

TANJA ZAST

METHODEN DER ANALYSE VON SOZIALEN NETZWERKEN
(GENERIERUNG VON SOZIAL NETZWERK ÄHNLICHEN STRUKTUREN)



universität
uulm

METHODEN DER ANALYSE VON SOZIALEN NETZWERKEN

TANJA ZAST BACHELOR THESIS

Generierung von sozial Netzwerk ähnlichen Strukturen



Institute of Information Resource Management
Faculty of Engineering, Computer Science and Psychology
Ulm University

Mai 2022

Prof. Dr.-Ing. Dr. h.c. Stefan Wesner
Dr. Dipl.-Inf. Lutz Schubert

Tanja Zast: *Methoden der Analyse von Sozialen Netzwerken*, Generierung von sozial Netzwerk ähnlichen Strukturen, © Mai 2022

ZUSAMMENFASSUNG

Diese Arbeit handelt von sozialen Netzwerken, ihrer Generierung und anschließender Analyse. Es werden Methoden vorgestellt, die zur Analyse benötigt werden und zudem die mathematischen Verteilungen der Ergebnisse dieser angewendeten Methoden betrachtet. Am Ende folgt ein Ausblick, über weitere Verbesserungsmöglichkeiten der Generierung und Analyse von sozialen Netzwerken.

INHALTSVERZEICHNIS

I EINFÜHRUNG IN DIE THEORIE

1	EINLEITUNG	2
1.1	Zielsetzung	2
2	EINFÜHRUNG IN DIE SOZIALEN NETZWERKE	3
2.1	Ziele der Analyse	3
2.2	Einführung in die Grundstruktur von sozialen Netzwerken	4
3	KERNAKTOREN EINER SOZIALEN NETZWERKANALYSE	7
3.1	Zentralitäten	7
3.2	Cliquen und Brücken	10
3.3	Soziale Netzwerk-Eigenschaften	12
3.4	Ein typisches soziales Netzwerk	13

II DER PRAKTISCHE TEIL

4	DER GRAPHEN GENERATOR	17
4.1	Generierung eines sozialen Netzwerks	18
4.2	Die Analyse des generierten Graphen	23
4.3	Die Verteilung der Zentralitäten	27
4.4	Kurzes Recap	31
5	DER VERGLEICH MIT SOZIALEN NETZWERKEN	32
5.1	Der Datensatz und die Analyse	32
5.2	Anpassung des generierten sozialen Netzwerks	37
6	FAZIT UND AUSBLICK	40
	LITERATUR	42

ABBILDUNGSVERZEICHNIS

Abbildung 2.1	Links ist Netzwerk ₁ als Graph dargestellt und rechts Netzwerk ₂	5
Abbildung 3.1	Graph mit den Cliques (1, 2, 3) und (3, 4, 5, 6)	11
Abbildung 3.2	Graph mit den Brücken (1, 3) und (3, 9)	11
Abbildung 3.3	Game of Thrones social Network, Quelle: https://predictivehacks.com/social-network-analysis-of-game-of-thrones/ , Stand: 28.03.2022	14
Abbildung 4.1	Erste Versuche eines Sozialen Netzwerks, selbst erstellt	17
Abbildung 4.2	Zufällig erstellte Graphen mit 25 Knoten entsprechend der jeweiligen Methoden	19
Abbildung 4.3	Zufälliger sozialer Graph mit höchster Grad-Zentralität als Verbindungsknoten	22
Abbildung 4.4	Zufälliges soziales Netzwerk mit realistischeren Verbindungen	24
Abbildung 4.5	Verteilung der Grad-Zentralität des Graphen (b)	28
Abbildung 4.6	Zufälliges soziales Netzwerk und realistischeren Verteilungen	29
Abbildung 5.1	Game of Thrones Graph 2.0, selbst erstellt	33
Abbildung 5.2	Game of Thrones Verteilung der Zentralitäten	34
Abbildung 5.3	Facebook Graph mit den Datensätzen aus [6]	35
Abbildung 5.4	Facebook Graph Distribution	36
Abbildung 5.5	Final optimierter Graph	38

TABELLENVERZEICHNIS

Tabelle 3.1	Werte Abbildung 3.2	12
Tabelle 3.2	Eigenschaften eines sozialen Netzwerks	13
Tabelle 3.3	Werte GOT Graph	15
Tabelle 4.1	Werte oberer Graph	25
Tabelle 4.2	Weitere Eigenschaft eines sozialen Netzwerks	27

Teil I

EINFÜHRUNG IN DIE THEORIE

Um das Thema zu verstehen und vor allem die spätere Interpretation, ist es nun von Bedeutung, eine Einführung in die Theorie durchzuführen.

1

EINLEITUNG

Der Begriff *soziales Netzwerk* oder auf Englisch *social network* weckt seit vielen Jahrzehnten das Interesse zahlreicher Sozial- und Verhaltenswissenschaftler*innen [11]. Neben diesen, weckt es zudem das Interesse von unzähligen Unternehmen, um gezielter auf das Kundenverhalten einzugehen und dadurch den Gewinn zu maximieren [10]. Doch vor allem nicht zu vergessen, sind es heutzutage letztendlich die Nutzer*innen der Social Media-Plattformen wie Twitter, Facebook und Instagram, welche von diesem Begriff vor allem betroffen sind und auf diese Weise könnte die Liste noch lange fortgeführt werden.

Um jedoch eine klare Aussage zu treffen, spezialisieren sich vor allem Sozial- und Verhaltenswissenschaftler*innen, ebenso Unternehmen, auf die Analyse sozialer Netzwerke [10, 11]. Diese fokussieren sich weitestgehend auf Beziehungen zwischen sozialen Einheiten, sowie die Muster und Implikationen, welche diesen Beziehungen zugeschrieben werden [16].

1.1 ZIELSETZUNG

Wie ist ein *soziales Netzwerk* definiert und wie kann eine Analyse dieses Netzwerks aussehen? Was zeichnet die einzelnen Methoden zur Analyse aus und welche gelten als besonders aussagekräftig? Ziel der Arbeit ist es, einen Generator für *soziale Netzwerke* zu erstellen, welcher selbst generierte oder vorgegebene Testdaten visuell darstellt und zudem automatisch einige Methoden zur Analyse durchführt, beziehungsweise anwendet. Dafür muss zunächst ein Verständnis entwickelt werden, was ein *soziales Netzwerk* auszeichnet und von zufälligen Netzwerken unterscheidet. Diese Arbeit wird daher in zwei Bereiche unterteilt. Zum Einen in die Einführung von *sozialen Netzwerken* und die Erörterung der verschiedenen Zentralitäten und Eigenschaften von Cliques und Brücke, die Aufschluss darüber geben, wie die Einheiten miteinander verbunden sind, beziehungsweise zusammenhängen. Anschließend wird im zweiten Teil dieser Arbeit der Generator entwickelt, welcher Soziale Netzwerke so gut wie möglich nachstellt. Zum Schluss werden mehrere Analysen durchgeführt und die Verteilungen der Zentralitätswerte genauer betrachtet.

Diese Arbeit distanziert sich von dem Begriff *social networking*, welcher bei Recherchen zahlreichst auftaucht, aber lediglich den Vorgang oder Zustand beschreibt, dass Menschen über soziale Netzwerke durch beispielsweise gemeinsame Interesse zueinanderfinden.

2

EINFÜHRUNG IN DIE SOZIALEN NETZWERKE

Um zu verstehen, wie Soziale Netzwerke nachgebildet und analysiert werden können, sollte zunächst die Frage geklärt werden, was ein *soziales Netzwerk* ist. Hierfür existieren zwei Definitionen, eine gehört in den Bereich der Soziologie und die andere in den Bereich des Internets. In der Soziologie, ist ein *soziales Netzwerk* eine soziale Struktur, welche zwischen Akteuren besteht. Ein Akteur kann entweder von einer Einzelperson oder von Organisationen repräsentiert werden. Ein soziales Netzwerk zeigt die Art und Weise, wie Menschen und Organisationen durch soziale Vertrautheiten verbunden sind, die von zufälligen Bekanntschaften bis hin zu engen familiären Bindungen reichen [18]. Im Bereich des Internets ist der Begriff des Sozialen Netzwerks erst mit dem Web 2.0 entstanden. Der Begriff bezeichnet eine virtuelle Gemeinschaft. Diese wird überwiegend über eine Internetplattform gepflegt und aufrechterhalten. Soziale Netzwerke variieren in ihren Funktionen. Beispiele hierfür sind themenorientierte Netzwerke, siehe Twitter, oder Netzwerke, die überwiegend der zwischenmenschlichen Kommunikation dienen, siehe Facebook [19]. Das heißt, die Soziologie bezeichnet ausschließlich die soziale Struktur, wohingegen im Internet die virtuelle Gemeinschaft bezeichnet wird.

2.1 ZIELE DER ANALYSE

Der Fokus der *Sozialen Netzwerkanalyse* liegt auf der Interpretation und Analyse sozialer Beziehungen, genauer gesagt, auf den Beziehungen zwischen zwei sozialen Einheiten. Forscher haben erkannt, dass die Netzwerkperspektive neue Erkenntnisse und Möglichkeiten zu der Beantwortung sozial- und verhaltenswissenschaftlicher Forschungsfragen bietet. Dies ist möglich, da die *soziale Netzwerkanalyse* das soziale Umfeld als Muster oder Regelmäßigkeiten in Beziehungen zwischen Einheiten ausdrückt, beziehungsweise darstellen kann. Das regelmäßige Muster in den Beziehungen kann auch als Struktur bezeichnet werden [20]. Die Analysen, welche im Folgenden behandelt werden messen diese Strukturen, wodurch genauere Aussagen oder auch Vermutungen über die Beziehungen getroffen werden können. Die Beziehungen in sozialen Netzwerken können unterschiedlicher Art sein, beispielsweise wirtschaftlich oder politisch, welche nur zwei von vielen weiteren möglichen Beziehungstypen sind. Um die Muster oder Strukturen zu erkennen erfordert es Methoden oder analytische Konzepte. In den letzten Jahrzehnten haben sich die Methoden zur Analyse von sozialen Netzwerken als großer Bestandteil der Fortschritte in der Sozialtheorie erwiesen. Die Analyse sozialer Netzwerke besteht aus einer Reihe von mathematischen und grafischen Verfahren, beziehungsweise Techniken, welche Indizes zwischen Einheiten verwenden, um soziale Strukturen kompakt und systematisch darzustellen. Die Netzwerkanalyse verfolgt mehrere Ziele. Das erste Ziel ist die visuelle Darstellung von Beziehungen, was in Form eines Netzwerks oder Graphen möglich ist. Ein weiteres Ziel ist die Darstellung von Informationen. Dies soll es Benutzer*innen ermöglichen, die Beziehungen zwischen den Akteuren auf einen Blick zu erkennen. Zusätzlich verfolgt die Analyse das Ziel, grundlegende Eigenschaften von Beziehungen in einem Netzwerk zu untersuchen. Dies sind

Eigenschaften wie beispielsweise die Dichte und Zentralität. Ein weiteres Ziel besteht darin, Hypothesen über die Struktur der Verbindungen zwischen den Akteuren zu testen. Analysten sozialer Netzwerke können die Auswirkungen von Beziehungen auf die Einschränkung oder Verbesserung des individuellen Verhaltens oder der Netzwerkeffizienz untersuchen. Ein großer Vorteil von diesem Ansatz besteht darin, dass er sich auf die Beziehungen zwischen Akteuren konzentriert. Diese sind in ihren sozialen Kontext eingebettet.

Soziale Netzwerkanalyse kann in vier Schritte unterteilt werden. Erstens in die Definition eines Netzwerks, zweitens Messung der Beziehungen, drittens Darstellung der Beziehungen und viertens die Analyse der Beziehungen [20]. Um diese Einteilung sinnvoll durchführen zu können, ist es von Vorteil, wenn die Netzwerke eine gewisse Grundstruktur aufweisen.

2.2 EINFÜHRUNG IN DIE GRUNDSTRUKTUR VON SOZIALEN NETZWERKEN

Ein Graph G , besteht aus disjunkten Mengen (V, E) . Dabei bezeichnet V eine Menge von Knoten, und E stellt die sogenannten Kanten oder Bögen dar.

In dieser Arbeit werden ausschließlich ungerichtete Netze betrachtet, d.h. für jede Verbindung, die von einem Paar i nach j geht, gibt es eine Verbindung j nach i . Diese Verbindungen werden als Kanten bezeichnet. Gerichtete Verbindungen hingegen werden als Bögen bezeichnet. Netzwerkkanten können auch Gewichte haben, die z.B. die Stärke der Interaktion zwischen zwei Knoten angeben. Soziale Netzwerke können entweder als Graphen oder Matrizen dargestellt werden. Eine Netzwerkmatrix ist eine quadratische Anordnung von Werten, die das Vorhandensein oder Fehlen von Kommunikationsverbindungen zwischen Akteuren darstellen [4]. Das Vorhandensein wird mit einer "1" und das Nichtvorhandensein mit einer "0" beschrieben. Netzwerkmatrizen geben Verbindung zwischen den Knotenpunkten an.

Im Folgenden sind diese Matrizen zu betrachten:

Netzwerk 1:

$$\begin{pmatrix} & A & B & C & D & E \\ A & 0 & 0 & 0 & 1 & 1 \\ B & 1 & 0 & 1 & 1 & 1 \\ C & 0 & 1 & 0 & 1 & 0 \\ D & 1 & 1 & 0 & 0 & 1 \\ E & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Netzwerk 2:

$$\begin{pmatrix} & A & B & C & D & E \\ A & 0 & 1 & 0 & 1 & 1 \\ B & 0 & 0 & 1 & 0 & 1 \\ C & 1 & 1 & 0 & 0 & 0 \\ D & 0 & 0 & 0 & 0 & 1 \\ E & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Die erste Spalte und die erste Zeile der beiden Matrizen, stellen die Knoten innerhalb des Netzwerks dar. In sozialen Netzwerken ist es eher untypisch, dass Knoten auf sich selbst abbilden. Das würde beispielsweise heißen, dass eine Person eine Verbindung zu sich selbst aufweist, sich selbst folgt, oder die eigenen Beiträge liked, was üblicherweise nicht der Fall ist. Daher stehen in den beiden oberen Matrizen in den Diagonalen immer die Ziffer 0 [20].

Jedoch war die Rede davon, dass soziale Netzwerke nicht nur in Form von Matrizen dargestellt

werden können, sondern auch als Graphen. Die Matrizen oben bieten sich dafür idealerweise an. Die Graphen würde in diesem Fall wie folgt aussehen:

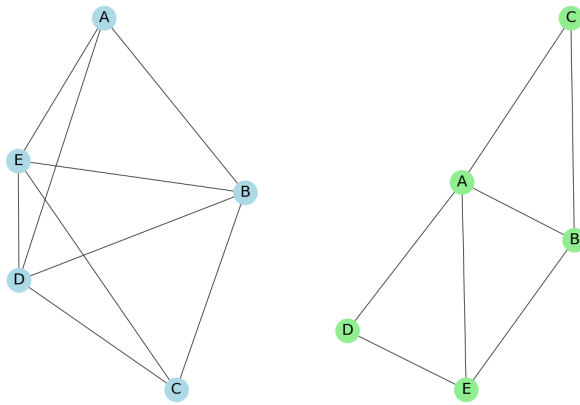


Abbildung 2.1: Links ist Netzwerk₁ als Graph dargestellt und rechts Netzwerk₂

Daher können für jegliche Netzwerkanalysen beide Varianten verwendet werden. Jedoch werden in dieser Arbeit überwiegend Graphen zur visuellen Veranschaulichung und Matrizen für jegliche Berechnungen verwendet, da es leichter ist auf die Werte einer Matrix zuzugreifen.

Ein soziales Netzwerk ist eine soziale Struktur, die zwischen Akteuren (Einzelpersonen oder Organisationen) besteht. Es zeigt die Art und Weise, wie Menschen und Organisationen durch verschiedene soziale Vertrautheiten verbunden sind, die von zufälligen Bekanntschaften bis hin zu engen familiären Bindungen reichen. Soziale Netzwerke bestehen aus Knotenpunkten und Verbindungen. Die Person oder Organisation, die am Netzwerk teilnimmt, wird als Knoten dargestellt/ bezeichnet. Bindungen sind die verschiedenen Arten von Verbindungen zwischen diesen Knotenpunkten. Bindungen werden nach ihrer Stärke bewertet. Lockere Verbindungen, wie bloße Bekanntschaften, werden als schwache Verbindungen bezeichnet. Starke Verbindungen, wie z. B. Familien oder Cliquen, werden als starke Bindungen bezeichnet [18]. Beispiele für soziale Netzwerke sind unsere Gesellschaft, das Internet, unser Gehirn und zelluläre Interaktionen. Doch welche grundsätzlichen Eigenschaften muss ein Netzwerk erfüllen, um als *soziales Netzwerk* bezeichnet zu werden? Sozialwissenschaftler*innen haben drei Arten von Netzwerken untersucht: egozentrische, soziozentrische und systemoffene Netzwerke. Egozentrische Netze sind Netze, die mit einem einzigen Knoten oder einer einzigen Person verbunden sind [8]. Um als Netze zu gelten, müssen diese Verbindungen nicht nur Auflistungen von Personen oder Organisationen sein, sondern müssen auch Informationen über die Verbindungen zwischen diesen Personen oder Organisationen enthalten. Im allgemeinen Sprachgebrauch, insbesondere, wenn von sozialer Unterstützung die Rede ist, wird jede Liste (dynamische Datenstruktur in der Informatik) als Netzwerk betrachtet. Eine Person, die eine große Anzahl guter Freunde hat, auf die sie sich verlassen kann, besitzt daher im Sprachgebrauch ein großes "Netzwerk".

Soziozentrische Netzwerke sind, wie Russell Bernard definiert, Netzwerke in einer Box. Netze mit offenen Systemen hingegen sind Netze, bei denen die Grenzen nicht klar sind, sie liegen nicht in einer Box, zum Beispiel die Verbindungen zwischen Unternehmen, oder die Kette an Auswirkungen, die eine Entscheidung oder eine Erneuerung in beispielsweise technischen Prozessen nachzieht. In gewisser Weise sind dies die interessantesten Netzwerke. Sie sind jedoch am schwierigsten zu untersuchen [5].

3

KERNFAKTOREN EINER SOZIALEN NETZWERKANALYSE

In komplexen Netzwerken können einige Knoten als wichtiger angesehen werden als andere. In einem sozialen Netzwerken zeichnen sich wichtige Knoten durch vergleichsweise mehr Verbindungen als andere Knoten aus. Auf das Beispiel *Instagram* bezogen, können solche Knoten Informationen gut verbreiten, sogenannte Influencer. Daher werden diese Knotenpunkte als zentral oder sozial wichtig interpretiert. Die Interpretation der Zentralität ist jedoch nicht eindeutig [3]. Zum Beispiel im Linienvorkehr, gilt eine Linie als zentral, wenn sie von großen Menschenmengen genutzt und stärker frequentiert wird, als andere Linien. Die Definition der Zentralität ist also nicht allgemein und hängt von der Anwendung ab. Da es keine allgemeine Definition von Zentralität gibt, wurden mehrere Maße entwickelt, die jeweils spezifische Konzepte berücksichtigen. Die Zentralität ist deshalb eine Schlüsseleigenschaft komplexer Netzwerke. Sie kann unter anderem das Verhalten dynamischer Prozesse wie beispielsweise eine epidemische Ausbreitung erklären, modellieren und abschätzen, jedoch nicht beschreiben, da es oftmals schwierig ist exakte Aussagen zu tätigen, wenn unbekannte Faktoren in den Datensätzen enthalten sind. [7]. Zudem kann die Zentralität Informationen über die Organisation komplexer Systeme und unsere Gesellschaft liefern. Es gibt viele Metriken zur Quantifizierung der Knotenzentralität in Netzwerken [13].

3.1 ZENTRALITÄTEN

Die *Grad-Zentralität* ist die am einfachsten zu berechnende Zentralität. Sie ist definiert durch die Anzahl der *direkten* Verbindungen eines Knoten. Mit der Adjazenzmatrix wird der Grad der Zentralität berechnet, indem die Summe der Elemente der betroffenen Zeile i berechnet wird. Mathematisch formuliert, wird folgende Formel verwendet:

$$k_i = \sum_{j=1}^N A_{ij} \quad (3.1)$$

Wobei A die Adjazenzmatrix beschreibt, N die Anzahl an Knoten darstellt und i, j die Knoten.

Da es sich bei der *Grad-Zentralität* um die einfachste Zentralität handelt, wird meist davon ausgegangen, dass Knoten mit vielen Verbindungen, daher mit einer hohen Zentralität, sich visuell betrachtet im Zentrum eines Netzwerkers befinden. Dies hat jedoch einige Nachteile, denn Knoten mit der höchsten *Grad-Zentralität* können sich auch visuell am Rand des Netzes befinden, was dazu führt, dass die *Grad-Zentralität* nicht als lokales Maß betrachtet wird. Zudem sollte hervorgehoben werden, dass bei der *Grad-Zentralität* nur ein- beziehungsweise ausgehende Kanten gezählt werden. Dies sagt zwar aus, dass ein solcher Knoten, auf das soziale Netzwerk bezogen, eine beliebte oder sehr bekannte Person ist, doch es ist dadurch keine Aussage über die Macht oder den Einfluss der Person möglich. Als extremes Beispiel, warum die *Grad-Zentralität* nicht immer optimal zur Netzwerkanalyse ist, diene ein Netzwerk mit einer

großen, dichten Gruppen von Knoten. Als dichte Gruppe ist hierbei eine Ansammlung von Knoten zu verstehen, welche sich alle nah beieinander befinden. Diese machen den größten Teil des Graphen aus, welcher auch als Kern des Netzes bezeichnet wird. Jedoch kann (visuell betrachtet) weit außerhalb des Kerns, entlang einer Kette von Knoten mit niedrigem Grad, ein Knoten liegen, welcher mit einer großen Anzahl von Knoten verbunden ist. Ein solcher Knoten hätte einen hohen Grad an Zentralität, obwohl er weit vom Kern des Netzes und den meisten Knoten entfernt ist [7]. Um solche Faktoren mit berücksichtigen zu können, wird ein weiterer Faktor in die Berechnung integriert, nämlich die Weglänge.

Diese spielt eine wichtige Rolle bei der *Nähe-Zentralität*, denn die Knotenzentralität kann auch anhand der kürzesten Wege definiert werden. Der Abstand zwischen zwei Knoten i und j ist gegeben durch die Anzahl der Kanten, welche sie möglichst direkt verbindet. Ein zentraler und daher wichtiger Knoten liegt, bezogen auf den Abstand, nahe an allen anderen Knoten des Netzes. Dieser Gedanke ist im Maß der sogenannten *Nähe-Zentralität* oder *Closeness-Centrality* enthalten. Diese wird durch den durchschnittlichen Abstand eines jeden Knotens zu allen anderen Knoten definiert. Mathematisch wird die Formel wie folgt beschrieben:

$$C_i = \frac{N}{\sum_{j=1, j \neq i}^N d_{ij}} \quad (3.2)$$

Dabei ist mit d_{ij} der kürzeste Weg zwischen i und j gemeint und mit N erneut die Anzahl an Knoten im Netzwerk [7]. Die *Nähe-Zentralität* ist vor allem dann sehr geeignet, wenn Prozesse über kurze Wege charakterisiert werden sollen. Beispielsweise kann der hierarchische Aufbau eines Unternehmens in einem sternförmigen Graphen dargestellt werden. In der Mitte des Graphen befindet sich der Vorstand, der in engem Kontakt mit den jeweiligen Abteilungsleitern steht. Die Abteilungsleiter sind, neben dem Vorstand, wiederum in sehr nahem Kontakt mit ihren jeweiligen Mitarbeitern ihrer Abteilung. Wenn nun ausschließlich anhand der *Grad-Zentralität* argumentiert wird, sind die Abteilungsleiter die wichtigsten Knoten im Graphen. Jedoch haben diese nicht die niedrigste *Nähe-Zentralität*, denn der Vorstand hat, da sich dieser Knoten visuell in der Mitte des Graphen befindet, zu allen anderen Knoten entweder einen oder zwei Kanten Abstand. Die einzelnen Abteilungsleiter haben aber, im worst-case Fall, zu anderen Angestellten aus anderen Abteilungen zwei bis drei Kanten Abstand. Dementsprechend ist es nicht ausreichend nur eine Zentralität bei der Analyse von *sozialen Netzwerken* zu betrachten. Bei der *Nähe-Zentralität* weisen die meisten komplexen Netze eine geringe durchschnittliche Länge des kürzesten Weges auf. Dies ist damit zu begründen, dass die durchschnittliche Entfernung mit dem Logarithmus über die Anzahl der Knoten zunimmt. Daher liegt das Verhältnis zwischen dem größten und dem kleinsten Abstand in der Größenordnung $\log(N)$, da der minimale Abstand gleich eins ist. In den meisten real existierenden Netzwerken beträgt dieses Verhältnis etwa 6 oder weniger. Es kann also mehrere Knoten mit der gleichen Zentralität geben, obwohl sie bei der Informationsverbreitung unterschiedliche Rollen spielen können. Daher ist die *Nähe-Zentralität* besser geeignet für räumliche Netze, bei denen die Abstände zwischen den Knoten größer sind als in zufälligen Netzen mit der gleichen Anzahl von Knoten und Verbindungen [7].

Die *Betweenness*- oder *Zwischen-Zentralität* hingegen misst, wie wichtig ein Knoten für die kürzesten Pfade durch das Netz ist. Um diese Zentralität für einen Knoten N zu berechnen, wird in dieser Methode eine Gruppe von Knoten ausgewählt und alle kürzesten Wege zwischen diesen Knoten gesucht. Dann wird der Anteil dieser kürzesten Wege berechnet, die den Knoten N einschließen. Wenn es beispielsweise 7 kürzeste Wege zwischen einem Knotenpaar gibt und 5 davon durch den Knoten N führen, dann wäre der Anteil $5/7 = 0.714$. Dieser Vorgang wird für jedes Knotenpaar im Netz wiederholt. Anschließend werden die berechneten Brüche addiert, wodurch die *Zwischen-Zentralität* des Knotens N generiert wird. Mathematisch formuliert sieht die Formel dann wie folgt aus:

$$B_i = \sum_{(a,b)} \frac{\eta(a,i,b)}{\eta(a,b)} \quad (3.3)$$

Hierbei bezeichnet $\eta(a,i,b)$ die Anzahl der kürzesten Wege zwischen den Knoten a und b die durch den Knoten i führen. Zudem stellt $\eta(a,b)$ die Gesamtzahl der kürzesten Wege zwischen a und b dar. Diese Zentralität, basierend auf dem *random walk-Algorithmus* und ist gegeben durch die erwartete Anzahl der Besuche von jedem Knoten i während einer zufälligen Schrittfolge durch den Graphen:

$$B_i = \sum_{a=1}^N \sum_{b=1}^N w(a,i,b) \quad (3.4)$$

dabei ist $w(a,i,b)$, wie oben bereits beschrieben für $\eta(a,i,b)$, die Anzahl der kürzesten Wege zwischen den Knoten a und b , die durch den Knoten i führen. Die Lösung wird daher nur angenähert. Die *Zwischen-Zentralität* ist eines der am häufigsten verwendeten Zentralitätsmaße. Sie gibt an, wie wichtig ein Knoten für den Informationsfluss von einem Knoten des Netzes zu einem anderen ist. In gerichteten Netzwerken kann die *Zwischen-Zentralität* mehrere Bedeutungen haben [7]. Einem Nutzer *Anton* mit hoher *Zwischen-Zentralität* folgen möglicherweise viele andere Nutzern, welche jedoch nicht denselben Personen folgen wie der Nutzer *Anton* selbst. Dies würde darauf hindeuten, dass der Nutzer *Anton* viele Anhänger oder Follower hat. Es kann aber auch sein, dass der Nutzer *Anton* weniger Follower hat, diese aber dafür mit vielen Knoten verbindet, die ansonsten weit entfernt sind. Daher ist es enorm wichtig die Richtung der Kanten eines Knotens zu kennen, um die Bedeutung der Zentralität zu verstehen.

Die *Eigenvektor*-oder *Eigenwert-Zentralität* misst die Bedeutung eines Knotens, wobei die Bedeutung seiner Nachbarn berücksichtigt wird. Deshalb wird sie manchmal verwendet, um den Einfluss eines Knotens im Netzwerk zu messen. Sie wird durch eine Matrixberechnung ermittelt, um den so genannten *Haupteigenvektor* anhand der Adjazenzmatrix zu bestimmen. Mathematisch betrachtet ist die *Eigenvektor-Zentralität* die komplizierteste, der in dieser Arbeit betrachteten Zentralitäten.

Wird nun die Tatsache betrachtet, dass ein Akteur zentraler ist, wenn er in Beziehung zu weiteren Akteuren steht, die selbst zentral sind, so kann argumentiert werden, dass die Zentralität eines Knotens nicht nur von der der Anzahl seiner Nachbarknoten abhängt, sondern auch von deren Zentralitätswert. Beispielsweise definiert Bonacich (1972) die Zentralität $c(v_i)$ eines Knotens v_i als positives Vielfaches der Summe der benachbarten Zentralitäten. Als Formel mathematisch dargestellt sieht dies folgendermaßen aus:

$$\lambda c(v_i) = \frac{1}{\lambda} \sum_{j=1}^N a_{ij} c(v_j) \forall i \quad (3.5)$$

oder umformuliert:

$$c(v_i) = \sum_{j=1}^N a_{ij}c(v_j) \forall i \quad (3.6)$$

Hierbei repräsentiert a_{ij} die Werte der Adjazenzmatrix A und λ einen konstanten Faktor. In Matrixschreibweise mit $c = (c(v_1), \dots, c(v_n))$ bedeutet dies auch:

$$Ac = \lambda c \quad (3.7)$$

Diese Art von Gleichung wird durch die Eigenwerte und Eigenvektoren von A gelöst. Aus der gesamten Menge an verschiedenen Eigenvektoren, soll es nur eine geeignete Lösung geben. Dieser Eigenvektor kann dann direkt als Zentralitätsmaß dienen. Da A die Adjazenzmatrix eines ungerichteten (zusammenhängenden) Graphen ist, ist A nicht negativ und aufgrund des Satzes von Perron-Frobenius, existiert ein Eigenvektor des maximalen Eigenwerts mit ausschließlich nicht negativen, also positiven, Einträgen [14].

3.2 CLIQUEN UND BRÜCKEN

Eine Clique ist laut Definition ein Teilgraph, aus mindestens drei Knoten bestehend, die zudem alle benachbart zueinander sind. Streng bezeichnet handelt es sich bei einer Clique um eine zusammenhängende Untergruppe. Eine Clique kann ebenso als Ansammlung von Akteuren gesehen werden, die sich gegenseitig wählen, jedoch wählt kein anderer Akteur dieser Gruppe weitere Akteure aus weiteren Gruppen und wird auch nicht von anderen Akteuren gewählt. Es ist zu beachten, dass sich Cliques in einem Graphen auch überlappen können, also derselbe Satz von Knoten zu mehr als nur einer Clique gehören kann. Zu beachten ist, dass eine vollständige Clique nicht in einer anderen Clique enthalten sein kann, denn sonst wäre die kleinere Clique nicht mehr maximal. Die Cliquendefinition ist vor allem nützlich, für die Untersuchung der Eigenschaften einer Untergruppe beziehungsweise eines Subgraphen [20]. Was genau damit gemeint ist, ist in folgendem Plot zu sehen:

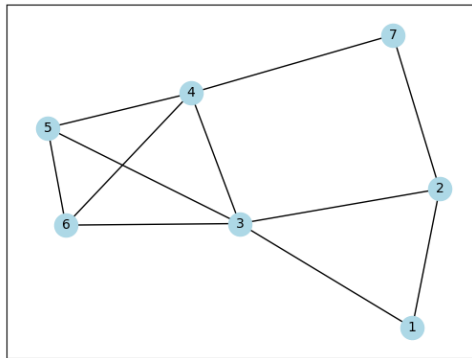


Abbildung 3.1: Graph mit den Cliques (1, 2, 3) und (3, 4, 5, 6)

Wichtig ist hierbei, dass es sich bei (2, 3, 4, 7) um keine Clique handelt, da keine Verbindung zwischen den Knoten **4** und **2** und ebenso keine Verbindung zwischen den Knoten **3** und **7** besteht. Neben den Cliques sind auch Brücken eine wichtige Diskussions- und Analysegrundlage für Graphen beziehungsweise in unserem Fall für *soziale Netzwerke*. Wenn von Brücken (bzw. englisch Bridge) die Rede ist, sind Verbindungen zwischen zwei Knoten gemeint. Jedoch handelt es sich um die einzige Verbindung zwischen diesen Knoten und deren Kontakten [15]. Ein Beispiel für Brücken im Graphen liefert folgender Plot:

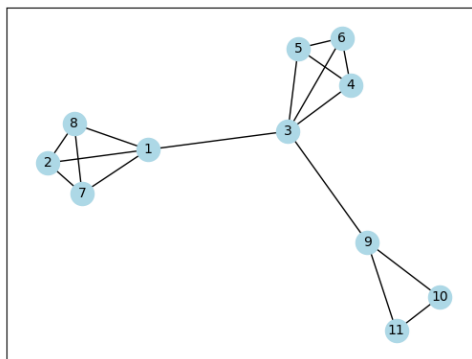


Abbildung 3.2: Graph mit den Brücken (1, 3) und (3, 9)

Hierbei ist gut zu erkennen, dass die drei Subgraphen durch *Brücken* miteinander verbunden sind. Neben den Brücken sind in Abbildung 3.2 die Cliques (1,2,7,8), (3,4,6,5) und (9,10,11) enthalten. Zu beachten ist, dass es sich bei beispielsweise (1,7,8) oder (4,5,6) um keine Cliques handelt. Wenn nun diese Brücken und Cliques im Zusammenhang mit den Zentralitäten betrachtet werden und die oben aufgeführten Formeln der Zentralitäten auf den Graphen 3.2 angewendet werden, erhält man die unten stehende Tabelle 3.1. Eine Berechnung für Abbildung

3.1 ist nicht nötig, da in Abbildung 3.2 ebenfalls Cliques und zusätzlich Brücken enthalten sind:

Tabelle 3.1: Werte Abbildung 3.2

Node	Grad-Zentr.	Nähe-Zentr.	Betweenness-Zentr.	Eigen.-Zentr.
3	0.5	0.666667	0.733333	0.470733
1	0.4	0.555556	0.466667	0.387253
4	0.3	0.454545	0	0.340195
6	0.3	0.454545	0	0.340195
5	0.3	0.454545	0	0.340195
2	0.3	0.4	0	0.279871
7	0.3	0.4	0	0.279871
8	0.3	0.4	0	0.279871
9	0.3	0.5	0.355556	0.184986
10	0.2	0.357143	0	0.0776041
11	0.2	0.357143	0	0.0776041

Direkt fällt auf, dass die Werte spaltenweise sehr ähnlich zueinander sind. Bei der *Gradzentralität* sind die Knoten 3 und 1 mit einem Wert von 0.5 und 0.4 am höchsten. Interessant, denn dabei handelt es sich um die Knoten, die unsere *Brücke* bilden. Bei den Knoten 3 und 1 fällt des weiteren auf, dass diese Knoten bei der *Nähe-*, *Zwischen-* und *Eigenvektor-Zentralität* ebenfalls am höchsten sind. Das heißt, die Vermutung liegt nahe, dass die Knoten eines Graphen, die Cliques bilden, relativ ähnliche *Zentralitätswerte* aufweisen beziehungsweise die Varianzen geringer sind. Aber vor allem erwähnenswert ist, dass in der Tabelle 3.1 lediglich bei den Knoten, welche die *Brücke* bilden, Werte ungleich *Null* in der Spalte *Betweenness-Zentr.* auffindbar sind. Dies sollte für den weiteren Teil der Arbeit in Erinnerung bleiben.

3.3 SOZIALE NETZWERK-EIGENSCHAFTEN

Die wichtigsten Eigenschaften eines sozialen Netzwerks sind die folgenden: Natürlich gibt es deutlich mehr Faktoren als die in Tabelle 3.2 dargestellten. Jedoch sind diese die primären Eigenschaften, welche in dieser Arbeit berücksichtigt werden. Nun sind die wichtigsten Eigenschaften der, in dieser Arbeit betrachteten und verwendeten, Metriken wie *Clique*, *Größe* oder *Brücken* bekannt und eingeführt. Manche Zentralitäten wurden oberflächlicher erklärt als andere, weil sie weniger relevant für die Untersuchung der sozialen Netzwerke sind. In Zukunft wird in dieser Arbeit bei *typischen Eigenschaften* von Netzwerken stets auf Tabelle 3.3 verwiesen.

Tabelle 3.2: Eigenschaften eines sozialen Netzwerks

Eigenschaft	Beschreibung
Cluster	Ein soziales Netzwerk sollte aus mehreren Cluster oder Subgraphen bestehen. Diese können in ihrer Größe und Anzahl stark variieren
Brücke	Die einzelnen Cluster sind über Brücken miteinander verbunden
Clique	In den Cluster sollten Cliques vorzufinden sein, d.h mindestens drei Knoten existieren die untereinander alle miteinander verbunden sind
Grad-Zentralität	Die einzelnen Knoten der Cluster sollten unterschiedliche Grad-Zentralitäten haben. Hohe Zentralitäten bedeuten, dass es sich um wichtige Knoten handelt, niedrige Werte, dass es weniger wichtige Knoten sind. Wichtig ist jedoch, dass es nicht ausschließlich wichtige oder ausschließlich unwichtige Knoten gibt.
Nähe-Zentralität	Auch hier sollen die Knoten im Cluster unterschiedliche Wert aufweisen. Hohe Werte bedeuten, die Knoten sind nah beieinander, haben kurze Wege zueinander. Niedrige Werte bedeuten, dass die Knoten weite Entfernungen zueinander haben.
Zwischen-Zentralität	Hier wird die Wichtigkeit der Nachbar-Knoten in Relation bewertet. Hohe Werte bedeuten, dass diese Knoten oft für den kürzesten Weg verwendet werden. Niedrige Werte, dass diese Knoten nicht für die kürzesten Wege relevant sind.

3.4 EIN TYPISCHES SOZIALES NETZWERK

Nachdem nun alle Zentralitäten, deren Berechnungen und weitere wichtige Eigenschaften von sozialen Netzwerken bekannt sind, ist es an der Zeit ein Musterbeispiel für ein soziales Netzwerk zu betrachten. Das bekannteste Netzwerk ist natürlich Facebook. Bei dieser sozialen Plattform ist die geeignetste Darstellung ein *ungerichteter* Graph. Bei Instagram hingegen, ein *gerichtet* Graph. Denn hier gibt es neben Leuten, denen wir folgen, die eigenen Follower [9]. Die Knoten sind sogenannte *Nutzer* und die *Kanten* sind Verbindungen zwischen ihnen. Zu beachten ist, dass sowohl *Knoten* als auch *Kanten* Attribute zugewiesen werden können. Knotenattribute in Facebook können zum Beispiel *Geschlecht*, *Ort*, *Alter* usw. sein, und Kantenattribute können *Datum der letzten Unterhaltung zwischen zwei Knoten*, *Anzahl der Likes*, *Datum der Verbindung* usw. sein [12]. Im folgenden wird ein, auf den ersten Blick und nach den Eigenschaften von Tabelle 3.3 typisch erscheinendes, soziales Netzwerk betrachtet. Es muss jedoch stets klar sein, dass es sich hierbei um den Datensatz eines fiktiven Fantasy Drama handelt [12] und es daher zu Unstimmigkeiten bei den Ergebnissen und der Analyse kommen kann.

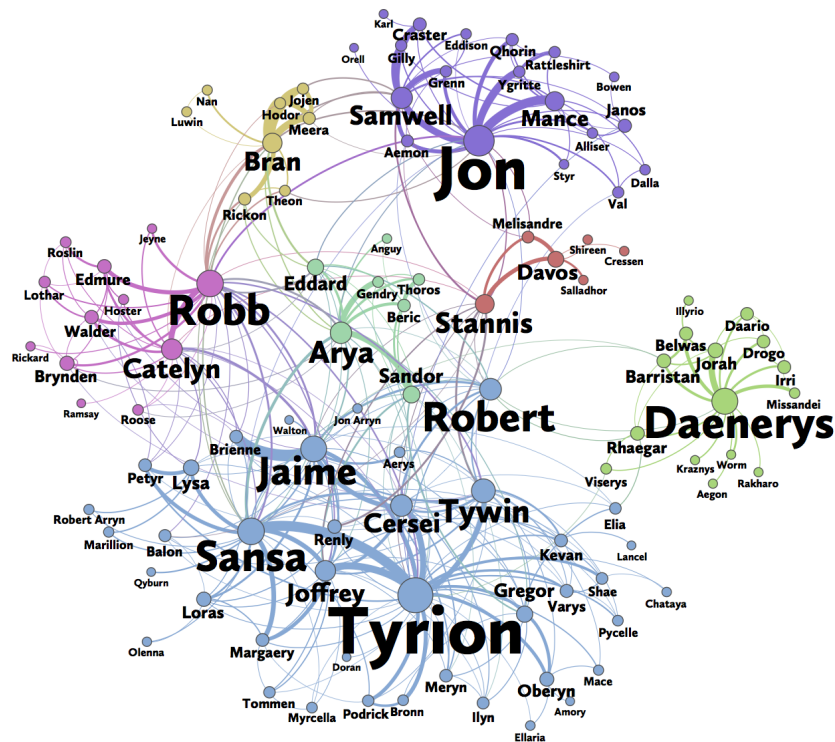


Abbildung 3.3: Game of Thrones social Network,
 Quelle: <https://predictivehacks.com/social-network-analysis-of-game-of-thrones/>,
 Stand: 28.03.2022

Das Netzwerk besteht aus 796 Knoten und 2823 Kanten. Insgesamt daher aus 796 Charakteren aus *Game of Thrones* (GOT). In dieser *sozialen Netzwerk Analyse* tauchen auch bisher unbekannte Messungen auf, die aber im Interpretations-Teil dieser Arbeit ebenfalls aufgegriffen werden. Beispielsweise beträgt der *Durchmesser* des GOT Graphen 9. Das heißt, wenn die kürzeste Pfadlänge von jedem Knoten zu allen anderen Knoten berechnet ist, ist der Durchmesser die längste aller berechneten Pfadlängen. Die durchschnittlich kürzeste Pfadlänge beträgt 3.41. Diese wird aber zu einem späteren Zeitpunkt analysiert. In Abbildung 3.3 ist gut zu erkennen, welche Knoten eine zentrale Rolle in diesem spielen. Hierfür wird mit der Knoten-Größe in der Abbildung 3.3 variiert. Große Knoten implizieren, dass es sich um einen wichtigen Knoten in dem Teilgraphen handelt und kleine, dass es sich um weniger relevante Knoten handelt [12]. Wenn diese Knoten in der Abbildung 3.3 gesucht werden, ist visuell direkt ersichtlich, dass es sich hierbei um die Knoten mit den meisten Verbindungen handelt. Oftmals ist leider bei den abgebildeten Graphen nicht eindeutig erkennbar, ob es sich hierbei um Kanten handelt, welche zwei Knoten direkt miteinander verbinden, oder die Kanten lediglich am Knoten vorbei verlaufen. Deshalb ist es wichtig, die Werte aus der Tabelle 3.3 zu analysieren. Hier fällt bei den Spalten *Charakter* auf, dass *Tyrion -Lannister* in allen aufgeführt wird. Das heißt, dass dieser Knoten im Graphen (visuell betrachtet) sowohl zentral liegen, und zudem kurze Abstände zu den anderen Knoten nachweisen muss. Zudem müssen über diesen Knoten die häufigsten kürzesten Wege verlaufen. Bei der Abbildung 3.3 fällt ebenfalls auf, dass der Knoten, beziehungsweise

Tabelle 3.3: Werte GOT Graph

Charakter	Grad-Zentr.	Charakter	Nähe-Zentr.	Charakter	Betweenness-Zentr.
Tyrion Lannister	0.1535	Tyrion Lannister	0.4763	Jon Snow	0.1921
Jon Snow	0.1434	Robert Baratheon	0.4593	Tyrion Lannister	0.1622
Jaime Lannister	0.1270	Eddard Stark	0.4558	Daenerys Targaryen	0.1184
Cersei Lannister	0.1220	Cersei Lannister	0.4545	Theon Greyjoy	0.1113
Stannis Baratheon	0.1119	Jaime Lannister	0.4520	Stannis Baratheon	0.1101

Charakter, *Tyrion* heraus sticht. Er ist von den meisten Knoten und Kanten umgeben. Da drei der fünf wichtigsten Knoten in der Spalte *Grad – Zentr.* den gleichen zweiten Namen tragen, liegt die Vermutung nahe, dass es sich hier um Knoten handelt, die auch visuell betrachtet sehr nah beieinander liegen müssen. Beim Betrachten des Graphen bestätigt sich diese Vermutung direkt, denn alle drei Knoten befinden sich im blauen Teilgraphen. Zudem handelt es sich bei dem Namen "*Lannister*" um ein Adelshaus in der US-amerikanischen Fantasy-Fernsehserie *Game of Thrones* was die starke Verbindung und Nähe zueinander begründet. Außerdem fällt sofort auf, dass drei der fünf Charaktere in der Spalte *Nähe-Zentr.* die selben sind, wie die wichtigsten Charaktere bezüglich der *Grad-Zentr.* Wieder bedeutet das, dass diese Charaktere sowohl zentral im Graphen liegen müssen als auch die kürzesten Wege zu anderen Knoten besitzen. Die Betrachtung von Abbildung 3.3 bestätigt dies sofort. Zudem weist der Graph auch einige Cliques auf. Die relevanteste und vor allem größte Clique befindet sich im blauen, grünen, ein Knoten im roten und zwei Knoten im pinken Teilgraphen. Aus dem Kapitel über Brücken und Cliques 3.2 ist bekannt, dass die Knoten mit den höchsten Zentralitäten höchstwahrscheinlich eine Clique darstellen und es sich vor allem bei den Knoten mit hohen *Zwischen-Zentralität* um Brücken handelt. Jedoch wird die Analyse dieses sozialen Netzwerks nicht weitergeführt, sondern auf die Analyse des selbst generierten sozialen Netzwerks fokussiert. Auch auf die Frage, welcher mathematischen bzw. stochastischen Verteilung die Zentralitäten entsprechen und warum eine solche Untersuchung sinnvoll ist, wird zu einem späteren Zeitpunkt eine Antwort gegeben.

Teil II

DER PRAKTISCHE TEIL

Nun folgt der Teil der Arbeit, in dem soziale Netzwerke selbst generiert und untersucht werden. Handelt es sich bei den generierten Netzwerken tatsächlich um soziale Netzwerke und erfüllen sie alle Ansprüche bezüglich der Zentralitäten und sonstigen Eigenschaften von sozialen Netzwerken aus Tabelle 3.2? Dies sind einige Fragen, die in diesem zweiten Teil der Arbeit beantwortet werden.

4.1 GENERIERUNG EINES SOZIALEN NETZWERKS

Bei einer endlichen Anzahl von Knoten n gibt es auch eine endliche Anzahl von möglichen Graphen, die aus diesen Knoten erzeugt werden können. Hierbei wächst die Anzahl der Graphen mit n Knoten exponentiell. Ein Zufallsgraph ist nur einer dieser Graphen, der durch einen Zufallsprozess erzeugt werden kann. Wenn von *Zufallsgraphen* die Rede ist, wird in den meisten Fällen das *Erdős-Rényi-Modell* als Graphengenerator verwendet (benannt nach den Mathematikern Paul Erdős und Alfréd Rényi). Eine wichtige Eigenschaft von, auf diese Weise erzeugten Zufallsgraphen ist, dass alle Konstellationsmöglichkeiten des Graphen gleichverteilt erzeugt werden [1]. Neben dem Erdős-Rényi-Modell, gibt es noch viele weitere Methoden zur zufälligen Netzwerkmodellierung [1].

- Die *dense_gnm_random_graph-Modellierung* liefert einen Zufallsgraphen. Bei dem Modell wird ein Graph gleichmäßig zufällig aus der Menge aller Graphen mit einer gegebenen Anzahl an Knoten und Kanten ausgewählt.
- Bei der *Newman-Watts-Strogatz small-world graph-Modellierung* wird zunächst ein Ring mit n Knoten erzeugt. Dann wird jeder Knoten im Ring mit seinen k nächsten Nachbarn verbunden (oder $k - 1$ Nachbarn, wenn k ungerade ist). Anschließend wird für jede Kante im Ring mit k nächsten Nachbarn, mit der Wahrscheinlichkeit p , eine neue Kante hinzugefügt.
- Die *random_regular_graph-Modellierung* gibt einen zufälligen regulären Graphen mit n Knoten zurück. Das heißt, alle Knoten besitzen gleich viele Nachbarn als somit den selben Grad. Der resultierende Graph hat keine Selbstschleifen oder parallele Kanten.
- Die *barabasi_albert_graph-Modellierung* hingegen liefert einen Zufallsgraphen nach dem Barabási-Albert-Präferenzmodell. Ein Graph mit n Knoten wird durch Anhängen neuer Knoten mit jeweils m Kanten erzeugt, die bevorzugt an bestehende Knoten mit hohem Grad angehängt werden.
- Die *powerlaw_cluster_graph-Modellierung* ist im Wesentlichen das Barabási-Albert-Wachstumsmodell mit dem zusätzlichen Schritt, dass für jede zufällige Kante die Chance besteht, dass ebenfalls eine Kante zu einem seiner Nachbarn besteht (und damit ein Dreieck entsteht) [1].

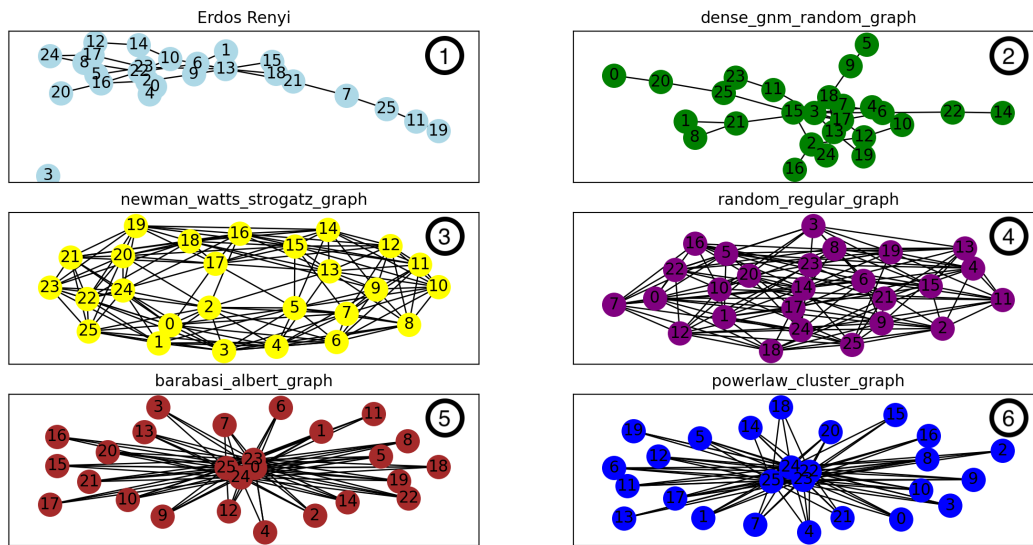


Abbildung 4.2: Zufällig erstellte Graphen mit 25 Knoten entsprechend der jeweiligen Methoden

Bei den Graphen in Abbildung 4.2 wurde lediglich eine visuelle Interpretation durchgeführt und nicht die Graphen mit den jeweiligen Zentralitäten analysiert. Auf den ersten Blick erkennen wir, dass bei allen sechs Modellen Unstimmigkeiten zu typischen Eigenschaften von *sozialen Netzwerken* aus Tabelle 3.2 auftreten. Beispielsweise bei dem *Barabasi Albert Graph* und dem *Powerlaw cluster graph* (5 und 6) sind einzelne zentrale Knoten zu erkennen. Diese zentrale Knoten sind von vielen weiteren Knoten umgeben, die alle wiederum mit diesen zentralen Knoten verbunden sind. Es sind also keine Cliquen und auch keine Cluster zu erkennen. Auch der *newman watts strogatz graph* und der *random regular graph* (3 und 4) entsprechen nicht den erwünschten *sozialen Netzwerken*. Bei beiden Plots scheint es, als sei jeder Knoten mit jedem weiteren Knoten verbunden, was erneut eine untypische Eigenschaft ist. Es existiert also nur lediglich eine große Clique. Nun bleiben noch die beiden Plots des *Erdos Renyi-Graphen* und *dense gnm random graph* (1 und 2), welche ebenfalls nicht unseren Erwartungen entsprechen. Der Plot des *dense gnm random graph* weist zwar einzelne Äste auf, doch generell wenige Cliquen enthalten und keine Cluster aufweist, weshalb dieses Modell ebenfalls nicht brauchbar ist. Bei dem *Erdos Renyi Modell* besteht die gleiche Problematik wobei hier noch das Problem hinzu kommt, dass ein isolierter Knoten existiert. Ein isolierter Knoten ist ein Knoten der keinen Nachbarn besitzt, also Grad 0 aufweist. Dies würde beispielsweise auf Sozial Media, bezogen bedeuten, dass Nutzer*innen auf dieser Plattform angemeldet sind, die keinerlei Verbindungen besitzen. Dies kann durchaus der Fall sein, es ist aber sehr unwahrscheinlich, dass Menschen auf solchen Plattformen angemeldet sind und keinerlei Freunde haben oder andere Nutzer*innen folgen.

Schließlich kommt bei den oberen zwei Modellen noch dazu, dass bereits visuell betrachtet kaum Cliques und auch keine Brücken auffindbar sind. Deshalb muss auch bei diesem Modell kritisch hinterfragt werden, ob es sich bei den Graphen um ein *typisches soziales Netzwerke* handeln könnte. Deshalb liegt nahe, dass Anpassungen durchgeführt werden müssen.

Eine mögliche Adaption wird erzielt, indem von den zufälligen Graphen-Methoden, die im vorherigen Abschnitt eingeführt wurden, abgewichen wird. Eine weitere Überlegung wäre, alle Formeln selbstständig zu implementieren und nicht die bereits vordefinierten Funktionen zu verwenden. Zum Einen sind diese vordefinierten Funktionen intransparent und daher auch fehleranfälliger, aber auch der Zugriff auf diese ist nicht ganz einfach.

Für die Generierung eines *sozialen Netzwerks* wird unter anderem eine Methode benötigt, die einzelne zufällige Graphen erstellt. Diese zufälligen Graphen sollen am Ende die jeweiligen Cluster darstellen, welche durch Brücken miteinander verbunden sind. Dadurch wären die ersten Eigenschaften eines typischen sozialen Netzwerks aus Tabelle 3.2 erfüllt.

Algorithm 1 Random Adjazenzmatrix

```

1: procedure RANDOM ADJACENCY MATRIX
2:    $matrix \leftarrow$  zufällige Matrix der Größe  $(n,n)$  zufällig befüllt mit Werten zwischen 0 und 1
3:   for alle  $i$  in  $matrix$  do
4:     befülle die Diagonale der Matrix mit 1
5:     for alle  $i$  und  $k$  in  $matrix$  do
6:       setze die Wahrscheinlichkeit  $prob$  auf einen zufälligen Wert zwischen 0 und 1
7:       if  $matrix$  an der Stelle  $[i][k]$  größer als  $prob$  then
8:         setze  $matrix$  an dieser Stelle auf 0
9:       else
10:        setze diese Stelle auf 1
11:   for alle  $i$  in  $matrix$  do
12:     was für  $matrix$  an der Stelle  $[i][j]$  gilt, muss auch für  $[j][i]$  gelten
13:   RETURN  $matrix$ 

```

Der Algorithmus 1 erstellt zufällige Matrizen, die aber erst noch zu einer großen Matrix zusammengefügt werden müssen. Die zufälligen einzelnen Matrizen sind die Cluster beziehungsweise Teilgraphen des sozialen Netzwerks. Doch wollen wir diese Cluster nun zu einem großen Graphen beziehungsweise einer großen Matrix zusammenführen. Hierfür benötigt man die Methode *Graph appender*. Der Algorithmus 2 dieser Methode soll wie folgt aussehen:

Algorithm 2 alle Subgraphen zu einer Liste zusammenführen

```

1: procedure GRAPH APPENDER
2:    $graphs \leftarrow$  leeres Array
3:    $n \leftarrow$  Anzahl an Subgraphen / Matrizen
4:   for alle  $i$  zwischen 1 und  $n$  do
5:      $k \leftarrow$  zufälliger integer, Größe des Subgraphen
6:     goTo Algorithm 1 mit dem übergebenen Wert  $k$ 
7:     füge random Matrix in  $graphs$  ein
8:   RETURN  $graphs$ 

```

Die einzelne Matrizen werden der dynamischen Datenstruktur (Liste) hinten angehängt. Nachdem nun eine Liste mit vielen zufällig erzeugten Matrizen generiert ist, fehlt lediglich eine Methode, um die Graphen zusammenzuführen und sicherzustellen, dass die Teilgraphen miteinander durch Brücken verbunden sind. Wir wollen also insgesamt sicherstellen, dass durch die Algorithmen 1, 2 und 3 einzelne zufällige Subgraphen erstellt werden, die mithilfe des *graph appenders* zu einer Liste zusammengeführt werden und nun über Brücken Verbindungen zueinander gebildet werden. Der Algorithmus 3 sieht wie folgt aus:

Algorithm 3 Graphs zusammenführen

```

1: procedure UNITE GRAPHS
2:   graphs  $\leftarrow$  Graph aus Algorithm 2
3:   if graphs aus nur einem Element besteht then
4:     gebe graphs zurück
5:   dim  $\leftarrow$  0
6:   big graph  $\leftarrow$  Graph mit Nullen befüllt
7:   for alle i zwischen 0 und der Länge von graphs do
8:     Variable a  $\leftarrow$  zufälliger integer zwischen 0 und Länge von graphs
9:     Variable b  $\leftarrow$  zufälliger integer zwischen 0 und Länge von graphs
10:    for alle j und k zwischen 0 und graphs do
11:      l  $\leftarrow$  summierte Länge von graphs bis zur Stelle i
12:      graph  $\leftarrow$  graphs an der Stelle i
13:
14:      in den Zeilen 16 und 17 werden die einzelnen Cluster in big graph eingefügt
15:
16:      big graph an der Stelle [(l+j)][(l+k)]  $\leftarrow$  graph[j][k]
17:      big graph an der Stelle [(l+k)][(l+j)]  $\leftarrow$  graph[k][j]
18:
19:      in den Zeilen 21 und 22 werden die einzelnen Cluster durch Brücken verbunden
20:
21:      big graph an der Stelle [(l+a)][(l+b+graphs Länge an [i]) modulo dim]  $\leftarrow$  1
22:      big graph an der Stelle [(l+b+graphs Länge an [i]) modulo dim)][(l+a)]  $\leftarrow$  1
23:
24: nun wird der Knoten mit der höchsten Gradzentralität gesucht
25:
26:   counter 1  $\leftarrow$  0
27:   counter 2  $\leftarrow$  0
28:   Knoten  $\leftarrow$  0
29:   for i und j zwischen 0 und der Länge von graphs do
30:     if graphs an der Stelle [i][j] ungleich 0 then
31:       counter 1  $\leftarrow$  erhöhe um 1
32:       if counter 1 größer counter 2 then
33:         counter 1  $\leftarrow$  counter 2
34:       Knoten  $\leftarrow$  i
RETURN Knoten

```

Jetzt ist ein großer Graph generiert, bestehend aus vielen zufälligen kleinen Graphen, welche durch den Knoten mit den meisten ein- und ausgehenden Kanten mit einem weiteren Subgraphen verbunden sind. Dieses Vorgehen ist aber nicht in einem der drei Algorithmen 1, 2 oder 3 beschrieben, sondern im Git Repo [23] zu finden. Nach weiteren Überlegungen, wie es möglich wäre, den generierten Graph noch mehr sozialen Netzwerken ähneln zu lassen und die Eigenschaften aus Tabelle 3.2 zu erfüllen, ist zusätzlich die Idee entstanden eine Methode zu schreiben, die sicherstellt, dass der generierte Graph aus einer bestimmten Anzahl an Cliquen besteht. Mit diesem zusätzlich Faktor soll sichergestellt werden, dass der generierte Graph mehr Kanten besitzt als davor, um die Wahrscheinlichkeit für eine existierende Clique zu erhöhen. Der Cliquen-Methode, die ebenfalls im Git Repo [23] zu finden ist, soll hierfür eine fixe Zahl n übergeben und zusätzlich sichergestellt werden, dass stetig neue Graphen generiert werden, bis die Anzahl an Cliquen genau der fixen Zahl n entspricht. Durch die Algorithmen 1, 2 und 3 entsteht schließlich folgender Graph:

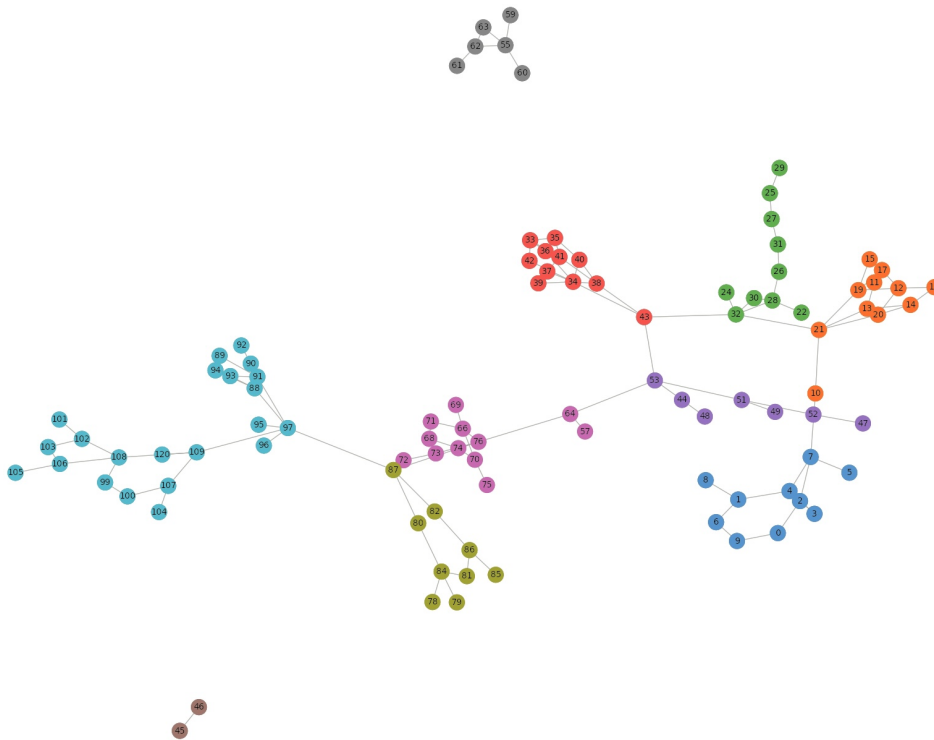


Abbildung 4.3: Zufälliger sozialer Graph mit höchster Grad-Zentralität als Verbindungsknoten

Nachdem Abbildung 4.3 visuell betrachtet durchaus *Sozialen Netzwerken* ähnelt, muss noch eine weitere Verbesserung durchgeführt werden. Bei einer genaueren Betrachtung der Abbildung fällt auf, dass die Teilgraphen wenige Verbindungen, Brücken, untereinander aufweisen. Dies liegt an der Idee von Algorithmus 3, den Knoten mit der höchsten Gradzentralität zu wählen und diesen dann mit einer beliebigen weiteren Gruppe zu verbinden. Doch in der Realität ist ein solches Phänomen sehr unwahrscheinlich und erfüllt nicht die Bedingungen aus Tabelle 3.2 für ein typisches soziales Netzwerk. Denn dies würde beispielsweise heißen, dass an der Universität Ulm alle Student(en)*innen der Fakultät für Ingenieurwissenschaften, Informatik und Psychologie untereinander in einer Weise miteinander verbunden sind, jedoch nur die Professor(en)*innen, welche die höchste Gradzentralität aufweisen, mit eine*m/r weiteren Professor*in einer anderen Fakultät verbunden sind. Dies ist aber nicht realistisch wenn bedacht wird, dass auch beispielsweise Student(en)*innen der Fakultät für Mathematik und Wirtschaftswissenschaften durchaus Kontakte zu der Fakultät für Ingenieurwissenschaften, Informatik und Psychologie haben können oder auch mit den jeweiligen Professor(en)*innen. Dementsprechend muss diese Eigenschaft ebenfalls in der Implementierung berücksichtigt werden. Das kann gewährleistet werden, indem jedem Knoten eine zufällige Wahrscheinlichkeit zugeschrieben wird, die angibt, ob eine Kante zwischen den Clustern oder Subgraphen existiert. Hierfür wird der Algorithmus 3 ab Zeile 17 ersetzt zu:

Algorithm 4 Verbindung Subgraphen

```

1: procedure CONNECTION SUBGRAPHS
2:    $prob \leftarrow$  zufällige Zahl, die sehr klein ist
3:   for alle  $i$  und  $j$  liegen in der Matrix big graph do
4:     befülle die Diagonale der Matrix mit 0
5:   for alle  $i$  und  $k$  liegen in der Matrix do
6:      $variable \leftarrow$  zufällige Zahl zwischen 0 und 1
7:     if  $variable$  kleiner  $prob$  then
8:       setze big graph  $[i][k]$  auf 1
9:   RETURN big graph
  
```

Mit Algorithmus 4 kann sichergestellt werden, dass die Subgraphen vermehrt miteinander verbunden sind und nicht von dem Knoten mit der höchsten Gradzentralität abhängen. Dadurch ist eine weitere Eigenschaft aus Tabelle 3.2 bezüglich der Existenz von mehreren Brücken, erfüllt.

4.2 DIE ANALYSE DES GENERIERTEN GRAPHEN

Mit den Überlegungen aus dem vorherigen Kapitel und den dort erläuterten Methoden, lassen sich *typische soziale Netzwerke* nach den Eigenschaften von Tabelle 3.2 generieren. Um zu beweisen, dass es sich tatsächlich um ein typisches Netzwerk handelt, soll ein neues generiert und eine Analyse damit durchgeführt werden. Ziel ist es zu zeigen, dass die mit dem Generator erzeugten Graphen tatsächlich näherungsweise *sozialen Netzwerken* entsprechen, welche die Bedingungen aus Tabelle 3.2.

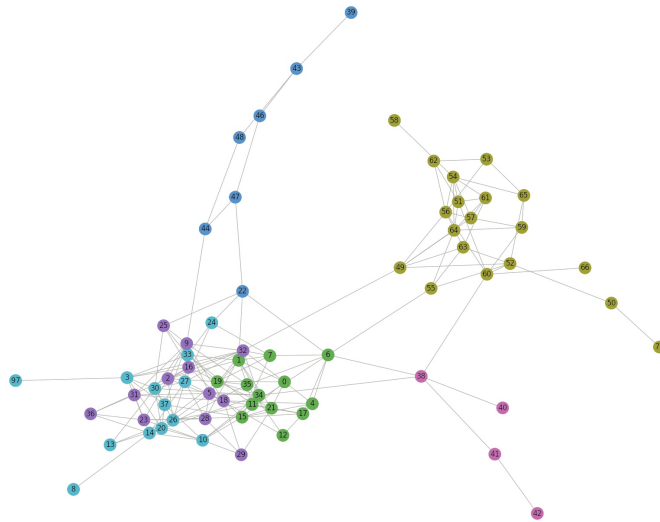


Abbildung 4.4: Zufälliges soziales Netzwerk mit realistischeren Verbindungen

Bei der visuellen Betrachtung der Abbildung 4.4 ähnelt die Struktur auf jeden Fall der, eines sozialen Netzwerks, siehe beispielsweise Abbildung 3.3. Doch um fundierte Aussagen treffen zu können, müssen auch die Zentralitäten genauer analysiert werden. Hierfür wird folgende Tabelle verwendet:

Tabelle 4.1: Werte oberer Graph

Knoten	Grad-Zentr.	Nähe-Zentr.	Between-Zentr.	Knoten	Grad-Zentr.	Nähe-Zentr.	Between-Zentr.
1	0.149254	0.389535	0.0429244	38	0.0746269	0.36612	0.154688
2	0.134328	0.370166	0.0366434	41	0.0298507	0.271255	0.0298507
3	0.119403	0.350785	0.0516569	43	0.0447761	0.198813	0.030303
5	0.119403	0.378531	0.0341306	44	0.0447761	0.295154	0.0773717
6	0.119403	0.385057	0.145038	46	0.0298507	0.219672	0.0205638
7	0.0895522	0.358289	0.0208983	47	0.0447761	0.27459	0.0520902
10	0.119403	0.341837	0.0240985	48	0.0298507	0.232639	0.0373285
11	0.104478	0.360215	0.0212421	49	0.0895522	0.36413	0.221288
14	0.119403	0.3350	0.0454434	50	0.0298507	0.241877	0.0298507
18	0.134328	0.340102	0.0283754	52	0.0895522	0.314554	0.0885577
22	0.0746269	0.348958	0.0740623	54	0.104478	0.254753	0.0327816
27	0.119403	0.360215	0.0342121	55	0.0597015	0.325243	0.0670173
30	0.149254	0.348958	0.0412278	56	0.104478	0.303167	0.0672381
32	0.179104	0.435065	0.266448	57	0.0746269	0.290043	0.0213757
34	0.134328	0.394118	0.112543	60	0.0895522	0.313084	0.0903114
35	0.104478	0.362162	0.0290967	64	0.0895522	0.304545	0.0530434

Bei dieser Tabelle handelt es sich um die 32 wichtigsten Knoten. Denn alle diese Knoten weisen eine höhere *Zwischen-Zentralität* als 0.02 auf. Dieser Wert entspricht ungefähr dem Mittelwert aller berechneten Zentralitäten. Die andern Knoten sind außen vor gelassen, da sie in Relation gesehen eher unwichtig für das Netzwerk sind. Bei der *Grad-Zentralität* aus der Tabelle 4.2 sehen wir, dass die meisten Knoten einen Wert höher als 0.1 aufweisen. Zudem weisen einige, wenige Knoten eine *Grad-Zentralität* höher als 0.13 auf. Genau genommen handelt es sich hier um Knoten 1 mit einem Wert von 0.149254, Knoten 2 mit dem Wert 0.134328, Knoten 18 mit dem Wert 0.134328, Knoten 30 mit einer Zentralität von 0.149254, zudem um Knoten 32 mit dem höchsten Wert von 0.179104 und schließlich Knoten 34 mit einer *Grad-Zentralität* von 0.134328.

Die aufgezählten Knoten sind, den Werten zu urteilen nach, zentral wichtig für den Graphen und befinden sich höchstwahrscheinlich, visuell gesehen, im Zentrum des Graphen in Abbildung 4.4. Bei der visuellen Analyse, kann diese Behauptung teilweise bestätigt werden, denn diese Knoten fallen direkt auf. Ein hoher Zentralitätswert bei einem Knoten sagt aus, dass es sich beispielsweise im realen Leben um eine vermutlich sehr berühmte / bekannte Person handeln wird. Daher besteht die Annahme, dass es sich bei diesem Knoten, würde er auf die Realität bezogen betrachtet werden, beispielsweise um einen Star, einen Influenzer oder eine, auf weitere Arten bekannte Person handelt. Doch ebenso ist es möglich, dass die Person viele weitere Personen kennt, oder von vielen weiteren Personen gekannt wird. Wichtig ist, dass nicht nur die *Grad-Zentralität* eine zentrale Rolle für die Analyse spielt. Im Weiteren betrachten wir auch die *Nähe-Zentralität*. Doch um auch bei diesem Aspekt nicht alle 32 Werte aufzuzählen, werden im Folgenden nur Knoten betrachtet, die einen Wert höher als 0.37 aufweisen, da es sich dabei

erneut um einen guten Mittelwert handelt. Diesen Grenzwert übertrifft der Knoten **1** mit einem Wert von **0.389535**, Knoten **2** mit dem Wert **0.370166**, zudem Knoten **5** mit dem Wert **0.378531**, zusätzlich Knoten **6** mit der Zentralität **0.385057**, und schließlich Knoten **32** mit dem höchsten Wert von **0.435065** und **34** mit der Zentralität von **0.394118**. Laut der Definition aus dem Abschnitt **3.1** gilt, je höher die Werte der *Nähe-Zentralität* eines Knoten ist, desto näher befindet sich dieser Knoten zu weiteren Knoten bzw. weist die durchschnittlich kürzesten Wege zu diesen auf.

Wird der Graph **4.4** anhand dieser Information betrachtet und werden so die Knoten mit der höchsten *Nähe-Zentralität* gesucht, ist visuell ersichtlich, dass sich diese im gleichen Bereich befinden, wie die Knoten mit der höchsten *Grad-Zentralität*. Die letzte zu untersuchenden Zentralität ist die *Betweenness-Zentralität*. Auch hier betrachtet man wieder die Knoten mit den höchsten Werten, und um nicht alle **32** Werte aufzuzählen, werden erneut nur Knoten mit einem Wert höher als **0.09** ausgewählt. Diese Voraussetzung erfüllen neben dem Knoten **6** mit dem Wert **0.145038** die Knoten **32** mit der höchsten Zentralität von **0.266448** und **34** mit einem Wert von **0.112543**, außerdem der Knoten **38** mit der Zentralität von **0.154688**, zudem der Knoten **49** mit dem Wert **0.221288** und schließlich der Knoten **60** mit dem Wert **0.0903114**.

Das bedeutet für unseren Graphen in Abbildung **4.4**, dass die kürzesten Wege anteilmäßig am öftesten über diese genannten Knoten verlaufen. Zudem kann vermutet werden, dass es sich bei diesen Knoten um *Brücken* handelt, was der Behauptung aus dem vorherigen Kapitel **3.2**, dass es sich bei hohen *Zwischen-Zentralitäten* um *Brücken* handelt, bestätigen würde. Bei der erneuten Betrachtung des Graphen, erkennt man nämlich, dass sich die Knoten in den grün, lila, hellblauen Teilgraphen befinden und links unten zentriert sind. Doch kommt bei der *Zwischen-Zentralität* hinzu, dass sich die Knoten **49** und **60** auch im gelbgrünen, rechts oben liegenden, Teilgraphen befinden. Außerdem ist auf den ersten Blick zu erkennen, dass es sich bei den 6 Knoten in Abbildung **4.4** visuell betrachtet, tatsächlich um *Brücken* handelt.

Leider ist an dieser Stelle zu erwähnen, dass die Anzahl an *Cliquen* des Graphen aus der Abbildung **4.4** nicht mehr exakt bekannt ist, da erst nach der weiteren Verbesserung des Codes Rücksicht drauf genommen wurde, die Cliquen-Größe und -Anzahl direkt vorzugeben beziehungsweise der Methode zu übergeben. Daher wird an dieser Stelle nur die Vermutung aufgestellt, dass alle Knoten aus der Tabelle **4.2** Teil von Cliquen sind. Wie viele es jedoch genau sind, lässt sich an dieser Stelle leider nur wage vermuten. Im nächsten Abschnitt wird zusätzlich die Betrachtung der Cliquen des Graphen mit einbezogen. Weitere Zentralitätswerte und Eigenschaften des Graphen werden an dieser Stelle nicht betrachtet. Nachdem alle Kriterien aus Tabelle **3.2** überprüft sind und erfolgreich festgestellt wurde, dass dieser Graph einem *sozialen Netzwerk* ähnelt, wird noch ein weiterer Graph generiert, um sicherzustellen dass es sich nicht um eine zufällige Übereinstimmung handelt. An dieser Stelle soll auch noch ein weiteres Kriterium untersucht werden, nämlich die Verteilung der Zentralitäten.

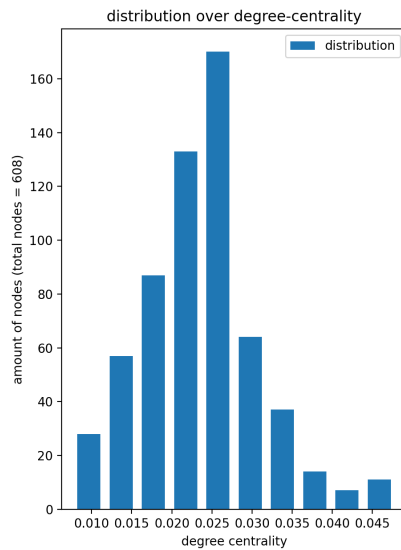
4.3 DIE VERTEILUNG DER ZENTRALITÄTEN

Nachdem im vorherigen Kapitel die Generierung eines *sozialen Netzwerkes* und die Analyse durchgeführt wurde, spielt im Folgenden die Verteilung der Zentralitätswerte eine wichtige Rolle. Im Laufe der Arbeit ist aufgefallen, dass sich die Werte der Zentralitäten, von den bisher generierten Graphen, oftmals in einem ähnlichen Wertebereich befinden. An dieser Stelle kommt die Frage auf, wie diese Werte verteilt sind und ob die Verteilung möglicherweise einer mathematischen Wahrscheinlichkeitsverteilung entspricht beziehungsweise ähnelt. Das heißt im Konkreten, es wird der Frage nachgegangen, ob alle Zentralitätswerte sozialer Netzwerke ähnliche Verteilungen nachweisen. Wenn sich die Vermutung bestätigt, können andere soziale Netzwerke anhand dieses Kriteriums verglichen werden. Erdös und Renyi (1960), Cliff und Ord (1973) und Friedkin (1981) arbeiteten bereits an Zufallsgraphen und haben die Definition erstellt, dass alle Graphen mit N Knoten und E Kanten dieselbe Wahrscheinlichkeiten haben, ausgewählt zu werden. Ein bekanntes Ergebnis von Erdös und Renyi (1960) ist zudem, dass die Zentralitäten, vor allem aber die Grad-Zentralität, hypergeometrisch verteilt ist, was durch eine Poisson-Verteilung angenähert werden kann [2]. Weitere, vor allem mathematische Beweise, können dem Buch [2] entnommen werden. Im weiteren Teil dieser Arbeit wird demnach untersucht, ob die Verteilungen der Zentralitäten der *Poisson-Verteilung* ähnelt. Zu der Tabelle 3.2 kommt also eine weitere, letzte Zeile hinzu:

Tabelle 4.2: Weitere Eigenschaft eines sozialen Netzwerks

Eigenschaft	Beschreibung
Verteilung der Zentralitäten	die Zentralitäten eines soziales Netzwerk sollten annähernd einer Poisson-Verteilung entsprechen [21]

Da der Graph in Abbildung 4.4 ein zufällig, einmalig erzeugter Graph ist, muss ein neuer Graphen mithilfe des Generators erzeugt werden, um die Verteilung der Zentralitäten zu betrachten. Dies wird sich nicht auf die Untersuchung auswirken, denn die Verteilungen der Zentralitäten unserer Graphen sollte stets gleich oder zumindest ähnlich sein. Bei der erneuten Generierung entsteht nun folgender Graph 4.5 und die zugehörige Verteilung der *Grad-Zentralität*:



(a)



(b)

Abbildung 4.5: Verteilung der Grad-Zentralität des Graphen (b)

Es wird ersichtlich, dass die *Grad-Zentralität* normal- beziehungsweise gaußverteilt ist. Natürlich ist zu erwähnen, dass keine perfekte Normal-Verteilung zu sehen ist, sondern eine etwas nach links verschobene Verteilung. Was die möglichen Gründe dafür sind, werden später betrachtet und korrigiert. Diese Verteilung erfüllt tatsächlich die Eigenschaft aus Tabelle 4.2, denn die Poisson-Verteilung wird für ein größer werdendes λ zu einer gaußschen Normalverteilung [22]. Nun ist aber auch noch zu untersuchen, ob sich die Eigenschaft, der poisson verteilten Zentralitäten, für die *Nähe-* und *Zwischen-Zentralität* ebenfalls beobachten lässt. Um zusätzlich zu beweisen, dass es sich bei der Gauß-Verteilung der Werte nicht um einen Zufall handelt, wird ein neuer sozialer Graph generiert und die Verteilung der *Grad-*, *Nähe-*, *Zwischen-* und *Eigenvektor-Zentralität* untersucht. Hierbei ist vor allem die Frage, ob die Verteilung dieser einer tatsächlichen Poisson-Verteilung entspricht und falls ja, warum dies der Fall ist, essentiell. Ansonsten wird die Frage gestellt, warum es keiner Poisson-Verteilung entspricht und ob es möglich ist, den Graphen zu verändern um eine solche zu erzielen. Bei der erneuten Generierung entstehen schließlich folgende Plots:

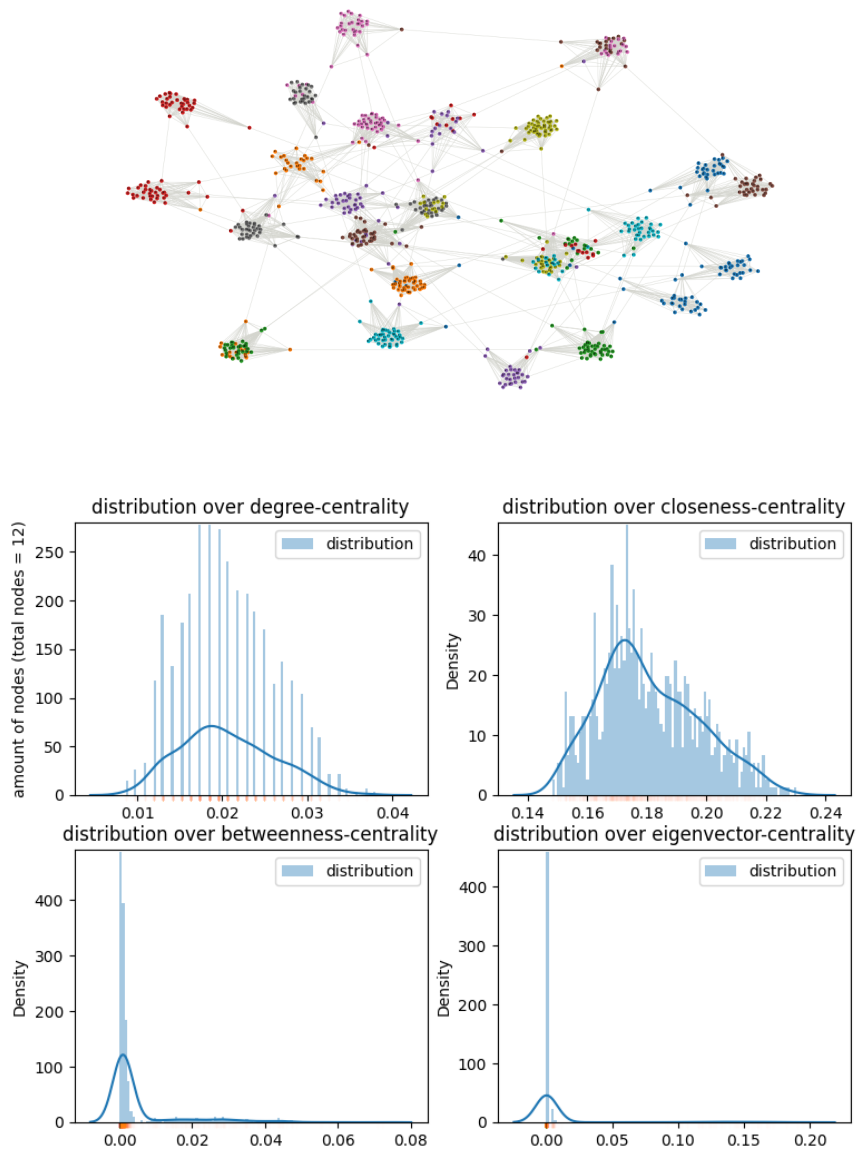


Abbildung 4.6: Zufälliges soziales Netzwerk und realistischeren Verteilungen

In der Abbildung 4.6 sieht man nun die Verteilungen der Zentralitäten von dem, sich darüber befindenden, sozialen Netzwerks. Die Tabelle mit den Zentralitäts-Werten des Netzwerks befindet sich als Datei in [23]. Oben links befindet sich die Verteilung der *Grad-Zentralität*, welche wie bereits zuvor festgestellt, nicht exakt normalverteilt ist, aber Ähnlichkeiten zu erkennen sind und demnach ebenfalls eine annähernde Poisson-Verteilung zu erkennen ist. Vor allem ist auffällig, dass der Balken bei 0.012 vergleichsweise sehr hoch ist. Über 50 Knoten

weisen diesen Wert auf. Danach geht der darauf folgende Balken nochmals zurück, denn nur noch etwas über 25 Knoten haben eine Zentralität von circa 0.03. Jedoch war zu erwarten, dass sich das Balkendiagramm symmetrisch verhält um der mathematischen Verteilung zu entsprechen, doch das Gegenteil tritt ein. Der genaue Grund hierfür ist mir nicht ersichtlich, aber besteht die Vermutung, dass es nicht weiter schlimm ist und es ausreicht, dass die Verteilungen lediglich annähernd der Poisson-Verteilung entsprechen. Über 150 Knoten weisen eine Zentralität von 0.0125 auf, daher sollten auch ebenso viele den Wert 0.025 besitzen. Hingegen ist positiv hervorzuheben, dass genau *ein* Peak erreicht wurde, wie auch zu erwarten war, um der mathematischen Verteilung zu entsprechen. Außerdem sind alle Balken vor dem Peak kontinuierlich aufsteigend und nach dem Peak kontinuierlich absteigend. Doch lediglich eine Unstimmigkeit sticht hier heraus, bei dem Zentralitätswert von 0.0357 den etwas unter 25 Knoten besitzen. Fraglich ist hier, warum der Balken erneut höher ist als sein Vorgänger. Denn im Regelfall sollten maximal ein bis drei Knoten gefunden werden, die diesen Wert aufweisen. Doch im Allgemeinen weist der Plot genau die Eigenschaft nach, die auch zu erwarten ist, nämlich dass die *Grad-Zentralität* annähernd poisson verteilt ist.

Das Balkendiagramm der *Nähe-Zentralität* weist einen ähnlichen Verlauf auf wie das der *Grad-Zentralität*. Wir erkennen erneut das erwartete Peak und weitere Balken, die im linken Bereich sehr schnell zum Peak hin ansteigen und rechts vom Peak vergleichsweise langsam abflachen. Auffällig ist erneut, dass der letzte Balken wider Erwartens höher ist als der Balken davor. Eine Aussage, welche auf jeden Fall getroffen werden kann ist, dass es erneut zu Unstimmigkeiten kommt, welche stets an anderen Stellen auftreten und nicht immer denselben Balken betreffen. Doch ist erneut eine Normal Verteilung zu erkennen, die der Poisson Verteilung für hohes λ entspricht [22]. Es kann im Allgemeinen zudem angenommen werden, dass je größer der Graph ist, umso eher sind die Zentralitäten von diesem poisson verteilt. Was daran liegt, dass dieser dann mehr Knoten besitzt und diese irgendwann zwangsläufig eine Regelmäßigkeit aufzeigen, da Zahlen in der Mathematik prinzipiell nicht *zufällig*, sondern normalverteilt sind. Grob zusammengefasst, kann die Existenz einer Kante als Binomialverteilung interpretiert werden und diese konvergiert mathematisch gesehen bei einer sehr großen Stichprobe (Anzahl an Knoten in unserem Fall) gegen eine Normal- bzw Gaußverteilung.

Bei der *Zwischen-* und *Eigenvektor-Zentralität* sind andere Verteilungen zu erkennen. Zum einen weisen die Balken wenige unterschiedliche Werte auf, zum anderen sind die Ausschläge nicht mehr mittig sondern direkt zu Beginn der Verteilung. Was zunächst verwunderlich erscheint, ist mit einer simplen Erklärung begründet. Die *Nähe-Zentralität* gibt bekanntlich an, wie oft ein Knoten anteilmäßig bei der Suche nach dem kürzesten Weg durch einen Graphen benutzt wird. Der Ausschlag ist daher die Folge davon, wenn viele kürzeste Wege stets über die gleichen Knoten verlaufen. Das heißt, es existieren keine bis wenige Alternativen und daher verlaufen die kürzesten Wege von beispielsweise Knoten 1 zu einem weiteren Knoten stets über gleiche, beziehungsweise ähnliche Knoten. Bei der *Eigenvektor-Zentralität* wird zwar die gleiche Beobachtung gemacht, doch sagt diese hier etwas anderes aus. Diese Zentralität gibt eine Einschätzung der Wichtigkeit des Knotens, im Bezug auf seine Nachbarn an, was bezogen auf die Balkendiagramm heißt, dass viele Knoten in diesen Graphen wichtig sind mit Einbeziehung der Nachbarn. Wobei auch vermuten werden darf, dass dies mit der hohen Anzahl an Knoten mit höher *Zwischen-Zentralität* zusammenhängt.

Das heißt im Umkehrschluss wiederum, dass mehr Cliques im Graph enthalten sind. Tatsächlich sind es **935** Knoten, **8952** Kanten und **10301** Cliques mit der maximalen Größe von acht Knoten in der Clique. Die Verteilungen sind dennoch aus der Mathematik bekannt, denn es handelt sich um die *Exponential-Verteilung*, welche ebenfalls in eine *Poisson-Verteilung* übergehen kann. Wie dies genau funktioniert, ist [17] zu entnehmen. Danach zu Urteilen handelt es sich bei dem Netzwerk in Abbildung 4.6 um ein typisches *soziales Netzwerk* nach Tabelle 3.2 und Tabelle 4.2.

4.4 KURZES RECAP

Nachdem zunächst überlegt wurde, wie soziale Netzwerke generiert werden, sind auch gleichzeitig die Probleme der Generierung aufgefallen. Daher wurde der Code fortlaufend verbessert, ein soziales Netzwerk erstellt und danach eine *soziale Netzwerk Analyse* durchgeführt. Schließlich konnte so die Bestätigung erhalten werden, dass die generierten Graphen die Anforderungen eines sozialen Netzwerks aus den Tabellen 3.2, 4.2 erfüllt. Danach ist zudem aufgefallen, dass die Zentralitäten regelmäßig sind und eine Poisson-Verteilung nachgewiesen werden kann. Doch muss im Folgenden noch die Frage beantwortet werden, wie die Verteilung der Zentralitäten bei anderen, bereits analysierten, Netzwerken aussieht.

5

DER VERGLEICH MIT SOZIALEN NETZWERKEN

Im vorherigen Teil der Arbeit haben wir uns damit beschäftigt, wie soziale Netzwerke bestmöglich und vor allem realitätsnah konstruiert werden können. Wir haben Analysen durchgeführt und festgestellt, dass die Werte der *Grad-* und *Nähe-Zentralität* näherungsweise poissonverteilt sind. Jedoch gilt dies bisher nur für die Netzwerke, die in dieser Arbeit generierten Graphen. Daher liegt es nahe, weitere sozialen Netzwerke und deren Analysen zum Vergleich heranzuziehen. Leitfragen sind hierbei, was zu erwarten ist, ob die Ergebnisse den Erwartungen entsprechen oder womöglich widersprechen und warum dies der Fall ist. Zusätzlich soll optimalerweise eine Möglichkeit erarbeitet werden, wie die Graphen bzw. die Generierung dieser angepasst werden könnte, um möglicherweise noch bessere Graphen oder bessere Verteilungen der Zentralitäten zu erhalten.

5.1 DER DATENSATZ UND DIE ANALYSE

Auf der Suche nach vergleichbaren sozialen Netzwerken, beziehungsweise Datensätzen, ist die Suche scheinbar endlos. Auf vielen Webseiten sind große Datensätze für alle Nutzer*innen frei zugänglich. Meistens als *comma separated values* (CSV) Datei, welche ideal zur Erstellung von Graphen mit unserem Generator geeignet sind. In diesem Teil der Arbeit betrachten wir mehrere Datensätze. Natürlich zu einem aufgrund der Tatsache, dass sie spannend sind aber auch zum anderen, um mehrere Vergleichswerte zu erhalten. Starten wir zunächst mit den Daten [12] von unserem *Game of Thrones* Graphen in Abbildung 3.3. Da bereits die Analyse der *Zentralitäten* und die generelle visuelle Analyse des Graphen durchgeführt ist, reicht nun lediglich die Verteilung der Zentralitäten zu betrachten.

Nachdem der Datensatz als CSV Datei in dem Generator eingelesen und anschließend geplottet wurde, wird folgender Graph konstruiert:

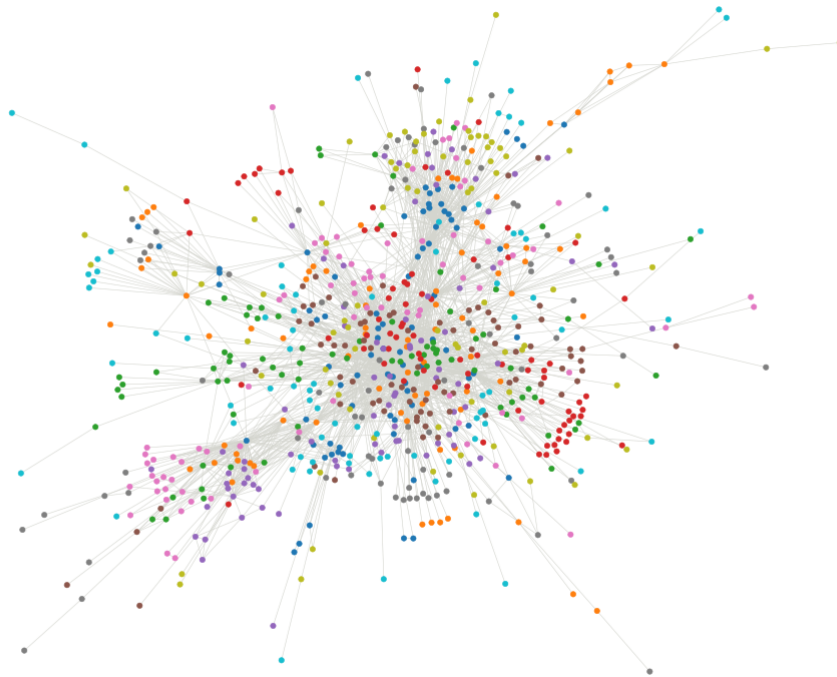


Abbildung 5.1: Game of Thrones Graph 2.0,
selbst erstellt

Dieser Plot bleibt absichtlich unkommentiert, da er lediglich zur Argumentation für die Verteilung der Zentralitäten benötigt wird und daher die visuelle Form des Graphen nur von zweitrangiger Bedeutung für diese Arbeit ist. Zudem ist zu vermerken, dass der eigentliche Datensatz gewichtet ist, und die bisher generierten Graph daher bereits schon visuell nicht dem Graphen aus Abbildung 3.3 ähnelt. Jedoch ist es sinnvoll die Gewichte außen vor zu lassen, da in dieser Arbeit ausschließlich ungewichtete Graphen nachgebildet, beziehungsweise behandelt werden. Nachdem die Daten im Generator eingelesen, die Zentralitäten berechnet und anschließend die Balkengraphen erstellt sind, ist folgender Plot 5.2 zu sehen:

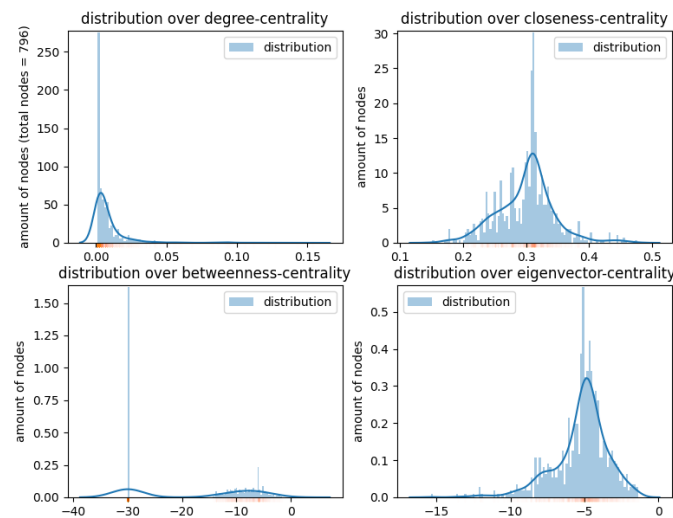


Abbildung 5.2: Game of Thrones Verteilung der Zentralitäten

Auf den ersten Blick wird bereits klar, die *Zwischen-* und *Eigenvektor-Zentralität* ähneln den Verteilungen in Abbildung 4.6 und es handelt sich um zwei annähernde *Normal Verteilungen* also demnach *Poisson-Verteilungen*. Beide Zentralitäten haben einen Ausschlag von mindestens einem Balken, was bereits im vorherigen Kapitel damit begründet wurde, dass es die Folge von vielen kürzesten Wegen ist, die stets über die gleichen Knoten verlaufen, daher keine Alternativen im Graph existieren.

Die *Grad-Zentralität* ähnelt einer Exponentialverteilung, welche ein Spezialfall der Poisson-Verteilung ist. Der Ausschlag der Balken ist erneut schnell erklärt. Es sind viele Knoten, in diesem Fall repräsentieren sie *Game of Thrones* Charaktere, die alle gleich wichtig für den Graphen sind. Diese Knoten sind daher mit vielen anderen Knoten verbunden, werden also von vielen anderen Charakteren gekannt oder kennen viele andere Charaktere. Im Allgemeinen sind die Balkendiagramme der Zentralitäten aus Abbildung 5.2 leider verglichen mit Abbildung 4.6 nicht zufriedenstellend, da sie sich untereinander nicht ähneln. Der Grund, warum die Ergebnisse stark abweicht ist vermutlich, dass es sich bei dem Graphen um fiktive Charaktere handelt. Dadurch kann es schnell zu Unstimmigkeiten kommen. Zudem war der Datensatz, bevor er im Zusammenhang dieser Arbeit verwendet wurde, gewichtet. Das kann durchaus zu anderen Werten bei der Berechnung der Zentralitäten führen. Doch wurde der Datensatz in dieser Arbeit ungewichtet betrachtet, um ihn besser mit den generierten Graphen zu vergleichen, welche ungewichtet sind. Dies kann einen Grund für Unstimmigkeiten darstellen. Zudem ist die Anzahl der geplotteten Balken stark erhöht und so fallen Unstimmigkeiten womöglich auch schneller auf. Doch wird erneut bei der Abbildung 5.2 deutlich, dass die Zentralitäten, bis auf die Zwischen-Zentralität, tatsächlich annähernd als Poisson-Verteilung interpretiert werden können. Warum dies bei der Zwischen-Zentralität nicht der Fall ist, erscheint mir un schlüssig.

Dennoch soll die Theorie, dass Zentralitäten Poisson-Verteilt sind, nicht verworfen werden und wir betrachten noch einen weiteren Datensatz. Der nächste Datensatz, der aus *Kreisen* (oder *Freundeslisten*) besteht, ist von Facebook veröffentlicht worden [6]. Die Daten wurden jedoch vor der Veröffentlichung von Facebook anonymisiert, daher ist lediglich bekannt, dass es sich bei dem Datensatz um politische Interessen handelt. So kann mit dem Datensatz festgestellt werden, dass zwei Nutzer die gleiche politische Zugehörigkeit haben, aber nicht, was ihre individuelle politische Zugehörigkeit bedeutet [6]. Nachdem die Daten wieder in eine .CSV Datei umgewandelt und anschließend geplottet wurden, ist folgender Graphen entstanden:



Abbildung 5.3: Facebook Graph mit den Datensätzen aus [6]

Der Graph ähnelt auf den ersten Blick keinem, der bisher generierten Graphen. Zudem fällt aber sofort auf, dass dieser Graph aus deutlich mehr Knoten besteht, zudem weniger Subgraphen, beziehungsweise Cluster, besitzt aber dennoch eine grundsätzlich ähnliche Struktur zu unserem Graphen in Abbildung 4.4 aufweist. Nun interessiert uns jedoch, wie diese Zentralitäten verteilt sind und ob dieser Graph die erwarteten Verteilungen nachweist. Die Zentralitäten wurden erneut berechnet und es sich folgende Verteilungen entstanden:

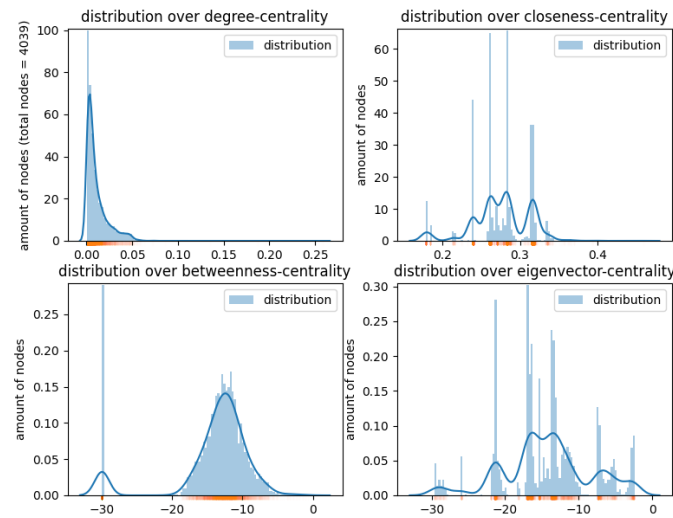


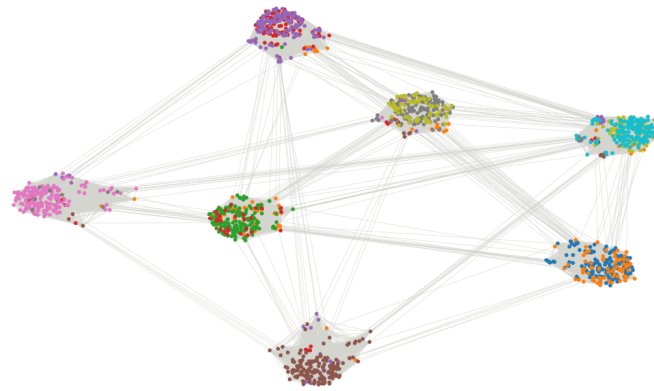
Abbildung 5.4: Facebook Graph Distribution

Die *Grad-Zentralität* fällt hier direkt auf, denn es handelt sich erneut um eine Exponentialverteilung. Die anderen Balkendiagramme der *Nähe-* und *Eigenvektor-Zentralität* ähneln jedoch den Verteilungen aus Abbildung 4.6 keinesfalls. Was zudem auffällt ist, dass die Diagramme, bis auf die Verteilung der *Zwischen-Zentralität*, an die Verteilungen von Abbildung 5.2 erinnern, welche im Endeffekt alle annähernd Poisson-Verteilt sind. Auch wenn diese Ergebnisse vielleicht ernüchternd sind, wollen wir uns überlegen, woran dies liegen kann. Bei der *Nähe-* und *Eigenvektor-Zentralität* ist eine starke Schwankung der Balken zu erkennen, wodurch mathematischen Wahrscheinlichkeitsverteilungen schwer erkennbar sind. Visuell fällt jedoch auf, dass der Graph in Abbildung 5.4 verglichen mit dem Plot des Graphen in Abbildung 4.4 durchaus Parallelen aufweist. Es sind deutliche Ansammlungen von Knoten erkennbar, die auch als Cluster bezeichnet werden können. Zwischen den Cluster sind, so wie bei Abbildung 4.4, einige Kanten zu erkennen, die die Cluster untereinander verbinden. Natürlich weist der obige Graph in Abbildung 5.3 deutlich mehr Kanten und Knoten auf, als die bisherigen Graphen. Unsere Graphen haben im Schnitt um die **950** Knoten und **8700** Kanten, daher also circa neun mal so viele Kanten wie Knoten. Auch existieren im Schnitt um die **10100** Cliques, welche maximal acht Knoten groß sind. Bei dem Facebook Graphen in Abbildung 5.3 hingegen **4093** Knoten und **88234** Kanten. Das heißt circa einundzwanzig mal so viele Kanten wie Knoten. Leider ist die Anzahl an Knoten und Kanten des Graphen in Abbildung 5.3 nur durch die Homepage [6] bekannt, denn der Datensatz ist zu groß, um die Analyse der Zentralitäten und die Untersuchung auf Cliques zu Ende zu führen. Daher ist auch die genaue Anzahl an Cliques dieses Graphen unbekannt, doch kann vermutet werden, dass diese möglicherweise höher sind als bei Abbildung 4.6, denn es existieren mehr Knoten mit ähnlich hohen Zentralitäten.

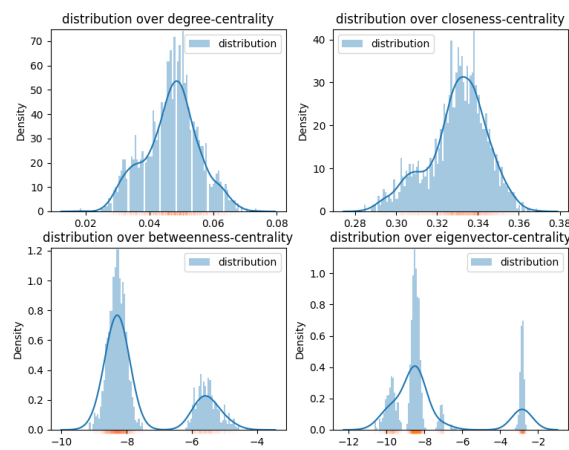
Schließlich wird doch noch ein letzter Versuch gestartet, und die Kanten und Knoten im Code des Graphen Generators werden erhöht, um damit die selben Relation zu erhalten wie in Abbildung 5.3. Dadurch wird womöglich gezeigt, dass alle vier untersuchten Zentralitäten annähernd Poisson verteilt sind. Es darf aber auf jeden Fall festgehalten werden, dass die *Betweenness*- und *Eigenvektor-Zentralität* bei allen untersuchten Datensätzen starke Parallelen zu den Verteilungen des generierten sozialen Netzwerk nachweisen. Jedoch ist nach wie vor die Verteilung der *Grad-Zentralität* verwunderlich. Daher ist es ratsam, den Code und den damit verbundenen Plot so anzupassen, dass es Abbildung 5.4 ähnelt. Anschließend kann die Verteilungen der Zentralitäten betrachtet und dadurch womöglich eine genauere Aussage erzielt werden.

5.2 ANPASSUNG DES GENERIERTEN SOZIALEN NETZWERKS

Der folgende Plot wird nun weitestgehend an Abbildung 5.4 angepasst. An dieser Stelle muss betont werden, dass es sich bei den bisher generierten sozialen Netzwerken keinesfalls um untypische oder falsche soziale Netzwerke handelt, siehe Tabelle 3.2 und Tabelle 4.2. In diesem Abschnitt wird lediglich eine bessere Vergleichsbasis hergestellt. Dies wird ermöglicht, indem zum einen die Anzahl an Cluster auf 7 Stück anpasst und die Anzahl der Knoten pro Cluster erhöht wird. Gleichzeitig sollen die jeweiligen Größen deutlich mehr variieren und vor allem die Kanten-Menge, also Anzahl an Verbindungen, stark erhöht werden. Schließlich erhält man folgende Graphen 5.5:



(a)



(b)

Abbildung 5.5: Final optimierter Graph

Es ist direkt ersichtlich, ohne die Werte genauer analysiert zu haben, dass kein zu Abbildung 5.4 identischer Graph, beziehungsweise Plot, erzeugt wurde. Allgemein sind aber die in dieser Arbeit generierten Graphen, auch wenn die Varianz der Graph-Größen best möglichst garantiert wird, auf den ersten Blick visuell gesehen ähnlich groß. Positiv zu erwähnen ist, dass bei der Verteilung der *Zwischen-* und *Eigenvektor-Zentralität* eine absolute Verbesserung erzielt, indem die Zentralitäten, bevor sie geplottet werden, logarithmiert werden. Dies wurde nachträglich auch bei allen vorherigen Verteilungen gemacht. Das ermöglicht es, die Verteilung auseinander zu zerren, da sich die Werte davor stets um 0.0 verteilt haben. Nun fällt auf, dass die Verteilungen mancher Zentralitäten doch sehr ähnlich sind. Die Verteilung der *Grad-* und *Nähe-Zentralität* erinnert beispielsweise wieder an eine Normalverteilung, welche wieder die Bedingung der Poisson-Verteilung aus Tabelle 4.2 erfüllt. Auch die *Zwischen-* und *Eigenvektor-Zentralität* weisen Parallelen auf, denn sie sehen wie jeweils zwei zusammengefügte Normalverteilungen aus.

Erneut können diese als annähernde Poisson-Verteilung gesehen werden. Das heißt, die Abbildung 5.5 erfüllt die Eigenschaft aus Tabelle 4.2 und visuell betrachtet auch die Eigenschaften aus Tabelle 3.2. Im Allgemeinen kann also gesagt werden, dass es sich um ein typisches soziales Netzwerk handelt.

Zwar entsprechen die Verteilungen von Abbildung 5.5 nicht den Verteilungen von 5.4, doch hat die Verbesserung des Graphen in diesem Kapitel dennoch viel gebracht. Unter anderem konnte bewiesen werden, dass mit wenigen Anpassungen des Codes, annähernd vergleichbare Verteilungen erhalten werden. Außerdem wurde auf diese Weise eine Abbildung 5.5 generiert, die visuell betrachtet durchaus der Abbildung 5.3 ähnelt. Um jedoch den idealen Vergleich herzustellen zu können, müssten größere Änderungen am Code vorgenommen werden. Das entspricht jedoch nicht mehr dem Umfang dieser Arbeit und wird daher nicht weiter fortgeführt. Jedoch kann als Fazit festgehalten werden, dass auch wenn die Verteilungen untereinander nicht perfekte Ähnlichkeiten aufweisen, sie im Endeffekt dennoch typische soziale Netzwerke sind und weitestgehend die Anforderungen aus Tabelle 3.2 und Tabelle 4.2 erfüllen.

6

FAZIT UND AUSBLICK

Nachdem in dieser Arbeit ausführlich die Generierung und Analyse sozialer Netzwerk behandelt wurde, werden nun die wichtigsten Erkenntnisse zusammengeführt.

Die Analyse sozialer Netzwerke besteht aus vielen Faktoren. Es gibt zahlreiche Methoden um eine Analyse durchzuführen und viele charakteristische Merkmale, die bei dieser von Bedeutung sind, siehe Tabelle 3.2 und Tabelle 4.2. In dieser Arbeit wurde lediglich ein Teil davon betrachtet, was es bei der Analyse von sozialen Netzwerken noch zu untersuchen gäbe. Zentralitäten spielen bei der Analyse beispielsweise eine wichtige Rolle, da sie in direktem Zusammenhang mit Cliquen und Brücken stehen. Die Arbeit hat gezeigt, dass Cliquen höhere *Grad-Zentralitäten* aufweisen, und hohe *Zwischen-Zentralitäten* bedeuten, dass dieser Knoten relevant für Brücken zwischen den Teilgraphen ist. In dieser Arbeit wurde zudem gezeigt, dass die Betrachtung der Cliquen und Brücken bei der visuellen Interpretation sehr aufschlussreich ist und bereits Vermutungen entstehen lässt.

Bei der Analyse der Daten ist zudem deutlich geworden, dass es hilfreich ist, die Verteilungen dieser zu betrachten. Denn oftmals werden Graphen analysiert mit einer großen Menge an Knoten und Kanten. Hier bietet die Betrachtung der Verteilung eine gute Möglichkeit, einen Überblick der Zentralitäten zu bekommen und parallel, ohne den Plot dazu gesehen zu haben, die Visualisierung zu erraten. Zwar sind die Verteilungen, die in dieser Arbeit erzielt wurden, nicht identisch gewesen doch konnten stets Parallelen zur gesuchten Poisson-Verteilung nachgewiesen werden. Gleichzeitig ist in dem Zusammenhang mit Zentralitäten zu beachten, dass soziale Netzwerke unterschiedlichste Thematiken darstellen können. Alleine eine kurze Suche im Internet präsentiert unzählige unterschiedliche Netzwerke. Diese weisen visuell gesehen starke Unterschiede auf oder womöglich keine Ähnlichkeiten, doch haben diese dennoch eine Gemeinsamkeit, sie sind alle auf ihre Weise soziale Netzwerke und erfüllen dennoch die Eigenschaften dafür. Wie beispielsweise mehrere Teilgraphen, Cliquen, Brücken und zudem ähnliche Verteilungen der Zentralitäten. Zwar kann die Interpretation dieser, sehr unterschiedlich aussehender Netzwerke, schwierig sein, doch können die untersuchten Merkmale trotzdem ähnliche oder sogar identische Ergebnisse erzielen.

Anhand dieser Arbeit wurde ebenfalls ersichtlich, dass bereits kleine Optimierungen im Quelltext des Generators, die Generierung visuell ähnlicher Graphen ermöglicht. Denn sobald Graphen gleiche oder ähnliche visuelle Grundstrukturen aufweisen, folgen direkt auch starke Gleichheiten in der Verteilung der Zentralitäten. Diese Arbeit lässt einige Punkte offen, die durchaus noch weiter optimiert werden können. Beispielsweise die generierten Plots noch besser an existierende Graphen anpassen, um die Verteilung bestmöglich nachzustellen. Der Generator in dieser Arbeit bildet lediglich mehrere Cluster und verbindet diese miteinander, wobei es in *sozialen Netzwerken* auch Knoten geben kann, die sich zwischen den Subgraphen befinden, siehe Abbildung 5.3. Dies könnte durch eine weitere Methode im Quelltext nachgestellt werden. Zudem wäre ein weiterer interessanter Faktor die Dichte in den Cluster zu untersuchen und festzustellen, ob die Knoten sehr nah beieinander liegen oder weit voneinander entfernt sind.

Dies würde sich wiederum auf die Werte der Zentralitäten und schließlich deren Verteilungen auswirken. Auch die Untersuchung größerer Datensätze, oder der Vergleich dieser wäre eine interessante Fortsetzung dieser Arbeit. Ein weiterer Ansatz für die Fortsetzung der Arbeit wäre die Untersuchung der zur Berechnung verwendeten Algorithmen. Der Frage diesbezüglich nachzugehen, ob die Algorithmen zur Berechnungen der Zentralitäten bereits optimiert sind oder ob nicht möglicherweise Verbesserungspotenzial besteht. Alle diese Ideen zeigen erneut, wie vielfältig soziale Netzwerke sind und die Analyse dieser ist und warum sie zahlreiche Wissenschaftler*innen seit Jahren beschäftigt. Schließlich kann diese Arbeit damit beendet werden, dass es unglaublich vielzählige Methoden zu Analyse von Netzwerken gibt. Welche die geeignetste ist, ob es möglicherweise viel bessere gibt, die wir womöglich nicht betrachtet haben, lässt sich nicht in einem Satz zufriedenstellend beantworten. Es kommt auf die Anzahl der Kanten und Knoten an, aber auch auf die zu untersuchende Thematik.

LITERATUR

- [1] NetworkX Developers. *Graph generators*. 2014-2022. URL: <https://networkx.org/documentation/stable/reference/generators.html> (besucht am 28.03.2022).
- [2] Ch. Donnering. "The distribution of centrality in social networks". In: *Social Networks* 8.2 (1986), S. 191–203. ISSN: 0378-8733. DOI: [https://doi.org/10.1016/S0378-8733\(86\)80003-X](https://doi.org/10.1016/S0378-8733(86)80003-X). URL: <https://www.sciencedirect.com/science/article/pii/S037887338680003X>.
- [3] Jennifer Golbeck. "Chapter 3 - Network Structure and Measures". In: *Analyzing the Social Web*. Hrsg. von Jennifer Golbeck. Boston: Morgan Kaufmann, 2013, S. 25–44. ISBN: 978-0-12-405531-5. DOI: <https://doi.org/10.1016/B978-0-12-405531-5.00003-1>. URL: <https://www.sciencedirect.com/science/article/pii/B9780124055315000031>.
- [4] Riddle M. Hanneman R. *Introduction to Social Network Methods (Hanneman)*. University of California, Riverside, 2019. URL: <https://math.libretexts.org/@go/page/7645>.
- [5] Charles Kadushin. "Introduction to Social Network Theory". In: (Jan. 2004).
- [6] By Jure Leskovec. *Social circles: Facebook*. 2012. URL: <https://snap.stanford.edu/data/ego-Facebook.html> (besucht am 28.03.2022).
- [7] Elbert E N Macau. *A mathematical modeling approach from nonlinear dynamics to complex systems*. Springer, 20198. URL: <https://www.worldcat.org/title/mathematical-modeling-approach-from-nonlinear-dynamics-to-complex-systems/oclc/1117866920>.
- [8] Peter Marsden. "Egocentric and Sociocentric Measures of Network Centrality". In: *Social Networks - SOC NETWORKS* 24 (Okt. 2002), S. 407–422. DOI: [10.1016/S0378-8733\(02\)00016-3](https://doi.org/10.1016/S0378-8733(02)00016-3).
- [9] Ruchi Nayyar. *Representing Graphs in Data Structures*. Oktober 2017. URL: <https://www.mygreatlearning.com/blog/representing-graphs-in-data-structures/> (besucht am 28.03.2022).
- [10] Christina Newberry. *How to Find and Target Your Social Media Audience (Free Template)*. 2020. URL: <https://blog.hootsuite.com/target-market/> (besucht am 28.03.2020).
- [11] Ioannis Panges. *Social Network Analysis. An Introduction*. GRIN Verlag, 2016. URL: <https://www.grin.com/document/371489>.
- [12] George Pipis. *Social Network Analysis Of Game Of Thrones In NetworkX*. September 2019. URL: <https://predictivehacks.com/social-network-analysis-of-game-of-thrones/> (besucht am 28.03.2022).
- [13] Francisco Rodrigues. "Network Centrality: An Introduction". In: März 2018. ISBN: 978-3-319-78511-0. DOI: [10.1007/978-3-319-78512-7_10](https://doi.org/10.1007/978-3-319-78512-7_10).
- [14] Britta Ruhnau. "Eigenvector-centrality — a node-centrality?" In: *Social Networks* 22 (Okt. 2000), S. 357–365. DOI: [10.1016/S0378-8733\(00\)00031-9](https://doi.org/10.1016/S0378-8733(00)00031-9).
- [15] John P. Scott und Peter J. Carrington. *The SAGE Handbook of Social Network Analysis*. Sage Publications Ltd., 2011. ISBN: 1847873952.

- [16] Laura Sheble, Kathy Brennan und Barbara Wildemuth. "Social network analysis". In: Jan. 2016, S. 250–339. ISBN: 978-1440839047.
- [17] hmas Stinfd. *Beispiele zur Poisson-Verteilung*. URL: https://mathepedia.de/Beispiele_Poisson-Verteilung.html (besucht am 15.05.2022).
- [18] Unknown. *Social Networks*. 2021, February 20. URL: <https://socialsci.libretexts.org/@go/page/8043> (besucht am 28.03.2022).
- [19] Unknown. *Web 2.0 and Social Media*. 2022, March 02. URL: <https://mitchell.libguides.com/c.php?g=529360&p=3620303> (besucht am 28.03.2022).
- [20] Stanley Wasserman und Katherine Faust. *Social network analysis: Methods and applications*. Bd. 8. Cambridge university press, 1994. URL: http://scholar.google.com/scholar.bib?q=info:gET6m8icitMJ:scholar.google.com/&output=citation&hl=en&as_sdt=0,5&as_vis=1&ct=citation&cd=0.
- [21] Wikipedia. *Exponentialverteilung*. 2008-2022. URL: https://de.wikipedia.org/wiki/Exponentialverteilung#Beziehung_zur_Normalverteilung (besucht am 20.04.2022).
- [22] Wikipedia. *Poisson-Verteilung*. 2022-04-20. URL: <https://de.wikipedia.org/wiki/Poisson-Verteilung> (besucht am 15.05.2022).
- [23] Tanja Zast. *Social Network Analysis*. 2022. URL: <https://github.com/TanjaZast/bachelor-thesis-sna> (besucht am 28.03.2022).

ERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Ausarbeitung selbst und ohne Verwendung anderer als der zitierten Quellen und Hilfsmittel verfasst habe. Wörtlich zitierte Sätze oder Satzteile sind als solche kenntlich gemacht; andere Hinweise zur Aussage und zum Umfang sind durch vollständige Angaben zu den betreffenden Publikationen gekennzeichnet. Die Ausarbeitung wurde in gleicher oder ähnlicher Form keiner Prüfungsstelle vorgelegt und ist nicht veröffentlicht worden. Diese Arbeit wurde noch nicht, auch nicht teilweise, in einer anderen Prüfung oder als Lehrveranstaltungsleistung verwendet.

Ulm, Mai 2022



Tanja Zast