseudocode in Listing 4 dargestellt.

```
public Vector ermittleDaten() {
    Vector liste = new Vector();
    for (int i=0; i<anzahlWebsites; i++) {
        Vector datensatz = new Vector();
        anzahlInLinks(i) = HoleAnzahlInLinks(i);
        anzahlOutLinks(i) = HoleAnzahlOutLinks(i);
        anzahlWebseiten(i) = HoleAnzahlWebseiten(i);
        datensatz.add(anzahlInLinks(i));
        datensatz.add(anzahlOutLinks(i));
        datensatz.add(anzahlWebseiten(i));
        liste.add(datensatz)
    }
    return liste;
}
```

#### Listing 4: Pseudocode zur Übertragung der Daten aus der Datenbank in einen Vektor

Als Nächstes muss der Vektor sortiert werden, damit nur ein paar Knoten für die Darstellung im Netzwerk herangezogen werden können und so die einzelnen Knoten im Netzwerk noch identifiziert werden können. In diesem Fall werden nur die Knoten mit den meisten eingehenden Links oder jene mit den meisten ausgehenden Links dargestellt.

Das Sortieren ist im folgenden Pseudocode aus Listing 5 dargestellt. Die Methode *ermittleMaximum(Vector liste)* bestimmt jeweils aus einer Liste von Elementen das Maximum und gibt dieses als Rückgabewert zurück. Danach wird dieses maximale Element auch gleichzeitig aus der Ausgangsliste entfernt. Der Rückgabewert dieser Methode ist zuerst das Element mit dem höchsten Wert für die ein- oder ausgehenden Links, dann jenes mit dem zweithöchsten Wert (weil das mit den höchsten bereits entfernt wurde), dann jenes mit dem dritthöchsten Wert und so weiter. In dieser Reihenfolge werden die Elemente dann einem sortierten Vektor hinzugefügt. Um bei einem großen Netzwerk Speicher zu sparen wird dieser Vektor in die Datenbank geschrieben, nachdem das Sortieren abgeschlossen ist.

```

int zaehler = liste.size();
Vector sortierteListe = new Vector();

for (int i=0; i<zaehler; i++) {
    max = ermittleMaximum(liste);
    sortierteListe.add(max);
}

public Vector ermittleMaximum(Vector liste) {
    int zaehler = liste.size();
    int max = 0;
    for (int i=0; i<zaehler; i++) {
        if (sortiereNach==inLinks) {
            if (max<liste(i)) {
                max = i;
            }
        }
        else if (sortiereNach==outLinks) {
            if (max<liste(i)) {
                max = i;
            }
        }
        else if (sortiereNach==inoutLinks) {
            if (max<liste(i)) {
                max = i;
            }
        }
    }
    remove max from liste;
    return max;
}

schreibeSortierteListeInDatenbank();

```

**Listing 5: Pseudocode zur Sortierung der Daten nach der Anzahl eingehender, ausgehender oder ein- und ausgehender Links**

Nach diesem Schritt können die Daten in eine Textdatei geschrieben werden, wobei die Textdatei folgende Form haben muss:

```

Startknoten, Zielknoten
Startknoten, Zielknoten
Startknoten, Zielknoten
.....
Startknoten, Zielknoten

```

In dieser Form entspricht jede Zeile einer Kantenverbindung im Netzwerk. Das Schreiben in die Textdatei geschieht wie in Listing 6 dargestellt.

Für jede Website wird folgende Prozedur durchgeführt: Ermittle die Anzahl der Datensätze für diese Website in der Tabelle mit den ausgehenden Links, ermittle den Namen der aktuellen Website und ermittle alle ausgehenden Links für diese Website. Für jede Relation wird dann eine Zeile in die Textdatei geschrieben.

```

public void schreibeTextDatei() {
    for (int i=0; i<anzahlWebsites; i++) {
        int anzahlDatensaetze = ermittleAnzahlOutLinks(i);
        int aktuelleWebsite = ermittleWebsiteNamen(i);
        Vector outLinks = ermittleOutLinksVonWebsite(i);
        for (int j=0; j<anzahlVectorOutLinks; j++) {
            schreibeInTextDatei (aktuelleWebsite,outLinks[j]);
        }
    }
}

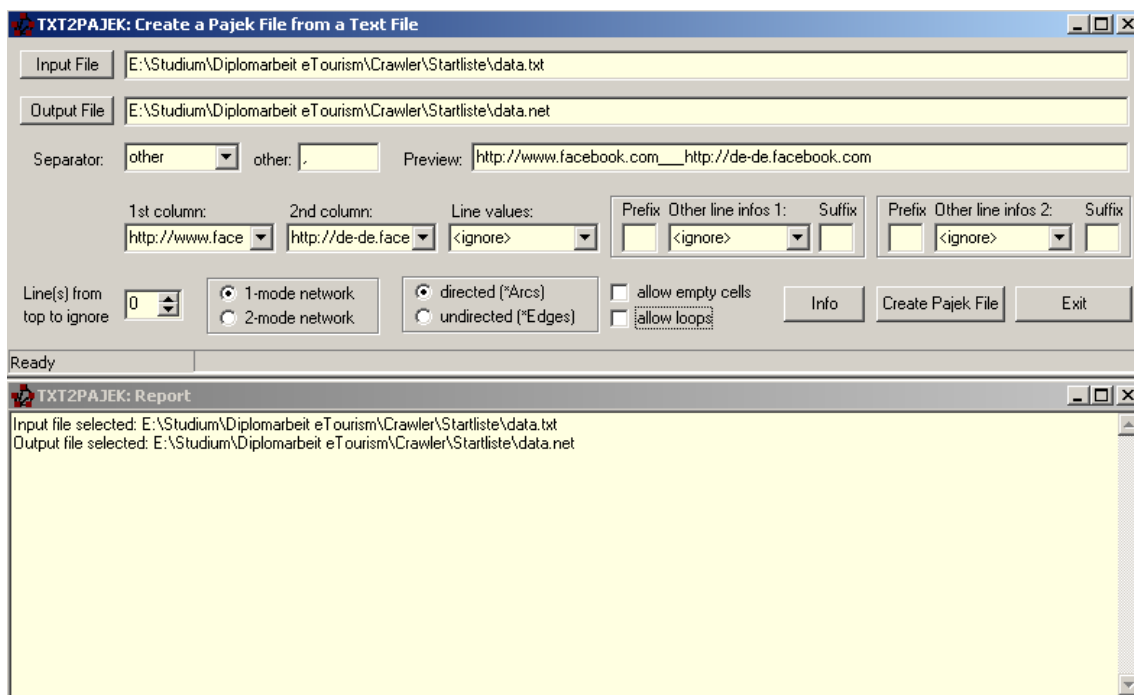
```

**Listing 6: Pseudocode für das Schreiben der Daten in die Textdatei**

## 5.2.2 Konvertierung der Textdatei in eine Pajek-Datei

Wenn die Daten in Form einer Textdatei vorliegen, können sie mit Hilfe eines externen Tools in eine *.net* Datei für Pajek umgewandelt werden. Das dafür zuständige Tool heißt, wie bereits erwähnt, txt2pajek und ist frei erhältlich. Die Referenz dazu ist in [59] angegeben; die Oberfläche des Programms ist in Abbildung 29 dargestellt.

Zur Oberfläche in Abbildung 29: Neben dem Knopf „Input File“ wird die umzuwandelnde Textdatei eingegeben. Darunter, neben dem Knopf „Output File“, wird die Datei, in welche die Textdatei umgewandelt werden soll, angegeben.



**Abbildung 29: Das Programm txt2pajek**

Bei „Separator“ kann ausgewählt werden, welches Trennzeichen zwischen zwei Knoten in der Textdatei steht. In diesem Fall ist das Trennzeichen ein Beistrich und es wird „other“ ausgewählt, weil der Beistrich nicht als Trennzeichen vordefiniert ist. Im Textfeld daneben muss deshalb ein Beistrich eingetragen werden. Eine wichtige Einstellung ist auch, ob es sich um ein 1-mode oder ein 2-mode-Netzwerk handelt. In diesem Fall geht es immer um ein 1-mode Netzwerk. Was genau das ist, wurde bereits in Abschnitt 3.2.2 beschrieben. Falls es sich um ein 1-mode-Netzwerk handelt, ist auch anzugeben, ob es ein gerichtetes oder ein ungerichtetes Netzwerk ist (directed, undirected). Hier handelt es sich natürlich um ein gerichtetes Netzwerk. Zusätzlich kann noch gewählt werden, wie viele Zeilen von ganz oben



ignoriert werden sollen (Line(s) from top to ignore), ob leere Zellen zugelassen werden (allow empty cells) und ob Verbindungen von einem Knoten zu sich selbst zugelassen werden sollen (allow loops).

Durch einen Klick auf den Knopf „Create Pajek File“ wird die Datei erzeugt

### 5.3 Import der Daten

Die Daten liegen nun in Form einer *.net* Datei für Pajek vor. Zum Import der Daten in UCINET muss im Menü Data > Import text file > Pajek gewählt werden. Dieser Vorgang wird in Abbildung 30 dargestellt.

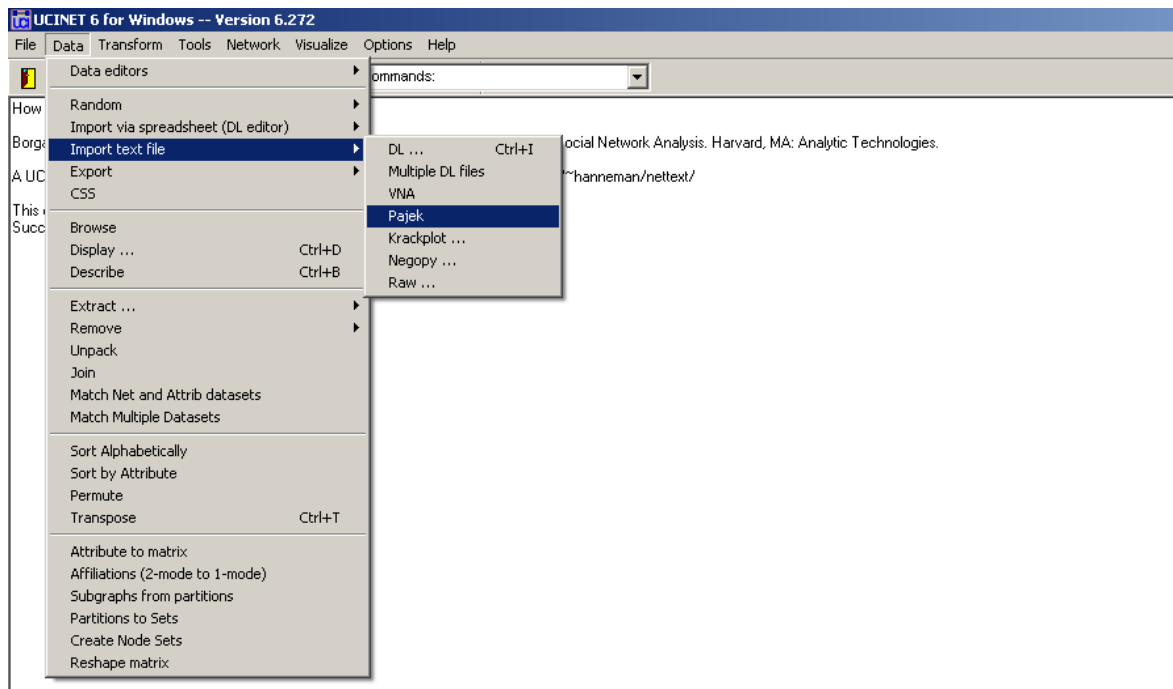


Abbildung 30: Fenster zum Import der Pajek-Datei in UCINET

Daraufhin wird der in Abbildung 31 dargestellte Dialog angezeigt.

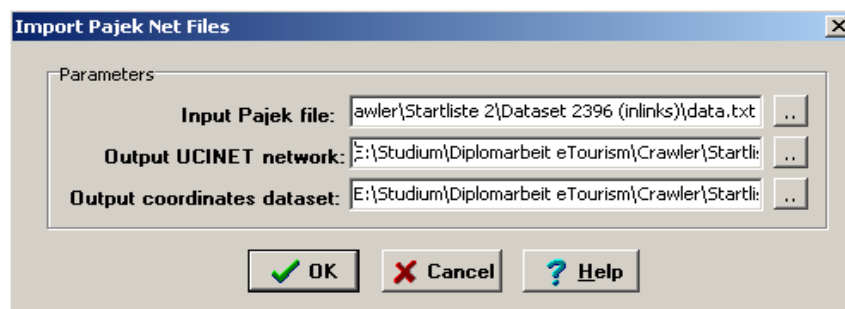


Abbildung 31: Fenster zur Wahl des Ortes wo die zu importierende Datei liegt

Dazu werden im Dialog, der in Abbildung 31 dargestellt ist, im Feld „Input Pajek file“ die zu importierende Datei und im Feld „Output UCINET network“ Ort und Name der erstellten UCINET-Datei angegeben. Zusätzlich kann im Feld „Output coordinates dataset“ der Dateiname für die Speicherung der Koordinaten des Netzwerks angegeben werden. Falls der Import der Datei erfolgreich war, wird eine Textdatei unter anderem mit den Knoten des Netzwerks angezeigt. Ein Beispiel einer solchen Datei ist in Abbildung 32 dargestellt.

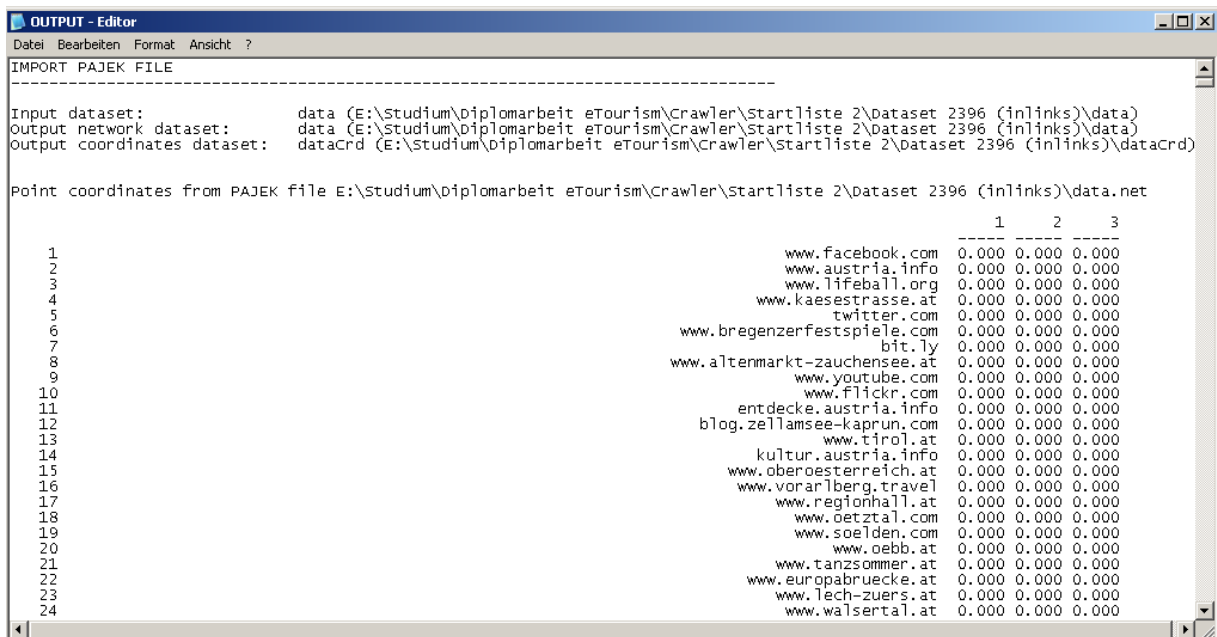


Abbildung 32: Textdatei nach erfolgreichem Import einer Pajek-Datei in UCINET

## 5.4 Darstellung der Daten

In diesem Abschnitt wird beschrieben, wie die Daten des Netzwerks graphisch dargestellt werden. Die Daten werden mit dem Programm NetDraw dargestellt, das in UCINET integriert ist. Um zu NetDraw zu gelangen, wird – wie in Abbildung 33 zu sehen ist – in UCINET im Menü Visualize > NetDraw gewählt.

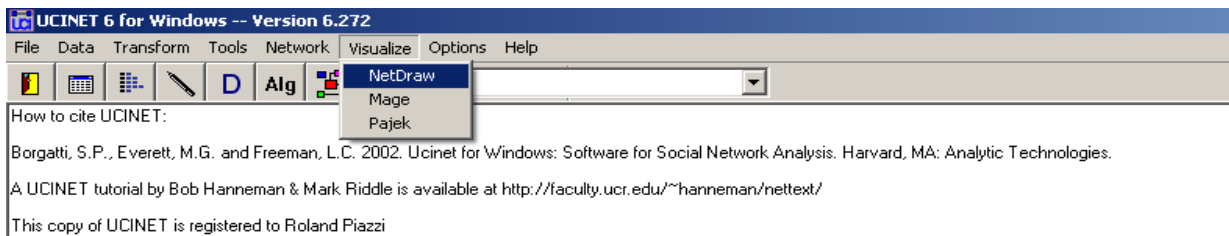


Abbildung 33: Darstellung des Aufrufs von NetDraw in UCINET

NetDraw sieht dann so aus, wie in Abbildung 34. Um ein Netzwerk darzustellen, muss im Menü File > Open > UCINET Data File > Network gewählt werden.

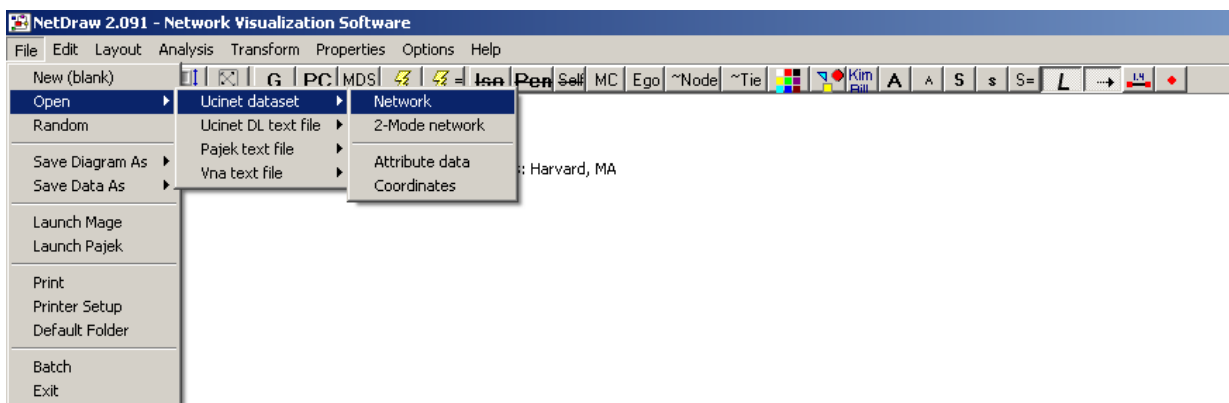
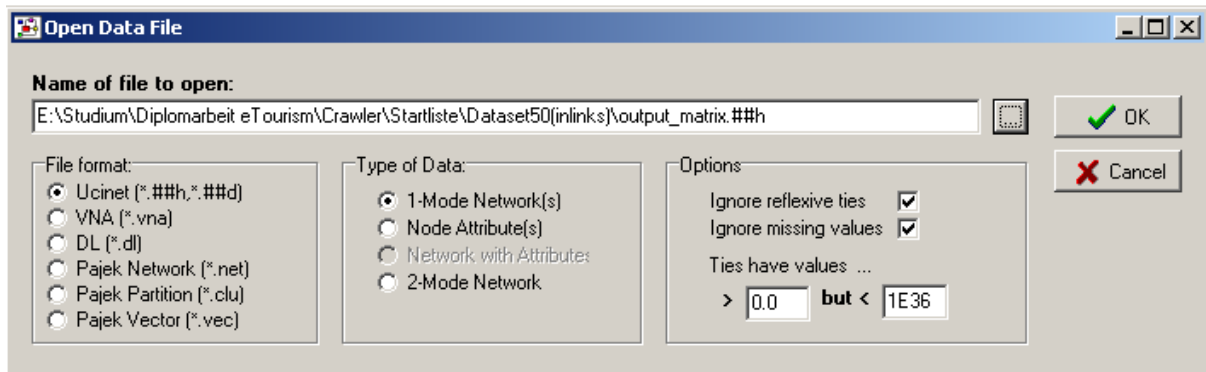


Abbildung 34: Öffnen einer UCINET-Datei in NetDraw

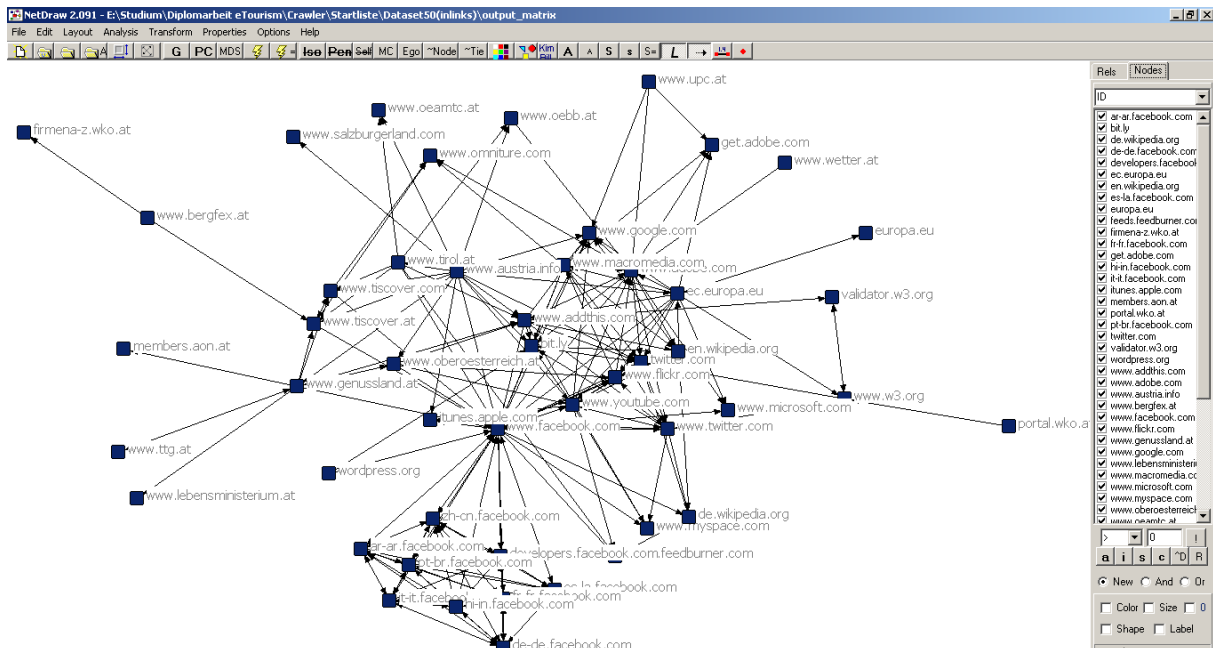
In Abbildung 34 ist zu sehen, dass es in NetDraw auch möglich ist, ein Netzwerk aus einer Pajek-Datei darzustellen. Leider kommt es bei der Auswahl der in dieser Arbeit erzeugten Pajek-Datei immer zu einer Fehlermeldung, weshalb die Pajek-Datei zuvor in UCINET importiert und als UCINET-Datensatz vorhanden sein muss.

Nach dem Klick auf den Menüeintrag „Network“ in Abbildung 34 erscheint der folgende, in Abbildung 35 dargestellte Dialog.



**Abbildung 35: Fenster in NetDraw zur Wahl der zu öffnenden UCINET-Datei**

Durch Klicken auf den Knopf links neben dem Knopf „OK“ kann die zu öffnende Datei ausgewählt werden. Links darunter wird das Format der Datei ausgewählt. In diesem Fall handelt es sich um eine UCINET-Datei. Daneben wird ausgewählt, welche Art von Netzwerk dargestellt werden soll. Es handelt sich in diesem Fall, wie bereits in Abschnitt 3.2.2 erklärt wurde, um ein 1-mode-Netzwerk. Die beiden Checkboxes unter „Options“, die rechts dargestellt sind, sowie die zwei Felder unter „Ties have values ...“ bleiben wie voreingestellt. Mit einem Klick auf den Knopf „OK“ wird das Netzwerk dargestellt. Abbildung 36 zeigt eine beispielhafte Darstellung.



**Abbildung 36: Beispielhafte Darstellung eines Netzwerks in NetDraw**

In NetDraw ist es dann möglich, Veränderungen an der Darstellung des Netzwerks durchzuführen. Für diese Arbeit speziell ist nur wichtig, auf welcher Position die Knoten

angeordnet werden und wie die Größe der Knotensymbole und Knotenbeschriftungen nach bestimmten Kriterien verändert werden kann. Die Position ist deshalb wichtig, weil Knoten mit einem hohen Zentralitätswert in Bezug auf eine bestimmte Zentralitätskennzahl auch optisch zentral erscheinen sollten. Das heißt sie sollten sich im Zentrum des Netzwerks befinden. Zusätzlich ist es wichtig die Knotensymbole und die Knotenbeschriftungen in verschiedenen Größen darzustellen um auch so Knoten mit höheren Zentralitätswerten herauszuheben.

Die am besten geeignete Methode zur Bestimmung der Anordnung bzw. der Position der Knoten, die auch in NetDraw implementiert ist, ist die so genannte Spring-embedding-Methode. Diese Methode ordnet die Knoten so an, dass die Länge der geodätischen Distanz zwischen zwei Knoten deren Entfernung im Netzwerk bestimmt. Die Methode ist iterativ. Sie ermittelt die geodätische Distanz zwischen zwei Knotenpaaren, ermittelt dann die Distanz im Netzwerk und rückt die Knoten näher zusammen, die eine geringe geodätische Distanz zueinander haben. Dieser Vorgang wird wiederholt so oft der Benutzer will – je öfter, desto genauer fällt die Darstellung letztendlich aus. In NetDraw wird diese Darstellung durch Klick im Menü auf Layout > Graph-theoretic Layout > Spring embedding erreicht. Es erscheint dann ein Fenster, in dem verschiedene Einstellungen zur Methode gemacht werden können. Im Feld „Layout criteria“ wird „Distances only“ gewählt. Im Feld „Starting positions“ wird „Current positions“ gewählt. Im Feld „No. Of Iterations“ wird 500 eingefügt. Das Feld „Distance between components“ bleibt bei 5 wie voreingestellt. Schließlich wird auch die Voreinstellung im Feld „Proximities“ beibehalten.

Die Größe der Knoten kann von Attributen der Knoten bestimmt werden. Da die Knoten noch keine Attributwerte besitzen, müssen erst welche hinzugefügt werden, wobei sich die Zentralitätsmaße dazu eignen. Dazu wird im Menü der Punkt Analysis > Centrality measures gewählt. Dann erscheint ein Fenster mit Auswahlmöglichkeiten. In diesem werden am besten alle voreingestellten Zentralitätsarten als Attribute für die Knoten übernommen. Jedoch muss bei der Richtung die gerichtete Version gewählt werden. Die restlichen Einstellungen können beibehalten werden. Die Attributswerte werden dann den Knoten zugefügt. Anschließend müssen die Attributswerte gespeichert und wieder geöffnet werden. Wenn kein Netzwerk geöffnet ist, muss zuerst ein Netzwerk gewählt werden und danach können zusätzlich die Attributswerte für das Netzwerk geöffnet werden. Wenn das erledigt ist, muss im Menü Properties > Nodes > Symbols > Size > Attribute-base gewählt werden. Im darauf folgenden Dialog wird gewählt, welches Attribut die Größe des Knotens bestimmen soll. In diesem Fall wird entweder In- oder OutDegree für den Ein- oder den Ausgangsgrad gewählt. Die restlichen Einstellungen bleiben gleich. Danach entspricht die Knotengröße dem Ein- oder Ausgangsgrad des Knotens. Dasselbe wird mit den Bezeichnungen (Labels) der Knoten gemacht. Dazu muss im Menü auf der Punkt Properties > Nodes > Labels > Size > Attribute-based gewählt werden. Es erscheint wiederum ein Fenster, in dem jenes Attribut gewählt werden muss, das die Größe der Bezeichnung bestimmt. Hier wird dasselbe wie bei den Knoten (entweder der Ein- oder der Ausgangsgrad) gewählt. Die restlichen Einstellungen bleiben gleich.

## **5.5 Berechnung der Kennzahlen**

Die Zentralitätskennzahlen werden in UCINET berechnet. Nachdem die Daten dort importiert worden sind, können verschiedenen Berechnungen zum Netzwerk durchgeführt werden. Im Folgenden wird beschrieben wie speziell für diese Arbeit die Zentralitätskennzahlen und ein paar zusätzlich wichtige Kennzahlen berechnet werden.

## 5.5.1 Berechnung Grad-basierter Zentralität

Für die Berechnung der Ein- und Ausgangsgrade wird im Menü Network > Centrality > Degree gewählt. Der dazugehörige Bildschirm ist in Abbildung 37 dargestellt.

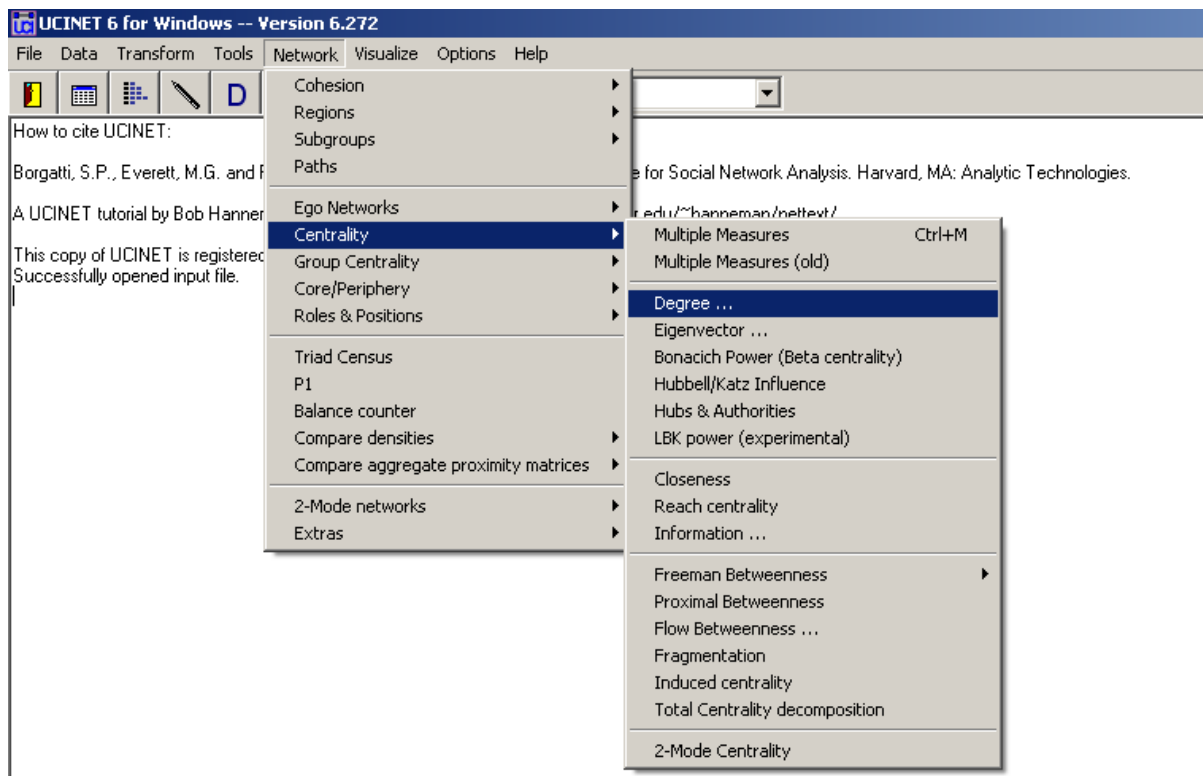


Abbildung 37: Menüauswahl bei der Berechnung des Ein- und Ausgangsgrades

Danach wird der in Abbildung 38 dargestellte Dialog angezeigt. In diesem Dialog muss der Datensatz, aus welchem die Kennzahlen berechnet werden sollen, ausgewählt werden. Zusätzlich kann bestimmt werden, ob es sich um symmetrische Daten handelt und ob die Werte in der Diagonale der Matrix auch einbezogen werden sollen. In diesem Fall handelt es sich weder um eine symmetrische Matrix, da der darunter liegende Graph gerichtet ist, noch werden die Diagonalwerte der Matrix einbezogen. Auch kann im Feld „Output dataset“ der Name der Datei in welchem die durchgeführten Berechnungen, in diesem Fall die Gradberechnungen, gespeichert werden, eingegeben werden.

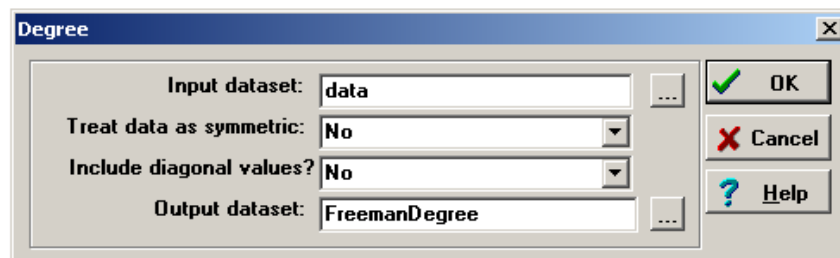


Abbildung 38: Wahl des Datensatzes bei der Berechnung des Ein- und Ausgangsgrades

Durch einen Klick auf den Knopf „OK“ wird das Ergebnis in einer Textdatei angezeigt. Ein Beispiel dafür wird in Abbildung 39 gezeigt.

OUTPUT.LOG7 - Editor

Datei Bearbeiten Format Ansicht ?

FREEMAN'S DEGREE CENTRALITY MEASURES

-----

Diagonal valid? NO  
 Model: ASYMMETRIC  
 Input dataset: data (E:\Studium\Diplomarbeit eTourism\Crawler\Startliste\data)

		1	2	3	4
		OutDegree	InDegree	NrmOutDeg	NrmInDeg
1378	http://www.austria.info	880.000	70.000	10.574	0.841
2586	http://www.alpenstadt.com	793.000	33.000	9.529	0.397
2251	http://www.ausflugstipps.at	440.000	26.000	5.287	0.312
613	http://www.addthis.com	416.000	167.000	4.999	2.007
49	http://www.youtube.com	361.000	355.000	4.338	4.266
1	http://www.facebook.com	270.000	616.000	3.244	7.402
3758	http://www.nationalparksaustria.at	239.000	3.000	2.872	0.036
2007	http://www.naturparke.at	187.000	13.000	2.247	0.156
1423	http://www.wsoe.at	164.000	4.000	1.971	0.048
2452	http://www.golf-alpin.at	158.000	17.000	1.899	0.204
1133	http://ec.europa.eu	157.000	74.000	1.887	0.889
1708	http://www.genussland.at	151.000	40.000	1.814	0.481
5672	http://www.urlaubsempfehlung.at	128.000	2.000	1.538	0.024
2462	http://www.bergfex.at	128.000	34.000	1.538	0.409
4950	http://www.packages.at	124.000	5.000	1.490	0.060
1418	http://www.mondial.at	123.000	21.000	1.478	0.252
1393	http://www.tripsbytips.de	123.000	7.000	1.478	0.084
50	http://bit.ly	115.000	72.000	1.382	0.865
1024	http://www.oebb.at	107.000	116.000	1.286	1.394
1239	http://www.ec.europa.eu	102.000	1.000	1.226	0.012
3772	http://www.urlaubstreif.at	94.000	25.000	1.130	0.300
1382	http://blog.austria.info	77.000	22.000	0.925	0.264
4584	http://www.urlauburlaub.at	69.000	3.000	0.829	0.036
1650	http://www.golf.at	67.000	20.000	0.805	0.240
1546	http://www.tin209.at	62.000	10.000	0.745	0.120

Abbildung 39: Darstellung der Ergebnisse der Ein- und Ausgangsgrade

Die in Abbildung 39 dargestellte Tabelle wird nun kurz erklärt: Die Werte in der Spalte mit der Bezeichnung „OutDegree“ entsprechen der Anzahl der ausgehenden Links eines Knotens. Jene in der Spalte mit der Bezeichnung „InDegree“ entsprechen der Anzahl der eingehenden Links eines Knotens. Die Werte in der Spalte mit der Bezeichnung „NrmOutDeg“ stehen für den Prozentsatz, der von diesem Knoten ausgehenden Links im Vergleich zur Anzahl der Links, die es geben würde, wenn alle Knoten des Netzwerks mit diesem Knoten verbunden wären. Die Werte in der Spalte mit der Bezeichnung „NrmInDeg“ sind ebenfalls eine Prozentzahl und können analog dazu berechnet werden. Dies entspricht dem relativen Ein- bzw. Ausgangsgrad, welcher in Abschnitt 3.2.4.1 beschrieben wurde, multipliziert mit 100.

### 5.5.2 Berechnung der Nähe-basierten Zentralität

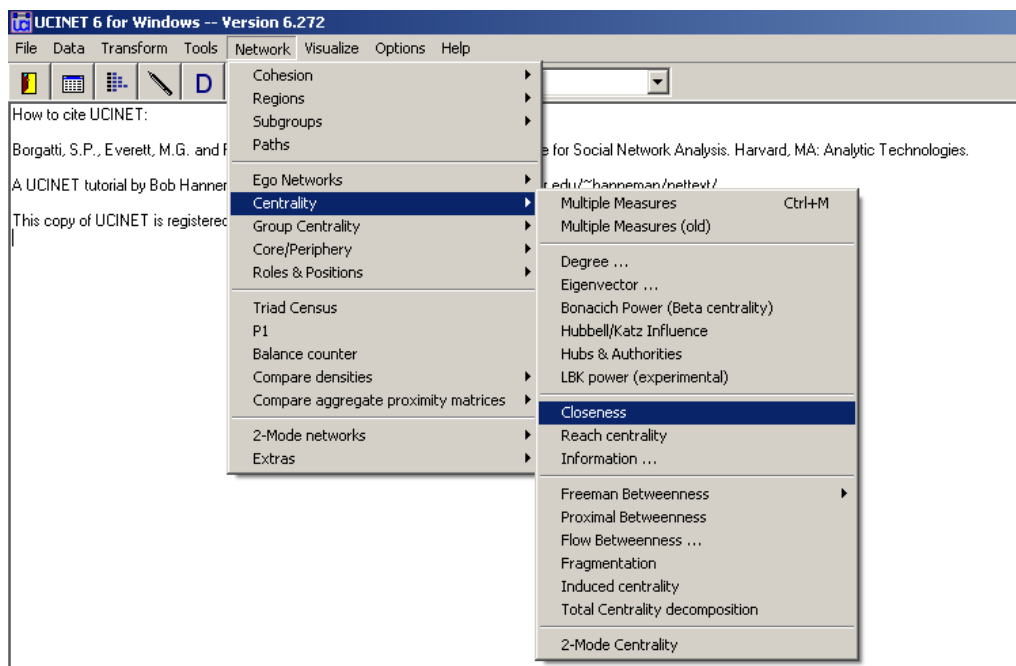


Abbildung 40: Menüauswahl zur Berechnung der Nähe-basierten Zentralität

Um die Nähe-basierte Zentralität zu berechnen, wird im UCINET-Menü Network > Centrality > Closeness gewählt. (Siehe Abbildung 40)

Danach wird wieder – gleich wie in Abschnitt 5.5.1 – bei der Berechnung des Ein- und Ausgangsgrades, ein Dialog zur Auswahl des Datensatzes angezeigt (siehe Abbildung 38). In diesem Dialog muss auch die Art und Weise der Berechnung der Nähe-basierten Zentralität ausgewählt werden. Dies geschieht im Untermenü „Type“. Hier muss normalerweise „Sum of geodesic distances (Freeman)“ gewählt werden, wenn die Berechnung, so wie in Abschnitt 3.2.4.2 beschrieben, als Summe der geodätischen Distanzen nach Freeman (der bereits in Abschnitt 3.2.4 genannt wird) erfolgen soll. Für das Netzwerk des österreichischen E-Tourismus-Marktes muss allerdings die etwas abgeänderte Variante vom Typ „Sum of reciprocal distances“ gewählt werden. Mehr zu dieser Variante und warum sie gewählt wurde wird bei den Ergebnissen für die Nähe-basierte Zentralität in Abschnitt 6.3.2 beschrieben.

The screenshot shows a text editor window titled 'OUTPUT.LOG6 - Editor' with a menu bar (Datei, Bearbeiten, Format, Ansicht, ?). The text content includes file paths and a table of results. The table has four columns: 'inCloseness', 'outCloseness', 'NinCloseness', and 'NoutCloseness'. The rows list various websites and their corresponding values.

	1 inCloseness	2 outCloseness	3 NinCloseness	4 NoutCloseness
www.facebook.com	943.332	1151.668	39.470	48.187
www.tiscover.at	838.081	1166.668	35.066	48.815
www.youtube.com	834.648	964.922	34.923	40.373
www.oebb.at	816.382	720.241	34.158	30.136
www.tiscover.com	742.930	961.484	31.085	40.229
twitter.com	731.148	1005.836	30.592	42.085
www.austria.info	730.032	1698.335	30.545	71.060
www.addthis.com	727.581	789.490	30.443	33.033
www.oberoesterreich.at	721.165	1190.824	30.174	49.825
www.tiro1.at	708.647	1441.868	29.650	60.321
www.flickr.com	700.099	774.909	29.293	32.423
de.wikipedia.org	698.949	874.155	29.245	36.576
www.linz.at	673.482	1086.331	28.179	45.453
www.oamt.c.at	662.998	770.992	27.741	32.259
maps.google.at	659.332	0.000	27.587	0.000
www.bahn.de	658.498	959.589	27.552	40.150
www.alpenverein.at	655.532	895.903	27.428	37.485
www.innsbruck.info	649.932	942.317	27.194	39.427

Abbildung 41: Darstellung der Ergebnisse für die Nähe-basierten Zentralität

Die Ergebnisse der Nähe-basierten Zentralität werden wie immer in einer Textdatei dargestellt. Ein Beispiel dafür ist in Abbildung 41 zu sehen. In der Tabelle dieser Abbildung sind die Werte in den Spalten 1 und 2, jeweils für die eingehende und die ausgehende Richtung, die Summen der reziproken Werte der Längen der kürzesten Wege (geodätische Distanzen) zwischen einem Knoten und allen anderen. Die Werte in den Spalten 3 und 4 sind die dazugehörigen in einer Prozentzahl normierten (relativen) Werte. Wie letztere Werte zu interpretieren sind wird in der Hilfe von UCINET nicht erwähnt. Wichtig hierbei ist, dass die Knoten welche in Bezug auf die Nähebasierte Zentralität am zentralsten liegen, bei dieser Berechnung auch die höchsten Werte erhalten.

### 5.5.3 Berechnung der Zwischenzentralität

Für die Berechnung der Zwischenzentralität muss im Menü Network > Centrality > Freeman Betweenness > Node Betweenness gewählt werden. Dieser Vorgang wird wie in Abbildung 42 angezeigt.

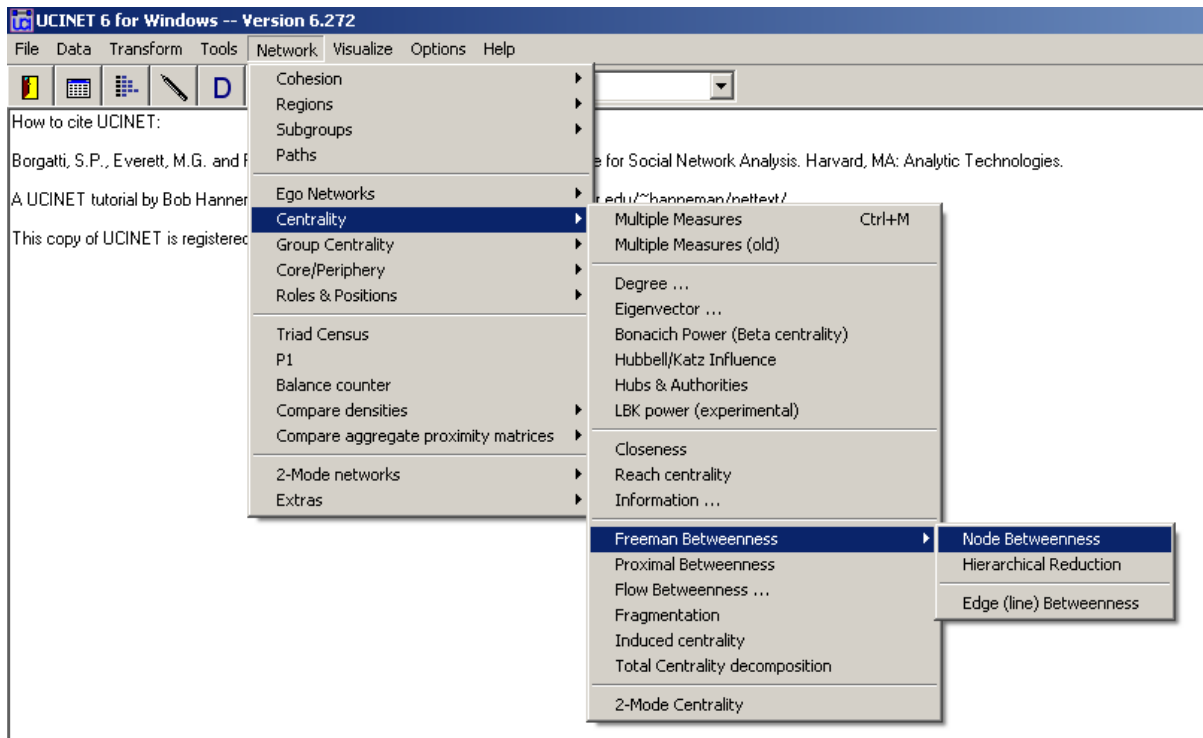


Abbildung 42: Menüauswahl zur Berechnung der Zwischenzentralität

Bei Auswahl des Menüeintrags „Node Betweenness“ erscheint – wie in den beiden vorigen Kapiteln auch schon beschrieben wurde – wieder ein Dialog, in welchem der Datensatz zur Berechnung ausgewählt werden muss. Anschließend erscheinen die Ergebnisse wieder in einer Textdatei. Ein Beispiel dafür ist in Abbildung 43 zu sehen.

	Betweenness	Normalized Betweenness
2	1374966.750	24.081
1	849694.500	14.882
181	780058.750	13.662
13	771091.875	13.505
67	519618.969	9.101
15	311591.813	5.457
9	146201.297	2.561
5	127363.025	2.231
16	126151.922	2.209
78	118708.758	2.079
55	72790.969	1.275
73	71827.984	1.258
137	70202.977	1.230
61	67826.844	1.188
126	57305.004	1.004
998	56598.594	0.991
105	50626.115	0.887
104	50483.258	0.884
60	49380.027	0.865
127	49345.816	0.864
109	48573.902	0.851

Abbildung 43: Darstellung der Ergebnisse für die Zwischenzentralität

Die Werte in Spalte 1 sind die absoluten Werte und die Werte in Spalte 2 die normalisierten (relativen) Werte für die Zwischenzentralität, so wie sie üblicherweise verwendet werden und wie sie auch in Abschnitt 3.2.4.3 beschrieben wurden.



## 5.5.4 Berechnung der Dichte

Zur Berechnung der Dichte wird im Menü Network > Cohesion > Density > Old density procedure gewählt. Im Programm sieht das wie in Abbildung 44 aus.

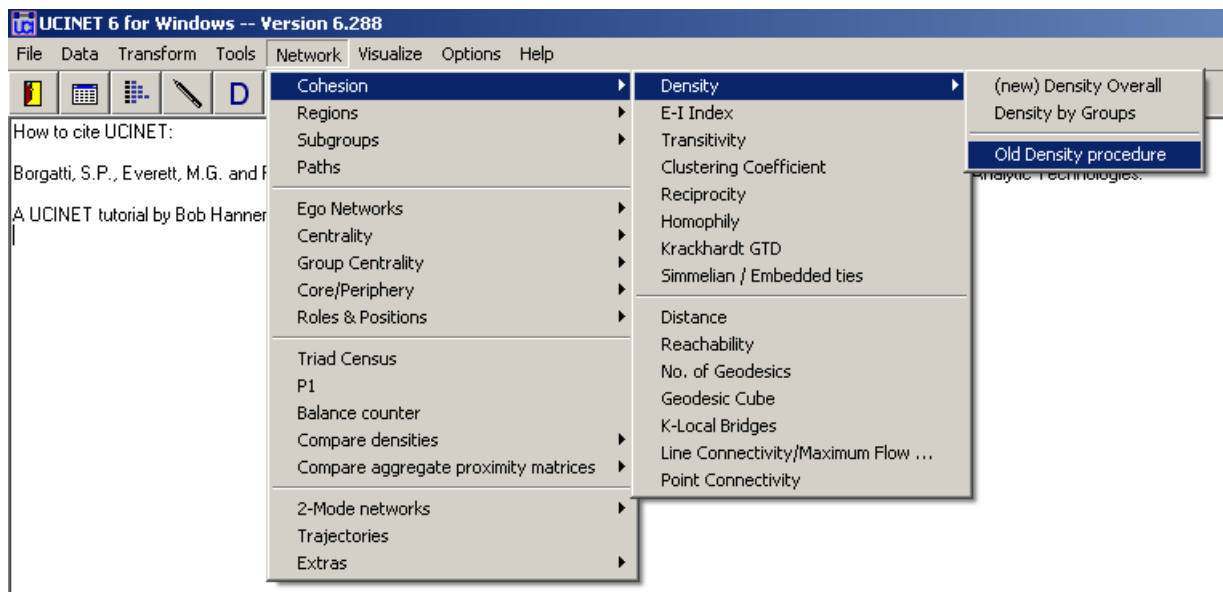


Abbildung 44: Menüauswahl bei der Berechnung der Dichte

Danach wird ein Dialog angezeigt, in dem der Datensatz, für den die Dichte berechnet werden soll, ausgewählt werden muss. Die zusätzlichen Einstellungen, die in diesem Dialog möglich sind, bleiben wie voreingestellt. Das Ergebnis der Dichte erscheint wie bei den vorherigen Berechnungen wieder in einer Textdatei, welche in Abbildung 45 dargestellt ist.

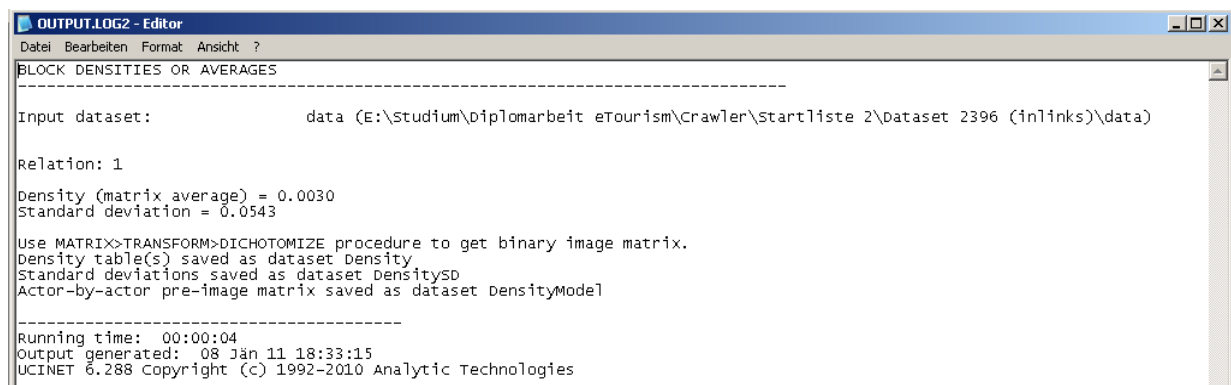
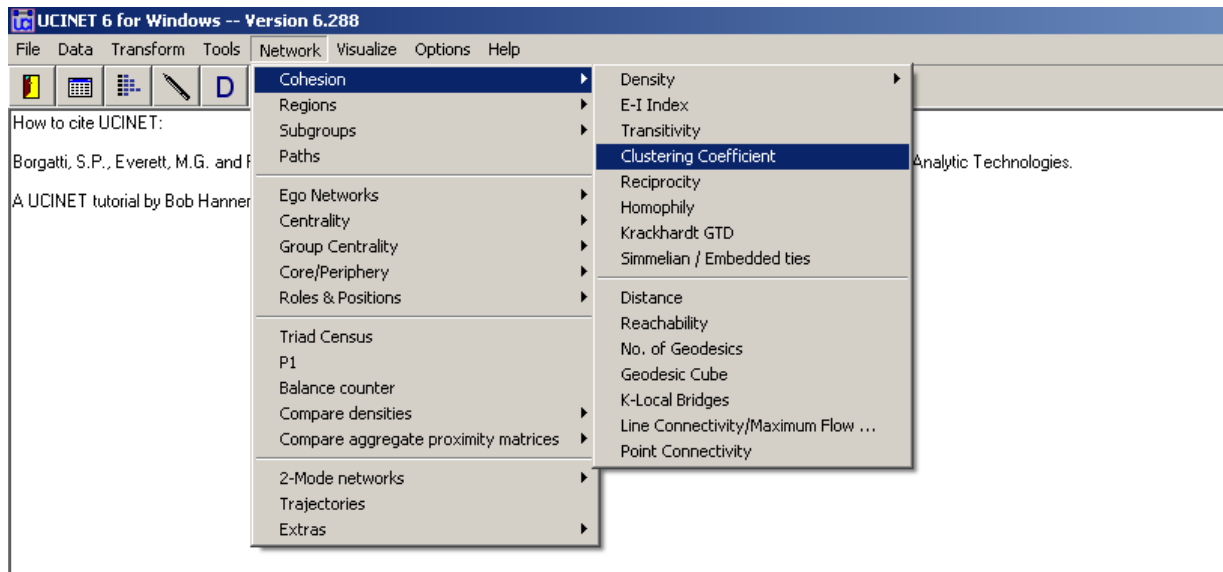


Abbildung 45: Ergebnis für die Dichte des Netzwerks

Die Dichte des Netzwerks steht in der Textdatei die in Abbildung 45 dargestellt ist rechts neben „Density (matrix average)“ und beträgt in diesem Fall 0,0030. Die Berechnung entspricht der Dichteberechnung aus Abschnitt 3.1.2.

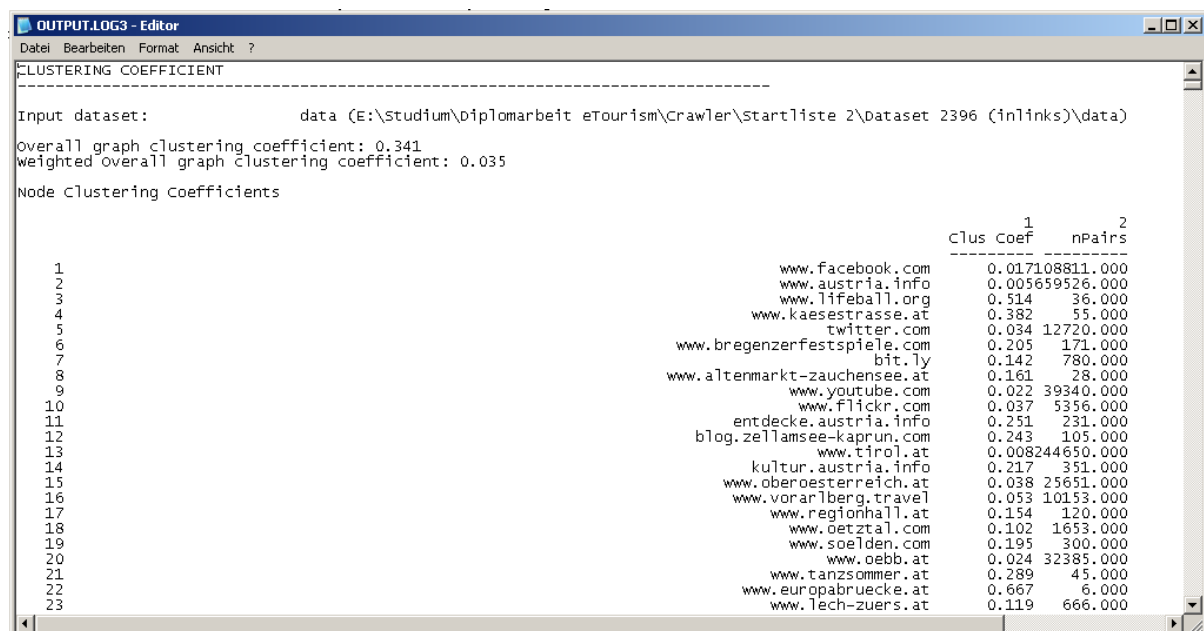
## 5.5.5 Berechnung des Clusterkoeffizienten

Zur Berechnung des Clusterkoeffizienten wird im Menü Network > Cohesion > Clustering Coefficient gewählt. (Siehe Abbildung 46)



**Abbildung 46: Menüauswahl bei der Berechnung des Clusterkoeffizienten**

Nach dem Klick auf den Menüeintrag „Clustering Coefficient“ erscheint, gleich wie bei den vorherigen Berechnungen, ein Dialog, in welchem der Datensatz, für welchen der Clusterkoeffizient berechnet werden soll, ausgewählt werden muss. Die restlichen Einstellungen bleiben wie voreingestellt. Nach der Berechnung erscheint eine Textdatei (siehe Abbildung 47) mit dem Ergebnis, welches den Clusterkoeffizienten für jeden Knoten sowie den durchschnittlichen Clusterkoeffizienten für das gesamte Netzwerk zeigt.



**Abbildung 47: Darstellung der Ergebnisse für den Clusterkoeffizienten des Netzwerks**

Der Clusterkoeffizient für das gesamte Netzwerk steht in der Textdatei in Abbildung 47 neben „Overall graph clustering coefficient“ und beträgt in diesem Fall 0,341. Die Berechnung erfolgt wie in Abschnitt 3.2.4.4 angegeben. Zusätzlich dazu wird noch der gewichtete Clusterkoeffizient des gesamten Netzwerkes bei der Beschriftung „Weighted overall graph clustering coefficient“ berechnet. Dies entspricht dem Clusterkoeffizienten für das gesamte Netzwerk, wobei bei dessen Berechnung der Clusterkoeffizient eines jeden Knotens mit dem Knotengrad gewichtet wird. Darunter sind die Clusterkoeffizienten für jeden Knoten in Spalte

1 angegeben. Was die Zahlen in Spalte 2 aussagen wird aus der Hilfe von UCINET nicht deutlich. Es steht zwar, dass dies angeblich der Knotengrad sein soll, was aber nicht stimmt. Für diese Arbeit wichtig ist aber nur der Clusterkoeffizient für das gesamte Netzwerk.

### 5.5.6 Berechnung der durchschnittlichen Weglänge

Um die durchschnittliche Weglänge innerhalb des Netzwerks zu berechnen muss im Menü Network > Cohesion > Distance gewählt werden. Dies ist in Abbildung 48 dargestellt.

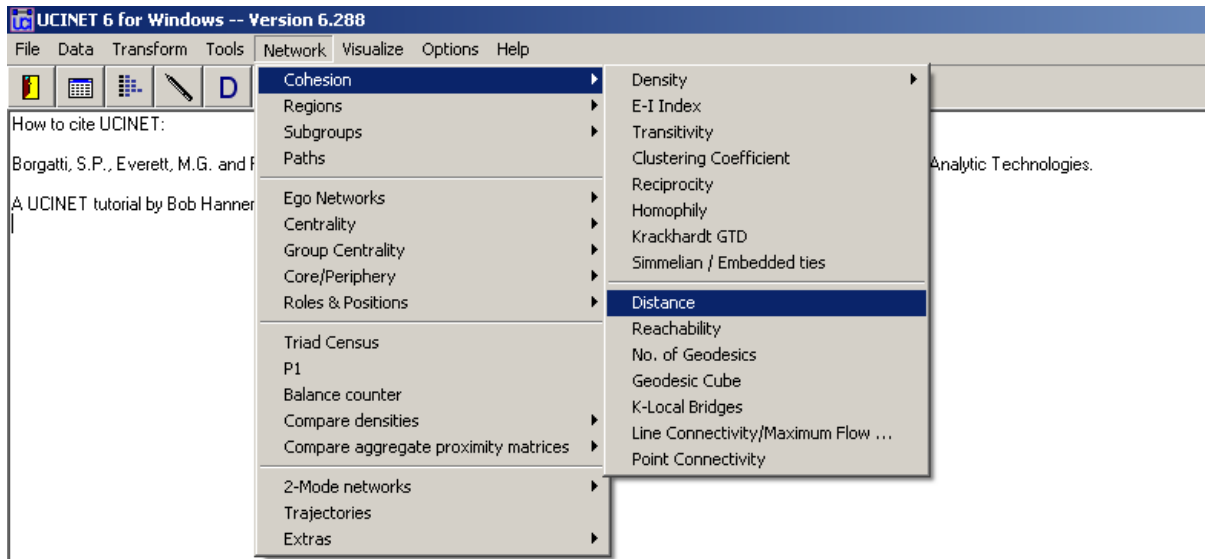


Abbildung 48: Menüauswahl bei der Berechnung der durchschnittlichen Weglänge

Danach muss in einem Dialog der gewünschte Datensatz, für den die durchschnittliche Weglänge berechnet werden soll, ausgewählt werden. Alle anderen Einstellungen in diesem Dialog bleiben für dieses Netzwerk wie voreingestellt.

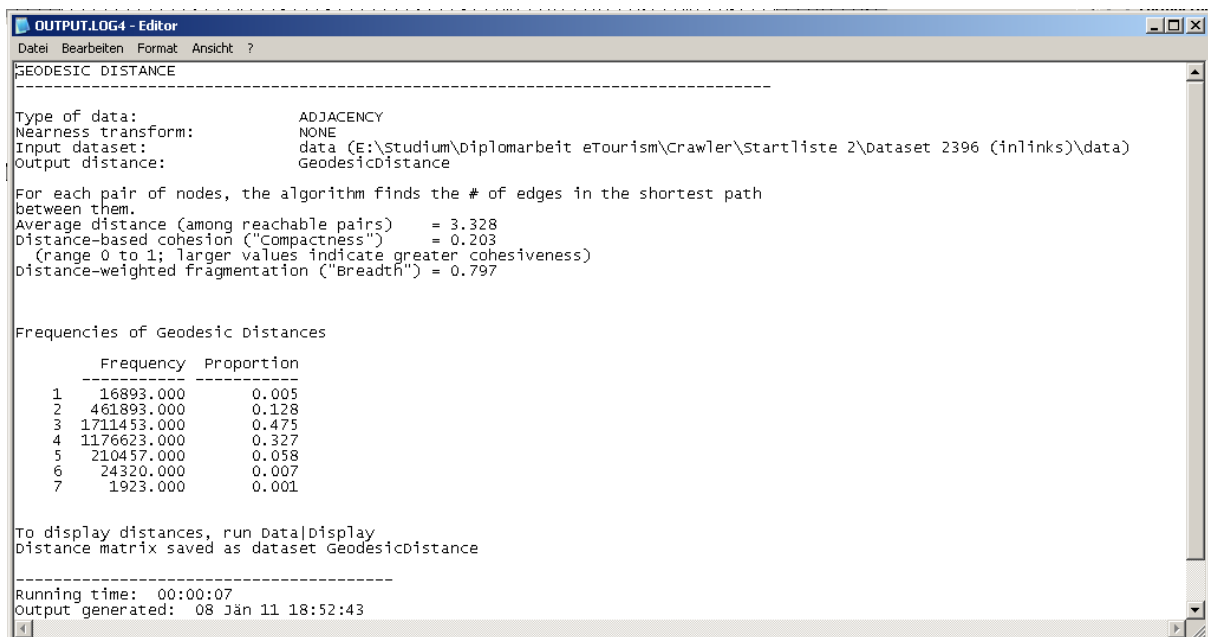


Abbildung 49: Ergebnis für die durchschnittliche Weglänge

Das Ergebnis ist wiederum in einer Textdatei (siehe Abbildung 49) dargestellt und liefert neben der durchschnittlichen Weglänge aller erreichbaren Knoten auch eine Frequenz und eine Proportion der unterschiedlichen Weglängen.

Die durchschnittliche Weglänge steht in Abbildung 49 neben „Average distance (among reachable pairs)“ und beträgt in diesem Fall 3,328. Die Frequenzen und Proportionen für jede Weglänge sind in der kleinen Tabelle dargestellt. In diesem Fall bedeutet dies zum Beispiel für die Weglänge 3, dass 47,5 % aller kürzesten Wege im Netzwerk die Länge 3 haben. Die Frequenz ist die Summe aller kürzesten Wege oder geodätischen Distanzen mit dieser Länge. Die restlichen Ergebnisse die in Abbildung 49 dargestellt sind werden in der Hilfe von UCINET nicht weiter erklärt und sind für diese Arbeit auch nicht von Bedeutung.

## 6 Die Ergebnisse

In diesem Kapitel werden die Ergebnisse dieser Arbeit präsentiert. Zuerst wird versucht die Qualität des Crawlers anhand der erhaltenen Daten zu überprüfen. Zusätzlich werden die Ergebnisse eines externen Crawlers für eine Qualitätsbeurteilung herangezogen. Dies deshalb um die Ergebnisse des entwickelten Crawlers zu unterstreichen. Dafür wurde der am ehesten für eine derartige Analyse geeignete Crawler herangezogen. Trotzdem wird dabei festgestellt werden, dass dieser externe Crawler weder für die Untersuchung des österreichischen E-Tourismus-Marktes geeignet gewesen wäre, noch für einen Vergleich der Ergebnisse mit dem selbst entwickelten Crawler taugt, weil dessen Ergebnisse einerseits andersartig als die des eigenen Crawlers waren und andererseits, aufgrund unzureichender Dokumentation, auch nicht eindeutig interpretiert werden konnten. Am Ende dieses Abschnittes über die Qualität des Crawlers wird auch die Qualität der vier durchgeführten Suchdurchläufe ermittelt. Dabei wird jenes von den vier Suchdurchläufen erzeugte Netzwerk ermittelt, welches die Struktur des österreichischen E-Tourismus am besten widerspiegelt, um anhand dieses Netzwerks die notwendigen Analysen durchführen zu können.

Danach werden die Netzwerkdiagramme für das ausgewählte Netzwerk in graphischer Form präsentiert und beurteilt. Anschließend werden zunächst die Werte für die drei wichtigen Zentralitätskennzahlen, die für dieses Netzwerk berechnet wurden, aufgelistet und beurteilt. Danach wird ein Vergleich mit einem Netzwerk der Websites der Tourismusanbieter der Insel Elba in Italien sowie dem Netzwerk des World Wide Webs anhand wichtiger charakteristischer Eigenschaften von Netzwerkmodellen sowie der Fliegenstruktur des World Wide Webs, durchgeführt.

Um die zentralen Knoten des Netzwerks ermitteln zu können, wird in Abschnitt 6.4 versucht die für die Analyse eines Teilnetzwerks des World Wide Webs relevanteste Zentralitätskennzahl herauszufinden um die zentralen Knoten des Netzwerks bestimmen zu können, da die drei Zentralitäten Zentralität in unterschiedlicher Weise definieren und deswegen auch unterschiedliche Ergebnisse liefern die nicht miteinander kombinierbar sind. Hierbei wird mit der, von zum Beispiel einem Transport- oder Personennetzwerk unterschiedlichen Natur eines Netzwerks des World Wide Webs, sowie mit zusätzlichen wichtigen Kennzahlen (Web Impact Factor und HITS) zur Bewertung von Webseiten im World Wide Web argumentiert. In Abschnitt 6.5 werden die nach der relevantesten Zentralitätskennzahl ermittelten zentralen Knoten des Netzwerks beschrieben und Gründe dafür angegeben, weshalb diese Knoten an zentraler Stelle stehen könnten. Zum Schluss werden in Abschnitt 6.6 noch ein paar zusätzliche Ergebnisse und Beobachtungen angeführt, welche in den vorherigen Analysen nicht zur Geltung gekommen sind.

### **6.1 Vergleiche zur Beurteilung der Crawler-Qualität und den Suchdurchläufen**

In diesem Abschnitt wird zuerst versucht, die Qualität des Crawlers durch einen Vergleich der Ergebnisse zu beurteilen. Danach werden die Ergebnisse dieses Crawlers mit jenen eines externen Crawlers verglichen. Trotzdem, dass der Crawler selbst entwickelt wurde, war ein Vergleich mit einem externen Crawler wünschenswert um die Ergebnisse des eigenen Crawlers zu unterstreichen, auch wenn sich dies als nicht möglich herausstellen sollte. Am Ende dieses Abschnittes werden die vier Suchdurchläufe und deren Ergebnisnetzwerke miteinander verglichen, um das am besten auf den österreichischen E-Tourismus-Markt zutreffende Netzwerk zu ermitteln.

### 6.1.1 Beurteilung der Crawler-Qualität

Die Qualität des Crawlers kann dadurch beurteilt werden, dass die Ergebnisse zweier verschiedener Suchdurchläufe miteinander verglichen werden, wobei es um die jeweils ermittelte Anzahl der ausgehenden Links geht. Dazu müssen die Geschwindigkeit des Crawlers und die maximale Anzahl von zu durchsuchenden Webseiten pro Website gleich sein. Die Geschwindigkeit, welche durch die Wartezeiten der Threads gesteuert wird (siehe Abschnitt 4.5.2), muss deswegen gleich sein, weil wenn sie zum Beispiel zu hoch ist, schafft es der Thread nicht den gesamten Such- und Speichervorgang durchzuführen, wird deswegen zu früh abgebrochen, weil ein neuer Thread benötigt wird und somit werden nicht alle Webseiten durchsucht bzw. gespeichert, weswegen das Ergebnis unvollständig sein kann. Auch die maximale Anzahl von zu durchsuchenden Webseiten pro Website muss bei den verglichenen Suchdurchläufen gleich sein, weil bei einer unterschiedlichen Anzahl mehr oder weniger ausgehende Links gefunden werden können. Da diese aber zur Qualitätsbeurteilung miteinander verglichen werden, muss die maximale Anzahl an zu durchsuchenden Webseiten gleich sein, um den Vergleich aussagekräftig zu machen. Alternativ zum Einsatz von Wartezeiten wäre vielleicht eine Vergrößerung des Thread-Pools möglich um genug Threads zu schaffen, sodass kein Thread abgebrochen werden muss. Dies konnte aber in dieser Arbeit nicht angewandt werden, da der PC auf dem der Crawler lief zu wenig Speicher dafür hatte.

Diese Voraussetzungen waren bei den Suchdurchläufen 1 und 2 gegeben. Wenn nun die Anzahl der ausgehenden Links derjenigen Websites, welche in beiden Suchdurchläufen durchsucht wurden, gleich oder ähnlich ist, so arbeitet der Crawler korrekt und seine Qualität kann als gut eingeschätzt werden. Der Vergleich wurde anhand von 300 gleichen Websites in beiden Suchdurchläufen durchgeführt – das dabei erzielte Resultat ist in Abbildung 50 in Form eines Diagramms dargestellt.

In Abbildung 50 ist zu erkennen, dass die Anzahl der gefundenen ausgehenden Links je Website zwar nicht immer gleich groß ist, die beiden Kurven insgesamt aber doch ziemlich ähnlich sind.

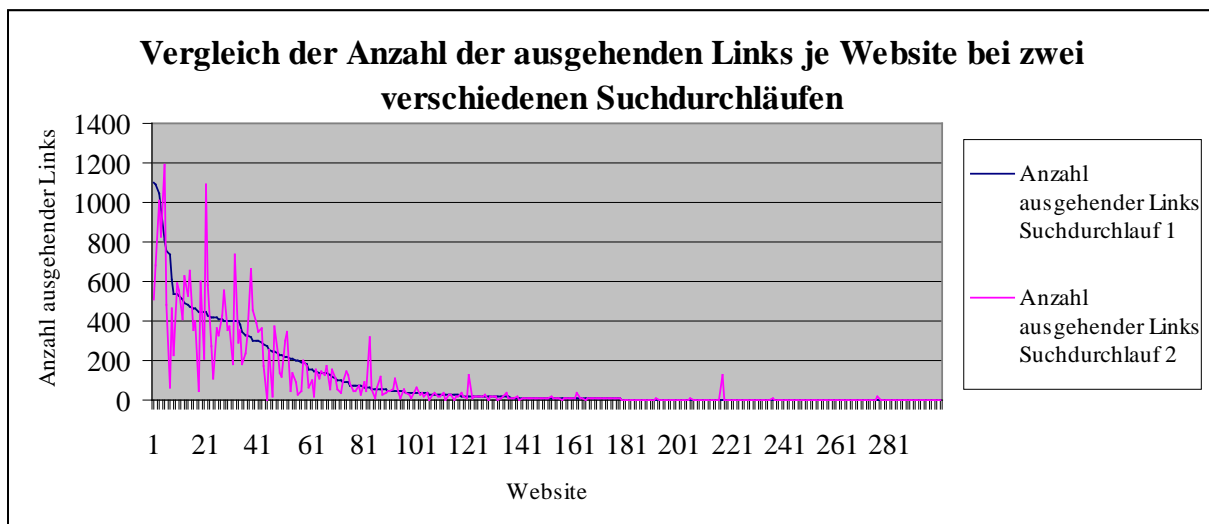
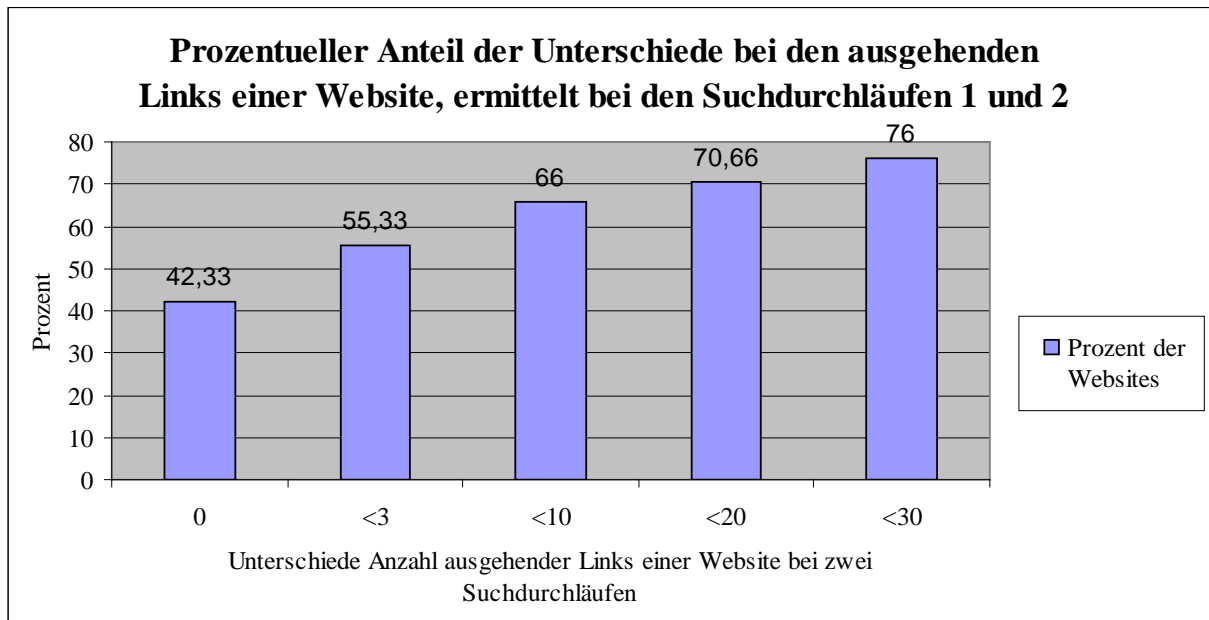


Abbildung 50: Vergleich der Anzahl der ausgehenden Links je Website in zwei Suchdurchläufen

Die durchschnittliche Anzahl der ausgehenden Links beläuft sich bei Suchdurchlauf 1 auf 98,63 und bei Suchdurchlauf 2 auf 89,11. Das bedeutet, dass bei Suchdurchlauf 1 im Durchschnitt mehr ausgehende Links pro Website gefunden wurden als bei Suchdurchlauf 2. Die durchschnittliche Differenz der ausgehenden Links je Website beträgt ungefähr 39. Hierbei wurde die Differenz der Anzahl der ausgehenden Links je Website gebildet, diese

Differenzen wurden für alle verglichenen Websites zusammen gezählt und dann der Durchschnitt gebildet (im Gegensatz zur Differenz 98,63 und 89,11 welche der Differenz des Durchschnitts entspricht). Diese Zahl scheint verhältnismäßig hoch und würde auf eine eher schlechte Qualität des Crawlers schließen lassen. Jedoch muss dazu gesagt werden, dass die meisten verglichenen Websites einen geringen Unterschied aufweisen und die durchschnittliche Differenz aufgrund von ein paar sehr hohen Differenzen derart hoch ist. Dies ist auch in Abbildung 51 zu erkennen.



**Abbildung 51: Prozentueller Anteil der Unterschiede (Differenzen) bei den ausgehenden Links einer Website, ermittelt bei den Suchdurchläufen 1 und 2**

Abbildung 51 zeigt die prozentuellen Anteile der Unterschiede bei den ausgehenden Links einer Website welche bei den zwei verschiedenen Suchdurchläufen 1 und 2 ermittelt wurden. Bei 42,33% der Websites war kein Unterschied bei der Anzahl der ausgehenden Links festzustellen. Bei 55,33% war der Unterschied kleiner als 3 ausgehende Links. 66% unterschieden sich in weniger als 10, 70,66% in weniger als 20 und 76% in weniger als 30 ausgehenden Links.

Diese Ergebnisse zeigen, dass die Werte insgesamt sehr ähnlich sind. Daraus kann geschlossen werden, dass der Crawler korrekt funktioniert. Gründe für die voneinander abweichenden Ergebnisse der beiden Suchdurchläufe könnten einerseits der Zeitunterschied zwischen den beiden Suchdurchläufen und eine damit verbundene Veränderung der Websites sein. Bei einer solchen Veränderung können nicht nur ausgehende Links, sondern auch einzelne Webseiten hinzugefügt oder entfernt werden, sodass der Suchdurchlauf des Crawlers beeinflusst wird und deshalb verschiedene Webseiten durchsucht werden. Andererseits können die unterschiedlichen Startadressen der Grund für die verschiedenen Ergebnisse sein.

Warum dies so ist, kann folgendermaßen erklärt werden: Angenommen, es gibt zwei verschiedene Startadressen. Beide zeigen auf die gleiche Ziel-Website, jedoch auf unterschiedliche Webseiten der Website. Zum Beispiel verweist die erste Startadresse auf <http://www.austria.info/hotels/tirol.php> und die zweite auf <http://www.austria.info/hotels/steiermark.php>. Wenn die Website *austria.info* in der Folge durchsucht wird, startet der Crawler jeweils bei einer anderen Webseite der Website. Wenn der Crawler aber bei verschiedenen Webseiten startet, verändert sich sein Suchlauf.

Unterschiedliche Webseiten einer Website werden so durchsucht und deshalb vielleicht auch unterschiedlich viele ausgehende Links gefunden.

Um die gute Qualität des Crawlers zu unterstreichen wurden für Testzwecke zusätzlich zwei eigene Suchdurchläufe – diesmal jeweils mit gleicher Startadresse – nur für eine bestimmte Website direkt hintereinander durchgeführt und die so gefundenen Webseiten sowie die Anzahl der ausgehenden Links miteinander verglichen. Die Startadresse war *http://www.touristik.at*. Bei beiden Suchdurchläufen wurden 2769 Webseiten gefunden. Die Anzahl der gefundenen ausgehenden Links war mit einer minimalen Abweichung von 1 nahezu gleich. In einem Suchdurchlauf wurden 32, im anderen 33 ausgehende Links gefunden. Das heißt der Crawler erzielte bei beiden Suchdurchläufen die gleichen Ergebnisse, da genau gleich viele Webseiten und nahezu gleich viele ausgehende Links gefunden wurden.

### **6.1.2 Vergleich der Ergebnisse mit einem externen Crawler**

Der einzige externe Crawler, der annähernd geeignete Ergebnisse für diese Analyse (zum Beispiel die Anzahl der eingehenden Links) auch in geeigneter Form lieferte, war der so genannte „IssueCrawler“. Dieser Crawler läuft nicht auf dem eigenen Computer, sondern auf einem niederländischen Server und wurde von der nicht profitorientierten Organisation *Govcom.org* entwickelt. Forscher und Interessierte können diesen Crawler kostenlos benutzen. [60] In diesem Unterkapitel werden die Ergebnisse des „IssueCrawlers“ mit jenen des eigenen Crawlers verglichen und es wird gezeigt, warum sich der „IssueCrawler“ nicht gut eignet, um das Netzwerk des österreichischen E-Tourismus-Marktes zu finden.

Der „IssueCrawler“ führt eine so genannte „Co-Link-Analyse“ durch. Dabei startet der Crawler zum Beispiel mit zwei Websites, ermittelt die ausgehenden Links dieser beiden Websites und sucht dabei jene Websites heraus, welche von beiden Startwebsites verlinkt wurden. Für den Vergleich mit dem eigenen Crawler wurden dem „IssueCrawler“ die ersten 100 durchsuchten Websites aus Suchdurchlauf 3 als Startknoten gegeben. Für diese Websites sollte der „IssueCrawler“ eine „Co-Link-Analyse“ durchführen. Parallel dazu wurde auch eine „Co-Link-Analyse“ mit dem eigenen Crawler durchgeführt, wozu ein Programm in Java geschrieben wurde. Aus den Ergebnissen von Suchdurchlauf 3 waren die ausgehenden Links dieser 100 Websites bereits bekannt. Diese mussten in der Folge nur noch miteinander verglichen werden, um Ergebnisse zu erhalten, die mit jenen des „IssueCrawlers“ vergleichbar sind.

Der „IssueCrawler“ liefert zum Beispiel als Ergebnis eine Liste, der durch die „Co-Link-Analyse“ gefundenen Websites, welche die meisten eingehenden Links enthalten. Dies sollten jene Websites sein, die von den eingegebenen Startportalen aus am öftesten verlinkt wurden. Diese Liste enthielt 644 Websites. Anschließend wurde der Vergleich mit den ersten 644 Websites, die durch die „Co-Link-Analyse“ mit dem eigenen Crawler gefunden wurden, gemacht. Von diesen 644 Domains waren 202 Domains in den Ergebnissen beider Crawler enthalten. Die 20 wichtigsten sind in Tabelle 11 aufgelistet.

Wie aus dem Ergebnis und aus der Tabelle 11 ersichtlich ist, sind zwar ein Drittel der bei der „Co-Link-Analyse“ gefundenen Knoten bei beiden Crawlern gleich, jedoch ist die Anzahl der erhaltenen eingehenden Links sehr unterschiedlich. Dies kann vielleicht darauf zurückgeführt werden, dass der IssueCrawler die Links von einer Website zu einer anderen mehrfach zählt, wenn der Link öfters innerhalb der ausgehenden Website und auf verschiedenen Webseiten dieser Website angeführt ist. Der eigene Crawler zählt jedoch die Links von einer Website zur anderen immer einfach. Andererseits heißt die Datei, in der diese Liste abgespeichert ist, „Ranked Actor List by inlink count from total network (by site)“, d.h. die Links würden nicht mehrfach gezählt, denn es gibt gleichzeitig eine Ergebnisdatei die „Ranked Actor List by inlink count from total network (by page)“ heißt. In dieser Datei ist die Anzahl der



eingehenden Links der Websites um einiges höher, was darauf zurückgeführt werden kann, dass eben die Links zwischen zwei Websites mehrfach gezählt werden. Werden die eingehenden Links in der erstgenannten Datei aber nur einfach gezählt, so würde die Behauptung, dass eine „Co-Link-Analyse“ nur für die angegebenen Websites durchgeführt wird, nicht stimmen. In diesem Fall könnte der Knoten *oebb.at* in Tabelle 11 nicht 112 eingehende Links erhalten, wenn nur 100 Websites untersucht würden. Das würde zu dem Schluss führen, dass nicht nur die angegebenen 100 Websites durchsucht werden, sondern auch die durch sie gefundenen zusätzlichen Websites.

Url	Anzahl eingehender Links Eigener Crawler	Anzahl eingehender Links „IssueCrawler“
<i>oebb.at</i>	5	112
<i>addthis.com</i>	6	80
<i>oberoesterreich.at</i>	9	76
<i>austria.info</i>	6	72
<i>steiermark.com</i>	5	58
<i>ttg.at</i>	3	57
<i>tiscover.com</i>	6	49
<i>facebook.com</i>	7	43
<i>salzburgerland.com</i>	4	39
<i>genussland.at</i>	4	38
<i>gesundheitsurlaub.at</i>	3	37
<i>salzburg-airport.com</i>	1	36
<i>radfahren.at</i>	4	35
<i>kultururlaub.at</i>	4	34
<i>ausflugstipps.at</i>	3	34
<i>tagung.info</i>	2	34
<i>biken.at</i>	5	32
<i>wandern.at</i>	4	32
<i>pferdeurlaub.at</i>	3	31
<i>oberoesterreich-tourismus.at</i>	3	31

**Tabelle 11: Vergleich der von den zwei verschiedenen Crawlern (eigener und externer) gefundenen 20 Knoten mit den meisten eingehenden Links**

Der „IssueCrawler“ liefert als Ergebnis auch Netzwerkdiagramme, die ansprechend gestaltet sind. Ein solches Netzwerkdiagramm des „IssueCrawlers“ für die obige Untersuchung ist in Abbildung 52 dargestellt. Es ist jedoch unklar, woraus diese Knoten entnommen worden sind. Es scheint so, als wären dies die Knoten mit den meisten eingehenden Links per Webseite. Doch in dieser Liste ist der Knoten *tiscover.com* einer der größten Knoten, in der Abbildung 52 aber ist dieser Knoten nur sehr klein.

Wahrscheinlich hängt die Knotengröße der erhaltenen eingehenden Links nur von den in diesem Netzwerkdiagramm gezeichneten Knoten ab. In diesem Fall könnte die Darstellung angemessen sein. Es muss allerdings gesagt werden, dass die Abbildung 52 – obwohl darin hauptsächlich Websites von Tourismustreibenden enthalten sind – in keinem Fall einem Netzwerk des österreichischen E-Tourismus entspricht. Denn die Daten, aus denen sie gewonnen wurden, waren nur für den Vergleich mit den Ergebnissen des eigenen Crawlers von Nutzen.

Leider konnte aus den Beschreibungen auf dem Portal des „IssueCrawlers“ keine Klarheit darüber gewonnen werden, wie der „IssueCrawler“ genau funktioniert. Dies ist auch ein

Grund, weshalb sich der „IssueCrawler“ weder für die Analyse der Netzwerkstruktur des österreichischen E-Tourismus-Marktes noch für einen Vergleich mit den Ergebnissen des eigenen Crawlers eignet. Ein weiteres Problem des „IssueCrawlers“ ist zum Beispiel die Begrenzung auf maximal 500 Webseiten pro Website und auf maximal 60000 Webseiten insgesamt bei einem Suchvorgang. Problematisch ist auch, dass von den Betreibern des „IssueCrawlers“ empfohlen wird, es sollten hauptsächlich die Links enthaltenden Webseiten einer Website, welche in den meisten Websites einen eigenen Bereich darstellen (zum Beispiel <http://www.lai.at/links>), als Startadressen angegeben werden. Um möglichst genaue Informationen über die Netzwerkstruktur zu bekommen, sollten aber nicht nur die Links auf den Link-Webseiten der Website, sondern alle externen Links innerhalb einer Website gefunden werden. Es stellt sich also generell die Frage, wie die „Co-Link-Analyse“ beim „IssueCrawler“ eigentlich durchgeführt wird. Werden nur bestimmte Webseiten innerhalb einer Website (zum Beispiel jene mit den Links) angesteuert? Werden – entgegen der Beschreibung – zusätzlich zu den Startadressen weitere Websites durchsucht?



Abbildung 52: Netzwerkdiagramm des „IssueCrawlers“

Außerdem stellt der „IssueCrawler“ auch eine UCINET-kompatible Datei mit Ergebnissen zur Verfügung. Wenn anhand dieses Datensatzes die Grad-basierten Zentralitäten berechnet werden, ist das Ergebnis für den Eingangsgrad bzw. die eingehenden Links wieder ein

anderes als in den beiden Textdateien, welche vorher angesprochen worden sind. Außerdem ist bei diesem Ergebnis die Zentralität des gesamten Netzwerks, welche später auch für die Ergebnisse des eigenen Crawlers besprochen wird, außergewöhnlich gering. Schließlich erwartet der „IssueCrawler“ bereits die Eingabe der zu einer bestimmten Gruppe gehörenden Knoten. Diese müssen aber mit dem Crawler oder andernfalls erst händisch gefunden werden. Auch ist die Anzahl bei der Eingabe dieser Knoten auf 500 begrenzt. Weil der österreichische E-Tourismus-Markt aber aus viel mehr Knoten besteht, wäre dies eine hinderliche Begrenzung.

Aufgrund dieser Unverständlichkeiten und Probleme eignet sich der „IssueCrawler“ weder für die Analyse des Netzwerks des österreichischen E-Tourismus-Marktes noch für einen Vergleich mit den Ergebnissen des eigenen Crawlers.

### 6.1.3 Vergleich der vier Suchdurchläufe

Suchdurchlauf 2 (bestehend 1385 Knoten) mit der Startadresse *tiscover.com* lieferte mit Sicherheit die schlechtesten Ergebnisse. Von den gefundenen Websites, gehören nur wenige zur *.at*-Domain bzw. sind nur wenige Portale von Tourismusanbietern. Der Grund dafür ist, dass *tiscover.com* fast keine Tourismusbetriebe verlinkt, da die Informationen zu diesen Betrieben im Portal integriert sind und die Buchung auf *tiscover.com* erfolgt und nicht über die Seite des Anbieters.

Bei Durchsicht der Ergebnisse von Suchdurchlauf 3, welcher mit den in Anhang 1 angeführten Startadressen ausgeführt wurde, fällt auf, dass auch die Ergebnisse dieses Suchdurchlaufs nicht für die weitere Analyse geeignet sind. Bei diesem Suchdurchlauf wurden zwar die meisten Websites untersucht (8276), jedoch enthält das Ergebnis viele nicht touristische Portale, teils aus dem In- und Ausland, sowie viele touristische Portale, welche aber nicht zum österreichischen Tourismus gerechnet werden können. Dazu gehören vor allem Portale aus den Ländern Deutschland, Schweiz und Italien. Der Grund dafür ist, dass viele der als Startadressen angegebenen Portale nur wenige ausgehende Links haben oder auch Tourismusbetriebe aus dem Ausland vertreten, indem sie zum Beispiel den gesamten Alpenraum bedienen.

Die Suchdurchläufe 1 und 4 sind diejenigen, bei denen das Ergebnisnetzwerk am ehesten einem Netzwerk des österreichischen E-Tourismus-Marktes entspricht. Beide Suchdurchläufe erzeugen, wie bei einer Durchsicht festgestellt wurde, Netzwerke, welche das Netzwerk des österreichischen E-Tourismus-Marktes annähernd gut repräsentieren. Bei beiden Suchdurchläufen wurden in etwa gleich viele Websites durchsucht (2116 bei Suchdurchlauf 1 und 2395 bei Suchdurchlauf 4). Jedoch sind die Ergebnisse von Suchdurchlauf 4 besser zu bewerten. Beide Suchdurchläufe starten zwar beim Portal *austria.info*, jedoch wurden bei Suchdurchlauf 4 weitere Startadressen aus dem Tourismusbereich hinzugefügt. Dies führte sicher zu einem besseren Ergebnis, da *austria.info* nur ca. 1000 Portale verlinkt. Die danach durchsuchten Portale bei Suchdurchlauf 1 sind jene, die mit den in *austria.info* verlinkten Portalen verbunden sind und können sich deshalb schon etwas von der Tourismusdomäne entfernt haben. Bei Suchdurchlauf 4 hingegen sind die Ergebnisse wegen der zusätzlichen Startadressen breiter gefächert und entsprechen dadurch eher einem Netzwerk des österreichischen E-Tourismus-Marktes. Die hier gewählten Startadressen sind das Ergebnis der Optimierung aus den drei vorherigen Suchdurchläufen und bestehen aus Portalen mit vielen ausgehenden Links, welche zu einem großen Teil lokale und regionale Destinationsmanagementportale sind. Durch die vielen ausgehenden Links sind diese Portale zur Auffindung des Netzwerks des österreichischen E-Tourismus-Marktes am besten geeignet. Zudem sind Destinationsmanagementportale eher dazu geeignet als private Portale, da sie möglichst alle Tourismusanbieter vertreten und diese auch eher verlinken. Private

Portale neigen eher dazu, nur bestimmte Anbieter zu vertreten, diese Anbieter in ihr System zu integrieren und auch Buchungen anzubieten, ohne die Anbieter zu verlinken.

Tabelle 12 zeigt eine Zusammenfassung der Ergebnisse der vier Suchdurchläufe. Dabei wurden als Probe 1000 Websites herangezogen und diese in Kategorien unterteilt, welche nach der Tabelle 12 beschrieben werden. Dafür wurde zuerst die Liste der gefundenen Websites, sortiert nach eingehenden Links herangezogen. Dann wurden jeweils 425 Websites mit den wenigsten und 425 Websites mit den meisten eingehenden Links ermittelt und kategorisiert. Die restlichen 150 von 1000 verbleibenden Websites wurden innerhalb der sortierten Liste aus der „Mitte“ heraus genommen und kategorisiert. Dies sollte dazu führen dass die Stichprobe möglichst aussagekräftig ist.

		Suchdurch- lauf 1	Suchdurch- lauf 2	Suchdurch- lauf 3	Suchdurch- lauf 4
<b>Touristische inländische Websites</b>	<i>Anzahl</i>	580	435	372	692
	<i>Prozent</i>	58%	43,5%	37,2%	69,2%
<b>Indirekt touristische inländische Websites</b>	<i>Anzahl</i>	70	67	73	158
	<i>Prozent</i>	7%	6,7%	7,3%	15,8%
<b>Nicht touristische inländische Websites</b>	<i>Anzahl</i>	105	108	278	67
	<i>Prozent</i>	10,5%	10,8%	27,8%	6,7%
<b>Touristische ausländische Websites</b>	<i>Anzahl</i>	22	80	50	18
	<i>Prozent</i>	2,2%	8%	5%	1,8%
<b>Indirekt touristische ausländische Websites</b>	<i>Anzahl</i>	22	22	51	9
	<i>Prozent</i>	2,2%	2,2%	5,1%	0,9%
<b>Nicht touristische ausländische Websites</b>	<i>Anzahl</i>	173	265	151	14
	<i>Prozent</i>	17,3%	26,5%	15,1%	1,4%
<b>Fehlerhafte Websites</b>	<i>Anzahl</i>	28	23	25	42
	<i>Prozent</i>	2,8%	2,3%	2,5%	4,2%
<b>Summe</b>	<i>Anzahl</i>	1000	1000	1000	1000
	<i>Prozent</i>	100%	100%	100%	100%

**Tabelle 12: Vergleich der Ergebnisse der vier Suchdurchläufe durch Unterteilung der Websites in Kategorien**

Die definierten Kategorien enthalten folgende Websites:

- *Touristische inländische Websites*: Hierzu zählen alle Websites deren Unternehmen Tourismus-Akteure (nach Abschnitt 2.3.3) sind und die sich hauptsächlich in Österreich befinden. Von diesen wurden aber nur jene Unternehmen berücksichtigt die sich auf den Tourismus in Österreich konzentrieren. Beispiele sind österreichische Hotels und Gaststätten, Anbieter von Attraktionen, ÖBB, PKW-Leihfirmen, sowie die Desitnationsmanagement-Organisationen. In diese Kategorie fallen auch Websites von österreichischen Destinationsmanagement-Organisationen die sich auf ausländischen Domains befinden.

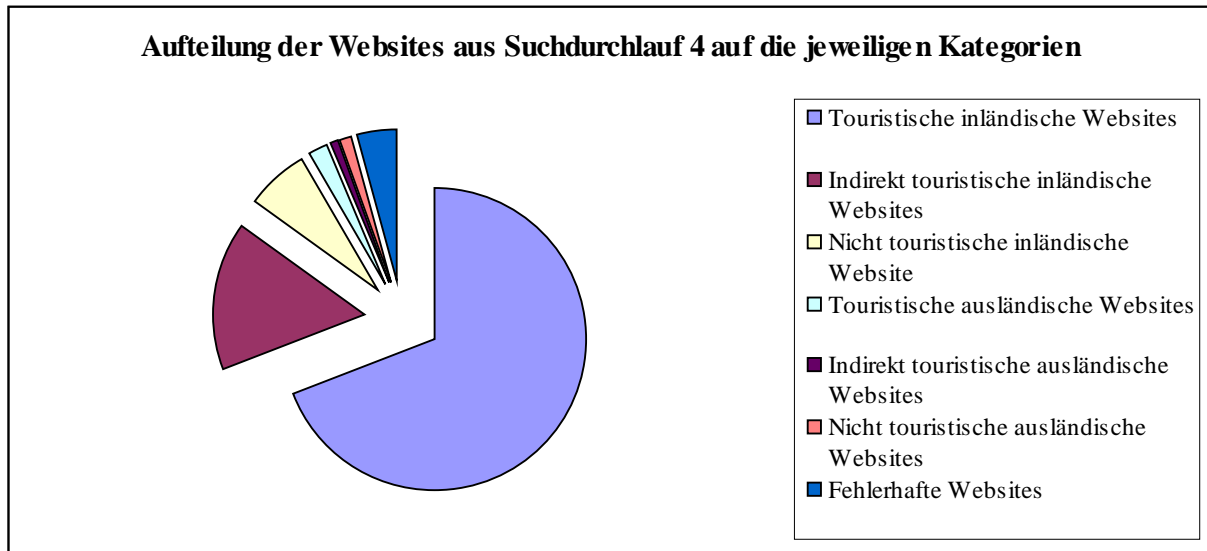
- *Indirekt touristische inländische Websites:* Sind alle Websites deren Inhaber sich in Österreich befinden (hauptsächlich .at-Domains, teilweise .com-Domains), die aber nur indirekt mit dem Tourismus zu tun haben. In diesem Fall wurden hierzu auch die Social Medias gezählt, weil sie indirekt mit dem Tourismus zu tun haben (siehe Abschnitt 3.2.2). Hauptsächlich finden sich in dieser Kategorie Websites von Unternehmen die typische österreichische Produkte (speziell Lebensmittel) anbieten oder Unternehmen die für den Tourismus unterstützende Dienstleistungen oder Produkte anbieten.
- *Nicht touristische inländische Websites:* Dazu gehören alle Websites deren Inhaber sich in Österreich befinden (hauptsächlich .at-Domains, teilweise .com-Domains), die aber nichts mit dem Tourismus zu tun haben. Hier hinein fallen unter anderem die Portale der öffentlich-rechtlichen Körperschaften.
- *Touristische ausländische Websites:* Dies sind all jene Websites deren Inhaber sich im Ausland befinden (größtenteils alle nicht at-Domains) und die hauptsächlich Tourismus im Ausland anbieten oder überwiegend International operieren, wobei die Inhaber zu den Tourismusakteuren gehören (siehe Abschnitt 2.3.3). Beispiele sind: Reiseportale wie HRS, Tripwolf, AccorHotels, Checkfelix oder Fluglinien bzw. Transportunternehmen wie Ryanair, Air Berlin, Lufthansa, Deutsche Bahn.
- *Indirekt touristische ausländische Websites:* In diese Gruppe fallen alle Websites deren Inhaber sich im Ausland befinden (größtenteils alle nicht at-Domains) oder deren Betätigungsfeld sich hauptsächlich auf den Tourismus im Ausland konzentriert, die aber indirekt mit dem Tourismus zu tun haben (zum Beispiel Reiseblogs, Foto-sharing Portale, Reiseempfehlungsportale). Sie analog dazu die Kategorien Touristische Inländische Websites und Indirekt Touristische Inländische Websites.
- *Nicht touristische ausländische Websites:* Zu dieser Kategorie gehören alle nicht touristischen Websites deren Inhaber sich im Ausland befinden (größtenteils alle nicht .at-Domains)
- *Fehlerhafte Websites:* Dies sind alle Websites deren Links falsch geschrieben wurden oder die nicht existieren.

Die Kategorisierung wurde nur zum Vergleich der Suchdurchläufe vorgenommen. Es ist nicht gesichert ab wann ein Tourismus-Akteur zum Beispiel zu Österreich gehört (ab einem bestimmten Anteil von Gästen die nach Österreich reisen, wenn er seine Domain in Österreich hat oder wenn er seinen Firmensitz in Österreich hat) bzw. wo die Grenze gezogen wird. Die dunkelgrau eingefärbte Zeile in Tabelle 12 enthält die Daten zu den Tourismusakteuren aus Österreich. Dies ist die wichtigste Spalte, da in dieser Arbeit das Netzwerk des österreichischen Tourismus-Marktes analysiert werden soll. Die hellgrau gefärbte Zeile in Tabelle 12 enthält Websites deren Inhaber mit dem Tourismus verwandt sind und die im Netzwerk des österreichischen E-Tourismus-Marktes auch eine kleine Rolle spielen.

Hier wird die oben getätigte Vermutung bestätigt, dass Suchdurchlauf 4 mit 69,2% der Knoten aus dem österreichischen Tourismus, das repräsentativste Netzwerk darstellt. Daneben enthält dieser Suchdurchlauf im Vergleich zu den anderen sehr wenige ausländische Websites. Die zweitgrößte Kategorie ist die der indirekten inländischen touristischen Websites mit 15,8%. Zählt man diese beiden Kategorien (mit den eingefärbten Zeilen) zusammen dann ergeben sich 85%. Das heißt das Netzwerk aus Suchdurchlauf 4 ist sehr repräsentativ für den österreichischen E-Tourismus-Markt.

Abbildung 53 zeigt noch einmal die prozentuelle Aufteilung der Websites aus Suchdurchlauf 4 auf die jeweiligen Kategorien. Hier wird noch mal verdeutlicht, dass der weitaus größte Teil

des Netzwerks aus direkt und indirekt für den Tourismus relevanten inländischen Websites besteht, wobei die Kategorie der österreichischen Tourismusakteure mit fast  $\frac{3}{4}$  die Mehrheit bildet.



**Abbildung 53: Anteil der Websites je Kategorie aus Suchdurchlauf 4. Beschreibung der Kategorien siehe oben**

Für die Analyse des Netzwerks werden also die Daten aus Suchdurchlauf 4 herangezogen. Im nächsten Abschnitt wird das Netzwerk zunächst grafisch analysiert.

## 6.2 Grafische Darstellung des Netzwerks

In diesem Abschnitt werden einige Netzwerkdiagramme, die mit Suchdurchlauf 4 ermittelt wurden, präsentiert. Neben einem Diagramm das alle Knoten des Netzwerks enthält, in welchem die einzelnen Knoten aber aufgrund der großen Anzahl nur mehr schwer voneinander unterschieden werden können, werden auch Diagramme mit nur 50 und 256 Knoten gezeigt. Um die Anzahl der gezeigten Knoten auf 50 oder 256 zu begrenzen, muss eine Auswahl getroffen werden. Diese Auswahl erfolgt durch Sortierung nach ein- oder ausgehenden Links.

Wie bereits in Abschnitt 5.4 erklärt wurde, wird für die Anordnung der Knoten das so genannte Spring-embedding-Verfahren angewendet. Bei diesem Verfahren werden die Knoten, die nahe beieinander liegen (d. h. über relativ kurze Wege miteinander verbunden sind) auch im Netzwerk nahe nebeneinander angeordnet. Dieses Verfahren stellt zwar nicht immer die Knoten mit den meisten ein- oder ausgehenden Links ins Zentrum. Trotzdem ist die Wahrscheinlichkeit groß, dass ein im Zentrum stehender Knoten auch viele ein- oder ausgehende Links hat. Denn sollte dies der Fall sein, so muss der Knoten, um den Abstand zu jedem Nachbarsknoten zu minimieren, im Zentrum stehen. Würde der Knoten eher am Rande stehen, so würde er zwar den Abstand zu einigen Nachbarsknoten die in der Nähe sind klein halten, gleichzeitig aber den Abstand zu den Nachbarsknoten am gegenüberliegenden Rand vergrößern.

Zusätzlich muss erwähnt werden, dass die dargestellten Netzwerke mit 50 und 256 Knoten etwas verzerrt sind, da jeweils nur die ein- und ausgehenden Links dieser Knoten zueinander, und nicht die Links zu allen Knoten des Netzwerks, gezählt werden.

### 6.2.1 Netzwerke nach eingehenden Links

In Abbildung 54 ist das Netzwerk mit den ersten 50 Knoten, sortiert nach eingehenden Links, angeführt. Die Knoten mit den meisten eingehenden Links, die an der Größe der Knoten erkannt werden können, sind *facebook.com*, *youtube.com* und *tiscover.at*, gefolgt von *oberoesterreich.at*, *tiscover.com*, *oebb.at* und *addthis.com*. Die zentralsten Knoten sind *tiscover.at*, *austriatourism.com*, *facebook.com*, *austria.info*, *linz.at* und *youtube.com*. Dies bestätigt die Vermutung, dass Knoten mit vielen ein- und ausgehenden Links eher zentral stehen, denn die aufgezählten zentralen Knoten besitzen fast alle in Summe die meisten ein- und ausgehenden Links.

Danach zeigt Abbildung 55 das Netzwerk mit 256 Knoten, auch nach eingehenden Links sortiert. Der eindeutig größte Knoten ist hier *facebook.com*, gefolgt von *youtube.com*, *tiscover.at*, *oebb.at*, *tiscover.com* und *oberoesterreich.at*. Bei den größten Knoten hat sich also gegenüber der Darstellung mit 50 Knoten in Abbildung 54 nichts geändert. Die Knoten, welche im Zentrum stehen, sind jedoch zum Teil andere. Hier sind zum Beispiel Knoten wie *zamg.ac.at* oder *tirol.gv.at* im Zentrum, welche bei der Summe der ein- und ausgehenden Links nicht an der Spitze stehen. Hier zeigt sich, dass die Vermutung, große Knoten (mit vielen ein- oder ausgehenden Links) müssten im Zentrum stehen nicht immer richtig sein muss. In diesem Fall haben die Knoten *zamg.ac.at* und *tirol.gv.at* wahrscheinlich nur die meisten ihrer Verbindungen zu Knoten, die auch im Zentrum stehen. Bei den großen Knoten wie *oebb.at*, *facebook.com* und *youtube.com* die auch zentral liegen, stimmt die vermutete Korrelation zwischen zentraler Lage und hoher Anzahl von Links wiederum. Wieder andere größere Knoten wie *austria.info*, *tiscover.com*, *tiscover.at* und *oberoesterreich.at* stehen jedoch, anders als vermutet eher außerhalb. Auffällig ist zum Beispiel, dass eher im oberen Teil des Netzwerks (von links nach rechts in der Mitte gesehen) einige zu Oberösterreich gehörende Knoten wie *oberoesterreich.at*, *genussland.at*, *donausteig.com*, *donauradweg.at*, *wandern.at* und *biken.at* ein lokales Zentrum bilden. Weil diese Knoten untereinander stark vernetzt sind, sind sie auch im Netzwerk nebeneinander angesiedelt. Ein zweites solches Zentrum gibt es auf der rechten Seite, von oben oder unten gesehen eher in der Mitte, mit zu Tirol gehörenden Websites.

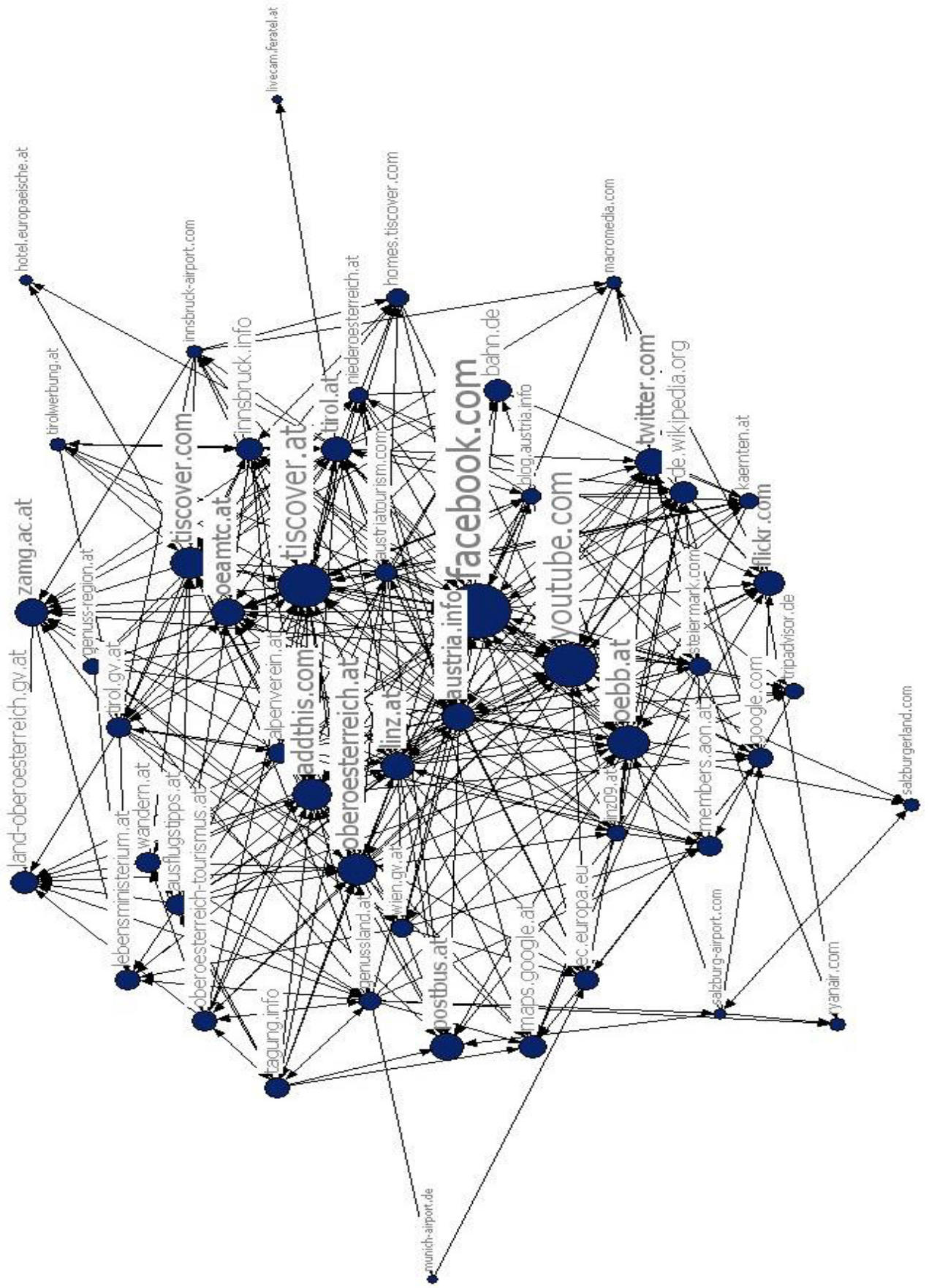


Abbildung 54: Darstellung des Netzwerks aus Suchdurchlauf 4 mit 50 Knoten nach eingehenden Links sortiert



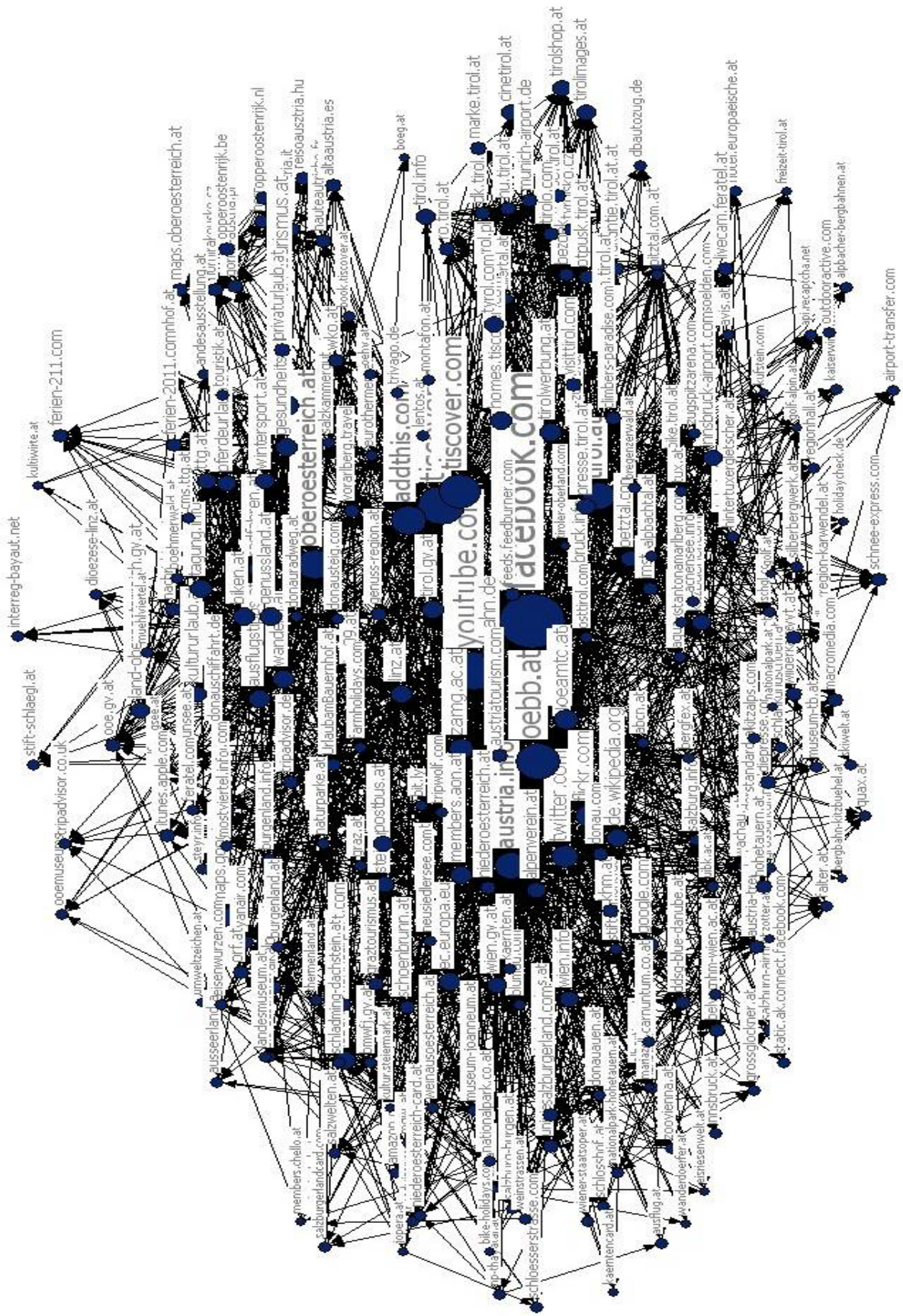


Abbildung 55: Darstellung des Netzwerks aus Suchdurchlauf 4 mit 256 Knoten nach eingehenden Links sortiert

## 6.2.2 Netzwerk nach ausgehenden Links

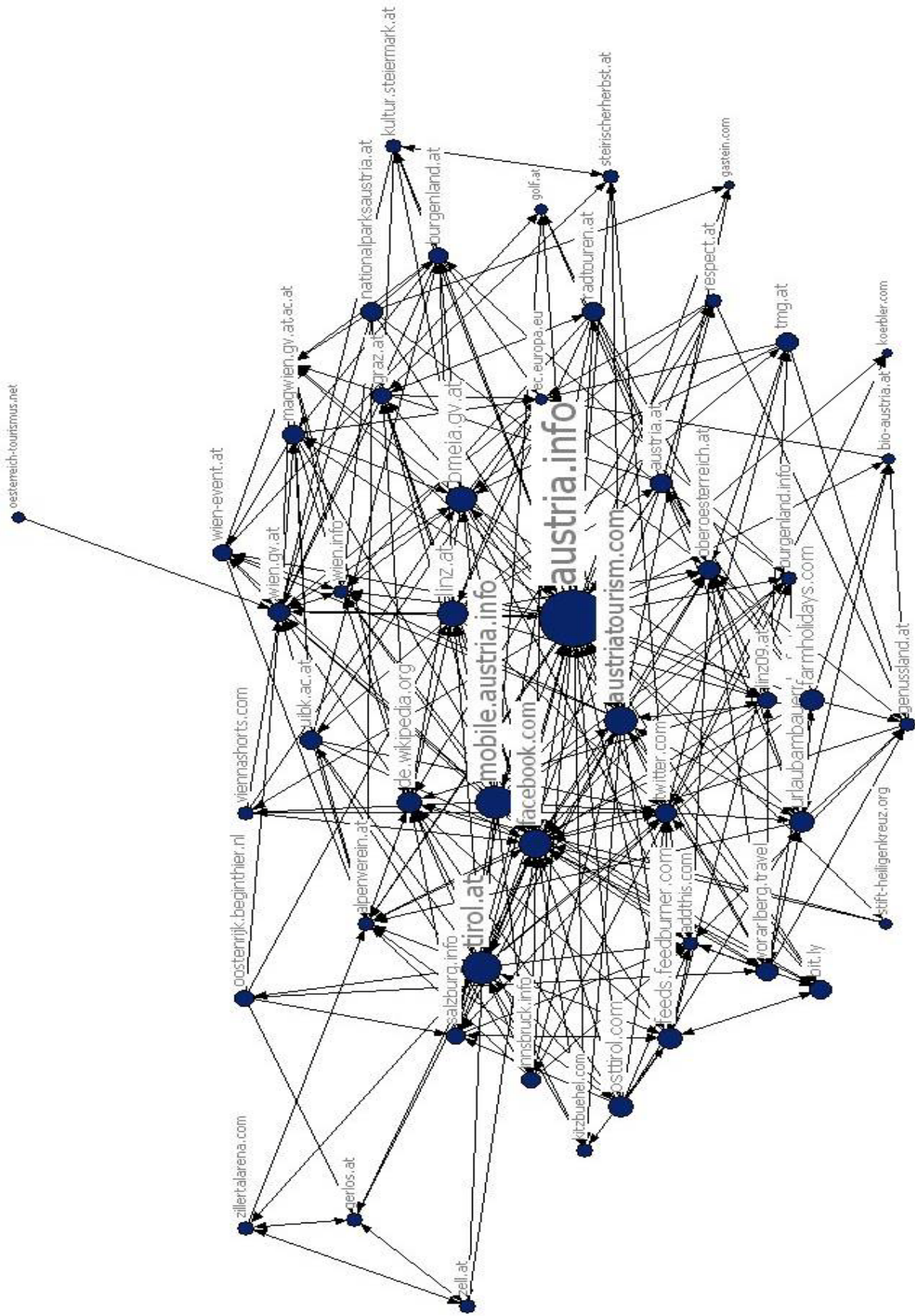
In diesem Kapitel werden zwei Netzwerkdiagramme präsentiert, welche die Knoten mit den meisten ausgehenden Links enthalten. In diesen zwei Diagrammen entspricht die Größe eines Knotens der Anzahl der ausgehenden Links.

In Abbildung 56 ist das Diagramm mit den 50 Knoten, welche die meisten ausgehenden Links haben, dargestellt. In diesem Diagramm kann der Knoten *austria.info* eindeutig als der größte Knoten identifiziert werden. Dies ist deshalb so eindeutig, weil etwa die Hälfte aller Websites, die durchsucht wurden, über einen ausgehenden Link von *austria.info* gefunden worden sind. Würden mehr Websites durchsucht, so würde der Größenunterschied zwischen dem Knoten *austria.info* und den anderen Knoten sicher verringert werden. Nach dem Knoten *austria.info* folgen der Größe nach *mobile.austria.info*, *tirol.at*, *facebook.com*, *austriatourism.com* und *bmeia.gv.at*. Auffällig ist, dass *facebook.com* nicht nur viele eingehende Links, sondern auch viele ausgehende Links in diesem Netzwerk hat. Das heißt, die meisten touristischen Websites setzen nicht nur einen Link auf *facebook.com*, sondern *facebook.com* setzt auch Links auf touristische Websites. Diese sind wahrscheinlich auf den spezifischen Seiten der Tourismusbetriebe in *facebook.com* zu finden. Dass diese spezifischen Seiten beim Durchsuchen des Portals *facebook.com* gefunden wurden, ist auch ein Zeichen für die gute Arbeit des Crawlers.

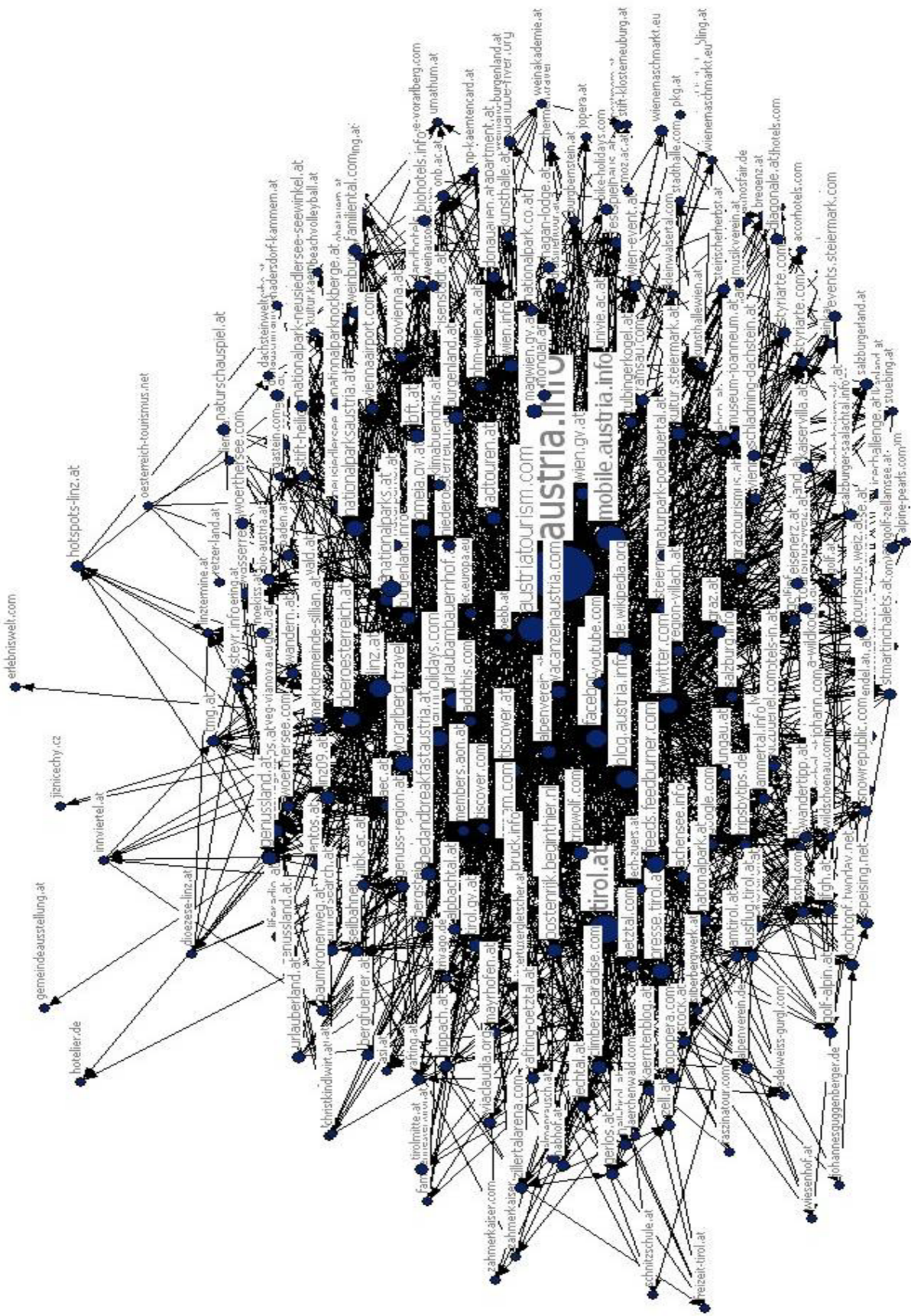
Allgemein fällt auf, dass in diesem Netzwerkdiagramm aus Abbildung 56 sehr viele Knoten aus dem öffentlichen Bereich stammen. Das heißt, entweder sind es Destinationsmanagementportale oder Seiten der öffentlichen Verwaltung. Durch die hohe Anzahl von ausgehenden Links, welche diese Seiten auf verschiedenste private – in diesem Fall Tourismusbetriebe – setzen, erfüllen sie durchaus ihren öffentlichen Auftrag, indem sie als Verteiler für die Tourismusbetriebe und Bewerber der Tourismusbetriebe auftreten.

Aufgrund der oben festgestellten, überragenden Stellung von *austria.info*, ist dieses Netzwerk auch zentraler ausgerichtet als etwa das Netzwerk in Abbildung 54 mit den 50, nach eingehenden Links sortierten Knoten. Denn in diesem Netzwerk gibt es mit dem Knoten *austria.info* genau ein Zentrum, welcher auch von der Position her global gesehen absolut zentral liegt. Auch die übrigen Knoten sind, im Vergleich zu Abbildung 54, gleichmäßiger um das Zentrum im Raum verteilt.

Darunter ist in Abbildung 57 das Netzwerk mit 256 Knoten, die nach ausgehenden Links sortiert sind, dargestellt. Auch dieses Netzwerk scheint gleichmäßiger verteilt als jenes in Abbildung 55 mit 256 Knoten, die jedoch nach eingehenden Links sortiert sind. *Austria.info* ist wieder eindeutig der größte Knoten und steht zusammen mit *austriatourism.com* im Zentrum. Auffällig ist zum Beispiel, dass ein sehr kleiner Knoten wie *oebb.at* sehr zentral liegt. *Oebb.at* hat zwar wenige ausgehende Links, aber sehr viele eingehende Links und damit viele direkte Verbindungen zu anderen Knoten, weshalb der Knoten auch im Zentrum liegt.



**Abbildung 56:** Darstellung des Netzwerks aus Suchdurchlauf 4 mit 50 Knoten nach ausgehenden Links sortiert



**Abbildung 57: Darstellung des Netzwerks aus Suchdurchlauf 4 mit 256 Knoten nach ausgehenden Links sortiert**

### 6.2.3 Gesamtes Netzwerk

In Abbildung 58 ist das Netzwerk mit allen 2395 Knoten dargestellt. Die Größe der Knoten entspricht in diesem Fall der Anzahl der eingehenden Links. Die größten Knoten sind wieder *facebook.com*, *tiscover.at*, *youtube.com*, *oebb.at*, *genussland.at*, *tiscover.com*, *twitter.com*, *tirol.at*, *austria.info* und *oberoesterreich.at*. Bis auf ein paar wenige Ausnahmen liegen diese Knoten eigentlich alle zentral. Das heißt, sie werden von allen übrigen Knoten gleichmäßig verlinkt. Die großen Knoten, welche nicht zentral liegen, sind *genussland.at* und *oberoesterreich.at*. Diese beiden Knoten werden hauptsächlich von oberösterreichischen Tourismusbetrieben verlinkt und es ist daher anzunehmen, dass um sie herum oberösterreichische Tourismusbetriebe angesiedelt sind. Auch *tirol.at* und *austria.info* liegen nicht ganz im Zentrum. Bei *tirol.at* ist das gleiche anzunehmen, wie bei den oberösterreichischen Websites eben beschrieben wurde.

Bei Betrachtung des gesamten Netzwerks fällt auf, dass das Netzwerk weder total zentral auf einen Knoten ausgerichtet noch total dezentral aufgebaut ist. Es gibt durchaus ein Zentrum mit einigen herausragenden Knoten, die sehr viel mehr eingehenden Links erhalten als alle anderen. Jedoch teilen sich diese Knoten das Zentrum nicht allein. Um sie herum sind einige mittelgroße Knoten wie *austria.info*, *tirol.at*, *oberoesterreich.at*, *maps.google.at*, *macromedia.com* und andere verteilt. Auch gibt es eher außerhalb des Netzwerks mehrere kleinere Zentren. Es ist anzunehmen, dass sich diese Zentren um regionale und lokale Destinationsmanagementsysteme und zum Teil private, themenspezifische Reiseportale bilden. Insgesamt liegen mehr Knoten in der Peripherie des Netzwerks als im Zentrum. Es gibt also insgesamt eine durchaus zentrale Ausrichtung des Netzwerks. Dies kann auch gemessen werden.

Laut [46b] definierte Freeman eine Maßzahl für die Zentralisierung eines ganzen Netzwerks. Eine kurze Beschreibung für die Zentralität eines gesamten Netzwerks wurde auch schon in Abschnitt 3.2.4 gegeben. Bei Berechnung der Graph- bzw. Netzwerkzentralität erfolgt ein Vergleich des Netzwerks mit dem zu 100% zentralisierten Stern-Netzwerk, das in Abbildung 11 dargestellt ist. Beim Stern-Netzwerk hat der zentrale Knoten eine direkte Verbindung zu allen anderen Knoten, erreicht also den höchstmöglichen Grad. Alle anderen Knoten sind nur mit dem zentralen Knoten verbunden und haben einen Grad von 1. In diesem Fall ist also das Netzwerk vollständig auf einen zentralen Knoten ausgerichtet. Die von Freeman entwickelte Maßzahl für die Zentralisierung eines Netzwerks ist ein Prozentsatz, zu welchem ein Netzwerk in Bezug auf die Graphzentralität, dem vollkommen zentralisierten Stern-Netzwerk entspricht. Beim in Abbildung 58 dargestellten Netzwerk ist diese Maßzahl für den Eingangsgrad 18% und für den Ausgangsgrad 47%.

Letztere Maßzahl ist ungerechtfertigt hoch, da die Startwebsite des Suchvorgangs *austria.info* 1129 ausgehende Links hatte und diese 1129 Websites, welche von der Startwebsite aus verlinkt wurden, etwa die Hälfte aller insgesamt durchsuchten Websites stellten. Würden beim Suchdurchlauf mehr Knoten durchsucht werden, so würde diese Maßzahl sicherlich sinken. Diese relativ hohe Zentralisierung des Netzwerks in Bezug auf den Ausgangsgrad und die spezielle Ausrichtung auf den Knoten *austria.info* hin war auch schon in den Netzwerkdiagrammen, die im vorigen Abschnitt dargestellt wurden, erkennbar.

Die 18% Zentralisierung beim Eingangsgrad sind „kein schlechter Wert“. Es kann behauptet werden, dass es durchaus eine Zentralisierung des Netzwerks gibt und damit einen größeren Einfluss von einem kleineren Teil der Knoten auf die übrigen Netzwerkteilnehmer. Andererseits ist diese Zentralisierung nicht sehr ausgeprägt, d. h. nur auf einen Knoten ausgerichtet, wenn dieses Netzwerk mit einem sternförmigen Netzwerk verglichen wird. Denn nur die Hälfte aller gefundenen Seiten besitzt einen Eingangsgrad von maximal 1. Im Stern-Netzwerk sind das jedoch alle Knoten, bis auf jenen im Zentrum.

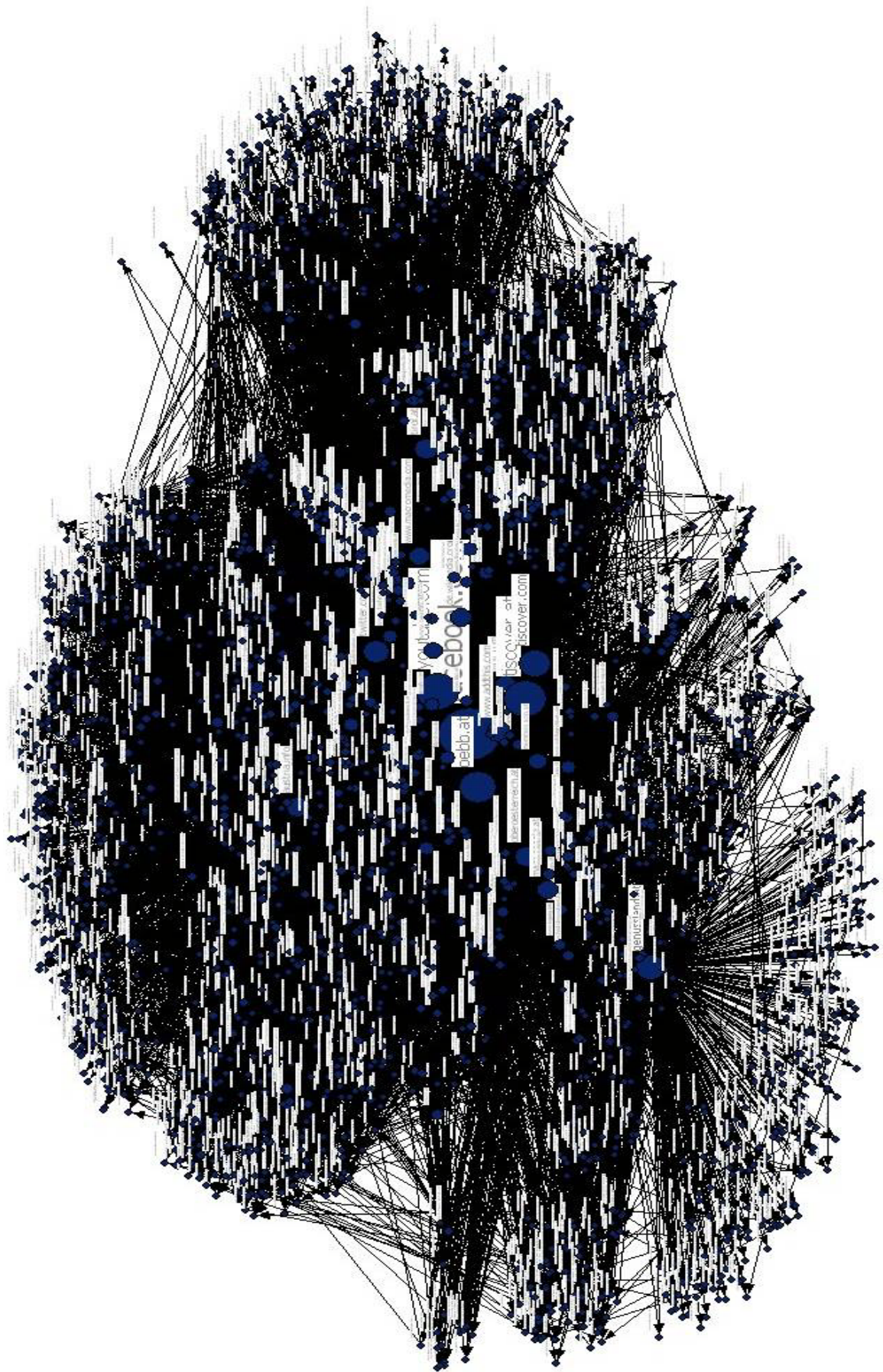


Abbildung 58: Gesamtes Netzwerk aus Suchdurchlauf 4 mit 2395 Knoten

### 6.3 Ergebnisse für die Zentralitätskennzahlen und Vergleich zusätzlicher Ergebnisse mit ähnlichen Netzwerken

In diesem Abschnitt werden zuerst die Ergebnisse für die berechneten Zentralitätskennzahlen aufgelistet und beurteilt. Danach werden zusätzliche Ergebnisse (Netzwerkdicke und Durchschnittsgrad, durchschnittliche Weglänge, Clusterkoeffizient, Potenzgesetzverteilung der Knotengrade, Fliegenstruktur und die Anzahl nicht verbundener Knoten) präsentiert und mit dem ähnlichen Netzwerk der Websites der Tourismusanbieter der Insel Elba in Italien sowie dem Netzwerk des World Wide Webs verglichen.

#### 6.3.1 Ergebnisse für die Grad-basierte Zentralität

Ein hoher Eingangsgrad zeugt von einer hohen Prominenz oder einem hohen Prestigewert eines Knotens innerhalb eines sozialen Netzwerks [46b]. Die Knoten mit den 20 höchsten Eingangsgraden sind in Tabelle 13 aufgelistet. Dabei wird zwischen absolutem und relativem Ein- bzw. Ausgangsgrad unterschieden. Der absolute Ein- bzw. Ausgangsgrad gibt an, mit wie vielen Knoten ein Knoten in eine bestimmte Richtung verbunden ist. Der relative Ein- bzw. Ausgangsgrad gibt an, wie viel Prozent diese absolute Zahl an verbundenen Knoten im Vergleich mit der Anzahl der Knoten des gesamten Netzwerks ausmacht.

Website	Eingangsgrad-basierte Zentralität	
	Absolut	Relativ
<i>facebook.com</i>	439	18,368
<i>tiscover.at</i>	287	12,008
<i>youtube.com</i>	280	11,715
<i>oebb.at</i>	250	10,460
<i>tiscover.com</i>	197	8,243
<i>genussland.at</i>	165	6,904
<i>twitter.com</i>	145	6,067
<i>addthis.com</i>	142	5,941
<i>oberoesterreich.at</i>	133	5,565
<i>tirol.at</i>	131	5,481
<i>austria.info</i>	129	5,397
<i>flickr.com</i>	102	4,268
<i>maps.google.at</i>	101	4,226
<i>macromedia.com</i>	100	4,184
<i>de.wikipedia.org</i>	99	4,142
<i>oeamtc.at</i>	94	3,933
<i>members.aon.at</i>	94	3,933
<i>bahn.de</i>	69	2,887
<i>zamg.ac.at</i>	65	2,720
<i>steiermark.com</i>	65	2,720

**Tabelle 13: Knoten mit den 20 höchsten Eingangsgraden des Netzwerks aus Suchdurchlauf 4**

In diesem Fall ist der Knoten *facebook.com* derjenige mit dem höchsten Eingangsgrad. Werden jedoch die Eingangsgrade für *tiscover.com* und *tiscover.at* zusammengezählt, so hat das Portal Tiscover den höchsten Eingangsgrad. Beide Knoten (Facebook und Tiscover)

zusammen sind mit Abstand die Knoten mit den meisten eingehenden Links. Sicherlich kann speziell in Zusammenhang mit diesen beiden Knoten und mit allen anderen in Tabelle 13 angeführten Knoten – auch wenn sie aus einem Webnetzwerk und nicht aus einem sozialen Netzwerk stammen – von hoher Prominenz gesprochen werden. Ein Knoten, der von vielen Websites aus verlinkt ist, wird von diesen Websites als wichtig erachtet und kann deswegen auch als prominent bezeichnet werden.

Ein hoher Ausgangsgrad eines Knotens zeugt hingegen von hohem Einfluss, den ein Knoten auf die anderen Knoten innerhalb eines sozialen Netzwerks hat [46b]. In Tabelle 14 sind die Knoten mit den 20 höchsten Ausgangsgraden aufgelistet.

Website	Ausgangsgrad-basierte Zentralität	
	Absolut	Relativ
<i>austria.info</i>	1.129	47,238
<i>tirol.at</i>	670	28,033
<i>genussland.at</i>	620	25,941
<i>mobile.austria.info</i>	306	12,803
<i>austriatourism.com</i>	258	10,795
<i>oberoesterreich.at</i>	157	6,569
<i>vorarlberg.travel</i>	133	5,565
<i>blog.austria.info</i>	112	4,686
<i>ausflugstipps.at</i>	106	4,435
<i>linz.at</i>	99	4,142
<i>tiscover.at</i>	94	3,933
<i>linz09.at</i>	87	3,640
<i>presse.tirol.at</i>	87	3,640
<i>austria.at</i>	80	3,347
<i>facebook.com</i>	76	3,180
<i>farmholidays.com</i>	74	3,096
<i>UrlaubamBauernhof.at</i>	74	3,096
<i>radtouren.at</i>	73	3,054
<i>twitter.com</i>	71	2,971
<i>innsbruck.info</i>	67	2,803

**Tabelle 14: Knoten mit den 20 höchsten Ausgangsgraden des Netzwerks aus Suchdurchlauf 4**

Der Knoten *austria.info* besitzt hier (Tabelle 14) mit Abstand den höchsten Ausgangsgrad. Dies ist – wie schon im Abschnitt 6.2 erkannt werden konnte – deswegen so, weil ungefähr die Hälfte aller Knoten, die mit dem Crawler durchsucht wurden, von *austria.info* aus verlinkt sind. Der relative Ausgangsgrad in der Tabelle zeigt dies gut. *Austria.info* erreicht einen relativen Ausgangsgrad von fast 50%. Der Knoten *austria.info* wäre also, nach der Interpretation der sozialen Netzwerkanalyse zufolge, der einflussreichste Knoten im Netzwerk. Jedoch stimmt das nur bis zu einem gewissen Grad auch für ein Netzwerk des World Wide Web, da es hier nicht um das Senden von Informationen geht. Zumindest übt dieser Knoten einen begrenzten Einfluss auf die Knoten aus, die er verlinkt, denn durch das Setzen eines Links unterstreicht der Knoten die Wichtigkeit des verlinkten Knotens und macht ihn gleichzeitig leichter auffindbar im World Wide Web. Obwohl Websites und Webseiten meist über Suchmaschinen gefunden werden, finden doch die Suchmaschinen



selbst die Webseiten über die Verfolgung der Links und wenn also eine Website oft verlinkt ist, so wird sie dadurch leichter auffindbar.

Sicher erfüllen dieser und alle anderen in Tabelle 14 aufgelisteten Knoten, die aus dem öffentlichen Bereich kommen, ihren öffentlichen Auftrag gut, indem sie Werbung für die ihnen unterstellten Tourismusbetriebe betreiben, deren Wichtigkeit unterstreichen und sie im World Wide Web leichter auffindbar machen. Wichtig dabei ist allerdings auch, dass der Knoten mit einem hohen Ausgangsgrad selbst angeklickt wird, damit sich seine unterstützende Wirkung für die anderen Knoten die er verlinkt erst entfalten kann. Damit der Knoten selbst oft angeklickt wird, muss er bekannt sein bzw. von einer Suchmaschine leicht gefunden werden können. Da diese Eigenschaften hier aber nicht gemessen werden, kann in den Begriffen der sozialen Netzwerkanalyse gesagt werden, der Knoten sollte prominent sein bzw. einen hohen Eingangsgrad haben.

Auch muss damit argumentiert werden, dass sicher auch ein Knoten mit hohem Eingangsgrad einen hohen Einfluss haben kann. Damit ist nicht gemeint, dass der Knoten durch das Setzen von ausgehenden Links Einfluss auf andere Knoten ausübt, so wie es vorher interpretiert wurde, sondern, dass der Knoten, durch seine Prominenz generell einen hohen Einfluss im Netzwerk hat.

Wichtig zu bemerken ist, dass die in Tabelle 14 angeführte Liste der Knoten mit den meisten ausgehenden Links etwas verzerrt ist. Bestimmte Knoten fehlen in dieser Liste, weil viele der Knoten, die sie durch ihre ausgehenden Links verlinken, aufgrund der begrenzten Größe des Netzwerks nicht durchsucht werden konnten. Diese wurden deshalb nicht erfasst, weil logischerweise nur jene ausgehenden Links zählen, die auch zu Knoten innerhalb des durchsuchten Netzwerks führen.

### **6.3.2 Ergebnisse für die Nähe-basierte Zentralität**

Nun soll die Nähe-basierte Zentralität berechnet werden. Die Berechnung der Ein- und Ausgangsgrade hat den Nachteil, dass sie nur die unmittelbaren Nachbarn eines Knotens mit einbezieht. Sie ist, wie bereits erwähnt wurde, ein lokales Zentralitätsmaß. Hingegen berücksichtigt die Nähe-basierte Zentralität alle Knoten des Netzwerks. In Abschnitt 3.2.4.2 wurde bereits erwähnt, dass für die Berechnung der Nähe-basierten Zentralität die kürzesten Wege von einem Knoten zu jedem anderen Knoten berechnet werden müssen. Mit dem Feststehen der Ergebnisse wurde aber festgestellt, dass das Netzwerk nicht vollständig verbunden ist. Das heißt, es ist nicht möglich von jedem Knoten aus jeden anderen Knoten egal in welcher Richtung zu erreichen und es kann somit nicht für jedes Knotenpaar ein kürzester Weg berechnet werden. Deshalb ist das in Abschnitt 3.2.4.2 beschriebene Verfahren zur Berechnung der Nähe-basierten Zentralität nicht geeignet und wenn versucht wird die Nähe-basierte Zentralität nach diesem Verfahren in UCINET zu berechnen, so ist das Ergebnis nicht aussagekräftig, da viele Knoten eine sehr hohe und gleiche Nähe-basierte Zentralität erhalten.

Um dieses Problem zu umgehen, bietet UCINET für diese Fälle einen anderen Weg zur Berechnung der Nähe-basierten Zentralität an. Laut [61] ist diese Variante für nicht verbundene Netzwerke, in denen es nicht von jedem Knoten zu jedem anderen einen Weg gibt, im Vergleich zur üblichen Variante der kürzesten Wege die bessere Variante zur Berechnung der Nähe-basierten Zentralität. Dabei werden die reziproken Werte der geodätischen Distanzen zwischen zwei Knoten verwendet. Für Knotenpaare für die es keinen verbindenden Weg gibt, wird eine unendlich große Wegdistanz angenommen und der reziproke Wert geht somit gegen Null. Da ein Knoten umso zentraler in Bezug auf die Nähe-basierte Zentralität liegt, je höher sein Zentralitätswert ist, führt diese Variante auch bei

Netzwerken in denen es nicht für jedes Knotenpaar einen verbindenden Weg gibt zu aussagekräftigen Ergebnissen.

Dieses Verfahren entspricht zwar nicht exakt dem von Freeman entwickelten Verfahren zur Berechnung der Nähe-basierten Zentralität, kann aber dennoch zur Berechnung der Nähe-basierten Zentralität herangezogen werden, da trotzdem das zentrale Konzept der Nähe-basierten Zentralität, nämlich die Entfernung zwischen zwei Knoten, gemessen wird und je kürzer die Entfernung, desto besser die Nähe-basierte Zentralität.

In Tabelle 15 sind die Knoten mit den 20 höchsten Eingangsnähe-basierten Zentralitätswerten aufgelistet. In diesem Fall spricht ein hoher Wert für eine bessere Nähe-basierte Zentralität, da durch die Reziprozität der Distanzen bei der Berechnung eine kürzere Distanz einen höheren Wert ergibt. Neben der absoluten Eingangsnähe-basierten Zentralität wird hier auch ein relativer Wert gegeben, welcher eine Prozentzahl darstellt und sich aus dem Verhältnis zur maximalen Eingangsnähe-basierten Zentralität (d.h. der nächst möglichen Entfernungen, siehe Abschnitt 3.2.4.2) ergibt.

Website	Eingangsnähe-basierte Zentralität	
	Absolut	Relativ
<i>facebook.com</i>	943,332	39,47
<i>tiscover.at</i>	838,081	35,066
<i>youtube.com</i>	834,648	34,923
<i>oebb.at</i>	816,382	34,158
<i>tiscover.com</i>	742,930	31,085
<i>twitter.com</i>	731,148	30,592
<i>austria.info</i>	730,032	30,545
<i>addthis.com</i>	727,581	30,443
<i>oberoesterreich.at</i>	721,165	30,174
<i>tirol.at</i>	708,647	29,650
<i>flickr.com</i>	700,099	29,293
<i>de.wikipedia.org</i>	698,949	29,245
<i>linz.at</i>	673,482	28,179
<i>oeamtc.at</i>	662,998	27,741
<i>maps.google.at</i>	659,332	27,587
<i>bahn.de</i>	658,498	27,552
<i>alpenverein.at</i>	655,532	27,428
<i>innsbruck.info</i>	649,932	27,194
<i>members.aon.at</i>	649,416	27,172
<i>genuss-region.at</i>	645,849	27,023

**Tabelle 15: Die Knoten mit den 20 höchsten Eingangsnähe-basierten Zentralitätswerten des Netzwerks aus Suchdurchlauf 4**

Allgemein bedeutet laut [46b] ein guter und deshalb niedriger Nähe-basierter Zentralitätswert in der sozialen Netzwerkanalyse viel Macht. Aufgrund der kurzen Wege zu allen anderen Netzwerkteilnehmern hat der Knoten mit einem niedrigen Nähe-basierten Zentralitätswert eine begünstigte Position und diese begünstigte Position kann auch als Machtfaktor ausgedrückt werden. Auch zeugt ein guter Nähe-basierter Zentralitätswert von hohem Einflussvermögen auf die gesamte Gruppe und ist vor allem dann wichtig, wenn es gilt, jedes Gruppenmitglied schnell zu erreichen [62]. Für die Unterscheidung der Bedeutung eines hohen Ein- oder Ausgangsnähe-basierten Zentralitätswerts bei gerichteten Netzwerken gibt es

in der Literatur keine spezielle Nennung. Es kann aber vermutet werden, dass in einem gerichteten Netzwerk ein guter Eingangsnähe-basierter Zentralitätswert analog zu einem hohen Eingangsgrad, einen hohen Prestige- oder Prominenzwert bedeutet. Jedoch bezieht sich dieser Prominenzwert eines Knotens nicht nur auf die unmittelbaren Nachbarknoten, sondern auf das gesamte Netzwerk.

In Tabelle 15 ist zu erkennen, dass die 20 Knoten mit den höchsten Eingangsnähe-basierten Zentralitätswerten fast die gleichen sind wie die 20 Knoten mit den höchsten Eingangsgraden, welche in Tabelle 13 dargestellt sind. Auch die Reihenfolge der Knoten ist sehr ähnlich. Die ersten fünf Knoten sind in derselben Reihenfolge aufgelistet.

In der unten angeführten Tabelle 16 sind schließlich die 20 Webseiten mit den höchsten Ausgangsnähe-basierten Zentralitätswerten aufgelistet. Auch hier gilt analog zur Interpretation des Ausgangsgrades, dass bei einem guten Ausgangsnähe-basierten Zentralitätswert auf hohen Einfluss des Knotens im Bereich des gesamten Netzwerkes geschlossen werden kann. Auch in dieser Tabelle ist die Reihenfolge der ersten fünf Seiten mit den höchsten Ausgangsnähe-basierten Zentralitätswerten die gleiche wie in der Tabelle 14 für die Ausgangsgrade. Insgesamt sind beide Tabellen wiederum sehr ähnlich. Der relative Ausgangsnähe-basierte Zentralitätswert entspricht wiederum einer Prozentzahl die das Verhältnis der absoluten Ausgangsnähe-basierten Zentralität zum maximalen Ausgangsnähe-basierten Zentralitätswert (d.h. der kürzest möglichen Distanz zu allen Knoten, siehe Abschnitt 3.2.4.2) widerspiegelt.

Website	Ausgangsnähe-basierte Zentralität	
	Absolut	Relativ
<i>austria.info</i>	1.698,335	71,06
<i>tirol.at</i>	1.441,668	60,321
<i>genussland.at</i>	1.273,426	53,281
<i>mobile.austria.info</i>	1.267,001	53,013
<i>austriatourism.com</i>	1.243,668	52,036
<i>scnem.com</i>	1.224,167	51,22
<i>region.austria.info</i>	1.216,333	50,893
<i>oberoesterreich.at</i>	1.190,824	49,825
<i>vorarlberg.travel</i>	1.180,991	49,414
<i>blog.austria.info</i>	1.169,167	48,919
<i>tiscover.at</i>	1.166,668	48,815
<i>farmholidays.com</i>	1.155,827	48,361
<i>UrlaubamBauernhof.at</i>	1.155,827	48,361
<i>facebook.com</i>	1.151,668	48,187
<i>austria.at</i>	1.149,667	48,103
<i>vacanzeinaustria.com</i>	1.136,500	47,552
<i>bedandbreakfaustria.at</i>	1.132,667	47,392
<i>tripwolf.com</i>	1.128,500	47,218
<i>osttirol.com</i>	1.127,501	47,176
<i>feeds.feedburner.com</i>	1.118,417	46,796

**Tabelle 16: Die Knoten mit den 20 höchsten Ausgangsnähe-basierten Zentralitätswerten des Netzwerks aus Suchdurchlauf 4**

Die Frage bei der Ein- und Ausgangsnähe-basierten Zentralität ist aber, inwieweit diese Kennzahl für ein Netzwerk des World Wide Webs etwas über die Prominenz bzw. den

Einfluss eines Knotens aussagt. Denn im World Wide Web bewegt sich der Benutzer üblicherweise nicht mehr über die Links von einer Website zur anderen; wenn doch, dann besucht er höchstens die nächste Seite, die mit der Ausgangsseite verlinkt ist, aber nicht die übernächste und so weiter, denn für gewöhnlich verwendet der Benutzer im World Wide Web eine Suchmaschine zum Auffinden bestimmter Websites.

Auch sind die Werte für die Nähe-basierte Zentralität wegen der Behandlung der Websites als Knoten etwas verzerrt. Weil ein Weg zwischen zwei Websites immer über die Links erfolgt, die Links aber zwei Webseiten verbinden, könnte es theoretisch vorkommen, dass der Eingangslink auf eine andere Webseite innerhalb der Website führt, als jener der Webseite, von der aus der Ausgangslink weitergeht. In diesem Fall müsste der Benutzer, wollte er dem Weg folgen, zuerst noch innerhalb der Website herumsuchen, um die Webseite zu finden, die den ausgehenden Link enthält.

### 6.3.3 Ergebnisse für die Zwischenzentralität

In Abschnitt 3.2.4.3 wurde die Zwischenzentralität beschrieben. Wie in diesem Abschnitt gezeigt wurde, ist die Zwischenzentralität ein Maß dafür, wie oft ein Knoten auf dem kürzesten Weg zwischen zwei anderen Knoten liegt. Ein Knoten mit hoher Zwischenzentralität nimmt also eine Art Vermittlerrolle im Netzwerk ein und hat deswegen eine große Auswirkung auf den Fluss im Netzwerk und damit auch einen relativ hohen Einfluss auf jene Knotenpaare, deren verbindender Weg über ihn führt.

In Tabelle 17 sind die 20 Websites mit der höchsten Zwischenzentralität aufgelistet. Eindeutig an der Spitze steht hier die Seite *austria.info*. Diese Website liegt also am öftesten auf dem kürzesten Weg zwischen zwei Knoten

Website	Zwischenzentralität	
	Absolut	Relativ
<i>austria.info</i>	1.374.966,75	24,081
<i>facebook.com</i>	849.694,5	14,882
<i>genussland.at</i>	780.058,75	13,662
<i>tirol.at</i>	771.091,875	13,505
<i>tiscover.at</i>	519.618,969	9,101
<i>oberoesterreich.at</i>	311.591,813	5,457
<i>youtube.com</i>	146.201,297	2,561
<i>twitter.com</i>	127.363,023	2,231
<i>vorarlberg.travel</i>	126.131,922	2,209
<i>tiscover.com</i>	118.708,758	2,079
<i>linz.at</i>	72.790,969	1,275
<i>de.wikipedia.org</i>	71.827,984	1,258
<i>austriatourism.com</i>	70.202,977	1,230
<i>genuss-region.at</i>	67.826,844	1,188
<i>farmholidays.com</i>	57.305,004	1,004
<i>macromedia.com</i>	56.598,594	0,991
<i>linz09.at</i>	50.626,113	0,887
<i>bahn.de</i>	50.483,258	0,884
<i>blog.austria.info</i>	49.380,027	0,865
<i>UrlaubamBauernhof.at</i>	49.345,816	0,864

Tabelle 17: Die Knoten mit den 20 höchsten Zwischenzentralitäten aus Suchdurchlauf 4

Tabelle 17 unterscheidet, wie auch die Definition in Abschnitt 3.2.4.3 zwischen relativer und absoluter Zwischenzentralität. Beide lassen sich wie in Abschnitt 3.2.4.3 für ungerichtete Graphen angegeben berechnen, denn UCINET unterscheidet hier nicht zwischen gerichteten und ungerichteten Graphen. Trotzdem kann an der Höhe der Zwischenzentralitäten gesehen werden, welche Knoten zentraler liegen und welche eher nicht, auch wenn ein Vergleich mit den Zwischenzentralität eines anderen Netzwerks mit Vorsicht zu behandeln wäre. Die relativen Werte in Tabelle 17 sind in Prozent ausgedrückte relative Zwischenzentralitäten (nach der Definition in Abschnitt 3.2.4.3 für ungerichtete Graphen).

Aus der Tabelle ist zu erkennen, dass die Werte bis auf die ersten fünf Knoten, relativ niedrig sind. Insgesamt haben etwa zwei Fünftel aller Knoten einen Zwischenzentralitätswert von 0. Daraus folgt, dass die meisten Knoten entweder direkt oder gar nicht miteinander verbunden sind. Somit haben die meisten Vermittlerknoten nur eine begrenzte Macht über die Knotenpaare deren verbindende Weg über sie führen, mit Ausnahme der Knoten mit den höchsten Werten wie *austria.info*, *facebook.com*, *genussland.at* und *tirol.at*.

Allgemein kann noch über die Bedeutung der Zwischenzentralität für ein Netzwerk wie das World Wide Web nachgedacht werden. Wie bereits in Zusammenhang mit der Nähe-basierten Zentralität erwähnt, benützt ein User im World Wide Web meist eine Suchmaschine um eine Website zu finden, anstatt sich über die Links von Website zu Website zu klicken. Deshalb ist fraglich, ob eine hohe Zwischenzentralität einen Knoten im Netzwerk des World Wide Webs zentral oder wichtig macht.

Auch muss hier wieder, gleich wie bei der Nähe-basierten Zentralität, die Behandlung der Websites und nicht der Webseiten als Knoten beachtet werden. Da es bei der Zwischenzentralität auch um Wege im Netzwerk geht, ist dieser Zentralitätswert ebenfalls etwas verzerrt. Denn es kann vorkommen, dass der Weg über eine Website führt, deren eingehender Link auf eine andere Webseite verweist als auf jene Webseite, auf welcher der ausgehende Link liegt. In diesem Fall verliert der bestehende Weg an Bedeutung, da der Benutzer auf einer bestimmten Website erst die richtige Webseite finden muss, um den Weg fortsetzen zu können.

### **6.3.4 Berechnung und Vergleich zusätzlicher Ergebnisse**

In diesem Abschnitt werden eine Berechnung weiterer Kennzahlen des Netzwerks sowie ein Vergleich dieser Kennzahlen mit denen ähnlicher Netzwerke durchgeführt. In Tabelle 18 sind die Ergebnisse dieser Berechnungen über das Netzwerk des österreichischen E-Tourismus-Marktes im Vergleich mit einer ähnlichen Studie über das Netzwerk der Websites der Tourismusanbieter der Insel Elba in Italien sowie den Ergebnissen für das gesamte Netzwerk des World Wide Web angeführt. Die letztgenannten Ergebnisse wurden aus [64b] entnommen. Diese Untersuchung fasst die wichtigsten bisher erzielten Ergebnisse über das Netzwerk des World Wide Web aus unterschiedlichen Quellen zusammen.

Nicht ganz klar wird aus [64b], warum der Wert der Dichte für die Untersuchung aus Spalte 6 der Tabelle 18 0,000041 beträgt. Würde die Dichte (mit der Formel aus Abschnitt 3.1.2) aus der hier genannten Anzahl von Knoten und Kanten berechnet werden, so würde dies den Wert 0,000013768 ergeben, was fast drei Mal so klein wie die angegebene Dichte ist. Dieser Unterschied ist zwar nicht allzu groß, es muss aber darauf hingewiesen werden, auch wenn der Grund für diesen Unterschied im gesamten Dokument [64] nicht gefunden werden kann. Zusätzlich muss bei der Analyse der Werte aus Tabelle 18 beachtet werden, dass bei zwei Untersuchungen über das World Wide Web, anstatt, wie in dieser Arbeit, Websites, Webseiten als Knoten verwendet wurden. Auch sind die Ergebnisse aus [64b] zum Teil unvollständig sind. Aufgrund dessen und weil bei zwei Untersuchungen des World Wide Webs eben Webseiten als Knoten verwendet wurden, ist ein genauer Vergleich nicht absolut

möglich bzw. muss die Interpretation nicht notwendigerweise zutreffen. Das Netzwerk des österreichischen E-Tourismus-Marktes wurde – wie bereits in Abschnitt 5.1.2.1 beschrieben – gleich wie jenes der Tourismusanbieter der Insel Elba auf Basis von Websites als Knoten aufgebaut.

Insgesamt ist zu sagen, dass ein Vergleich der unterschiedlichen Netzwerke durchaus angebracht ist, dass es aber auch, wie in Abschnitt 3.4 bereits bemerkt wurde, aufgrund der unterschiedlichen Strategie eines Crawlers, der unterschiedlichen Startadressen und der unterschiedlichen Teilgraphen des World Wide Webs zu unterschiedlichen Ergebnissen kommen kann.

	Österreichischer E-Tourismus-Markt (Website)	Tourismusanbieter der Insel Elba (Website) <sup>a</sup>	World Wide Web (Website) <sup>b</sup>	World Wide Web (Website) <sup>b</sup>	World Wide Web (Webseite) Domain nd.edu <sup>b</sup>
<b>Knoten</b>	2.395	468	200.000.000	153.127	269.504
<b>Kanten</b>	16.893	507	2.000.000.000	Nicht verfügbar <sup>e</sup>	1.000.000
<b>Dichte</b>	0,003	0,002	0,0000001	Nicht verfügbar <sup>e</sup>	0,000041
<b>Durchschnittsgrad</b>	7,065	2,12	10,5	35,2	5,6
<b>Durchschnittliche Weglänge</b>	3,328	4,5	16,18	3,1	11,27
<b>Clusterkoeffizient</b>	0,341	0,003	Nicht verfügbar <sup>e</sup>	0,108	0,29
<b>Wert des Parameters der Potenzgesetzverteilung der Eingangsgrade</b>	2,47 <sup>c</sup>	2,96	2,1	1,94	2,1
<b>Wert des Parameters der Potenzgesetzverteilung der Ausgangsgrade</b>	2,63 <sup>d</sup>	1,89	2,7	Nicht verfügbar <sup>e</sup>	2,4

**Tabelle 18: Vergleich der Netzwerke: Österreichischer E-Tourismus-Markt, Tourismusanbieter der Insel Elba und Netzwerk des World Wide Web.** <sup>a</sup> aus [64d], <sup>b</sup> aus [64b], <sup>c</sup> für den Parameter  $k > 12$ , <sup>d</sup> für den Parameter  $k > 29$ , <sup>e</sup> in [64b] nicht verfügbar

### 6.3.4.1 Dichte und Durchschnittsgrad

Die Dichte des Netzwerks des österreichischen E-Tourismus-Marktes beträgt – wie in Tabelle 18 zu sehen ist – 0,003. Insgesamt gibt es im Netzwerk 16.893 Kanten und 2395 Knoten. Das Ergebnis der Dichte für das Netzwerk der Insel Elba beträgt 0,002. Insgesamt ist das Ergebnis der Netzwerkdichte aus beiden Untersuchungen also relativ ähnlich. Daraus folgt, dass Tourismusanbieter im Allgemeinen relativ undicht bzw. leicht vernetzt sein könnten. Das Netzwerk des österreichischen E-Tourismus-Marktes kann aber insgesamt im Vergleich mit dem der Insel Elba als etwas dichter eingeschätzt werden – auch, weil es aus einer größeren Anzahl von Knoten besteht und die Dichte eines Netzwerks von Websites aber mit zunehmender Knotenanzahl sinken sollte. Denn je größer das Netzwerk ist, mit desto mehr Knoten müsste jeder neu dazukommende Knoten verbunden werden, um eine gleich hohe Dichte zu erreichen wie ein kleineres Netzwerk. Da die Anzahl der Verbindungen eines Knotens mit zunehmender Größe des Netzwerks aber nicht notwendigerweise steigt, sinkt die Dichte bei größer werdenden Netzwerken.

Die Betrachtung des Durchschnittsgrades bestätigt diese Behauptung. Der Durchschnittsgrad ist ein besseres Maß für den Vergleich der Dichte von zwei unterschiedlich großen

Netzwerken, weil die Dichte, wie bereits erwähnt, mit steigender Knotenanzahl sinkt. Der Durchschnittsgrad des Netzwerks des österreichischen E-Tourismus-Marktes ist ungefähr um fünf Grade höher als jener der Insel Elba und deswegen ist das Netzwerk des österreichischen E-Tourismus-Marktes dichter als jenes der Insel Elba.

Der Vergleich der Dichte des World Wide Webs mit der Dichte des Netzwerks des österreichischen E-Tourismus-Marktes lässt Ähnlichkeiten vermuten. Die gleichzeitige Betrachtung des Durchschnittsgrades für diese Netzwerke bestätigt diese Vermutung. Bis auf den Durchschnittsgrad des World Wide Webs mit den Websites als Knoten sind die Durchschnittsgrade auch hier ähnlich denen des österreichischen E-Tourismus-Marktes. Der sehr viel höhere Wert des World Wide Webs mit den Websites als Knoten ist wahrscheinlich, weil Websites insgesamt zumindest mehr ausgehende Links haben als Webseiten. Auch ist es möglich, dass der durchschnittliche Grad des Netzwerks des österreichischen E-Tourismus-Marktes bei größerem Netzwerk steigen würde; denn wahrscheinlich würden die Ausgangsgrade ziemlich ansteigen, da – wie schon in einem vorherigen Kapitel bemerkt wurde – die Knoten im Netzwerk in der Realität viel höhere Ausgangsgrade haben als angegeben. Dies ist deshalb so, weil hier nur jene Ausgangsgrade gezählt werden, die auch die im Netzwerk vorhandenen Knoten betreffen (2395). Würde die Knotenanzahl des Netzwerks allerdings steigen, so erhöht sich die Wahrscheinlichkeit, dass die durch ausgehende Links verbundenen Knoten auch im Netzwerk enthalten wären und damit diese ausgehenden Links auch mitgezählt werden könnten, weshalb die Anzahl der Ausgangsgrade steigen würde. Ebenso stellt sich, wie in Abschnitt 5.1.2 erklärt wurde, bei der Verwendung von Websites als Knoten die Frage, ob mehrere Links zwischen zwei gleichen Seiten auch mehrfach gezählt werden. Da dem bei der Untersuchung des österreichischen E-Tourismus-Marktes nicht so war, dies aber beim World Wide Web mit Websites als Knoten der Fall gewesen wäre, würde das den großen Unterschied beim Durchschnittsgrad erklären.

#### **6.3.4.2 Durchschnittliche Weglänge**

Bei der durchschnittlichen Weglänge ist der Wert, wie wiederum in Tabelle 18 zu sehen ist, für den österreichischen E-Tourismus-Markt um zirka einen Knoten geringer als der für das Netzwerk der Websites der Tourismusanbieter der Insel Elba. Dies lässt auf eine etwas bessere Verbundenheit des Netzwerks des österreichischen E-Tourismus-Marktes schließen, so wie es schon beim Dichtevergleich im vorherigen Kapitel festgestellt wurde, auch weil dieses Netzwerk mehr Knoten enthält als jenes der Insel Elba. Die durchschnittlichen Weglängen beider Netzwerke sind insgesamt sehr viel kleiner als jene für das gesamte World Wide Web, wenn dabei die Webseiten als Knoten betrachtet werden. Dies kann dadurch erklärt werden, dass das Netzwerk des World Wide Web sehr viel größer ist als die beiden Vergleichsnetzwerke und deshalb auch die Distanzen zwischen zwei Knoten sehr viel weiter sind. Werden allerdings die Websites des World Wide Webs als Knoten herangezogen, so sind die durchschnittlichen Weglängen ungefähr gleich. Der Unterschied in der durchschnittlichen Weglänge, der bei differenzierter Betrachtung von Websites und Webseiten des World Wide Webs zu Tage tritt, ist also einerseits auf die Größe des Netzwerks zurückzuführen, andererseits aber auch auf die Verkürzung der Wege im Netzwerk, wenn nur Websites betrachtet werden. Dadurch, dass es viel mehr Webseiten als Websites gibt und dadurch, dass mehrere Links zwischen zwei Websites immer nur einen Weg bedeuten, diese Links auf Webseitenebene aber mehrere Wege bedeuten und somit die Wahrscheinlichkeit, einen Weg zwischen zwei Knoten zu finden sinkt, werden die durchschnittlichen Wege länger, wenn die Webseiten als Knoten des Netzwerks betrachtet werden.

Die so genannte Kleine-Welt-Netzwerkcharakteristik verlangt, wie in Abschnitt 3.3.3 bereits erklärt wurde, ziemlich kurze durchschnittliche Weglängen. Genauer gesagt heißt das, dass

die durchschnittliche Weglänge viel kleiner als die Ordnung des Graphen sein muss und ungefähr wie der Logarithmus der Ordnung des Graphen anwachsen muss. Das heißt:  $L(G) \in O(\log(|V(G)|))$ , wobei  $L(G)$  der durchschnittlichen Weglänge und  $V(G)$  der Ordnung des Graphen entspricht. [65] Die Ordnung des Graphen ist die Anzahl der Knoten, in diesem Fall 2395, der  $\log_{10}$  davon ist 3,38 und der  $\ln$  davon ist 7,78. Damit liegt die durchschnittliche Weglänge im Netzwerk des österreichischen E-Tourismus-Marktes klar im Bereich des Logarithmus der Ordnung des Graphen und weist somit – gleich wie das World Wide Web und das Tourismusnetzwerk der Insel Elba – eine solche so genannte Kleine-Welt-Netzwerkcharakteristik auf.

### 6.3.4.3 Der Clusterkoeffizient

Beim Clusterkoeffizienten sind die Werte zwischen dem Netzwerk des österreichischen E-Tourismus-Marktes und jenem der Tourismusanbieter der Insel Elba in Tabelle 18 allerdings ziemlich unterschiedlich.

Der Clusterkoeffizient ist für Kleine-Welt-Netzwerke relativ hoch (vgl. dazu Abschnitt 3.3.3). Der Clusterkoeffizient für das Netzwerk des österreichischen E-Tourismus-Marktes weist, wie in Tabelle 18 zu sehen ist, mit 0,341 einen relativ hohen Wert auf. Dies ließe auf ein Kleine-Welt-Netzwerk schließen, in dem der Clusterkoeffizient auch unabhängig von der Ordnung des Netzwerks ist. Jedoch sinkt der Clusterkoeffizient des Netzwerks des österreichischen E-Tourismus-Marktes mit zunehmender Ordnung. Dies wird dann deutlich, wenn der Clusterkoeffizient für das in Suchdurchlauf 3 ermittelte Netzwerk berechnet wird, welches viermal so viele Knoten enthält wie das hier untersuchte Netzwerk aus Suchdurchlauf 4. Hier ergibt sich ein Wert von 0,264 und dieser ist somit kleiner als jener in Tabelle 18. Dadurch kann beim vorliegenden Netzwerk hinsichtlich des Clusterkoeffizienten auf eine Mischung der Charakteristiken von Kleine-Welt-Netzwerken und skalenfreien Netzwerken geschlossen werden. Der Wert des Clusterkoeffizienten für das World Wide Web ist, wie bereits in Abschnitt 3.4 beschrieben, ähnlich hoch. Einen Unterschied gibt es hier bei den Werten aus Spalte 5 und 6 in Tabelle 18. Es ist zu erkennen, dass der Clusterkoeffizient für die Webseiten der Domain *nd.edu* etwas höher als jener für das gesamte World Wide Web und fast gleich hoch wie jener des österreichischen E-Tourismus-Marktes ist. Dies lässt den Schluss zu, dass zum Beispiel organisations- oder themenbezogene Teilgraphen des World Wide Webs, bestehend zum Beispiel aus den Webseiten oder Websites von nur einer bestimmten Organisation oder zu nur einem bestimmten Thema, einen höheren Clusterkoeffizienten aufweisen als beliebige Teilgraphen des World Wide Webs ohne Beschränkung der Knoten auf ein Themengebiet oder eine Organisation, weil die Webseiten oder Websites eher verwandt miteinander sind. Diese Folgerung würde somit auch auf den hohen Clusterkoeffizienten des österreichischen E-Tourismus-Marktes zutreffen. Der sehr kleine Wert des Clusterkoeffizienten des Netzwerks der Insel Elba wird in [64e] dadurch erklärt, dass es wenig Zusammenarbeit und eine hohe Abgeschlossenheit der Knoten gibt.

Allgemein fällt beim Netzwerk des österreichischen E-Tourismus-Marktes auf, dass je höher der Knotengrad, desto geringer die Clusterkoeffizienten der einzelnen Knoten (siehe Abschnitt 3.2.4.4) sind. Dies trifft für die meisten Knoten zu. Der Grund dafür ist einerseits, dass Knoten mit einem niedrigen Knotengrad auch wenige Knoten in deren Nachbarschaft haben. Dadurch erhöht sich aber die Wahrscheinlichkeit für einen relativ hohen Clusterkoeffizienten. Zum Beispiel soll der Clusterkoeffizient für den Knoten X berechnet werden. Dieser Knoten X hätte zwei Nachbarknoten Y und Z. In diesem Fall müsste nur noch, zusätzlich zu den Verbindungen zwischen Knoten X und Knoten Y, sowie zwischen Knoten X und Knoten Z, die Verbindung zwischen den Knoten Y und Z bestehen und Knoten X hätte bereits den höchst möglichen Clusterkoeffizienten von 1.

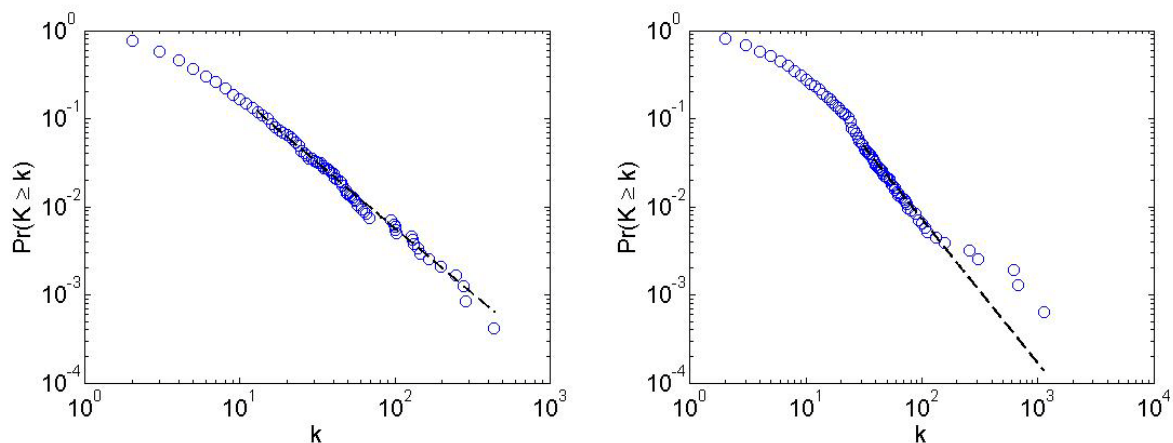


Andererseits haben Knoten mit einem hohen Grad, aufgrund der Potenzgesetzverteilung der Grade (siehe den folgenden Abschnitt 6.3.4.4), viele Verbindungen zu Knoten mit einem sehr niedrigen Grad. Dadurch, dass die Knoten mit niedrigem Grad aber wenige Verbindungen zu anderen Knoten – und somit höchstwahrscheinlich auch keine Verbindungen zu den Nachbarsknoten des Knotens mit hohem Grad – haben, bleibt die Dichte und somit der Clusterkoeffizient im Umfeld eines Knotens mit hohem Grad relativ gering. Weil der Clusterkoeffizient eines Knotens mit hohem Knotengrad nur dann auch hoch sein würde, wenn die Nachbarsknoten (mit meist niedrigem Knotengrad) dieses Knotens auch untereinander verbunden wären.

Insgesamt spricht der hohe Clusterkoeffizient des Netzwerks des österreichischen E-Tourismus-Marktes also für eine stärkere Verbundenheit im lokalen Bereich, als im gesamten Netzwerk. Die schwache Verbundenheit im gesamten Netzwerk konnte bereits aufgrund der insgesamt gesehen relativ niedrigen Dichte im Netzwerk (siehe Abschnitt 6.3.4.1) festgestellt werden. Diese relativ starke Verbundenheit in den lokalen Nachbarschaften des Netzwerks passt auch gut mit der Feststellung aus Abschnitt 6.2.3 zusammen, in der gesagt wurde, dass es, im Gegensatz zu einem großen Knoten im Zentrum, mehrere kleinere Zentren im Netzwerk gibt, in welchen die Dichte relativ hoch ist.

#### 6.3.4.4 Die Verteilung der Ein- und Ausgangsgrade und deren Parameter

Die wichtigste Charakteristik von skalenfreien Netzwerken ist die Existenz von Potenzgesetzverteilungen der Grade der Knoten des Netzwerks (vgl. Abschnitt 3.3.4). Diesem Verteilungsgesetz gehorcht auch das Netzwerk des World Wide Webs. Die Parameterangaben in Tabelle 18 beziehen sich auf den Parameter  $x$  der Formel:  $P(k) = k^{-x}$ , welche die Verteilung beschreibt, wobei  $k$  dem Knotengrad und  $P(k)$  der Wahrscheinlichkeit entspricht, mit der ein Knoten den Grad  $k$  hat.



**Abbildung 59: Verteilung der Knotengrade im Netzwerk des österreichischen E-Tourismus-Marktes. Links die Verteilung der Eingangsgrade und rechts die Verteilung der Ausgangsgrade**

In Abbildung 59 sind die empirischen Wahrscheinlichkeiten der Knotengrade graphisch und mit logarithmierten Werten dargestellt. Die blauen Kreise entsprechen den empirischen Wahrscheinlichkeiten der Grade im Netzwerk des österreichischen E-Tourismus-Marktes. Die gestrichelte Linie in den Diagrammen der Abbildung 59 stellt die am besten an die empirischen Daten angenäherte Form der Potenzgesetzverteilung dar, die nach der Vorgangsweise von [103] mit Hilfe der von [104] zur Verfügung gestellten Matlab-Funktionen [105] ermittelt wurde. Diese am besten geeignete Form der Potenzgesetzverteilung erreicht für die Eingangsgrade einen Parameterwert von 2,47, für alle  $k > 12$  und für die Ausgangsgrade einen Parameterwert von 2,63 für alle  $k > 29$ . Für diese spezifische

Annäherung der Potenzgesetzverteilung an die empirischen Daten gilt, dass eine Potenzgesetzverteilung für alle  $k > 12$  für die Eingangsgrade sowie für alle  $k > 29$  für die Ausgangsgrade nach den ersten zwei Schritten des Verfahrens aus [103] nicht ausgeschlossen werden kann. Das heißt, für diese Werte der empirischen Daten kann annähernd mit Sicherheit eine Potenzgesetzverteilung angenommen werden.

Der Verteilungsparameter für die Eingangsgrade dieses Netzwerks ist etwas höher als jener des World Wide Webs und etwas niedriger als jener des Tourismusnetzwerks der Insel Elba. Ein hoher Verteilungsparameter zeugt von einer eher steileren Geraden und ein niedriger Verteilungsparameter von einer eher flacheren Geraden. Bei einem hohen Verteilungsparameter ist demnach die verhältnismäßige Anzahl von Knoten mit niedrem Grad noch größer und die der Knoten mit höchstem Grad nicht so groß wie bei einem niederen Verteilungsparameter. Die erzielten Verteilungsparameter der beobachteten Netzwerke sind demnach durchaus logisch nachvollziehbar. Denn im ziemlich kleinen Netzwerk der Tourismusanbieter der Insel Elba gibt es sicher nur ein paar wenige Knoten mit hohem Grad und die meisten der restlichen Knoten haben nur einen sehr niedrigen Grad. Deswegen hat dieses Netzwerk den höchsten Verteilungsparameter. Im Netzwerk des österreichischen E-Tourismus-Marktes gibt es schon mehr Knoten und die Anzahl der Knoten mit einem höheren Grad ist auch deshalb höher, weil es ein größeres touristisches Angebot und somit mehr Knoten gibt, die von vielen anderen verlinkt werden (zum Beispiel die Destinationsmanagementportale). Deshalb ist diese Verteilung eher flacher. Im World Wide Web hingegen steigt die Diversität noch weiter an, da das Netzwerk größer und die Themengebiete zahlreicher sind, was auch zu mehr Knoten führt, die von vielen anderen verlinkt werden. Insgesamt weisen jedoch alle drei Netzwerke eine Potenzgesetzverteilung der Eingangsgrade auf, mit dem speziellen Charakteristikum, dass es viele Knoten mit einem geringen Grad und wenige Knoten mit einem hohen Grad gibt. Damit verhält sich das Netzwerk des österreichischen E-Tourismus-Marktes in Bezug auf die Eingangsgradverteilung ähnlich wie die beiden anderen Netzwerke.

Der Verteilungsparameter für die Ausgangsgrade des Netzwerks des österreichischen E-Tourismus-Marktes ist ungefähr gleich groß wie jene des World Wide Webs und um einiges größer als jener des Netzwerks der Insel Elba. Auch hier bestimmt die Höhe des Parameters wieder die Steigung der Geraden. Der sehr niedere Wert beim Netzwerk der Insel Elba lässt – im Gegensatz zu den beiden anderen Netzwerken, deren Ausgangsgrade ungefähr gleich verteilt sind – auf eine gleichmäßigere Verteilung der Ausgangsgrade schließen. Ein Grund dafür könnte sein, dass aufgrund der vergleichsweise geringen Anzahl an touristischen Betrieben auf der Insel Elba, die Knoten mit einem hohen Ausgangsgrad auch nur wenigere andere Knoten verlinken können, als in den beiden anderen Netzwerken. Dadurch ist der Unterschied dieser Knoten zu jenen mit einem geringen Ausgangsgrad nicht mehr so hoch und die Verteilung deswegen gleichmäßiger. Insgesamt kann auch hier festgestellt werden, dass sich das Netzwerk des österreichischen E-Tourismus-Marktes bezogen auf die Ausgangsgradverteilung ungefähr gleich wie die beiden anderen Netzwerke verhält.

Es kann also gesagt werden dass die Verteilung der Knotengrade im österreichischen E-Tourismus-Markt, wenn auch nur annähernd, einer Potenzgesetzverteilung folgt und damit durchaus mit der Verteilung der Knotengrade im World Wide Web vergleichbar ist, weshalb dieses Ergebnis auch dadurch unterstrichen werden kann. Obwohl die in Abschnitt 3.4.1 genannten Literaturquellen, zumindest für die Eingangsgrade, nicht von einer nur annähernd gegebenen Potenzgesetzverteilung im World Wide Web sprechen, bestätigt [103], dass dies doch der Fall ist. Für diese Quelle gilt die Potenzgesetzverteilung für das World Wide Web, so wie auch für viele andere in der Realität existierenden Netzwerke, erst ab einem relativ hohen Knotengrad und sie wird als so genannte abgeschnittene Potenzgesetzverteilung

(Power Law with cut-off) bezeichnet. Dies trifft auch für das Netzwerk des österreichischen E-Tourismus-Marktes zu.

### ***Kolmogorov-Smirnov-Test der Verteilungen für den Bereich vor dem Knotengrad ab dem die Potenzgesetzverteilung gesichert ist***

Da die Potenzgesetzverteilung der Knotengrade erst ab einem bestimmten Knotengrad gilt, die Anzahl der davor liegenden Knoten aber einen nicht unbedeutenden Anteil der gesamten Knotenanzahl darstellt, wurde zusätzlich versucht die Verteilung der Daten in diesem Bereich zu bestimmen.

Mit dem Kolmogorov-Smirnov-Test kann geprüft werden, ob eine vorliegende Stichprobe  $x_1, \dots, x_n$  die Realisation einer Zufallsvariablen mit der Verteilungsfunktion  $F_0(x), x \in \mathbb{R}$  ist. Dabei ist  $F_0$  fest vorgegeben und stetig. Mithilfe einer empirischen Verteilungsfunktion  $F_n(x), x \in \mathbb{R}$  der Stichprobe wird die K-S-Teststatistik gebildet:

$$d_n = \sup |F(x) - F_0(x)|$$

$d_n$  entspricht dem maximalen y-Abstand der beiden Verteilungen. Die Hypothese  $H_0$ , dass die vorliegende Stichprobe die Realisation der verglichenen Verteilungsfunktion ist, wird verworfen, wenn  $d_n > k_{n,1-\alpha}$  ist. Die Quantile  $k_{n,\gamma}$  können einer Tabelle entnommen werden. [109]

Dieser Test wurde mit Hilfe des Programms Matlab [105] durchgeführt. Dazu wurden für die jeweils vermuteten theoretischen Wahrscheinlichkeitsverteilungen, jeweils 100 zufällige Datensätze (Stichproben) erzeugt, welche diesen Wahrscheinlichkeitsverteilungen entsprechen, und nach dem Kolmogorov-Smirnov-Test mit den empirischen Daten verglichen. Getestet wurde auf Exponential-, Pareto-, Potenzgesetz-, Log-Normal-, Poisson- und gedehnte Exponentialverteilung. Obwohl die Nullhypothese von gleicher Wahrscheinlichkeitsverteilung für keine der vermuteten Wahrscheinlichkeitsverteilungen angenommen werden konnte, konnte doch der durchschnittliche (über die 100 Datensätze ermittelte) maximale Abstand  $d_n$  der empirischen Verteilungsfunktionen der beiden verglichenen Daten anhand des Kolmogorov-Smirnov-Tests ermittelt und somit diejenige Wahrscheinlichkeitsverteilung mit dem geringsten maximalen Abstand als die möglichste Wahrscheinlichkeitsverteilung anerkannt werden. Für die Ausgangsgrade war dies immer noch eine Potenzgesetzverteilung, jedoch mit einem Verteilungsparameter von  $k = 1,52$ , was sehr viel niedriger ist als der oben ermittelte Wert und eine eher flachere Verteilungsgerade bedeutet. Für die Eingangsgrade war die annähernd beste Wahrscheinlichkeitsverteilung eine Poisson-Verteilung mit einem Parameterwert  $\lambda = 3,1$ .

### **6.3.4.5 Feststellung der Fliegen-Form**

Zusätzlich zu den oben durchgeführten Vergleichen gilt es noch festzustellen, ob das Netzwerk des österreichischen E-Tourismus-Marktes die so genannte Fliegenstruktur, welche in Abschnitt 3.4.4 für das World Wide Web festgestellt werden konnte, aufweist. Diese Untersuchung wurde auch in [64] für das Netzwerk der Insel Elba durchgeführt. Für diese Analyse konnte aber das in dieser Arbeit verwendete Analysetool UCINET nicht verwendet werden, weswegen hierfür das Programm zur Analyse von großen Netzwerken mit dem Namen Pajek [106] genutzt wurde.

Tabelle 19 zeigt die Aufteilungen der Knoten auf die Komponenten der Fliegen-Form für die verschiedenen Netzwerke, wobei für das Netzwerk des österreichischen E-Tourismus-Marktes zusätzlich zum Netzwerk, das mit Suchdurchlauf 4 festgestellt wurde, auch jenes, welches

sich aus Suchdurchlauf 3 ergeben hat, angeführt ist. Auch muss gesagt werden, dass die Tubes-Komponente hier deswegen nicht angeführt wird, weil sie aufgrund ihrer geringen Größe vernachlässigt werden kann (siehe Abschnitt 3.4.4).

	Hauptkomponente <sup>a</sup>	Eingangskomponente <sup>a</sup>	Ausgangskomponente <sup>a</sup>	Ranken <sup>a</sup>	Nicht verbundene Komponente <sup>a</sup>
Österreichischer E-Tourismus-Markt / Suchdurchlauf 4	63,06%	0,04%	36,9%	0%	0%
Österreichischer E-Tourismus-Markt / Suchdurchlauf 3 <sup>b</sup>	30,55%	28,9%	38,27%	1,83%	0,26%
World Wide Web <sup>c</sup>	28%	21%	21%	22%	9%
Insel Elba <sup>d</sup>	3%	2%	52%	16%	25%

**Tabelle 19: Aufteilung der Knoten der Netzwerke auf die verschiedenen Komponenten der Fliegen-Form.** <sup>a</sup> siehe zur Erklärung des Begriffs Abschnitt 3.4.4, <sup>b</sup> es fehlen 0,19% bei der „Tubes“-Komponente, <sup>c</sup> siehe Tabelle 8 in Abschnitt 3.4.4 <sup>d</sup> aus [64d], es fehlen 2% bei der „Tubes“-Komponente.

In Tabelle 19 fällt auf, dass bei Suchdurchlauf 4, dessen Netzwerk in dieser Arbeit hauptsächlich verwendet wurde, die zwei Komponenten Ranken und nicht verbundene Komponenten gar keine Knoten enthalten. Dafür sind die meisten Knoten dieses Netzwerks in der Hauptkomponente zu finden, was im Vergleich zu den anderen Netzwerken zusätzlich auffällt. Dies ist eigentlich ein Zeichen für eine hohe Verbundenheit des Netzwerks, denn die Knoten der Hauptkomponente können jeden anderen Knoten dieser Komponente über einen Weg erreichen. Die geringen 3% Anteil der Hauptkomponente im Netzwerk der Insel Elba zeugen, wie bereits erwähnt, von der schlechten Verbundenheit dieses Netzwerks.

Eine Begründung für diesen Unterschied im Netzwerk aus Suchdurchlauf 4 ist vielleicht die hohe Verbundenheit innerhalb eines bestimmten Themenbereichs im World Wide Web. Dies ist durchaus nachvollziehbar, denn die Websites eines bestimmten Themenbereichs sind eher stärker miteinander verbunden als die Websites im gesamten World Wide Web. Die ausgeglichene Aufteilung im Netzwerk aus Suchdurchlauf 3 würde diese Vermutung bestätigen, denn wie in Abschnitt 6.1.3 bei der Auswahl des repräsentativsten Netzwerks festgestellt wurde, stellt das Netzwerk aus Suchdurchlauf 3 keine gute Repräsentation eines Netzwerks des österreichischen E-Tourismus-Marktes dar, weil viel zu viele, nicht zum Tourismus gehörende Websites darin enthalten sind. Dies würde also eher dem gesamten World Wide Web entsprechen, weshalb sich die mehr ausgeglichene Aufteilung auf die Komponenten der Fliegen-Form erklären ließe.

Allgemein ist zu sagen, dass nicht zu erwarten ist, dass die Aufteilung der Knoten auf die Komponenten bei den verschiedenen Netzwerken gleichmäßig oder zumindest ähnlich sein muss, was bereits in Abschnitt 3.4.4 festgestellt wurde. Wie bereits erklärt, gibt es andere Untersuchungen des World Wide Webs, welche andere als die in Tabelle 19 angegebenen Werte liefern. Zusätzlich sind auch hier die verwendete Strategie des Crawlers sowie die Unvollständigkeit der gefundenen Netzwerke ausschlaggebende Faktoren dafür, dass nicht dieselben Ergebnisse erzielt werden können.

### **6.3.4.6 Nicht verbundene Knoten**

Ein letzter Vergleich bezieht sich auf die nicht verbundenen Knoten im Netzwerk. Im Tourismusnetzwerk der Insel Elba sind laut [64d] 21% der Knoten überhaupt nicht mit dem Netzwerk verbunden. Nicht verbundene Knoten können jedoch mit einem Crawler nicht gefunden werden, weshalb diese Anzahl beim Netzwerk des österreichischen E-Tourismus-Marktes nicht festgestellt werden konnte. Die Untersuchung der Insel Elba hingegen wurde teils mit einem Crawler, teils manuell durchgeführt.

Es kann nicht vermutet werden, dass im Netzwerk des österreichischen E-Tourismus-Marktes eine derart hohe Anzahl nicht verbundener Knoten besteht. Ein Grund dafür könnte sein, dass Österreich im Bereich der Informationstechnologien allgemein – und damit auch speziell im Bereich des Internets – um einiges fortgeschrittener ist als Italien. Dies wird von [63] bestätigt. Auch könnte sich die Anzahl der nicht verbundenen Knoten im Netzwerk der Tourismusanbieter der Insel Elba inzwischen, aufgrund der vergangenen Zeit und dem damit einhergehenden technologischen Fortschritt, verringert haben.

## **6.4 Bestimmung einer Zentralitätskennzahl zur Ermittlung der zentralen Knoten**

Im vorhergehenden Abschnitt wurden die zentralsten Knoten jeweils unter Berücksichtigung der drei vorgestellten Zentralitätskennzahlen ermittelt. Für die Bestimmung der zentralen Knoten gibt es also mehrere Möglichkeiten. Um die zentralsten Knoten des Netzwerks zu bestimmen, muss zwischen einer der vorgestellten Varianten gewählt werden. Es gilt festzustellen, inwieweit diese wichtigen Kennzahlen der sozialen Netzwerkanalyse auch die Zentralität von Knoten im Netzwerk des World Wide Webs, bestehend aus Websites, die durch Hyperlinks miteinander verbunden sind, bestimmen können.

### **6.4.1 Betrachtungen zur Relevanz der Zwischenzentralität**

Die Zwischenzentralität gibt sicherlich wenig Aufschluss darüber, ob ein Knoten im Netzwerk des World Wide Webs zentral liegt oder nicht. Bei dieser Zentralitätsart werden Knoten bestimmt, die auf kürzesten Wegen zwischen zwei anderen Knoten liegen. Es geht aber im Netzwerk des World Wide Webs nicht um den Fluss von Informationen. Würde es darum gehen, so würde der Knoten über den die meisten kürzesten Verbindungen laufen ziemlich zentral sein. Der Benutzer kommt zu einer bestimmten Website meist nicht mehr über die Links von anderen Websites, sondern über eine Suchmaschine. Deswegen kann eine Website auch nur mehr selten eine Vermittlerrolle spielen und die Zwischenzentralität wird eher unwichtig, auch weil sie beim Ranking der Suchmaschinen keine Rolle spielt.

Allerdings kann ein hoher Zwischenzentralitätswert zusammen mit einer hohen Gradzentralität auftreten und somit bei der Bestimmung der Zentralität eines Knotens im Netzwerk des World Wide Webs trotzdem eine Rolle spielen. Denn um die Wahrscheinlichkeit eines hohen Zwischenzentralitätswerts zu erhöhen, muss ein Knoten sehr viele ein- und ausgehende Links haben, da der Knoten ja seine Vermittlerrolle dadurch ausübt, dass er eingehende Links von Websites erhält und ausgehende Links auf andere Websites setzt und somit Wege im Graphen eröffnet. Je mehr ein- und ausgehende Links ein Knoten hat, desto mehr Wege führen zwischen zwei Knotenpaaren über diesen Knoten und desto höher ist sein Zwischenzentralitätswert. Wenn allerdings ein Knoten viele ein- und ausgehende Links hat, kann er durchaus als zentral erachtet werden, wie später noch gezeigt werden wird.

## 6.4.2 Vergleich mit zwei weiteren Maßzahlen aus der Webometrie und der Netzwerkanalyse

Nun wird zusätzlich versucht, anhand zweier weiterer Maßzahlen zu zeigen, welche der beiden verbleibenden relevanten Zentralitätskennzahlen (Grad- und Nähe-basierte Zentralität), für die Bestimmung der zentralen Knoten in einem Netzwerk des World Wide Webs am relevantesten ist.

### 6.4.2.1 Der Web Impact Factor

In der Webometrie gibt es eine Maßzahl zur Bewertung von Websites die sich Web Impact Factor (WIF) nennt und welche von Peter Ingwersen im Jahr 1998 eingeführt wurde. [66] Diese Maßzahl wurde vom so genannten Impact Factor abgeleitet, welcher die Anzahl der Zitierungen von wissenschaftlichen Publikationen heranzieht um diese Publikationen zu bewerten. Generell gesehen ergibt sich diese Maßzahl aus der Anzahl der Zitierungen einer wissenschaftlichen Publikation in einer Periode  $T_1$  dividiert durch die Anzahl der Publikationen dieses Autors in einer Periode  $T_2$ . [115]

Der Web Impact Factor errechnet sich demnach wie folgt:

$$WIF_i = \frac{\text{Anzahl\_Linkseiten\_auf\_Website}_i}{\text{Anzahl\_Webseiten\_Website}_i}$$

Dies entspricht also der Anzahl der Webseiten welche einen Link auf eine Website  $i$  enthalten (inklusive der Webseiten die zu Website  $i$  gehören) gebrochen durch die Anzahl der Webseiten aus denen Website  $i$  besteht. Das Ergebnis des Web Impact Factors sagt etwas über die Attraktivität einer Website aus. [115] Das heißt, die Attraktivität einer Website in einem Netzwerk hängt zum Beispiel unter anderem von der Anzahl der eingehenden Links, die diese Website erhält, ab.

### 6.4.2.2 Die HITS-Methode

HITS steht für „Hyperlink Induced Topic Search“ und basiert auf einer Methode zur Extrahierung von Information aus der Verlinkungsstruktur im World Wide Web welche von Kleinberg im Jahre 1998 entwickelt wurde. Die spezielle Aufgabe von HITS ist es, so genannte autoritative Webressourcen (Authorities) für bestimmte Themengebiete, nach denen im World Wide Web gesucht wird, zu identifizieren. Zum Beispiel sucht jemand nach dem Begriff „TU Wien“, welcher vielleicht in mehreren tausend Webseiten im Web vorkommt. Nun ist es aber wichtig darunter die Website <http://www.tuwien.ac.at> zu finden, welche sozusagen die am meisten autorisierte Webressource für diesen Begriff ist. Diese autorisierten Webressourcen werden auch als die zentralsten in Bezug auf ein bestimmtes Themengebiet bezeichnet, was für die Bestimmung der relevantesten Zentralitätskennzahl in dieser Arbeit sehr wichtig ist. Das dieser Methode zugrunde liegende Modell des World Wide Webs geht davon aus, dass es neben der Gruppe der autorisierten Webseiten auch eine Gruppe von Webseiten gibt, die auf viele autorisierte Webseiten verweisen - so genannte Hubs - und, dass es zwischen diesen beiden Gruppen eine natürliche Beziehung gibt. [67] und [116]

Zunächst wird ein Teilgraph des World Wide Webs bestimmt anhand dessen dann die Hubs und Authorities ermittelt werden können. Anfangs wurden von Kleinberg nur die eingehenden Links gezählt und die Webseite damit bewertet. Jedoch wurde dabei festgestellt, dass auch Webseiten einen hohen Authorities-Wert erhielten, die nicht mit dem gesuchten Thema oder Suchwort verwandt sind. Aufgrund dessen wird angenommen, dass nicht nur die Authorities-Knoten thematisch zusammenhängen müssen, sondern auch die Hubs-Knoten. Daraus entwickelt sich eine gegenseitige Abhängigkeit die dazu führt, dass thematisch nicht passende Knoten aus dem Teilgraphen heraus fallen. Ein Knoten ist dann ein guter Authorities-Knoten

wenn er von vielen guten Hubs-Knoten verlinkt wird und umgekehrt ist ein Knoten ein guter Hubs-Knoten wenn er viele Authorities-Knoten verlinkt. [116]

Zur Berechnung der Authorities- und Hubs-Werte wird ein iterativer Algorithmus, welcher die Werte berechnet und immer wieder anpasst, angewandt. Jede Webseite im Teilgraphen hat eine Authorities- und eine Hubs-Gewichtung. Es sei  $x^{(p)}$  die Authorities-Gewichtung des Knotens  $p$  und  $y^{(p)}$  die Hubs-Gewichtung des Knotens  $p$ . Beide Gewichtungen werden wie folgt berechnet [116]:

$$x^{(p)} \leftarrow \sum_{q:(q,p) \in G} y^{(q)} \text{ und } y^{(p)} \leftarrow \sum_{q:(q,p) \in G} x^{(q)}$$

$q$  und  $p$  sind jeweils Knoten des vorher ermittelten Teilgraphen  $G$ .  $\sum_{q:(q,p) \in G}$  heißt, dass hier die Summe über alle Knoten  $q$  gebildet wird, die eine Relation (d.h. eine Kante oder einen Link) zu Knoten  $p$  haben. Diese Gewichtungen werden immer normalisiert dargestellt sodass folgendes gilt [116]:

$$\sum_{p \in G} (x^p)^2 = 1 \text{ und } \sum_{p \in G} (y^p)^2 = 1$$

Danach werden diese Gewichtungen iterativ in einem in [116] genannten Algorithmus solange berechnet, bis sich die Gewichtungen in einem „Gleichgewicht“ befinden, d.h. bis ihre Werte nur mehr geringfügig um einen bestimmten Wert herum schwanken bzw. zu diesem Wert hin konvergieren. Einfach gesagt wird für jeden Knoten die Gewichtung berechnet, diese wird dann normalisiert und dieser Vorgang wird solange angewandt bis sich die Werte in einem „Gleichgewicht“ befinden. Schließlich wird noch eine bestimmte Anzahl mit den höchsten Gewichtungen ermittelt, welche dann die Authorities- und Hubs-Knoten sind.

Diese Vorgangsweise zeigt also, dass die Gewichtung oder Bewertung von Webseiten oder Websites im World Wide Web unter anderem mit Hilfe der eingehenden Links erfolgt. Auch handelt es sich bei den Authorities-Knoten, wie in [116] genannt wird um zentrale Knoten eines bestimmten Teilgraphen. [67] nennt diese Knoten auch zu einem hohen Grad referenzierte Knoten. [116] stellt fest, dass das Setzen eines Links auf eine Webseite oft, wenn auch nicht immer, eine bestimmte Bewertung der verlinkten Seite bedeutet. Hier sei es dann auch schwierig das geeignete Gleichgewicht zwischen Popularität und Relevanz, welche beide im Authorities-Konzept eine Rolle spielen, zu finden. Nichtsdestotrotz deutet dies darauf hin, dass, wie auch im vorherigen Abschnitt über den Web Impact Factor festgestellt wurde, die eingehenden Links, und damit die Eingangsgrad-basierte Zentralität für die Ermittlung der zentralen Knoten des Netzwerks des österreichischen E-Tourismus-Marktes sehr relevant ist.

UCINET ermöglicht auch die Berechnung der Hubs- und Authorities-Knoten. In der nachfolgenden Tabelle 20 sind 20 Knoten mit den höchsten Authorities-Werten, ermittelt anhand der Daten aus Suchdurchlauf 4, dargestellt. Tabelle 20 zeigt fast die gleichen Knoten, die auch in den Tabellen 13 und 15 bei den Kennzahlen des Eingangsgrads und der Eingangsnahe-basierten Zentralität hohe Werte erhalten. Diese Ergebnisse zeigen wieder dass, wie bereits festgestellt, vor allem die Eingangsgrade - und aufgrund der Ähnlichkeit bei den Ergebnissen auch die Eingangsnahe-basierte Zentralität – einigermaßen relevant für die Bestimmung der zentralen Knoten dieses Netzwerk Ausgeschlossen werden können hiermit die Zentralitätskennzahlen in denen die ausgehenden Verbindungen eine Rollen spielen (Ausgangsgrad- und Ausgangsnahe-basierte Zentralität).

Website	Authorities-Wert
<i>facebook.com</i>	0.306
<i>tiscover.at</i>	0.232
<i>youtube.com</i>	0.227
<i>tiscover.com</i>	0.209
<i>oebb.at</i>	0.179
<i>addthis.com</i>	0.173
<i>oberoesterreich.at</i>	0.160
<i>austria.info</i>	0.157
<i>tirol.at</i>	0.139
<i>radfahren.at</i>	0.117
<i>twitter.com</i>	0.110
<i>oeamtc.at</i>	0.108
<i>wandern.at</i>	0.101
<i>tagung.info</i>	0.098
<i>flickr.com</i>	0.097
<i>gesundheitsurlaub.at</i>	0.096
<i>ausflugstipps.at</i>	0.093
<i>oberoesterreich-tourismus.at</i>	0.092
<i>biken.at</i>	0.092
<i>bahn.de</i>	0.091

**Tabelle 20: Reihenfolge Authorities-Knoten aus Suchdurchlauf 4 mit den 20 höchsten Werten**

### 6.4.3 Schlussbetrachtungen und logische Folgerungen zur Ermittlung der relevantesten Zentralitätskennzahl

Nun stellt sich allerdings die Frage, welche dieser beiden Maßzahlen für die Auswahl der zentralsten Knoten verwendet werden soll. Ist ein Knoten zentraler wenn er von seinen direkten Nachbarn mehr Links erhält oder wenn er insgesamt, im Vergleich zu allen Knoten im Netzwerk am nächsten liegt? Allgemein gesehen ist sicherlich die Maßzahl der Eingangsnähe-basierten Zentralität ein besserer Maßstab für Zentralität. Einerseits weil es sich hierbei um ein globales Zentralitätsmaß handelt und alle Knoten des Netzwerks berücksichtigt werden. Die Maßzahl scheint deswegen objektiver zu sein. Andererseits sollte ein Knoten mit hohem Eingangsnähe-basierten Zentralitätswert auch optisch zentraler im Netzwerk liegen, wenn die Distanzen zu allen anderen Knoten gering sein sollen.

Da aber bei den in diesem Abschnitt vorgestellten Maßzahlen des Web Impact Factors und der Hubs and Authorities nur die eingehenden Links der direkten Nachbarn gezählt werden, kann davon ausgegangen werden, dass die Maßzahl der Eingangsgradzentralität den wichtigeren Stellenwert bei der Ermittlung der Zentralität einer Website innerhalb eines Netzwerks des World Wide Webs hat.

Nun werden noch vier einleuchtende Gründe genannt wieso das so ist:

1. Wenn der Link zu einer viel besuchten Website Y auf einer bestimmten Website X gesetzt werden soll, so verweist dieser Link direkt von der Website X auf die Website Y und nicht zuerst auf eine dritte Website Z, welche dann erst auf Website Y verweist. Anders als in der Netzwerktheorie, wo es oft um Transport- oder Informationswege von einem Knoten zum anderen geht und wo es noch einen Unterschied macht, ob ein Knoten zwei oder acht Kanten entfernt ist, zählt in einem Netzwerk aus Websites die direkte Verlinkung am meisten. Denn der Benutzer des World Wide Webs hat ja keinen Plan des Netzwerks vor sich, der ihm zeigt, über welche Websites er zu einer



bestimmten Website kommt. Um eine Website zu finden, klickt sich der Benutzer üblicherweise nicht von Seite zu Seite durch, sondern benützt eine Suchmaschine und in der Suchmaschine findet er dann die Website selbst oder eine andere Website, die darauf seinen Link hat.

2. Wenn sich der Benutzer eine Website anschaut und die Links, welche auf andere Websites verweisen sieht, kann er nur die direkten Links wahrnehmen. Der Benutzer denkt dann möglicherweise, dass die Website, die hier verlinkt ist, für ihn von Bedeutung sein könnte. Da der Benutzer aber nur die direkten Links sehen kann, kann er auch nur die direkt verlinkten Websites als für ihn bedeutend erachten, und nicht jene Websites, die zwei oder drei Links entfernt sind.
3. Zusätzlich kann es, wie schon festgestellt wurde, durch die Benennung der Websites als Knoten anstatt der Webseiten oft schwierig sein, die Wege zu anderen Websites zu finden, weil der eingehende Link innerhalb der Website auf eine andere Webseite zeigt als auf jene, von welcher der ausgehende Link wegführt. Auch deshalb liegt eine Website im World Wide Web zentraler wenn sie viele eingehende Links erhält als wenn die Wegdistanz von allen Knoten zu ihr minimal ist.
4. Auch die Suchmaschine Google verwendet laut [68] unter anderem die Anzahl der eingehenden Links, um eine Website zu bewerten.

Es gilt also für diese Arbeit als gesichert, dass die Werte der Eingangsgrade am besten bestimmen, ob ein Knoten im Netzwerk des World Wide Web zentral liegt oder nicht. Als Auswahl der zentralsten Websites, die im nächsten Abschnitt beschrieben werden, wurden deshalb die ersten sechs Websites mit den höchsten Eingangsgraden gewählt. Diese sind:

- *facebook.com*
- *tiscover.com* oder *.at*
- *youtube.com*
- *oebb.at*
- *genussland.at*
- *twitter.com*

## **6.5 Beschreibung der zentralen Knoten**

In diesem Abschnitt folgt nun die Beschreibung der zentralsten Websites des Netzwerks des österreichischen E-Tourismus-Marktes. Zum einen werden das hinter der Website stehende Unternehmen sowie die Funktionen, welche die Website anbietet, beschrieben, und zum anderen wird versucht zu ergründen, weshalb gerade diese Website im Netzwerk an zentraler Stelle steht.

### **6.5.1 Facebook**

Facebook ist ein Portal für die Bildung und Unterhaltung von sozialen Netzwerken. Ursprünglich wurde Facebook für Studenten der Harvard Universität in den USA entwickelt. Von dort aus breitete sich das Portal auf andere Gesellschaftsgruppen und andere Länder aus. Facebook zählte im Juli 2010 weltweit 500 Millionen Nutzer. [69]

Die in der Folge aufgelisteten Funktionen die das Portal anbietet, können auf der Website von Facebook [70] gefunden werden:

- Jeder Benutzer kann ein Profil anlegen, auf dem er sich vorstellen und Fotos sowie Videos hoch laden kann.

- Jeder Benutzer kann andere Benutzer auf Anfrage als „Freunde“ deklarieren und sie einer Liste von Freunden hinzufügen.
- Auf einer Pinnwand des Profils kann der Nutzer selbst oder andere Benutzer Nachrichten hinterlassen, Fotos hoch laden und Nachrichten kommentieren.
- Zu diesen Nachrichten kann jeder Benutzer sagen, ob er sie mag oder nicht.
- Durch eine Beobachtungsliste können Nutzer Neuigkeiten auf den Pinwänden der „Freunde“ sehen.
- Zusätzlich kann der Nutzer privat an andere Benutzer Nachrichten schicken oder mit ihnen chatten.
- Andere Benutzer können zu Gruppen und Events eingeladen werden
- Facebook verfügt auch über einen Marktplatz, auf dem Benutzer Kleinanzeigen aufgeben können.

Es gibt mehrere Gründe, weshalb Facebook eine der zentralen Webseiten des österreichischen E-Tourismus Netzwerkes ist.

Wegen seiner zahlreichen Nutzer hat Facebook einen hohen Bekanntheitsgrad. Dies lässt sich sehr gut für Werbe- und Marketingzwecke nutzen. In Facebook können nicht nur Privatpersonen, sondern auch Unternehmen wie etwa Hotels einen Benutzer anlegen. Auf der Hotelwebsite können die Anbieter dann auf die dazu gehörige Facebook-Fanseite des Hotels verwiesen werden und den Benutzer einladen, ein „Freund“ dieser Institution zu werden. Auf dieser Fanpage kann sich der Betrieb oder das Hotel vorstellen. Der Betrieb hat somit Zugang zu einer breiten Masse und kann durch das Gewinnen von „Freunden“ auch an Beliebtheit dazu gewinnen. Wenn nun ein Facebook-Nutzer sieht, dass seine „Freunde“ zu den „Freunden“ der Fanpage des Betriebs zählen, so wird das diesen Nutzer auch positiv beeinflussen.

Auch kann jeder Facebook-Benutzer vorgefertigte Informationen über den Betrieb oder das Hotel durch einen Klick auf seiner eigenen Facebook-Profilseite veröffentlichen. Dies erhöht das Werbepotential des Betriebes, da jeder Facebook-Benutzer dadurch Werbung für den Betrieb oder das Hotel machen kann.

Zusätzlich kann ein Facebook-Nutzer mit einem Klick auf den „Like“-Knopf von Facebook auf der Website des Hotels oder des Unternehmens äußern, ob er diese Seite bzw. den dazugehörigen Betrieb auf Facebook mag. Diese Möglichkeit kann die Popularität eines Hotels oder Unternehmens steigern oder auch verringern.

All diese getätigten Aktionen werden sofort auf der Profilseite jedes Benutzers und somit auch für seine „Freunde“ sichtbar. Jedes Unternehmen hat durch Facebook eine zusätzliche potentielle Werbequelle, die nichts kostet und bei 500 Millionen potentiellen Kunden sehr beliebt ist. Dadurch, dass Facebook so viele Benutzer hat und allgemein so beliebt ist, ist es auch für die Unternehmen im Tourismusbereich ein Muss sich damit zu vernetzen, um zu zeigen, dass man „up-to-date“ ist. Gleichzeitig besteht für das Unternehmen auch die Gefahr, seine Beliebtheit vor vielen Menschen zu verlieren, wenn es sich schlecht präsentiert oder die Dienstleistung, die es anbietet, nicht zufrieden stellend ist und das Unternehmen deswegen von den Nutzern schlecht bewertet wird. Die Präsenz auf Facebook und das Potential, das daraus resultiert, ist für das Unternehmen jedoch wichtiger als die möglichen Verluste, die durch eine negative Bewertung entstehen könnten.

## 6.5.2 Tiscover

Tiscover ist eines der weltweit ersten Tourismus-Webportale. Das 1991 gegründete Unternehmen, das 1995 online ging, hat sich vom Tourismus-Informationssystem für Tirol zu einem Reiseportal für den gesamten Alpenraum entwickelt und zählt mittlerweile weltweit zu den erfolgreichsten Portalen seiner Art. Tiscover beinhaltet ca. 25.000 Unterkünfte in den alpinen Gegenden der Länder Österreich, Deutschland, Schweiz und Italien und hatte im Jahr 2008 23 Millionen Nutzer, 420 Millionen Seitenaufrufe und über eine Million Anfragen und Buchungen. [71]

Tiscover bietet folgende Funktionen an:

Für Konsumenten, zu finden auf der Website des Unternehmens [72]:

- Informationen zu Urlaubsregionen, Urlaubsorten und Unterkünften
- Buchungsmöglichkeiten für Unterkünfte mit der Option, erst später direkt in der Unterkunft zu bezahlen
- Informationen zu Schigebieten und Badeseen
- Online Routenplaner
- Aktuelle Wetterinformationen
- Livecams
- Bildergalerien
- MyTiscover mit der Speicherung von persönlichen Daten für die Buchung, einer Merkliste und Speicherung von Formulardaten zur Erhöhung der Benutzerfreundlichkeit

Für Anbieter, zu finden auf der eigenen Website für Firmenkunden [71]:

- Online Präsentation der Unterkunft, Destination, Hotelgruppe, des Schigebietes oder Campingplatzes mit Buchungsmöglichkeit
- Möglichkeit der selbstständigen Wartung der Inhalte (Texte und Bilder) für den Kunden
- Möglichkeit zur Werbung für den Kunden, sowohl auf der Portalseite als auch extern bei Partnern von Tiscover
- Entwicklung einer eigenen Online-Präsenz mit Buchungsmöglichkeit

Die Gründe für die Zentralität des Portals *Tiscover.com* im Netzwerk des österreichischen E-Tourismus-Marktes sind auch hier wieder vielfältig.

Die oben genannten Zahlen zeigen dass Tiscover eines der wichtigsten, wenn nicht das wichtigste Tourismusportal für den Alpenraum ist.

Viele Unternehmen setzen einen Link auf Tiscover, weil ihre gesamte Website oder die Buchungsmöglichkeit von Tiscover entwickelt wurde, oder aber die Buchung nur über Tiscover möglich ist und deswegen ein Link dorthin zu finden ist.

Der hohe Bekanntheitsgrad von Tiscover und die damit zusammenhängende hohe Kundenanzahl sind sicher ein Grund, weshalb sich Unternehmen mit Tiscover verbinden. Dadurch, dass Tiscover eines der ersten Reiseportale weltweit war, hat das Unternehmen inzwischen aber auch sehr viel Erfahrung und Wissen über Tourismusbereich gesammelt. Neben der Beliebtheit des Portals ziehen also auch die Erfahrung und das Wissen von Tiscover viele Kunden an.

Mit der Zeit hat Tiscover die Funktionalität des Portals derart erweitert, dass das Unternehmen damit fast alle Bedürfnisse des Reisenden und der Tourismusbetreiber abdeckt. Dies ist sicher auch ein Grund dafür, dass das Unternehmen die hohe Beliebtheit halten und sogar noch ausbauen kann.

Es ist also der Mix aus Erfahrung, Wissen, Service und bereits bestehendem Bekanntheitsgrad, der Tiscover ins Zentrum des österreichischen E-Tourismus-Netzwerks rückt. Der Vorteil gegenüber lokalen, regionalen und nationalen Destinationsmanagement-Organisationen liegt sicherlich in der länderübergreifenden Tätigkeit und der damit zusammenhängenden Größe und dem Bekanntheitsgrad des Portals sowie dem Umfang der angebotenen Funktionalität.

### **6.5.3 Youtube**

Youtube ist ein Internet-Videoportal, das 2005 in Kalifornien gegründet wurde und welches mittlerweile die beliebteste Plattform für Online-Videos ist. Jeder Internet-Nutzer kann auf Youtube Film- und Fernsehausschnitte, Musikvideos und selbst gedrehte Filme anschauen und hoch laden. Im November 2006 wurde Youtube von Google übernommen. Heute werden täglich hunderttausende von Videos auf Youtube hoch geladen und etwa 2 Milliarden Mal ein Video von Youtube abgespielt.

Die Funktionen von Youtube sind folgende:

- Videos mit maximal 15 Minuten Länge hoch laden
- Videos ansehen
- Mit einem Nutzerkonto kann ein Profil angelegt und eine Favoritenliste erstellt werden, die von anderen Benutzern eingesehen werden kann

Die Informationen zur Geschichte, Nutzerzahlen und Funktionalität von Youtube wurden der Website des Unternehmens [73] entnommen.

Viele Hotels und Tourismusportale haben Videos von ihrer Unterkunft oder der Region auf Youtube hoch geladen und auf ihre Website einen Link gesetzt, der auf das Video verweist.

Ein Grund, wieso sich viele Tourismusbetreiber mit Youtube verlinken ist sicherlich der hohe Bekanntheitsgrad des Videoportals. Da Youtube sehr viele Nutzer hat, die auch potentielle Kunden sind, ist es vorteilhaft sein Video dort hin zu stellen. Dies ermöglicht wieder – ähnlich wie bei Facebook – eine Art Gratis-Werbung für das Unternehmen. Durch diese Form der Werbung können vor allem Kunden gewonnen werden, welche nicht die Website eines Tourismus-Unternehmens besuchen. Durch die Bewertungs- und die Kommentarmöglichkeit von Videos besteht wieder ein Werbepotential durch die Nutzer selbst. Und durch die Anzahl der Aufrufe kann der Benutzer auch sehen wie beliebt das Video und somit auch der Beherbergungsbetrieb oder die Region ist.

Wie bei Facebook, kann dies natürlich auch zu einem Nachteil werden, wenn das Video negativ bewertet wird oder nur wenige Nutzer es sich ansehen.

Auch zeigt der Tourismusbetrieb durch das hoch laden eines Videos auf Youtube, dass er in der „Social Community“ vertreten ist. Da diese sozialen Netzwerkanwendungen wie eben Facebook, Youtube, Twitter und andere so beliebt sind, ist es fast ein Muss in diesem Bereich vertreten zu sein.

### **6.5.4 ÖBB**

Die Website *oebb.at* ist die Seite der österreichischen Bundesbahnen und bietet folgende Funktionen an:

- Fahrplanabfrage
- Online-Ticket
- Informationen zu diversen Aktionen der ÖBB
- Webshop mit Werbeartikeln der ÖBB
- Streckeninformationen (Baustellen und Verkehrsbehinderungen)
- Links zu den diversen Bereichen oder Untergruppen der ÖBB (Technische Services, Immobilienmanagement, Infrastruktur AG, Personenverkehr AG usw.)
- Einen Bereich für Jobs bei der ÖBB
- Einen B2B Bereich
- Eine Suchfunktion
- Kundensupport

Dies sind die wichtigsten Funktionen, welche die Website der ÖBB bietet. Sie wurden von der Website des Unternehmens [74] entnommen.

In diesem Fall sind die Funktionen der Website nicht ausschlaggebend für die Zentralität der Website im Netzwerk des österreichischen E-Tourismus. Die Zentralität der Website hängt mit der Wichtigkeit der Bahn als Transportmittel zusammen. Da die meisten Gäste in Österreich aus Österreich selbst kommen, sind die Transportmittel Auto und Bahn wichtiger als das Flugzeug. Eine Anfahrt mit dem Bus ist eher unwesentlich, da mit dem Bus oft nur kürzere Distanzen bewältigt werden, die Urlaubenden aber meist nicht aus derselben Region kommen.

Eine sehr wichtige Information auf der Website eines Tourismusbetriebes ist dessen Erreichbarkeit. Um möglichst alle Gäste anzusprechen, müssen auch alle möglichen Transportmöglichkeiten angegeben werden. Die Bahn steht zum Beispiel wegen der oben genannten Gründe zentraler im Netzwerk als die Website eines Flughafens oder eines öffentlichen Busbetreibers. Als Hilfe zur Erreichbarkeit mit dem Auto können Websites von Routenplanern angegeben werden. Manchmal werden entweder nur statische Karten oder sonst mehrere verschiedene Routenplaner angegeben, wogegen es nur einen Anbieter für das Transportmittel Zug gibt. Deswegen steht die Website der ÖBB zentraler im Netzwerk als die anderen Anbieter aus dem Transportbereich, allen voran die Routenplaner.

Meistens verweist ein externer Link von der Seite des Tourismusbetriebs auf die Seite der österreichischen Bundesbahnen oder direkt auf die Fahrplanabfrage auf dieser Seite. Die wichtigste Funktion der Website der ÖBB ist also zweifelsohne die Fahrplanabfrage, weil der Gast wissen möchte, wann und wie er zum Urlaubsort reisen kann.

### **6.5.5 Genussland.at**

*Genussland.at* ist neben anderen Websites wie *oberoesterreich.at*, *gesundheitsurlaub.at*, *radfahren.at*, *wandern.at*, *biken.at*, *kultururlaub.at*, *pferdeurlaub.at*, *wintersport.at* u.a. eine Website der Destinationsmanagement-Organisation für Oberösterreich. Sie bietet folgende Funktionen an:

- Zimmersuche mit Buchungsmöglichkeit
- Diverse Informationen zu Reisemöglichkeiten, Ausflugstipps, Gastronomie, Veranstaltungen, Kochrezepte und Kochkurse

- Eine interaktive Karte, wo der Benutzer Gastgeber, Gastronomiebetriebe, Veranstaltungen und Ausflugstipps finden kann
- Newsletter-Funktion
- Prospektbestellung
- Videos

Dies sind die wichtigsten Funktionen, welche die Website *genussland.at* bietet. [75]

Nun gilt es die Frage zu beantworten, weshalb diese Website im Netzwerk sehr zentral liegt. Diese Website erhält, wenn die Kennzahl des Eingangsgrads in Tabelle 10 betrachtet wird, 30 eingehende Links mehr als die Website *oberoesterreich.at*. Der Unterschied ist nicht allzu groß, wenn beachtet wird, dass *genussland.at* 167 eingehende Links erhält, während es bei *oberoesterreich.at* nur 138 sind. Das heißt möglicherweise, dass Oberösterreich eine wichtige Tourismusregion ist, muss aber nicht unbedingt so sein, denn das gefundene Netzwerk ist nicht vollständig und die beiden Adressen *oberoesterreich.at* und *genussland.at* stehen auf der Startliste der Adressen beim Durchsuchen mit dem Crawler weit vorne. Das heißt, die Websites, die damit verlinkt worden sind, wurden eher durchsucht als Websites, die mit Adressen verlinkt sind, welche in der Startliste weiter hinten bzw. gar nicht darauf stehen. Die Frage der Wichtigkeit der Tourismusregion Oberösterreich, unter Beachtung der Zahl der Nächtigungen, wird von einer Studie, die in [76] genannt wird, beantwortet: Darin ist Oberösterreich nur eine mittelmäßig besuchte Region.

Die Anzahl der eingehenden Links, welche eine Destinationsmanagementsite erhält, muss allerdings nicht unbedingt allein die touristische Popularität der vertretenen Region widerspiegeln. Es spielen vor allem auch die Größe der Region und die Anzahl der angeschlossenen Beherbergungsbetriebe eine Rolle. Je größer eine Region, desto mehr Beherbergungsbetriebe kann es geben und desto mehr Betriebe können einen Link auf die ihnen zugeordnete Destinationsmanagement-Organisation setzen.

Wenn nun der Unterschied in der Anzahl der erhaltenen eingehenden Links zwischen den Websites *genussland.at* und *oberoesterreich.at* betrachtet und erklärt werden soll, so kann folgendes behauptet werden: Das Konzept und die Website von *genussland.at* ist besser als jenes von *oberoesterreich.at*. Während *oberoesterreich.at* nur eine bestimmte Region vertritt, verbindet *genussland.at* eine Region mit einer Art Lebensphilosophie: dem Genuss. Diese hedonistische Konnotation scheint heutzutage sehr attraktiv zu sein. Alles auf der Seite *genussland.at* wird in Verbindung mit Genuss gebracht. Dies zeugt von einer hohen Integrität. Auch enthält die Seite *genussland.at* im Vergleich zu *oberoesterreich.at* weniger Menüpunkte und -unterpunkte. Das macht die Navigation einfacher und zieht die Besucher eher an.

Die Anzahl der eingehenden Links bzw. die Zentralität einer Destinationsmanagement-Site kann zwar von hoher Popularität der Website bzw. der vertretenen Destination zeugen, es muss aber nicht so sein. Hier spielen, wie oben beschrieben, weitere Faktoren eine Rolle (mehr dazu auch in Abschnitt 6.6). In diesem Fall wird davon ausgegangen, dass die hohe Zentralität der Website *genussland.at* – und in diesem Zusammenhang auch jene von *oberoesterreich.at* – im Vergleich zu den Websites anderer Destinationsmanagement-Organisationen vor allem mit der Wahl der Startadressen und der Unvollständigkeit des Netzwerkes zusammenhängt. Aufgrund der Unvollständigkeit des Netzwerkes wird angenommen, dass zum Beispiel mehr Websites von Tourismusbetrieben aus Oberösterreich als aus der Steiermark vertreten sind und *genussland.at* deswegen zum Beispiel zentraler liegt als etwa *steiermark.com*.

## 6.5.6 Twitter

Twitter ist eine Anwendung zum Mikroblogging und zum verbreiten von Nachrichten. Gegründet wurde Twitter 2006 in den USA, sein Marktwert betrug 2009 über eine Milliarde Dollar. Im April 2010 hatte Twitter weltweit 105,7 Millionen Nutzer. Neben der privaten Anwendung ist Twitter hauptsächlich zu einem Werkzeug für Marketing und Werbung von Unternehmen geworden. Die Nachrichten, die in Twitter verfasst werden, heißen „Tweets“. Ein Twitter-Nutzer kann sowohl „Follower“ als auch „Followed“ sein. Zum „Follower“ wird ein Nutzer, indem er die Beiträge eines anderen Nutzers abonniert. Zum „Followed“ wird ein Nutzer, wenn andere Nutzer seine Beiträge abonnieren. [77]

Twitter bietet folgende Funktionen:

- Nachrichten mit maximal 140 Zeichen verfassen und weitergeben – sie können dann von den „Followern“ gelesen werden
- Anderen Nutzern folgen, das heißt deren Nachrichten abonnieren
- Nachrichten anderer kommentieren
- Leute suchen und finden
- Listen mit abonnierten Beiträgen erstellen

Die Funktionen können der Website des Unternehmens entnommen werden. [78]

Nun stellt sich die Frage, wieso Twitter im Netzwerk der österreichischen Tourismustreibenden sehr zentral liegt.

Durch Twitter können Tourismusunternehmen Informationen über ihre Produkte an Kunden weitergeben, Werbung betreiben und mit Hilfe des Feedbacks Produkte leichter an Kundenwünsche anpassen. Nach Einsicht in die Twitter-Profile von mehreren Tourismusanbietern gehören zu den übermittelten Nachrichten vor allem Informationen über den Urlaubsort, Veranstaltungen, Aktionen oder Informationen darüber, dass auf der zum Betrieb gehörenden Facebook-Seite Fotos hochgeladen wurden. Der Tourismusanbieter verspricht sich dadurch – ähnlich wie bei Facebook und Youtube – seinen potentiellen Kundenkreis zu erweitern. Denn auf Twitter kann der Benutzer zufällig auf das Profil des Betriebs stoßen oder es in einer der angebotenen Kategorien finden. Auch kann der Tourismusanbieter seine bestehenden Kunden besser halten und stärker an sich binden, indem er sie ständig mit Produktinformationen auf dem Laufenden hält.

Trotz der angepriesenen Beliebtheit von Twitter ist die Anzahl der „Follower“, also jener Personen, die das Profil des Betriebes auf Twitter abonniert haben und somit die Nachrichten lesen können, bei den meisten Betrieben eher gering. In den meisten Fällen gibt es etwa 200 „Follower“, nur *tirol.at* erreicht ein Maximum von etwa 1100 „Followern“. Das heißt, dass das Interesse der Kunden nicht allzu groß ist.

Die niedrige Zahl der „Follower“ von Betrieben aus dem Tourismusbereich führt zu der Frage nach dem Nutzen von Twitter und demzufolge auch nach dem Grund für die Zentralität von Twitter im Netzwerk. Der Aufwand für die Pflege der Kundenbeziehungen via Twitter ist nicht groß und kann sich schon beim erreichten Kauf von wenigen Produkten lohnen. Auch eignet sich das Netzwerk sehr gut zur Weitergabe von Produktinformationen und Neuigkeiten, da ja anders als in Facebook, einfach Nachrichten weitergegeben werden. Möglicherweise handelt es sich auch einfach nur um eine Modeerscheinung, die der allgemeinen Beliebtheit der sozialen Netzwerkanwendungen zu verdanken ist. Schließlich gibt es auch Bestätigungen für die Wirksamkeit von Twitter. Laut [79] spricht die Firma Dell von einem Einkommenszuwachs von einer Million Dollar innerhalb eines Jahres, der durch die Anpreisung ihrer Produkte via Twitter erfolgt sein soll.

## 6.6 Zusätzliche Ergebnisse und Beobachtungen

Im gesamten Netzwerk sind zusätzlich zu *oebb.at* noch weitere Knoten vertreten, die den Transportbereich betreffen. Dazu gehören unter anderen Google Maps, diverse Flughäfen, der Postbus und die deutsche Bahn. Nach der ÖBB liegt Google Maps am zentralsten. Der Grund, wieso Google Maps nicht zentraler liegt, ist, dass nicht alle Betriebe einen Routenplaner auf ihrer Website und einige einen alternativen Routenplaner haben. Wieso *oebb.at* von den Transportunternehmen am zentralsten liegt wurde bereits in Abschnitt 6.5.4 erklärt. Nach *oebb.at* und Google Maps steht bei den Transportunternehmen die Site *bahn.de* an dritter Stelle. Der Grund dafür ist, dass nach den Österreichern die Gäste aus Deutschland die zweitgrößte Zielgruppe für den österreichischen Tourismus sind.

Zu den Destinationsmanagement-Organisationen wie *austria.info* und anderen ist zu sagen, dass diese gleich hinter den Top 6 zentralsten Knoten des Netzwerks liegen. Auf Platz 8, 9 und 10 (siehe Tabelle 10) folgen *oberoesterreich.at*, *tirol.at* und *austria.info*. Danach folgen die übrigen regionalen Destinationsmanagement-Organisationen. Außerdem handelt es sich bei einem der zentralsten Knoten, nämlich *genussland.at*, auch um eine Destinationsmanagement-Organisation aus Oberösterreich. Es kann also gesagt werden, dass die Seiten dieser Destinationsmanagement-Organisationen schon recht zentral liegen und sie deshalb auch eine wichtige Rolle im Netzwerk spielen. Sie liegen nicht noch zentraler bei den eingehenden Links, weil ihre Aufgabe hauptsächlich darin besteht, die Benutzer auf die untergeordneten Tourismusbetriebe durch viele ausgehende Links zu verweisen und nicht umgekehrt. Wenn allerdings Tourismusanbieter für die Buchung auf andere Tourismusportale verweisen, so liegt hier Tiscover vor den Destinationsmanagementsystemen. Der Grund dafür wurde bereits in Abschnitt 6.5.2 genannt. Wieso einige regionale Destinationsmanagement-Portale vor *austria.info*, dem bundesweiten Tourismusportal, liegen, ist vielleicht dadurch zu erklären, dass trotz aller Ähnlichkeiten innerhalb des Tourismuslandes Österreich, doch auch große regionale Unterschiede bestehen und es deshalb für die Tourismusanbieter besser ist, sich mit dem regionalen Destinationsmanagement-Portal zusammenzuschließen. Dadurch, dass die meisten Gäste ja aus Österreich oder aus Deutschland kommen, sind die österreichischen Bundesländer bekannt und der Gast identifiziert sich eher mit einer bestimmten Region und deren Eigenheiten. Es ist aber anzunehmen, dass mit der Zunahme von internationalen Gästen auch das gesamtösterreichische Portal *austria.info* noch wichtiger werden wird und deshalb mehr ins Zentrum des Netzwerks rücken könnte. Um allerdings eine noch genauere Reihung der Destinationsmanagementportale zu erreichen, müsste das gesamte Netzwerk des österreichischen E-Tourismus bekannt sein. Außer für *genussland.at* gilt für die vorher beschriebenen zentralsten Knoten mit Sicherheit, dass sie auch bei Vorhandensein des vollständigen Netzwerks im Zentrum stehen würden, da die Websites der sozialen Netzwerkanwendungen sowie die der ÖBB und Tiscover für alle Tourismusanbieter von großer Bedeutung sind. An der Stelle von *genussland.at* könnte bei Bekanntheit des vollständigen Netzwerks jedoch ein anderes Destinationsmanagement-Portal oder ein alternativer Knoten stehen, da die Verteilung der Knoten nach Bundesländern im bestehenden Netzwerk aufgrund der Begrenztheit und der Wahl der Startadressen ungleichmäßig sein könnte.

Wenn die Statistik der ausgehenden Links in Tabelle 13 betrachtet wird, so wird klar, dass die Portale der Destinationsmanagement-Organisationen vor allem auch viele ausgehende Links haben. Am besten ist dies am Beispiel von *austria.info* zu erkennen. Diese Seite hat Links zu allen Regionen, Sub-Regionen und sogar den kleinsten Tourismusbetrieben. Diese Organisationen erfüllen hier also neben der Informationsdienstleistung über die Region oder das Land auch ihren Werbezweck für die einzelnen Betriebe die in dieser Region oder in dem Land tätig sind. Sie verstärken damit die Bedeutung einer Region oder eines Landes auf einer Seite und leiten den Touristen dann durch Links an die untergeordneten Einheiten weiter.



In der vorliegenden Untersuchung hat sich auch heraus gestellt, dass es schwierig ist, das gesamte Netzwerk mit allen im Tourismusbereich tätigen Unternehmen, die eine Website haben, mit einem Crawler zu erreichen. Denn es haben zwar alle Websites der Unternehmen Links auf die zentralen Knoten, jedoch umgekehrt nicht immer die zentralen Knoten auch Links zu den weniger zentralen Knoten. Um aber alle Websites finden und durchsuchen zu können, braucht es Startseiten, welche Links zu allen Websites des österreichischen Tourismus haben, weil der Crawler nur über die Links von einer Website zur anderen wandert. Wenn es also von diesen Startseiten aus keine Links zu den anderen Websites gibt, können diese auch nicht gefunden und durchsucht werden. *Austria.info* ist sicherlich am besten als Startseite geeignet. Warum sind aber nicht alle Tourismusbetriebe als Links in den Seiten übergeordneter (Destinationsmanagement-)Organisationen vertreten, obwohl sie doch von diesen vertreten werden sollten? Vielfach sind die Betriebe zwar nicht durch Links, aber doch über eine Buchungsapplikation mit den Seiten der Destinationsmanagement-Organisationen oder anderen übergeordneten Portalen verbunden.

## 7 Zusammenfassung und Ausblick

In dieser Arbeit wurde eine Analyse der Netzwerkstruktur des österreichischen E-Tourismus-Marktes durchgeführt. Hauptsächliches Ziel der Arbeit war dabei die Bestimmung der zentralen Knoten dieses Netzwerks mit Hilfe der drei wichtigsten Zentralitätskennzahlen der sozialen Netzwerkanalyse (Grad- und Nähe-basierte Zentralität, sowie Zwischenzentralität). Die Daten über das Netzwerk wurden mit Hilfe eines eigens entwickelten Crawlers gesammelt. Da es grundsätzlich nicht möglich ist, das vollständige Netzwerk des österreichischen E-Tourismus-Marktes mit den Websites aller Tourismustreibenden in Österreich mit Hilfe eines Crawlers zu bestimmen (siehe Abschnitt 5.1.1), wurde deshalb aus mehreren Suchdurchläufen mit dem Crawler das repräsentativste Teilnetzwerk ausgewählt, anhand dessen dann die Analyse durchgeführt werden konnte. Die Analyse dieses Netzwerks bzw. die Ermittlung der zentralen Knoten wurde neben anderen Kennzahlen der Netzwerkanalyse vor allem mit Hilfe der drei wichtigsten Zentralitätskennzahlen der sozialen Netzwerkanalyse durchgeführt. Diese Zentralitätskennzahlen waren die Grad-basierte Zentralität (Degree centrality), die Nähe-basierte Zentralität (Closeness centrality) und die Zwischenzentralität (Betweenness centrality). Da diese Zentralitätskennzahlen Zentralität in unterschiedlicher Art und Weise definieren und auch unterschiedliche Ergebnisse liefern, die nicht miteinander kombiniert werden können, sollte zur Bestimmung der zentralen Knoten die für ein Teilnetzwerk des World Wide Webs relevanteste Kennzahl ermittelt werden. Nach eingehender Analyse der Bedeutung dieser Kennzahlen wurde festgestellt, dass die Grad-basierte Zentralität bzw. speziell die Eingangsgrad-basierte Zentralität am besten zur Bestimmung der zentralen Knoten des Netzwerks geeignet ist. Dies deshalb, weil die beiden anderen Kennzahlen (Nähe-basierte Zentralität und Zwischenzentralität) auf der Grundlage von Wegen im Netzwerk berechnet werden, diese in einem Netzwerk des World Wide Webs aber, im Vergleich zu den direkten Verbindungen, welche bei der Grad-basierten Zentralität berücksichtigt werden, nur nebensächlich sind. Auch spielt der Eingangsgrad bei zwei anderen wichtigen Kennzahlen (Web Impact Factor und HITS) zur Bewertung von Websites oder Webseiten im World Wide Web eine wichtige Rolle.

Als die zentralsten Knoten im Netzwerk wurden sechs Websites bestimmt. Diese sind *facebook.com*, *tiscover.com* oder *.at*, *youtube.com*, *oebb.at*, *twitter.com* und *genussland.at*. Die sozialen Netzwerke wie Facebook, Youtube und Twitter stehen deswegen im Zentrum des Netzwerks, weil sie durch ihre hohen Nutzerzahlen und die damit zusammenhängende Popularität sowie den öffentlichen Zugang ein wichtiges Feld für das Marketing von Tourismusbetreibern darstellen. Das Portal der Firma Tiscover steht deshalb im Zentrum, weil die mit Abstand meisten österreichischen Tourismusanbieter (vor allem Hotels) dessen Dienste in Anspruch nehmen und es gleichzeitig auch bei den Konsumenten eines der am meisten genutzten Portale des Tourismus im Alpenraum ist. Die Website der ÖBB steht an zentraler Stelle, weil neben dem Auto die Bahn das am meisten benützte Transportmittel im österreichischen Tourismus ist. Der Grund dafür ist, dass die größte Zielgruppe im österreichischen Tourismus noch immer die Gäste aus Österreich selbst sind. Als sechster zentraler Knoten gilt die Website *genussland.at*. Dieses für Oberösterreich stehende Destinationsmanagement-Portal könnte aufgrund seiner Konzeption und Gestaltung im Gegensatz zu anderen Destinationsmanagement-Systemen eher im Zentrum stehen. Wegen der geringen Unterschiede zwischen diesem und anderen Destinationsmanagement-Systemen beim Eingangsgrad und der Unvollständigkeit des untersuchten Netzwerks könnte jedoch bei einem vollständigen Netzwerk ein anderes Destinationsmanagementsystem oder ein alternativer Knoten zentraler liegen.

Für das gesamte Netzwerk wurde auch ein Vergleich von zusätzlichen Kennzahlen der Netzwerkanalyse, die für bestimmte Netzwerkmodelle typisch sind, einerseits mit dem

Netzwerk der Websites einer anderen touristischen Destination [64] und andererseits mit dem Netzwerk des World Wide Webs durchgeführt. Dieser Vergleich hat gezeigt, dass die Netzwerke insgesamt gesehen ähnliche Charakteristiken aufweisen. Zum Beispiel entspricht die Wahrscheinlichkeitsverteilung der Knotengrade ungefähr einer Potenzgesetzverteilung, so wie es auch bei den Verteilungen der beiden anderen Netzwerke und allgemein bei den skalenfreien Netzwerken der Fall ist. Auch weist das Netzwerk des österreichischen E-Tourismus-Marktes typische Eigenschaften eines Kleine-Welt-Netzwerks (geringe durchschnittliche Distanz und relativ hoher Clusterkoeffizient) auf. Auch in diesen Bereichen unterscheidet sich das untersuchte Netzwerk nur geringfügig von den beiden Vergleichsnetzwerken. Diese Ähnlichkeiten unterstreichen einerseits die Ergebnisse dieser Arbeit, andererseits zeigen sie auch, dass ein Netzwerk eines Teilbereichs des World Wide Webs, nämlich des Tourismus, sich ähnlich verhält bzw. ähnlich strukturiert ist wie das gesamte World Wide Web.

Im Rahmen des Vergleichs mit einem externen Crawler wurde am Schluss der Arbeit noch einmal eine bestimmte Website mit dem eigenen Crawler durchsucht. Dabei wurde im Vergleich zum hier ausgewählten Netzwerk – zumindest für die untersuchte Website – ein besseres Ergebnis mit mehr gefundenen ausgehenden Links erzielt, weshalb davon ausgegangen werden kann, dass die Ergebnisse der Untersuchung noch optimiert werden können. Um dies zu erreichen, wäre es deshalb in Zukunft notwendig, den Crawler noch mehrere Male und auch für längere Zeit laufen zu lassen, auch um ein vollständigeres Netzwerk zu erhalten. Auch die Wahl der Startadressen könnte durch weitere Durchläufe optimiert werden und dadurch zu einem vollständigeren Netzwerk beitragen. Schließlich könnte der Crawler auf einen leistungsfähigeren Rechner betrieben werden, um die Geschwindigkeit zu erhöhen. Geringere Verbesserungen bei der Geschwindigkeit könnten zusätzlich durch Optimierungen im Code des Crawlers erreicht werden. Des Weiteren könnte noch intensiver nach geeigneten Crawlern gesucht bzw. diese getestet werden, um die hier erzielten Ergebnisse zu stützen.

Schließlich gilt es zur Bestätigung der hier erzielten Ergebnisse, weitere Forschungsarbeiten in diesem Bereich abzuwarten. Leider gibt es noch keine gleichen Untersuchungen, weder für Österreich noch international. Einzige Untersuchungen aus dem Bereich Netzwerkanalyse und Tourismus sind in Quelle [64], [80], [81], [82] und [83] zu finden. Sie beschäftigen sich mit der Netzwerkstruktur des Tourismus auf der Insel Elba in Italien, mit jener in Australien und mit der Netzwerkstruktur von Websites von Reisebüros in Italien. Allerdings konnte nur die Studie über die Insel Elba für einen allgemeinen Vergleich der Netzwerkstruktur herangezogen werden, ohne jedoch die zentralen Knoten mit Hilfe der hier vorgestellten Zentralitätskennzahlen zu ermitteln und somit einen Vergleich auch in diesem Bereich zu ermöglichen. Die beiden anderen Studien sind für einen Vergleich mit dieser Arbeit nicht geeignet. Die Studie über Australien beschäftigt sich ausschließlich mit der sozioökonomischen Netzwerkstruktur der Tourismusanbieter und die Studie über die Netzwerkstruktur der Websites der italienischen Reisebüros ermittelt einzig die Wahrscheinlichkeitsverteilung der Knotengrade im Sinne einer Potenzgesetzverteilung, ohne auf zusätzliche Netzwerkeigenschaften näher einzugehen oder die zentralsten Reisebüros zu bestimmen.

Am Ende bleibt noch zu sagen, dass es zum Beispiel interessant wäre die zentralen Knoten dieses Netzwerks nach anderen Kriterien als den Zentralitätskennzahlen der sozialen Netzwerkanalyse zu bestimmen. Von speziellem Interesse wäre es hierbei jene Unternehmen (Knoten des Netzwerks) zu ermitteln die eine führende Rolle - zum Beispiel in Bezug auf die Anzahl der Besuche oder Klicks auf deren Website, die Anzahl der durchgeführten Buchungen, dem Marktanteil oder dem generierten Umsatz - im österreichischem E-Tourismus-Markt spielen. Diese Aspekte können mit einer sozialen Netzwerkanalyse nicht

ermittelt werden, da es sich in diesem Fall vielfach um große internationale Buchungsportale (wie zum Beispiel das Hotelreservierungsservice von HRS oder eine Website eines führenden GDS-Anbieters) handelt welche nicht von den kleineren Tourismusbetrieben verlinkt werden und somit nicht mit einer sozialen Netzwerkanalyse ermittelt werden können. Diese Fragestellungen sollten in zukünftigen Arbeiten berücksichtigt werden um den österreichischen E-Tourismus-Markt besser kennen zu lernen.

## Literaturverzeichnis

- [1] Homepage eTourism Fitness für Oberbayern und Salzburg: eTourism Grundlagen. Online im Internet unter: [http://www.etourism-fitness.com/wiki/index.php?title=1.\\_eTourism\\_Grundlagen](http://www.etourism-fitness.com/wiki/index.php?title=1._eTourism_Grundlagen) (19.01.2011).
- [2] Homepage Web-Tourismus - Tourismus Forschung & Beratung: Tourismus-Definitionen. Online im Internet unter: <http://www.web-tourismus.de/tips/tourismus-definitionen.aspx> (19.01.2011).
- [3] VisionLink Consulting (2002): Market Trends in Tourism. Online im Internet unter: [http://www.taskbc.bc.ca/documents/MktResearchDiscussionPaper\\_000.pdf](http://www.taskbc.bc.ca/documents/MktResearchDiscussionPaper_000.pdf) (19.01.2011).
- [4] Markus Gratzner, Werner Winiwarter, Hannes Werthner (2002): State of the Art in eTourism. Online im Internet unter: <http://homepage.univie.ac.at/werner.winiwarter/secec.pdf> (19.01.2011).
- [5] World Travel & Tourism Council (2011): Economic Impact Data and Forecasts – World – Key Facts at a Glance. Online im Internet unter: [http://www.wttc.org/eng/Tourism\\_Research/Economic\\_Research/](http://www.wttc.org/eng/Tourism_Research/Economic_Research/) (19.01.2011).
- [6] Roman Egger, Jakob Hörl, Brigitte Jellinek, Mario Jooss (2007): Virtual Tourism Content Network TANDEM – A Prototype for the Austrian Tourism Industry. In: Marianna Sigala, Luisa Mich, Jamie Murphy (Hg.): Information and Communication Technologies in Tourism 2007, Seiten 175-184. Wien: Springer Verlag.
- [7] Österreich Werbung (2010): 2009: Insgesamt sinkende Nachfrage, Inlandstourismus im Höhenflug. Online im Internet unter: [http://www.austriatourism.com/xxl/\\_site/int-de/\\_area/465219/\\_subArea/465248/\\_subArea2/481356/\\_id/1258591/marktdaten.html](http://www.austriatourism.com/xxl/_site/int-de/_area/465219/_subArea/465248/_subArea2/481356/_id/1258591/marktdaten.html) (19.01.2011).
- [8] World Travel & Tourism Council (2011): Key Facts at a Glance. Online im Internet unter: [http://www.wttc.org/eng/Tourism\\_Research/Economic\\_Research/Country\\_Reports/Austria/](http://www.wttc.org/eng/Tourism_Research/Economic_Research/Country_Reports/Austria/) (19.01.2011).
- [9] Douglas Pearce, Raewyn Tan, Christian Schott (2004): Tourism Distribution Channels in Wellington, New Zealand. In: International Journal of Tourism Research 6. Jahrgang, Heft 6, Seiten 397-410.
- [10] Douglas G. Pearce, Christian Schott (2005). Tourism Distribution Channels: The Visitors' Perspective. In: Journal of Travel Research 44. Jahrgang, Heft 1, Seiten 50-63.
- [11] Katrin Theos (2006): Multi-Channel-Distribution im Tourismus, Seiten 5-10. GRIN Verlag.
- [12] Hannes Werthner, Stefan Klein (1999): Information Technology and Tourism – A Challenging Relationship, Seiten 42-49. Wien: Springer Verlag.
- [13] Dimitrios Buhalis, Mike Peters (2006): SMEs in Tourism. In: Dimitrios Buhalis, Carlos Costa (Hg.): Tourism management dynamics – trends, management and tools, Kapitel 13, Seite 118. Oxford: Elsevier Butterworth-Heinemann Verlag.
- [14] Stephen Wanhill (2000): Small and Medium Tourism Enterprises. In: Annals of Tourism Research 27. Jahrgang, Heft 1, Seiten 132-147.
- [15] Nevenka Cavlek (2002): Business in Tourism: SMEs Versus MNCs. In: Zagreb International Review of Economics & Business 5. Jahrgang, Heft 2, Seiten 39-48.
- [16] Stephen Wheatcroft (1998): The Airline Industry and Tourism. In: Dimitri Ioannides und Keith G. Debbage (Hg.): The economic Geography of the Tourist Industry – A supply side analysis, Kapitel 9, Seiten 159-161. London: Routledge Verlag.
- [17] Thomas Bieger, Andreas Wittmer (2006): Air transport and tourism – Perspectives and challenges for destinations, airlines and governments. In: Journal of Air Transport Management 12. Jahrgang, Heft 1, Seiten 40-46.
- [18] Aleix Calveras (2003): Incentives of international and local hotel chains to invest in environmental quality. In: Tourism Economics 9. Jahrgang, Heft 3, Seiten 297-306.

- [19] Colin Johnson, Maurizio Vanetti (2005): Locational strategies of international hotel chains. In: *Annals of Tourism Research* 32. Jahrgang, Heft 4, Seiten 1077-1099.
- [20] Paul Slattery, Ian Gamse and Angela Roper (2008): The development of international hotel chains in Europe. In: Michael Olsen, Jinlin Zhao (Hg.): *Handbook of Hospitality Strategic Management*, Kapitel 3, Seiten 41-67. Butterworth-Heinemann Verlag..
- [21] Sofronis Clerides, Paris Nearchou, Panos Pashardes (2008): Intermediaries as quality assessors: Tour operators in the travel industry. In: *International Journal of Industrial Organization* 26. Jahrgang, Heft 1, Seiten 372-392.
- [22] Nevenka Cavlek (2002): Tour Operators and Destination Safety. In: *Annals of Tourism Research* 29. Jahrgang, Heft 2, Seiten 478-496.
- [23] Rita D. Medina-Munoz, Diego R. Medina-Munoz, Juan M. Garcia-Falcon (2003): Understanding European tour operators' control on accomodation companies: an empirical evidence. In: *Tourism management* 24. Jahrgang, Heft 2, Seiten 135-147.
- [24] Adriana Budeanu (2005): Impacts and responsibilities for sustainable tourism: a tour operator's perspective. In: *Journal of Cleaner Production* 13. Jahrgang, Heft 2, Seiten 89-97.
- [25] Martin Barnett und Craig Standing (2001): Repositioning travel agencies on the Internet. In: *Journal of Vacation Marketing* 7. Jahrgang, Heft 2, Seiten 143-152.
- [26] Matthias Kuom und Britta Oertel (1999): Virtual travel agencies. In: *Netnomics* 1. Jahrgang, Heft 2, Seiten 225-235.
- [27] Dong Jin Kim, Woo Gon Kim, Jin Soo Han (2007): A perceptual mapping of online travel agencies and preference attributes. In: *Tourism Management* 28. Jahrgang, Heft 2, Seiten 591-603.
- [28] Nelson Granados, Alok Gupta und Robert J. Kauffman (2003): Orbitz, Online Travel Agents and market structure changes in the presence of technology-driven market transparency. Online im Internet unter [http://misrc.csom.umn.edu/workingpapers/fullpapers/2003/0326\\_070903.pdf](http://misrc.csom.umn.edu/workingpapers/fullpapers/2003/0326_070903.pdf) (19.01.2011).
- [29] Jennifer Xiaoqiu Ma, Dimitrios Buhalis, Haiyan Song (2003): ICTs and Internet adoption in China's tourism industry. In: *International Journal of Information Management* 23. Jahrgang, Heft 6, Seiten 451-467.
- [30] Norbert Vanhove (2005): Globalisation of tourism demand, global distribution systems and marketing. In: Salah Wahab und Chris Cooper (Hg.): *Tourism in the age of globalisation*, Kapitel 6, Seiten 137-139. London: Routledge Verlag. Veröffentlicht in der Taylor & Francis e-Library.
- [31] Tom Bornhorst, J.R. Brent Ritchie, Lorn Sheehan (2010): Determinants of tourism success for DMOs and destinations: An empirical examination of stakeholders' perspective. In: *Tourism Management*, 31. Jahrgang (2010), Heft 5, Seiten 572-589.
- [32] Silvia Sussmann und Michael Baker (1996): Responding to the electronic marketplace: Lessons from Destination Management Systems. In: *International Journal of Hospitality Management* 15. Jahrgang, Heft 2, Seiten 99-112.
- [33] Larry Dwyer, Deborah Edwards, Nina Mistilis, Carolina Roman, Noel Scott (2009): Destination and enterprise management for a tourism future. In: *Tourism Management* 30. Jahrgang, Heft 1, Seiten 63-74.
- [34] Dimitrios Buhalis (1998): Strategic use of information technologies in the tourism industry. In: *Tourism Management* 19. Jahrgang, Heft 5, Seiten 409-421.
- [35] Hannes Werthner, Stefan Klein (1999): ICT and the Changing Landscape of Global Tourism Distribution. In: *Electronic Markets* 9. Jahrgang, Heft 4, Seiten 256-262.
- [36] Dimitrios Buhalis (2004): eAirlines: strategic and tactical use of ICTs in the airline industry. In: *Information & Management* 41. Jahrgang, Heft 7, Seiten 805-825.
- [37] Luis Monteiro, Stuart Macdonald (1996): From efficiency to flexibility: the strategic use of information in the airline industry. In: *The journal of strategic information systems* 5. Jahrgang, Heft 3, Seiten 169-188.

- [38] Alexandros Paraskevas (2005): The impact of technological innovation in managing global value chains in the tourism industry. Im Rahmen der OECD-Conference on Global Tourism Growth: A Challenge for SMEs (2005). Online im Internet unter [http://oxfordbrookes.academia.edu/documents/0009/2938/The\\_Impact\\_of\\_Technological\\_Innovation\\_in\\_Managing\\_Global\\_Value\\_Chains\\_in\\_the\\_Tourism\\_Industry.pdf](http://oxfordbrookes.academia.edu/documents/0009/2938/The_Impact_of_Technological_Innovation_in_Managing_Global_Value_Chains_in_the_Tourism_Industry.pdf) (19.01.2011).
- [39] Laddawan Kaewkitipong (2010): Disintermediation in the Tourism Industry: Theory vs. Practice. In: Matthew L. Nelson, Michael J. Shaw, Troy J. Strader (Hg.): Sustainable e-Business Management, Seiten 160-171. Berlin Heidelberg: Springer Verlag. (Herausgegeben im Rahmen der Reihe: Lecture Notes in Business Information Processing)
- [40] Dimitrios Buhalis, Rob Law (2008): Progress in information technology and tourism management: 20 years on and 10 years after the Internet – The state of eTourism research. In: Tourism Management 29. Jahrgang, Heft 4, Seiten 609-623.
- [41] Susan Briggs (2001): Successful Web Marketing for the Tourism and Leisure Sectors, Kapitel 4, Seiten 36-38. London: Kogan Page Limited Verlag.
- [42] Dimitrios Buhalis, Maria Cristina Licata (2002): The future eTourism intermediaries. In: Tourism Management 23. Jahrgang, Heft 3, Seiten 207-220.
- [43] Auliana Poon (2003): Competitive Strategies for a new Tourism. In: Chris Cooper (Hg.): Classical Reviews in Tourism, Kapitel 8, Seiten 130-132. Channel View Publications.
- [44a] Stanley Wasserman, Katherine Faust (2008): Social network analysis - Methods and Applications, Seiten 31-35. Cambridge: Cambridge University Press.
- [44b] Stanley Wasserman, Katherine Faust (2008): Social network analysis - Methods and Applications, Seiten 71-72. Cambridge: Cambridge University Press.
- [44c] Stanley Wasserman, Katherine Faust (2008): Social network analysis - Methods and Applications, Seiten 93-94. Cambridge: Cambridge University Press.
- [44d] Stanley Wasserman, Katherine Faust (2008): Social network analysis - Methods and Applications, Seiten 129. Cambridge: Cambridge University Press.
- [44e] Stanley Wasserman, Katherine Faust (2008): Social network analysis - Methods and Applications, Seiten 133-134. Cambridge: Cambridge University Press.
- [44f] Stanley Wasserman, Katherine Faust (2008): Social network analysis - Methods and Applications, Seiten 150-153. Cambridge: Cambridge University Press.
- [45a] Manfred Nitzsche (2009): Graphen für Einsteiger – Rund um das Haus vom Nikolaus, Seite 2. Vieweg+Teubner Verlag.
- [45b] Manfred Nitzsche (2009): Graphen für Einsteiger – Rund um das Haus vom Nikolaus, Seite 21. Vieweg+Teubner Verlag.
- [45c] Manfred Nitzsche (2009): Graphen für Einsteiger – Rund um das Haus vom Nikolaus, Seiten 46-47. Vieweg+Teubner Verlag.
- [45d] Manfred Nitzsche (2009): Graphen für Einsteiger – Rund um das Haus vom Nikolaus, Seiten 133-134. Vieweg+Teubner Verlag.
- [46a] Robert A. Hanneman, Mark Riddle (2005): Introduction to social network methods, Kapitel 8. Online im Internet unter <http://www.faculty.ucr.edu/~hanneman/nettext/> (14.10.2010).
- [46b] Robert A. Hanneman, Mark Riddle (2005): Introduction to social network methods, Kapitel 10. Online im Internet unter <http://www.faculty.ucr.edu/~hanneman/nettext/> (19.01.2011).
- [46c] Robert A. Hanneman, Mark Riddle (2005): Introduction to social network methods, Kapitel 11. Online im Internet unter <http://www.faculty.ucr.edu/~hanneman/nettext/> (19.01.2011).
- [47a] John Scott (2000): Social network analysis – a handbook, Seiten 7-8. London: SAGE Verlag.
- [47b] John Scott (2000): Social network analysis – a handbook, Seiten 82-90. London: SAGE Verlag.

- [47c] John Scott (2000): Social network analysis – a handbook, Seite 115. London: SAGE Verlag.
- [48] Linton C. Freeman (2004): The development of social network analysis – A study in the sociology of science, Seiten 23-30. Vancouver: Empirical Press Verlag.
- [49] Stephen P. Borgatti, Ajay Mehra, Daniel J. Brass, Giuseppe Labianca (2009): Network Analysis in the social sciences. In: Science 323. Jahrgang, Heft 5916, Seiten 892-895.
- [50] Stanley Wasserman, Dawn Iacobucci (1991): Statistical modeling of one-mode and two-mode networks: Simultaneous analysis of graphs and bipartite graphs. Online im Internet unter <http://marketing.wharton.upenn.edu/ideas/pdf/Iacobucci/Dyadic/bjmsp91.pdf> (19.01.2011).
- [51] Stephen P. Borgatti (2005): Centrality and network flow. In: Social Networks 27. Jahrgang, Heft 1, Seiten 55-71.
- [52] Stephen P. Borgatti, Martin G. Everett (2006): A Graph-theoretic perspective on centrality. In: Social Networks 28. Jahrgang, Heft 4, Seiten 466-484.
- [53] Per Hage, Frank Harary (1995): Eccentricity and centrality in networks. In: Social Networks 17. Jahrgang, Heft 1, Seiten 57-63.
- [54] Debora Donato, Luigi Laura, Stefano Leonardi, Stefano Millozzi (2007): The Web as a Graph: How far we are. In: ACM Transactions on Internet Technology 7. Jahrgang, Heft 1, Artikel Nr. 4.
- [55] Antonios Kogias, Dimosthenis Anagnostopoulos (2006): A methodology for the evaluation of web graph models and a test case. In: L. Felipe Perrone, Barry G. Lawson, Jason Liu, Frederick P. Wieland (Hg.): Proceedings of the 38th Winter Simulation Conference, Seiten 2202-2209. Herausgeber: Winter Simulation Conference.
- [56] Dirk Lewandowski (2005): Die Struktur des Web. Online im Internet unter [http://www.durchdenken.de/lewandowski/web-ir/?32\\_Die\\_Struktur\\_des\\_Web.html](http://www.durchdenken.de/lewandowski/web-ir/?32_Die_Struktur_des_Web.html) (19.01.2011).
- [57] Oracle (2010): Beschreibung der Klasse Thread. Online im Internet unter <http://download.oracle.com/javase/1.4.2/docs/api/java/lang/Thread.html> (19.01.2011).
- [58] Oracle (2010): Beschreibung des Interfaces Executor. Online im Internet unter <http://download.oracle.com/javase/1.5.0/docs/api/java/util/concurrent/Executor.html> (19.01.2011).
- [59] Vladimir Batagelj (2004): How to convert text file datasets into Pajek format. Beschreibung eines Programms das von Jürgen Pfeffer (FAS.research Wien) entwickelt wurde. Online im Internet unter <http://vlado.fmf.uni-lj.si/pub/networks/pajek/howto/text2pajek.htm> (19.01.2011).
- [60] Govcom.org Foundation (2011): IssueCrawler. Die Website eines Crawlers. Online im Internet unter <http://www.issuecrawler.net/> (19.01.2011).
- [61] Diederik van Liere (2004): Interpretation of UCINET Output. Dokument erstellt im Rahmen der Essex Summer School. Online im Internet unter [www.analytictech.com/essex/Exercises/Interpreting\\_UCINET.doc](http://www.analytictech.com/essex/Exercises/Interpreting_UCINET.doc) (19.01.2011).
- [62] Tina Wey, Daniel T. Blumstein, Weiwei Shen, Ferenc Jordan (2008): Social network analysis of animal behaviour: a promising tool for the study of sociality. In: Animal Behaviour Jahrgang 75, Heft 2, Seiten 333-344.
- [63] Salzburg Research (2001): Indikatoren zur Informationsgesellschaft - Benchmarking zum Entwicklung der Informationsgesellschaft in Österreich (in der Schriftenreihe zur Informationsgesellschaft, Band 16). Online im Internet unter [http://wwwold.salzburgresearch.at/research/gfx/band16\\_indikatoren.pdf](http://wwwold.salzburgresearch.at/research/gfx/band16_indikatoren.pdf) (19.01.2011).
- [64] Rodolfo Baggio (2008): Network analysis of a tourism destination. Dissertation eingereicht an der Schule für Tourismus der Universität Queensland. Online im Internet unter [http://www.iby.it/turismo/papers/baggio\\_phd-thesis.pdf](http://www.iby.it/turismo/papers/baggio_phd-thesis.pdf) (19.01.2011).



- [64a] Rodolfo Baggio (2008): Network analysis of a tourism destination, Seiten 42-70  
Dissertation eingereicht an der Schule für Tourismus der Universität Queensland. Online im Internet unter [http://www.iby.it/turismo/papers/baggio\\_phd-thesis.pdf](http://www.iby.it/turismo/papers/baggio_phd-thesis.pdf) (19.01.2011).
- [64b] Rodolfo Baggio (2008): Network analysis of a tourism destination, Seite 70.  
Dissertation eingereicht an der Schule für Tourismus der Universität Queensland. Online im Internet unter [http://www.iby.it/turismo/papers/baggio\\_phd-thesis.pdf](http://www.iby.it/turismo/papers/baggio_phd-thesis.pdf) (19.01.2011).
- [64c] Rodolfo Baggio (2008): Network analysis of a tourism destination, Seiten 101-104.  
Dissertation eingereicht an der Schule für Tourismus der Universität Queensland. Online im Internet unter [http://www.iby.it/turismo/papers/baggio\\_phd-thesis.pdf](http://www.iby.it/turismo/papers/baggio_phd-thesis.pdf) (19.01.2011).
- [64d] Rodolfo Baggio (2008): Network analysis of a tourism destination, Seiten 211-214.  
Dissertation eingereicht an der Schule für Tourismus der Universität Queensland. Online im Internet unter [http://www.iby.it/turismo/papers/baggio\\_phd-thesis.pdf](http://www.iby.it/turismo/papers/baggio_phd-thesis.pdf) (19.01.2011).
- [64e] Rodolfo Baggio (2008): Network analysis of a tourism destination, Seite 238.  
Dissertation eingereicht an der Schule für Tourismus der Universität Queensland. Online im Internet unter [http://www.iby.it/turismo/papers/baggio\\_phd-thesis.pdf](http://www.iby.it/turismo/papers/baggio_phd-thesis.pdf) (19.01.2011).
- [65] Anthony Bonato (2005): A survey of models of the web graph. In: Alejandro Lopez-Ortiz, Angele Hamel (Hg.): Combinatorial and Algorithmic Aspects of Networking, Seiten 159-172. Berlin Heidelberg: Springer Verlag. (Herausgegeben im Rahmen der Reihe: Lecture Notes in Computer Science)
- [66] Judit Bar-Ilan (2008): Informetrics at the beginning of the 21st century – A review. In: Journal of Informetrics 2. Jahrgang, Heft 1, Seiten 1-52.
- [67] David Gibson, Jon Kleinberg, Prabhakar Raghavan (1998): Inferring Web communities from link topology. In: Proceedings of the ninth ACM conference on Hypertext and hypermedia: links, objects, time and space—structure in hypermedia systems, Seiten 225-234. New York: ACM.
- [68] Sergey Brin, Lawrence Page (1998): The anatomy of a large-scale hypertextual web search engine. Vorstellung der Suchmaschine Google. Online im Internet unter <http://infolab.stanford.edu/~backrub/google.html> (19.01.2011).
- [69] Facebook (2011): Pressebereich. Online im Internet unter <http://www.facebook.com/press/> (19.01.2011).
- [70] Facebook (2011): Das Portal. Online im Internet unter <http://www.facebook.com> (19.01.2011).
- [71] Tiscover (2011): AboutTiscover. Online im Internet unter <http://www.abouttiscover.com/> (19.01.2011).
- [72] Tiscover (2011): Das Portal. Online im Internet unter <http://www.tiscover.com> (19.01.2011).
- [73] Youtube (2011): Das Portal. Online im Internet unter <http://www.youtube.com> (19.01.2011).
- [74] ÖBB (2011): Das Portal. Online im Internet unter <http://www.oebb.at> (19.01.2011).
- [75] Oberösterreich Tourismus (2011): Das Portal *Genussland.at*. Online im Internet unter <http://www.genussland.at> (19.01.2011).
- [76] ORF (2006): Die erfolgreichsten Tourismusregionen. Als Quelle wird die Österreichische Hotelierversammlung (ÖVH) genannt. Online im Internet unter <http://oesterreich.orf.at/stories/83037/> (19.01.2011).
- [77] Eileen Fischer, A. Rebecca Reuber (2011): Social interaction via new social media: (How) can interactions on Twitter affect effectual thinking and behaviour? In: Journal of Business Venturing 26. Jahrgang, Heft 1, Seiten 1-18.
- [78] Twitter (2011): Das Portal. Online im Internet unter <http://www.twitter.com> (19.01.2011).
- [79] Andreas M. Kaplan, Michael Haenlein (2010): Users of the world, unite! The challenges and opportunities of Social Media. In: Business Horizons 53. Jahrgang, Heft 1, Seiten 59-68.

- [80] Noel Scott, Chris Cooper, Rodolfo Baggio (2008): Destination Networks. Four Australian Cases. In: *Annals of Tourism Research* 35. Jahrgang, Heft 1, Seiten 169-188.
- [81] Rodolfo Baggio (2006): Complex systems, information technologies, and tourism: a network point of view. In: *Information Technology & Tourism* 8. Jahrgang, Heft 1, Seiten 15-29.
- [82] Rodolfo Baggio (2007): The web graph of a tourism system. In: *Physica A: Statistical Mechanics and its Applications* 379. Jahrgang, Heft 2, Seiten 727-734.
- [83] Luciano da Fontoura Costa, Rodolfo Baggio (2009): The web of connections between tourism companies: Structure and dynamics. In: *Physica A: Statistical Mechanics and its Applications* 388. Jahrgang, Heft 19, Seiten 4286-4296.
- [84] yWorks – the diagramming company (2010): Eine Online-Applikation zum Zeichnen von Modellen. Online im Internet unter <http://live.yworks.com/graphity/> (19.01.2011).
- [85] Analytic Technologies (2011): UCINET Website. <http://www.analytictech.com/ucinet/> (19.01.2011).
- [86a] Reka Albert, Albert-Laszlo Barabasi (2002): Statistical mechanics of complex networks. In: *Reviews of modern Physics* 74. Jahrgang, Heft 1, Seite 48.
- [86b] Reka Albert, Albert-Laszlo Barabasi (2002): Statistical mechanics of complex networks. In: *Reviews of modern Physics* 74. Jahrgang, Heft 1, Seiten 54-58.
- [86c] Reka Albert, Albert-Laszlo Barabasi (2002): Statistical mechanics of complex networks. In: *Reviews of modern Physics* 74. Jahrgang, Heft 1, Seiten 67-76.
- [86d] Reka Albert, Albert-Laszlo Barabasi (2002): Statistical mechanics of complex networks. In: *Reviews of modern Physics* 74. Jahrgang, Heft 1, Seite 69 (Anfang Punkt 2).
- [86e] Reka Albert, Albert-Laszlo Barabasi (2002): Statistical mechanics of complex networks. In: *Reviews of modern Physics* 74. Jahrgang, Heft 1, Seite 70 (Anfang Punkt 3).
- [87] Peter Tittman (2003): Graphentheorie. Eine anwendungsorientierte Einführung, Seite 24. München Wien: Fachbuchverlag Leipzig im Carl Hanser Verlag.
- [88] Steven H. Strogatz (2001): Exploring complex networks. In: *Nature* 410 (6825), Seiten 268-276.
- [89] Peter Erdi, Gabor Csardi (2006): Network analysis in cell biology: a new tool in bioinformatics. Darin enthalten ist ein Artikel über reguläre Netzwerkstrukturen. Online im Internet unter <http://geza.kzoo.edu/~csardi/module/html/regular.html> (19.01.2011).
- [90] M. E. J. Newman (2002): Random graphs as models of networks. Online im Internet unter <http://arxiv.org/abs/cond-mat/0202208v1> (19.01.2011).
- [91] Brandes Ulrik, Erlebach Thomas (2005): Network Analysis, Kapitel 11, Seite 294. Berlin Heidelberg: Springer Verlag.
- [92] igraph (2005-2009): Softwarepaket zum Erzeugen und Manipulieren von ungerichteten und gerichteten Graphen. Darin ein Beispiel eines Erdos-Renyi-Zufallsgraphen. Online im Internet unter [http://igraph.sourceforge.net/images/screenshots/erdos\\_renyi\\_components.png](http://igraph.sourceforge.net/images/screenshots/erdos_renyi_components.png) (19.01.2011).
- [93] Universität von Kansas (2008): KU researcher suggests human language in a ‘small-world’ system. Darin eine Abbildung eines Kleine-Welt-Netzwerks. Online im Internet unter <http://www.news.ku.edu/2008/february/5/language.shtml> (19.01.2011).
- [94] Oliver Hein, Michael Schwind, Wolfgang König (2006): Scale-free networks – The impact of fat tailed degree distribution on diffusion and communication processes. In: *Wirtschaftsinformatik* 48. Jahrgang, Heft 4, Seiten 267-275.
- [95] Abteilung für molekulare, zelluläre und sich entwickelnde Biologie der Universität von Colorado (2011): Scale-Free examples – Non-biological. Online im Internet unter <http://mcdm.colorado.edu/courses/3280/lectures/class16-2.html> (27.01.2011).
- [96] M. Angeles Serrano, Ana Maguitman, Marian Boguna, Santo Fortunato, Alessandro Vespignani (2006): Decoding the structure of the www: facts versus sampling biases. Online im Internet unter <http://arxiv.org/abs/cs/0511035> (19.01.2011).

- [97] Albert-Laszlo Barabasi, Reka Albert, Hawoong Jeong (2000): Scale-free characteristics of random networks: the topology of the world-wide web. In: *Physica A: Statistical Mechanics and its Applications* 281. Jahrgang, Heft 1-4, Seiten 69-77.
- [98] Debora Donato, Stefano Leonardi, Stefano Millozzi und Panayiotis Tsaparas (2008): Mining the inner structure of the Web graph. In: *Journal of Physics A: Mathematical and Theoretical* 41. Jahrgang, Heft 22, 224017 (12 Seiten).
- [99] Stephen Dill, Ravi Kumar, Kevin S. Mccurley, Sridhar Rajagopalan, D. Sivakumar und Andrew Tomkins (2002): Self-Similarity in the Web. In: *Transactions on Internet Topology* 2. Jahrgang, Heft 3, Seiten 205-223.
- [100] Lada A. Adamic (1999): The small world web. In: Serge Abiteboul, Anne-Marie Vercoustre (Hg.): *ECDL'99 Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries*, Seiten 443-452. London: Springer Verlag. (Herausgegeben im Rahmen der Reihe: *Lecture Notes in Computer Science*)
- [101] Lada A. Adamic, Eytan Adar (2003): Friends and neighbors on the web. In: *Social Networks* 25. Jahrgang, Heft 3, Seiten 211-230.
- [102] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, Janet Wiener (2000): Graph structure in the web. In: *Computer Networks* 33. Jahrgang, Heft 1-6, Seiten 309-320.
- [103] Aaron Clauset, Cosma Rohilla Shalizi, M. E. J. Newman (2009): Power-law distributions in empirical data. Online im Internet unter <http://arxiv.org/abs/0706.1062> (20.01.2011).
- [104] Aaron Clauset (2000-2010): Power-law Distributions in Empirical Date. Online im Internet unter <http://tuvalu.santafe.edu/~aaronc/powerlaws/> (20.01.2011).
- [105] MathWorks (1994-2011): MATLAB – The Language of Technical Computing. Online im Internet unter <http://www.mathworks.com/products/matlab/> (20.01.2011).
- [106] Vladimir Batagelj, Andrej Mrvar (1999): Pajek - Program for Large Network Analysis. Online im Internet unter <http://vlado.fmf.uni-lj.si/pub/networks/pajek/> (20.01.2011).
- [107a] Koh Khee Meng, Dong Fengming, Tay Eng Guan (2007): Introduction to Graph Theory, Seiten 14-16. Singapur: World Scientific Publishing Co. Pte. Ltd.
- [107b] Koh Khee Meng, Dong Fengming, Tay Eng Guan (2007): Introduction to Graph Theory, Seiten 19-21. Singapur: World Scientific Publishing Co. Pte. Ltd.
- [107c] Koh Khee Meng, Dong Fengming, Tay Eng Guan (2007): Introduction to Graph Theory, Seiten 197-198. Singapur: World Scientific Publishing Co. Pte. Ltd.
- [108] Eric W. Weisstein: Complete Digraph. In: *MathWorld—A Wolfram Web Resource*. Online im Internet unter <http://mathworld.wolfram.com/CompleteDigraph.html> (20.01.2011).
- [109] Helmut Pruscha (2005): *Statistisches Methodenbuch: Verfahren, Fallstudien, Programmcodes*, Seite 25. Berlin Heidelberg: Springer Verlag.
- [110] Stanley Wasserman (2008): Mr. Holland's Networks: A Brief Review of the Importance of Statistical Studies of Local Subgraphs or One Small Tune in a Large Opus. Online im Internet unter <http://www.stat.indiana.edu/files/TR/TR-08-08.pdf> (20.01.2011).
- [111] Mark Newman, Albert-Laszlo Barabasi, Duncan J. Watts (2006): *The Structure and Dynamics of Networks*, Seiten 286-288. Princeton (New Jersey): Princeton University Press.
- [112] *Orgnet.com*(2011): Introduction to Social Network Analysis. Online im Internet unter <http://www.orgnet.com/sna.html> (21.01.2011).
- [113] Linton C. Freeman (1979): Centrality in Social Networks Conceptual Clarification. In: *Social Networks* 1. Jahrgang, Heft 3, Seiten 215-239.
- [114] Douglas R. Wight, Stephen P. Borgatti (1994): Betweenness centrality measures for directed graphs. In: *Social Networks* 16. Jahrgang, Heft 4, Seiten 335-346.
- [115] Peter Ingwersen (1998): The calculation of Web Impact Factors. In: *Journal of Documentation* 54. Jarhgang, Heft 2, Seiten 236-243.

[116] Jon M. Kleinberg (1999): Authoritative Sources in a Hyperlinked Environment. In: Journal of the ACM (JACM) 46. Jahrgang, Heft 5, Seiten 604-632.

## Anhang 1: Startliste für Suchdurchlauf 3

<http://www.austria.info>  
<http://www.tiscover.com>  
<http://www.touris-net.co.at>  
<http://www.touristinformation.at>  
<http://www.alpenstadt.com>  
<http://www.urlaubinoesterreich.net>  
<http://www.urlauboesterreich.com>  
<http://www.thermen.travel>  
<http://www.cusoon.at>  
<http://www.wienfuehrungen.com>  
<http://www.thermenland-urlaub.at>  
<http://www.gast.at>  
<http://www.tourism-db.co.at>  
<http://www.nationalparksaustria.at>  
<http://www.gaestefuehrer.at>  
<http://www.citytours.co.at>  
<http://oesterreich-tourismus.net>  
<http://www.servus.in>  
<http://www.vienna-hotel-guide.com>  
<http://www.verkehrsverein.at>  
<http://www.packages.at>  
<http://www.touristik.at>  
<http://www.tourismus-tirol.com>  
<http://www.reiselinks.at>  
<http://www.web-hotels.at>  
<http://www.urlaubsreif.at>  
<http://www.austriaguides.com>  
<http://www.oesterreich.tourismus.net>  
<http://www.tourismus-forum.at>  
<http://www.tirol-netz.de>  
<http://www.hoteladressen.at>  
<http://www.naturidyll.com>  
<http://www.tiscover.at>  
<http://www.donau.com>  
<http://www.ausflugstipps.at>  
<http://www.hochkoenig.at>  
<http://www.thermenland-steiermark.com>  
<http://oesterreichurlaub.blogg.de>  
<http://www.naturparke.at>  
<http://www.bergfex.at>  
<http://www.urlaubambauernhof.at>  
<http://www.farmholidays.com>  
<http://urlaub.bauernhof.at>  
<http://www.ferienbauernhof.com>  
<http://www.bauernhof.at>  
<http://www.austria.at>  
<http://www.bauernhof-hotel.at>  
<http://www.familienurlaub-steiermark.at>  
<http://www.bedandbreakfastaustria.at>  
<http://www.landurlaub.at>

<http://www.austria-ferienland.com>  
<http://www.landferien.at>  
<http://www.ferienbauernhof.org>  
<http://www.holiday-austria.com>  
<http://www.ferienland.at>  
<http://www.kinderbauernhofurlaub.at>  
<http://www.wellnessbauernhof.at>  
<http://www.holidaycheck.at>  
<http://www.almenland.at>  
<http://www.bauernhof-urlaub.at>  
<http://www.austria-ferienland.com>  
<http://www.romantik-urlaub.at>  
<http://www.topaustria.com>  
<http://www.skiamade.com>  
<http://www.schiurlaub.info>  
<http://www.urlauberland.at>  
<http://www.thermenland.at>  
<http://www.skihuetten-oesterreich.com>  
<http://www.almsommer.com>  
<http://www.landerlebnis.at>  
<http://www.urlaubsempfehlung.at>  
<http://www.wintersport.at>  
<http://www.urlauburlaub.at>  
<http://www.skiurlauboesterreich.com>  
<http://www.skiurlauboesterreich.org>  
<http://www.top-skiurlaub.at>  
<http://www.schifahren.net>  
<http://www.reisefalke.at>  
<http://www.beautyfarmhotel.at>  
<http://www.winterurlaub.at>  
<http://www.almenrausch.at>  
<http://www.skiurlaub.or.at>  
<http://www.familienwinter.at>  
<http://www.skihit.at>  
<http://www.kurz-mal-weg.at>  
<http://www.first-austrian-hotels.com>  
<http://www.Bergwelt.at>  
<http://www.ferien-mit-hund.at>  
<http://www.wandern-wandern.com>  
<http://www.traumhotel.at>  
<http://www.schlankundschoen.at>  
<http://www.berg-urlaub.at>  
<http://www.oesterreich-gastgeber.com>  
<http://www.oesterreichreisen.com>  
<http://www.urlaub-in-oesterreich.cc>  
<http://www.alpenregion.at>  
<http://www.gesundheitsurlaub.at>  
<http://www.golf-alpin.at>  
<http://www.weg.at>  
<http://www.completeurlaub.com>  
<http://www.austria-classic-hotels.at>

*<http://www.kultururlaub.at>*  
*<http://www.pferdeurlaub.at>*  
*<http://www.privaturlaub.at>*  
*<http://www.holiday.at>*  
*<http://www.thermensuche.at>*  
*<http://www.langlauf-urlaub.at>*  
*<http://www.aod.at>*  
*<http://www.gesundheitsurlaub.at>*  
*<http://www.robinson-autria.at>*  
*<http://www.urlaubsquartiere.at>*  
*<http://www.sommerwelt.at>*  
*<http://www.dertour.at>*  
*<http://www.seenguide.at>*  
*<http://www.campingfuehrer.at>*  
*<http://www.topcamping.at>*  
*<http://www.camping.at>*

## Anhang 2: Startliste für Suchdurchlauf 4

<http://www.austria.info>  
<http://blog.austria.info>  
<http://www.tirol.at>  
<http://www.genussland.at>  
<http://www.oberoesterreich.at>  
<http://www.vorarlberg.travel>  
<http://www.neusiedlersee.com>  
<http://www.oetztal.com>  
<http://www.innsbruck.info>  
<http://www.burgenland.info>  
<http://www.kitzbuehel.com>  
<http://www.seilbahnen.at>  
<http://www.salzburg.info>  
<http://www.osttirol.com>  
<http://www.lech-zuers.at>  
<http://www.imst.at>  
<http://www.austriatourism.com>  
<http://www.mondial.at>  
<http://www.blumau.com>  
<http://www.sanktjohann.com>  
<http://www.region-villach.at>  
<http://www.romantikstrasse.at>  
<http://www.dachstein-salzkammergut.at>  
<http://www.kaernten.at>  
<http://www.stgeorg.com>  
<http://www.grossglockner.at>  
<http://www.stift-rein.at>  
<http://www.salzburgerland.com>  
<http://www.alpbachtal.at>  
<http://www.stift-heiligenkreuz.org>  
<http://www.woerthersee.com>  
<http://www.bedandbreakfastaustria.at>  
<http://www.walsertal.at>  
<http://www.traunsee.at>  
<http://www.wien.info>  
<http://www.salzburgerland.at>  
<http://www.steiermark.com>  
<http://www.schlankundschoen.at>  
<http://www.alpenverein.at>  
<http://www.schoenbrunn.at>  
<http://oesterreich-tourismus.net>  
<http://www.austria.at>  
<http://www.urlaubsempfehlung.at>  
<http://www.urlaubambauernhof.at>  
<http://www.farmholidays.com>  
<http://www.urlauburlaub.at>  
<http://www.naturparke.at>  
<http://www.ausflugstipps.at>  
<http://www.nationalparksaustria.at>



*<http://www.urlauberland.at>*  
*<http://www.touristinformation.at>*