



BASL Forschungsakademie

Institut für Mathematik



Die Zufallsgraphen als Modell einer kleinen Welt

Wissenschaftliche Arbeit - Masterarbeit

(Gutachter): Mark M.

Betreuer: Alexander S.

Erstprüfer: Axel D.

Zweitprüfer: Alman S.

Vorgelegt von: Maxim Weissman

Matrikelnummer: 554 420

Adresse: Leibnizstraße 100,
10625 Berlin

E-Mail: support@bachelorarbeit-schreiben-lassen.com

Telefon: +49 (0) 30 52 014 186

Studienfach: Mathematik

Sommersemester 2022

Berlin, 20.07.2021

Inhaltsverzeichnis

1	Einleitung	3
2	Grundlegende Konzepte	5
2.1	Definitionen und Axiome der Graphentheorie	5
2.2	Stochastik	6
2.3	Zufallsgraphen	10
2.3.1	Definition eines Zufallsgraphen	10
2.3.2	Tripletts in Zufallsgraphen	11
2.3.3	Die wichtigsten Messgrößen des Modells des sozialen Netzwerks	12
2.3.4	Einige Bestimmungen zu Zufallsgraphen	14
3	Die Zufallsgraphen als Modell einer kleinen Welt	15
3.1	Der Zufallsgraph von Erdős-Rényi	15
3.1.1	Beschreibung und Eigenschaften des Modells	15
3.1.2	Das verallgemeinerte Erdős-Rényi-Modell und zufällige Abstandsgraphen	20
3.2	Der Graph von Watts und Strogatz	20
3.2.1	Die allgemeine Beschreibung	20
3.2.2	Analytische und numerische Ergebnisse für ein Kleine-Welt-Modell	22
3.2.3	Sonderfälle der Berechnung der Modellkennzahlen nach Watts-Strogatz	26
3.3	Song and Wang Modell	28
3.3.1	Die allgemeine Beschreibung	28
3.3.2	Definition des Modells	29
3.3.3	Netzwerkeigenschaften des alternativen Modells	32
3.3.4	Der Kleine-Welt-Effekt	36
4	Mathematische Modellierung	39
4.1	Modellierung	39
4.2	Program	42

5	Fazit	47
6	Literatur	49
7	Anhang	51
7.1	Code	51
7.2	Eigenständigkeitserklärung	56

1 Einleitung

Diese Bachelorarbeit widmet sich der Modellierung der sogenannten kleinen Welt. Natürlich geht es nicht darum, die reale Welt in einem Miniaturformat zu erschaffen. Eine bekannte Redewendung besagt, dass die Welt, in der wir leben, klein ist, so dass wir, egal wo wir sind, immer auf vertraute Menschen treffen. Der erste Versuch, ein mathematisches Modell zu konstruieren, das die Nähe der Beziehung zwischen Menschen untersucht, gehört Manfred Kochen (University of Michigan) und Ithiel de Sola Pool (MIT), die die Ergebnisse ihrer Forschung in dem Buch "Contacts and influence" dargelegt haben. Das Buch beschreibt mathematisch die Gesetze der Dynamik sozialer Netzwerke und die Interaktion von Menschen darin. Natürlich ging es hier um persönliche Kontakte und nicht etwa um Facebook. Das Werk wurde in akademischen Kreisen mit Interesse aufgenommen, erlangte aber keine große Bekanntheit. Das Buch wurde erst etwa zwanzig Jahre später veröffentlicht, als die Idee einer kleinen Welt dank Stanley Milgrams vielbeachtetem Experiment allgemein bekannt wurde. Der amerikanische Pädagoge und Sozialpsychologe Stanley Milgram unternahm 1967 einen Versuch, die Ideen von Cohen und Pool in der Praxis auszuprobieren, indem er ein Brief-Experiment durchgeführt hat. Der Forscher wählte zwei amerikanische Städte aus, die in maximaler Entfernung zueinander liegen: Omaha und Boston, und bot den Bewohnern der ersten Stadt an, an der Untersuchung teilzunehmen. Diejenigen, die dem Experiment zugestimmt haben, bekamen Briefe und Anweisungen. Wenn der Bewohner von Omaha den Empfänger kannte, sollte der Brief direkt an ihn geschickt werden. Andernfalls wurde vorgeschlagen, unter den Bekannten denjenigen zu wählen, der den Empfänger am ehesten kennen würde. Nicht jeder Zwischenempfänger erklärte sich bereit, an dem Experiment teilzunehmen, sodass nur etwas mehr als 20 % der Briefe ans Ziel kamen. Es brauchte durchschnittlich sechs Empfänger, um das Ziel zu erreichen. Der geläufigste Name für Milgrams Experiment wurde "The Six Handshakes Rule". Das Aufkommen neuer Kommunikationsmittel, wie E-Mail, Messenger-Kanäle, soziale Netzwerke, hat die Anzahl der Kontakte von fast jedem Bewohner des Planeten, der regelmäßig Zugang zum Internet hat, erweitert. Die Qualität der Kontakte hat natürlich abgenommen, was aber nicht durch die Messgrößen der betrachteten mathematischen Modelle gemessen wird. Durch die sozialen Netzwerke haben manche Menschen eine riesige Publikumsreichweite. Der reichweitenstärkste Twitter-Account (des ehemaligen US-Präsidenten Barack Obama) hat zum Beispiel (Stand 04.05.2021) mehr als 130 Millionen Follower. Dank solcher Informationskanäle pflegen nicht nur öffentliche Personen eine große Anzahl von oberflächlichen Bekanntschaften. Das Vorhandensein einer großen Anzahl von Oberflächenbekanntschaften reduziert die durchschnittliche Anzahl von Verbindungen im Netzwerk der kleinen Welt. So analysierten die Mitarbeiter von Microsoft im Jahr 2006 mehr als 30 Milliarden E-Mails und ermittelten eine durchschnittliche Pfadlänge zwischen zwei Empfängern von 6,6. Fünf Jahre später, im Jahr 2011, wurde die Studie auf Facebook durchgeführt. Die durchschnittliche Pfadlänge betrug nur 3,74, was eine Verkürzung um 40 % im Vergleich zu Microsofts Experiment darstellt. Eine erneute Stichprobe ein paar Jahre später, die mit doppelt so vielen Nutzern des sozialen Netzwerks durchgeführt wurde, ergab eine durchschnittliche Pfadlänge von 3,57. Das mathematische Modell einer kleinen Welt wird am häufigsten mit Hilfe des Graphen "Kleine Weltumgesetz". Die Haupteigenschaft des "Kleine Welt"-Graphen ist die folgende: Zwei zufällig gewählte Eckpunk-

te haben mit hoher Wahrscheinlichkeit keine gemeinsamen Kanten, aber der eine ist vom anderen über einen Pfad mit relativ geringer Länge erreichbar. Der “Kleine Welt”- Graph besitzt eine breite praktische Anwendung. Neben der direkten Modellierung von sozialen Netzwerken, sowohl realen als auch Computernetzwerken, wird der Graph in der Soziologie, Chemie, Medizin und anderen Bereichen der modernen Wissenschaft verwendet. Das Ziel der Arbeit ist es, ein mathematisches und computergestütztes Modell eines sozialen Netzwerks mit Hilfe von Graphen zu konstruieren. Im ersten Teil werden die allgemeinen Aussagen der Graphentheorie und der Wahrscheinlichkeitstheorie betrachtet, die für die weitere fortlaufende Darstellung, Konstruktion und Einschätzung der untersuchten mathematischen Modelle, den Vergleich der theoretischen Einschätzung der Eigenschaften mit den Ergebnissen der Computermodellierung notwendig sind. Das zweite Kapitel beschäftigt sich mit einer kurzen Untersuchung der Eigenschaften, die den Graphen eines sozialen Netzwerks charakterisieren. Dazu gehören “Kleine Welt” sowie die Gradverteilung und der Grad der Clusterung (Transitivität). Das Kapitel beschreibt mathematisch die grundlegenden Eigenschaften, die für eine realistische Modellierung sozialer Netzwerke erforderlich sind, sowie die Arten von Zufallsgraphen, die zur Modellierung des sozialen Netzwerks mit der einen oder anderen Approximationsqualität verwendet werden können. Das dritte und vierte Kapitel sind praxisbezogen. Das dritte Kapitel ist eine aufeinanderfolgende Darstellung der Konstruktion von Algorithmen und der Ermittlung von Messgrößen der erhaltenen sozialen Netzwerkgraphenmodelle unter Verwendung verschiedener Zufallsgraphen. Die wichtigsten Vorgaben für die Gestaltung der Computerimplementierung der beschriebenen Modelle werden ebenfalls kurz skizziert. Das Ergebnis der im dritten Kapitel beschriebenen Arbeit ist der Entwurf eines Computersystems zur Modellierung sozialer Netzwerke auf der Grundlage der mathematischen Modelle des zweiten Kapitels. Das vierte Kapitel ist eine Überprüfung und Analyse der Ergebnisse der Computermodellierung unter Verwendung der Software, die auf der Grundlage des im dritten Kapitel erstellten Projekts entwickelt wurde. Außerdem werden die Grundlagen der Computeranwendung kurz beschrieben. Das konstruierte Computermodell ermöglicht z. B. die Modellierung von Gemeinschaften innerhalb eines sozialen Netzwerks, die Informationsverbreitung durch die Veröffentlichung von Influencer-Posts sowie andere Informationsflüsse im Netzwerk, die Interaktion zwischen Teilnehmern des sozialen Netzwerks und die Vorhersage des Verhaltens von Teilnehmern des sozialen Netzwerks. Ein solches Modell besitzt einen praktischen Nutzwert nicht nur für die Arbeit von Soziologen, sondern ist auch für SMM-Spezialisten nützlich, da es die Möglichkeit bietet, die Reichweite und Effektivität bestimmter Aktionen vorherzusagen, die zur Vermarktung von Ideen, Produkten oder Marken in sozialen Netzwerken eingesetzt werden.

2 Grundlegende Konzepte

2.1 Definitionen und Axiome der Graphentheorie

In diesem Teil der Arbeit wollen wir die Grundbegriffe der Graphentheorie definieren, die für die weitere Ausführung notwendig sind. Dazu müssen wir zunächst definieren, was ein Graph ist, dazu verwende ich hauptsächlich Definitionen aus dem Buch von R. Diestel für die Graphentheorie.

Definition 2.1. [c, S.2] Ein einfacher Graph ist ein Tupel von Mengen $G = (V, E)$ wobei V die Menge der Knoten und $E \subseteq [V]^2$ die Menge der Kanten ist. Ein Element $e \in E$ entspricht einer zweielementigen Teilmenge von V also $e = \{v_1, v_2\} \subseteq V, v_1 \neq v_2$. Der Grad eines Graphen ist gegeben durch $\#V$

Ein Graph besteht also, anschaulich einfach, aus Knoten und den Kanten die sie verbinden. Die Kanten sind über eben jene Knoten definiert, die sie verbinden. Zu beachten ist, dass wir durch diese Definition direkt einen einfachen Graphen erhalten, da wir doppelte Verbindungen und Verbindungen mit sich selbst ausschliessen. Dies trifft sich gut, da wir im folgenden mit einfachen, ungerichteten und ungewichteten Graphen arbeiten werden. Die Eigenschaften ungerichtet und ungewichtet werden wir nicht näher erläutern, da sie die Einführung von Eigenschaften erfordern, welche wir nicht nutzen werden. Im Folgenden gehen wir auf einige bestimmte Typen von Graphen ein:

Definition 2.2. [c, S.4] Sei $G = (V, E)$ ein Graph. Dann ist $G' = (V', E')$ ein Teilgraph von G , falls gilt $V' \subseteq V$ und $E' \subseteq E$ wir können dann schreiben $G' \subseteq G$ wobei G der Supergraph von G' ist.

Teilgraphen sind wie Teilmengen insofern, dass sie nur Elemente - sei es Knoten oder Kanten - ihres Supergraphen enthalten können.

Definition 2.3. [c, S.6] Ein Pfad ist eine Folge von Kanten $\{e_1, e_2, \dots, e_n\}$ wobei jede Kante die Form $e_i = \{v_i, v_{i+1}\}$ hat mit $i \in I \in \mathbb{N}, 1 \leq i \leq n+1$ und $v_i \neq v_j \forall i, j \in I$. v_1 und v_{n+1} werden die Endpunkte des Pfades genannt, der Pfad verläuft zwischen ihnen.

Nicht mathematisch gesehen ist ein Pfad also ein Weg über eine Folge von Knoten, der sich nicht selbst kreuzt. Ein Pfad ist allgemein eine der einfachsten Formen der Teilgraphen und häufig nur als solcher genutzt. Pfade durch einen Graphen helfen uns Verbindungen zu erkennen und werden später noch von Bedeutung sein bei dem Zusammenhang von Graphen.

Definition 2.4. [c, S.7] Ein Kreis ist ein Pfad $G = (V, E)$ mit einer Kante $\{v_{n+1}, v_1\}$. Der Kreis von G ist also $G' = (V, E \cup \{v_{n+1}, v_1\})$

Einfach ausgedrückt ist ein Kreis also ein Pfad dessen Enden verbunden wurden. Anschaulich muss man hierbei aufpassen, da ein Dreieck - also 3 jeweils verbundene Knoten - in einem Graphen ebenfalls ein Kreis ist.

Definition 2.5. [c, S.10] Ein nicht leerer Graph $G = (V, E)$ ist zusammenhängend, falls für alle $v_i, v_j \in V$ gilt, dass ein Pfad existiert dessen Endpunkte v_i und v_j sind. Zwei Knoten sind verbunden vom Grad n , falls gilt, dass der kürzeste Pfad zwischen ihnen Grad n hat.

Diese Definition ist wichtig, damit wir ohne große Umschweife erklären können, ob unser Graph aus kleinen nicht verbundenen Stückchen besteht oder eben ein ganzes ist.

Definition 2.6. Ein Wald ist ein Graph in dem kein Kreis existiert. Ein Baum ist ein zusammenhängender Wald.

Auf den ersten Blick mag es absurd erscheinen, einen Baum in Bezug auf einen Wald zu definieren, um einen Baum zu definieren. Aber wenn man sich einen graphentheoretischen Wald anschaut, dann sieht man wirklich eine unverbundene Masse von Bäumen, wodurch das Denkmodell erhalten bleibt.

2.2 Stochastik

Ein paar einleitende Notationen:

- i) Sei $A \subset \Omega$ dann ist $A^c = \Omega \setminus A$ das Komplementär von A
- ii) Sofern nicht anders angegeben sei $n \in \mathbb{N}$
- iii) Sofern nicht anders angegeben sei I eine Indexmenge und $i \in I$
- iii) $\mathcal{P}(M)$ ist die Potenzmenge von M , also die Menge aller Teilmengen von M .

Das grundlegende Element der Stochastik ist der Wahrscheinlichkeitsraum, um diesen definieren zu können müssen wir zuerst seine logischen Vorgänger definieren

Definition 2.7. Das Tupel (Ω, Σ) ist ein messbarer Raum, wenn Ω eine nicht leere Menge ist und Σ eine σ -Algebra über Ω . Dies ist der Fall, wenn folgende Voraussetzungen erfüllt sind:

- i) $\Omega \in \Sigma$
- ii) Wenn $A \in \Sigma$ dann gilt $A^c \in \Sigma$
- iii) Für $A_1, A_2, \dots \in \Sigma \Rightarrow \bigcup_i A_i \in \Sigma$

Wir haben hier also eine Menge und die σ -Algebra, die ihrerseits aus Teilmengen von der ursprünglichen Menge besteht, nun brauchen wir noch ein Maß auf diesem Raum und damit kommen wir zur Definition:

Definition 2.8. Sei (Ω, Σ) ein messbarer Raum. Eine Abbildung $P : \Sigma \rightarrow [0, \infty]$ ist ein Maß, falls gilt:

- i) $P(\emptyset) = 0$
- ii) Für jede Folge paarweiser disjunkter Mengen $(A_n)_{n \in \mathbb{N}}$ gilt:

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

(Ω, Σ, P) wird allgemein Maßraum genannt. Dieser ist schon in großen Teilen so, wie wir uns unseren Wahrscheinlichkeitsraum vorstellen. Es fehlen nur noch ein paar kleine Einschränkungen:

Definition 2.9. Sei (Ω, Σ, P) ein Maßraum mit folgenden Eigenschaften:

- i) Ω ist eine abzählbare Menge
- ii) $\Sigma = \mathcal{P}(\Omega)$ die Potenzmenge von Ω
- iii) $P : \Sigma \rightarrow [0, 1]$
- iv) $P(\Omega) = 1$

Dann ist (Ω, Σ, P) ein diskreter Wahrscheinlichkeitsraum.

Wir werden im Folgenden nur diskrete Wahrscheinlichkeitsräume betrachten, sofern nicht anders angegeben. Deswegen sind die zusätzlichen Einschränkungen direkt gegeben. Als zusätzliche Notation sei festzuhalten, dass Ω unser Ergebnisraum mit den $\omega \in \Omega$ Elementarereignisse ist. Σ ist das Ereignissystem mit den Ereignissen $A \in \Sigma$ und zuletzt ist P unser Wahrscheinlichkeitsmaß, welches einem Ereignis eine Wahrscheinlichkeit von 0="absolut unmöglich" bis 1="absolut sicher" zuordnet. Als nächstes kommen wir zu den sogenannten Zufallsvariablen:

Definition 2.10. Sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum und (Ω', Σ') ein messbarer Raum. Eine Abbildung $X : \Omega \rightarrow \Omega'$ ist eine Zufallsvariable, wenn gilt, dass $X^{-1}(A') \in \Sigma \forall A' \in \Sigma'$. Wenn Ω' abzählbar ist und $\Sigma' = \mathcal{P}(\Omega')$ dann ist X eine diskrete Zufallsvariable.

Das erste Kriterium bedeutet, dass unsere Zufallsvariable eine messbare Funktion sein muss. Die Einschränkung auf diskrete Zufallsvariablen erfolgt direkt, da wir in dieser Arbeit uns insbesondere mit diesen beschäftigen werden, sofern nicht anders angegeben. Außerdem können wir sie wesentlich besser erfassen als ihre nicht diskreten Artgenossen. Kommen wir zu einem großartigen Beispiel diskreter Zufallsvariablen der Indikatorfunktion:

Definition 2.11. Sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum und $A \in \Sigma$ ein Ereignis. Die diskrete Zufallsvariable χ_A ist die Indikatorfunktion von Ereignis A , falls gilt:

$$\chi_A(\omega) = \begin{cases} 1, & \text{wenn } \omega \in A \\ 0, & \text{wenn } \omega \notin A \end{cases} .$$

Diese Indikatorfunktion erlaubt es uns ein Ereignis genauer zu betrachten wie zum Beispiel mit dem folgenden wichtigen Werkzeug der Stochastik, dem Erwartungswert:

Definition 2.12. Sei X eine diskrete Zufallsvariable auf dem diskreten Wahrscheinlichkeitsraum (Ω, Σ, P) . In diesem Fall ist der Erwartungswert von X gegeben durch:

$$E(X) = \sum_i x_i \cdot P(X = x_i)$$

Mit anderen Worten erhalten wir den Erwartungswert, indem wir die angenommenen Werte mit ihrer Wahrscheinlichkeit multiplizieren und das ganze am Ende aufsummieren, das bringt uns zu den Indikatorfunktionen:

Korollar 2.13. Sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum, $A \in \Sigma$ ein Ereignis und χ_A die Indikatorfunktion von A . $E(\chi_A) = P(A = \omega \in A)$

Beweis. Nach Definition 2.12 gilt, dass $E(\chi_A) = \sum_i x_i \cdot P(X = x_i)$. Da χ_A nur zwei Werte (0 und 1) annehmen kann erhalten wir: $E(\chi_A) = 1 \cdot P(A = \omega \in A) + 0 \cdot P(A^c = \omega \notin A) = 1 \cdot P(A = \omega \in A) = P(A)$ \square

Dies ist ebenso hilfreich wie offensichtlich, da wir somit mathematisch erwarten können, dass ein Ereignis mit eben seiner Wahrscheinlichkeit eintritt. Zum Schluss noch zu einer wichtigen Eigenschaft in der Stochastik, der Unabhängigkeit:

Definition 2.14. [d,Def 5.6] Sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum. Eine Gruppe von Ereignissen (A_i) , $A_i \in \Sigma$, $i \in I \neq \emptyset$ heißt stochastisch unabhängig, wenn für jede endliche Teilmenge $\emptyset \neq J \subset I$ gilt:

Definition 2.15 (Bernoulli-Schema). Eine weitere probabilistische Konstruktion, die in der Wissenschaft und ihren Anwendungen sehr weit verbreitet ist. Angenommen, es wird eine Münze auf einen Tisch geworfen. Nehmen wir an, dass $p \in [0, 1]$ – das ist die Wahrscheinlichkeit der Zahl. Entsprechend ist dann die Wahrscheinlichkeit für Kopf $q = 1 - p$., die Münze fällt nicht auf den Rand.

Dann werfen wir eine Münze auf den Tisch und vermerken jedes Mal die “geworfene” Seite. Wenn wir Zahl erhalten, schreiben wir 1 und sagen, dass wir erfolgreich waren; wenn wir Kopf bekommen, schreiben wir 0 und sagen, dass wir gescheitert sind. Insgesamt werden n “Tests” durchgeführt. Es handelt sich hierbei um das Bernoulli-Prüfschema. Der Ausgang ist eine Folge von Nullen und Einsen der Länge n . Diese Reihenfolge ist natürlich zufällig. Wie groß ist aber die Wahrscheinlichkeit, dass es genau die Folge (x_1, \dots, x_n) mit bestimmten x_1, \dots, x_n sein wird? Es liegt nahe, dass diese Wahrscheinlichkeit gleich dem Produkt der Wahrscheinlichkeiten der Zahlen x_1, \dots, x_n ist. Mit anderen Worten: Wenn $\omega = (x_1, \dots, x_n)$, dann ist

$$P(\omega) = p^{\sum_{i=1}^n x_i} q^{n - \sum_{i=1}^n x_i} \tag{1}$$

Es entsteht wieder ein Wahrscheinlichkeitsraum. Die Elementarereignisse darin sind eben die Folgen ω : Daraus folgt, dass Ω – der Raum der Elementarereignisse – aus allen möglichen Realisierungen des Bernoulli-Testschemas 2.15 besteht, welche, 2^n betragen. Die Ereignisse sind alle möglichen Teilmengen von Ω . Zum Beispiel wird das Ereignis in n Bernoulli-Tests gibt es genau k Erfolge aus C_n^k Folgen von Nullen und Einsen mit genau k Einsen gebildet. Die Menge der Ereignisse F hat die Potenz 2^{2^n} .

Die Wahrscheinlichkeit P ist eine Funktion, die für F definiert ist und Werte aus $[0, 1]$ annimmt. In Analogie zur klassischen Wahrscheinlichkeit wird für ein gegebenes $A \in F$ als gleich der Summe der Wahrscheinlichkeiten der Elementarereignisse angenommen, die A begünstigen: $P(A) = \sum_{\omega \in A} P(\omega)$. Die Wahrscheinlichkeit für das im vorigen Absatz beschriebene Ereignis ist zum Beispiel $C_n^k p^k q^{n-k}$. Sie wird als Binomialwahrscheinlichkeit bezeichnet.

Nehmen wir ein Beispiel im Zusammenhang mit dem Bernoulli-Schema und den binomischen Wahrscheinlichkeiten. Gegeben sei die Menge $\mathfrak{R}_n = 1, \dots, n$. Wir erzeugen das Bernoulli-Testschema und entnehmen daraus das Element i , sofern der i -te Test erfolgreich war; andernfalls gehen wir zum nächsten Test über. Das Ergebnis ist eine zufällige Teilmenge $U \subset_n$ (bestehend aus extrahierten Elementen). Auf ähnliche Weise konstruieren wir die Mengen V und W . Es wird die Frage gestellt: Wie groß ist die Wahrscheinlichkeit, dass $U \cap V \subseteq W \subseteq U \cup V$ Ein solches Ereignis bezeichnen wir mit A . Schreiben wir untereinander die Erkenntnisse der Bernoulli-Schemata, die die Mengen U, V, W : „erzeugt“ haben:

$$\begin{aligned} &(x_1, x_2, \dots, x_n) \\ &(y_1, y_2, \dots, y_n) \\ &(z_1, z_2, \dots, z_n) \end{aligned} \tag{2}$$

Wenn A eingetreten ist, dann ist leicht zu sehen, dass für ein beliebiges i die Spalte mit der Nummer i in der obigen Tabelle nicht aus den Elementen $x_i = 1, y_i = 1, z_i = 0$ und aus den Elementen $x_i = 0, y_i = 0, z_i = 1$ bestehen kann: Im ersten Fall würde sich herausstellen, dass das i -te Element von \mathfrak{R}_n zu U und V gehört, aber nicht zu W , was unmöglich ist, weil wir angesichts von $A, U \cap V \subseteq W$ haben; im zweiten Fall würde es der Bedingung $W \subseteq U \cup V$ widersprechen. Alle anderen Dreiergruppen aus Nullen und Einsen in Spalten sind möglich.

Die Wahrscheinlichkeit der Triplets $1, 1, 0$ ist offensichtlich p^2q ; die Wahrscheinlichkeit der Triplets $0, 0, 1$ ist q^2p . Nach den Eigenschaften der Wahrscheinlichkeit ist die Wahrscheinlichkeit, dass das „verbotene“ Triplet in der gegebenen Spalte nicht vorkommt, also $1 - (p^2q + q^2p) = 1 - pq(p + q) = 1 - pq$. Wir erhalten schließlich $P(A) = (1 - pq)^n$

Definition 2.16 (Allgemeiner endlicher Wahrscheinlichkeitsraum). Betrachten wir die probabilistischen Konstruktionen, die in den vorherigen Abschnitten besprochen wurden, als allgemein gültig. Nehmen wir an, Ω sei eine beliebige endliche Menge. Nennen wir es den Raum der Elementarereignisse. Nehmen wir an, dass F eine beliebige Menge von Teilmengen von Ω ist, die diese Eigenschaften besitzen:

1. für beliebige $A, B \in \mathfrak{S}$ ist $A \cup B \in \mathfrak{S}$ erfüllt;
2. für beliebige $A, B \in \mathfrak{S}$ ist $A \cap B \in \mathfrak{S}$ erfüllt;
3. für ein beliebiges $A \in \mathfrak{S}$ ist $\bar{A} = \Omega \setminus A \in F$ erfüllt;
4. $\Omega \in F\mathfrak{S}$

Die Elemente der Menge \mathfrak{S} bezeichnen wir als Ereignisse. Wozu sind die Eigenschaften da? Um korrekt über die „gleichzeitige Erfüllung von zwei beliebigen Ereignissen

sprechen zu können (sie zu kreuzen), über die Erfüllung von "mindestens einem" von zwei beliebigen Ereignissen (sie zu kombinieren) und über die "Nichterfüllung eines beliebigen Ereignisses (die Negation zu nehmen). In diesem Fall muss "mindestens etwas passieren", d. h. Ω ist Ereignis, und es gibt ein unmögliches Ereignis $\emptyset = \Omega$. In der Regel ist \mathfrak{S} jedoch die Menge aller Teilmengen des Elementarereignisraums, so dass die Eigenschaften bekanntlich erfüllt sind. P sei eine Funktion von \mathfrak{S} in $[0, 1]$ mit den Eigenschaften

1. Für jedes Ereignis $A \in \mathfrak{S}$ ist $P(A) \in [0, 1]$ erfüllt.
2. $P(\Omega) = 1, P(\emptyset) = 0$.
3. Unabhängig von $A, B \in \mathfrak{S}, P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Insbesondere, wenn $A \cap B = \emptyset$ dann ist $P(A \cup B) = P(A) + P(B)$.
4. Für jedes Ereignis $A \in \mathfrak{S}$ ist $P(A) = 1 - P(\bar{A})$ erfüllt.

Dann nennen wir sie eine Wahrscheinlichkeit oder ein Wahrscheinlichkeitsmaß und das Triplet $(\Omega, \mathfrak{S}, P)$ einen Wahrscheinlichkeitsraum. Wenn $\Omega = \omega_1, \dots, \omega_n$ und $P(\omega_1) = p_1, \dots, P(\omega_n) = p_n$, dann ist unabhängig von $A \in \mathfrak{S}, P(A) = \sum_{\omega_i \in A} p_i$ und natürlich, $p_1 + \dots + p_n = P(\Omega) = 1$ erfüllt. Wir haben eine direkte Verallgemeinerung der früheren Situationen.

Definition 2.17 (Martingals). Auf einem endlichen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{S}, P)$ ist ein Martingal eine beliebige Folge von Zufallsvariablen $\xi_0, \xi_1, \dots, \xi_m$, so dass für jedes i

$$M(\xi_i \mid \xi_0, \dots, \xi_{i-1}) = \xi_{i-1}. \quad (3)$$

Das Wort "Martingal" bedeutet im Französischen "Hilfszügel" (die baumelnde Spitze eines Gürtels, eines Riemens). Martingale spielen eine große Rolle in der Finanzmathematik. Sie sind auch für die Zufallsgraphen-Theorie wichtig. Zunächst einmal dienen sie als ein weiteres Werkzeug, um Ergebnisse über die dichte Konzentration von Zufallsvariablen in der Nähe ihrer mathematischen Erwartungen zu erhalten.

2.3 Zufallsgraphen

2.3.1 Definition eines Zufallsgraphen

Ursprünglich beschrieb die Graphentheorie reguläre Graphen, aber seit den 1950er Jahren begann man, komplexe Netzwerke ohne deterministische Konstruktionsprinzipien mit Hilfe der Zufallsgraphen-Theorie zu beschreiben, die als das am besten geeignete Modell für komplexe Netzwerke vorgeschlagen wurde. Das Festhalten an allen oder nur einigen der oben skizzierten Zufallsgraphen-Paradigmen hat zu mehreren Ansätzen zur Modellierung sozialer Netzwerke mittels Graphen geführt: Die Verwendung des klassischen Zufallsgraphen nach dem Erdős-Rényi-Modell; Das Modell "Kleine Welt", das von Watts und Strogatz erstellt wurde. Netzwerkgraphen

mit einem Verteilungspotenzgesetz - Netzwerke aus "Homonymen" zur Beschreibung eines skalenfreien Verhaltens.

Es stellt sich also die Frage: Was ist ein Zufallsgraph? Nun wir nehmen zuerst an, dass die Knoten festgelegt sind als eine Menge $V = \{1, 2, \dots, n\}$. Anschliessend wählen wir die Menge aller potenziellen Kanten zwischen diesen Knoten, wobei wir nur einfache ungerichtete Graphen im Sinn haben. Diese Menge ist nach kurzem Nachdenken erkennbar als $\frac{n(n-1)}{2} = \binom{n}{2}$. Es wird ab hier klar, dass ein Zufallsgraph in dem Sinne kein Graph sein kann wie wir ihn oben beschrieben haben. Stattdessen gilt mathematisch gesehen folgendes:

Definition 2.18. Sei (Ω, Σ, P) ein Wahrscheinlichkeitsraum. Ein Zufallsgraph mit der festgelegten Knotenmenge V ist eine messbare Abbildung $\mathcal{E} : \Omega \rightarrow \mathcal{P}(\binom{n}{2})$, wobei wir auf der endlichen Menge $\mathcal{P}(\binom{n}{2})$ die Potenzmenge $\mathcal{P}(\mathcal{P}(\binom{n}{2}))$ als σ -Algebra nehmen. Für jedes $\omega \in \Omega$ ist also $(V, \mathcal{E}(\omega))$ ein Graph mit n Knoten und einer Kantenmenge $\mathcal{E}(\omega) \in \binom{n}{2}$ und für alle $V \subset \binom{n}{2}$ ist $\{\omega \in \Omega : (V, \mathcal{E}(\omega)) = (V, E)\}$ eine messbare Menge. Das Wahrscheinlichkeitsmaß sei definiert über $P(A) := P(\mathcal{E} \in A) = P(\{\omega : \mathcal{E}(\omega) \in A\})$ $A \in \mathcal{P}(\mathcal{P}(\binom{n}{2}))$

Dies bedeutet, dass unser Zufallsgraph eigentlich eine Abbildung ist, die einen Wahrscheinlichkeitsraum abbildet, auf die Menge aller möglichen Graphen über eine von uns vorgegebene Anzahl von Knoten. Die Elementarereignisse $\omega \in \Omega$ werden hier also auf jeweils einen Graphen $G(\omega) = (V, \mathcal{E}(\omega))$ abgebildet. Wir werden diesen auch einen Zufallsgraphen nennen, auch wenn er schlicht ein Element des eigentlichen Zufallsgraphen ist.

Im folgenden werden wir das Ereignis $A_e = E \in \mathcal{P}(\binom{n}{2}) : e \in E$ nutzen, um die Existenz einer Kante in unserem gewählten Zufallsgraphen zu beschreiben, sollten wir hierbei besonders auf eine Kante zwischen zwei Knoten v_1, v_2 hinweisen wollen können wir auch schreiben A_{v_1, v_2}

2.3.2 Tripletts in Zufallsgraphen

Betrachten wir ein Beispiel für Probleme auf Zufallsgraphen am Beispiel der Verteilung von Tripletts darin. Man bezeichne mit $T_{3,n}$ eine Zufallsvariable im Raum $G(n, p)$, die der Anzahl der Tripletts in einem Zufallsgraphen entspricht. Zum Beispiel ist $T_{3,n}(K_n) = C_n^3$, und für jeden Graphen G mit zwei Kanten ist $T_{3,n}(G) = 0$. Fassen wir die wichtigsten Aussagen zusammen und kommentieren sie.

Satz 2.19. *Man betrachte α als eine beliebige Funktion eines natürlichen Arguments n , die bei $n \rightarrow \infty$ gegen Null tendiert. Angenommen, $p(n) = \frac{\alpha(n)}{n}$ für jedes $n \in N$. Dann ist mit großer Wahrscheinlichkeit $T_{3,n} = 0$ (d. h., der Graph enthält keine Tripletts).*

Satz 2.20. *Man betrachte $p(n) \propto c/(n)$, wobei $c > 0$ eine konstante Größe ist. Dann hat $T_{3,n}$ eine asymptotische Poisson-Verteilung mit dem Parameter $\lambda = \frac{c^3}{6}$.*

Satz 2.21. *Man betrachte ω als eine beliebige Funktion des natürlichen Arguments n , die bei $n \rightarrow \infty$ gegen unendlich tendiert. Nehmen wir an, $p(n) = \frac{\omega(n)}{n}$ für jedes $n \in N$. Dann ist es sehr wahrscheinlich, dass $T_{3,n} \geq 1$ ist (d. h., der Graph enthält Tripletts).*

Im Folgenden soll erklärt werden, warum die Sätze 2.19 – 2.21 zusammengenommen ein fast erschöpfendes Bild des Tripletproblems ergeben. Tatsächlich spricht Satz 2.19 von Funktionen $p(n)$, die infinitesimal in Bezug auf $c\frac{1}{n}$ sind (die Notation $p = o(\frac{1}{n})$ ist in der mathematischen Analyse üblich); in Satz 2.21 haben wir $p(n) \propto \frac{c}{n}$; in Satz 2.21 haben wir $\frac{1}{n} = o(p)$.

Es ist klar, dass sinnvolle Funktionen in etwa so angeordnet sind: Sie sind entweder klein im Vergleich zum Bruchteil $\frac{1}{n}$, oder sie unterscheiden sich von ihm in einer konstanten Zeit, oder sie sind groß im Vergleich dazu.

Es zeigt sich, dass es in der ersten Situation (wenn die Kantenwahrscheinlichkeit klein ist) keine Triplets gibt; in der dritten Situation (wenn die Kantenwahrscheinlichkeit groß ist) gibt es Triplets; und nur in der begrenzten - zweiten - Situation tendiert die Wahrscheinlichkeit, dass keine Triplets vorhanden sind (wie auch die Wahrscheinlichkeit, dass sie vorhanden sind), weder gegen Null noch gegen Eins. Genauer gesagt, in der zweiten Situation

$$P_{n,p}(T_{3,n} = 0) \propto \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-c^3/6} \in (0, 1) \quad (4)$$

In der Wissenschaft der Zufallsgraphen ist es üblich, sobald ein solches Bild entstanden ist, von einem Phasenübergang zu sprechen, und eine Funktion vom vorliegenden $1/n$ Typ wird gemeinhin als Schwellenfunktion bezeichnet.

2.3.3 Die wichtigsten Messgrößen des Modells des sozialen Netzwerks

Das Modell des Graphen eines sozialen Netzwerks hat drei Hauptmerkmale, die die moderne Sichtweise auf komplexe Netzwerke definieren.

Definition 2.22. Die erste dieser Eigenschaften, "Kleine Welt" genannt, wurde bereits oben skizziert - der Abstand zwischen zwei beliebigen, zufällig ausgewählten Knoten ist relativ klein. Formal kann dieses Merkmal dargestellt werden als

$$L \propto \log N \quad (5)$$

wobei L ein typischer Abstand (Pfad) zwischen zwei Knoten ist und N die Gesamtzahl der Knoten des betreffenden Graphen ist.

Allerdings erweist sich dies als unzureichend für die Darstellung eines sozialen Netzwerks.

Definition 2.23. Die zweite Schlüsseleigenschaft eines Graphen, der ein modellhaftes soziales Netzwerk darstellt, ist die Verteilung der Knotengrade. Diese Eigenschaft beschreibt die Art und Weise, wie die Knoten miteinander interagieren. Einerseits wird bei Zufallsgraphen der Grad der Knoten durch die Poisson-Verteilung beschrieben

$$P(\xi = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (6)$$

wobei ξ eine Zufallsvariable ist, die auf dem betrachteten Wahrscheinlichkeitsraum definiert ist und eine zählbare Anzahl von Werten annimmt. λ ist die durchschnittliche Anzahl der erfolgreichen Tests in einem gegebenen Bereich von möglichen Ergebnissen.

Andererseits zeigen Untersuchungen von realen sozialen Netzwerken, dass die Gradverteilung der Knoten in den meisten Fällen sich entsprechend des Potenzgesetzes

$$P(k) \sim k^{-\gamma} \quad (7)$$

Das Potenzgesetz beschreibt sehr gut die Gradverteilungen von Knoten in großen sozialen Netzwerken. Gleichzeitig sind diese Potenzgesetze skalenunabhängig, so dass ein Netzwerk mit einem Potenzgesetz der Verteilung als skalenfreies Netzwerk bezeichnet wird.

Die dritte Haupteigenschaft ist ein hohes Maß an Transitivität (Clustering). Diese Eigenschaft lässt sich in der Praxis einfach veranschaulichen: "Deine Freunde sind oft auch unter sich befreundet". In der Graphentheorie bedeutet das, dass zwei Eckpunkte, die durch Kanten mit einem dritten verbunden sind, wahrscheinlich auch miteinander verbunden sind.

Wenn die Kanten in einem Netzwerk (Graph) zufällig generiert werden, dann hat die Tatsache, dass zwei Knoten einen gemeinsamen Nachbarn haben, im Allgemeinen keinen Einfluss auf die Wahrscheinlichkeit, diese Knoten miteinander zu verbinden. Daher muss bei der Verwendung von Zufallsgraphen in der Modellierung von sozialen Netzwerken eine zusätzliche Bedingung - ein hohes Maß an Transitivität oder Clustering - aufgestellt werden. Die zugehörigen Werte der Transitivität T und der Clusterbildung C bestimmen die beschriebene Eigenschaft, daher werden sie als zweite Messgröße des Modells zusammen mit der durchschnittlichen Pfadlänge verwendet.

Definition 2.24. Die Transitivität T eines Graphen wird durch die relative Anzahl der Dreiecke im Graphen im Vergleich zur Gesamtzahl der in Dreiergruppen verbundenen Knoten des Graphen definiert.

$$T = \frac{3 \times \text{Anzahl der Dreiecke im Graphen}}{\text{Anzahl der in Dreiergruppen verbundenen Knoten}} \quad (8)$$

Der Faktor drei im Zähler berücksichtigt die Tatsache, dass jedes Dreieck zu drei verschiedenen zusammenhängenden Dreiecken im Graphen beiträgt, und zwar zu einem Knoten in jedem Dreiecksknoten. Basierend auf dieser Definition ist der Wert der Transitivität normiert, da er auf dem Segment $0 \leq T \leq 1$ aufliegt, $T = 1$ bedeutet, dass alle Knoten des betreffenden Graphen in Dreiergruppen definiert sind.

Die Transitivität eines Graphen ist eng mit dem Clusterkoeffizienten des Graphen verbunden, da beide die relative Häufigkeit von Dreiecken messen.

Definition 2.25. Der Clusterkoeffizient des Graphen basiert auf dem lokalen Clusterkoeffizienten der einzelnen Knoten.

$$C_i = \frac{\text{Anzahl der im Knoten } i \text{ verbundenen Dreiecke}}{\text{die Anzahl der mit dem Knoten } i \text{ verbundenen Dreiergruppen}} \quad (9)$$

wobei eine Dreiergruppe, die um den Knoten i zentriert ist, eine Menge von zwei Kanten ist, die mit dem Knoten i verbunden sind. Wenn der Grad von Knoten i 0 oder 1 ist, wird angenommen, dass $C_i = 0$

Der Clusterkoeffizient des Graphen als Ganzes ist definiert als das arithmetische Mittel der Knoten-Clusterkoeffizienten

$$C = \sum_i C_i \quad (10)$$

Der Clusterkoeffizient des Graphen ist ebenfalls normiert $0 \leq C \leq 1$.

2.3.4 Einige Bestimmungen zu Zufallsgraphen

Satz 2.26. *Es gelte $p = \frac{c \ln n}{n}$. Wenn $c > 1$ ist, dann ist der Zufallsgraph mit hoher Wahrscheinlichkeit verbunden. Wenn $c < 1$, ist, dann ist der Zufallsgraph höchstwahrscheinlich nicht verbunden.*

Führen wir den Phasenübergang wie folgt durch. Zunächst formulieren wir ein vereinfachtes Ergebnis. Dann werden wir es kommentieren.

Satz 2.27. *Es gelte $p = \frac{c \ln n}{n}$. Wenn $c \geq 3$ und $n \geq 100$, dann*

$$P_{n,p}(G \text{ ist verbunden}) > 1 - \frac{1}{n} \quad (11)$$

Satz 2.28. *Der Satz gibt eine vereinfachte Version der ersten Aussage von Theorem 2.27 wieder. Allerdings konkretisiert es den Ausdruck "fast wahrscheinlich": nicht nur die Wahrscheinlichkeit tendiert gegen Eins, sondern sie macht es mit einer bestimmten Geschwindigkeit*

Definition 2.29. Die chromatische Zahl eines Graphen $G = (V, E)$ ist ein Wert $\chi(G)$, der gleich der minimalen Anzahl von Farben ist, mit denen alle Eckpunkte eines Graphen so gefärbt werden können, dass die Enden einer beliebigen Kante unterschiedliche Farben haben. Mit anderen Worten

$$\chi(G) = \min \{ \chi : V = V_1 \sqcup \dots \sqcup V_\chi, \forall i \forall x, y \in V_i (x, y) \notin E \} \quad (12)$$

Hier wird durch das Zeichen " \sqcup " ont, dass sich die Mengen V_i nicht paarweise schneiden.

3 Die Zufallsgraphen als Modell einer kleinen Welt

3.1 Der Zufallsgraph von Erdős-Rényi

Zufallsgraphen wurden erstmals von den ungarischen Mathematikern Paul Erdős und Alfred Rényi untersucht.

3.1.1 Beschreibung und Eigenschaften des Modells

Definition 3.1. Sei $n \in \mathbb{N}$ und $p \in [0, 1]$. Wir nennen einen zufälligen Graphen \mathcal{E} einen Erdős-Rényi Zufallsgraphen nach folgenden Eigenschaften:

- i) $V = 1, 2, \dots, n$
- ii) Die Ereignisse $\{\mathcal{E} \in A_e\}_{e \in \binom{[n]}{2}}$ sind unabhängig
- iii) $P(A_e) := P(\mathcal{E} \in A_e) = p$

Den Graphen $G(V, \mathcal{E})$ schreiben wir als $G(n, p)$ Zufallsgraphen

Dies sind also unsere Erdős-Rényi Zufallsgraphen. Sie werden definiert über zwei Werte namentlich n der Anzahl ihrer Knoten als auch über die Kantenwahrscheinlichkeit p , welche die Wahrscheinlichkeit beschreibt mit der jede einzelne Kante im Graphen existiert. Einfach gesehen haben wir eine Menge an Knoten und bevor wir eine Kante zwischen zwei Knoten zeichnen werfen wir eine Münze mit einer Wahrscheinlichkeit von p für Kopf. Nur, wenn Kopf erscheint zeichnen wir diese Kante und dies führen wir für alle möglichen Kanten zwischen den Knoten genau einmal durch.

Während wir n natürlich frei wählen können liegt unser Hauptaugenmerk auf p . Welche p sind überhaupt sinnvoll? Um hier schon ein paar Einschränkungen zu klären zuerst einmal ein Lemma:

Lemma 3.2. Für zwei beliebige verschiedene Knoten $v_1, v_2 \in V$ ist die Wahrscheinlichkeit das kein Pfad vom Grad 2 zwischen ihnen existiert gleich $(1 - p^2)^{n-2}$.

Beweis. Sei $V' = \{V \setminus \{v_1, v_2\}\}$. Sei $v' \in V'$ ein beliebiges Element dieser Menge. Dann gilt $P(A_{v_1, v'}) = p = P(A_{v_2, v'})$. Da die Wahrscheinlichkeit der Kantenexistenz unabhängig ist, gilt für die Wahrscheinlichkeit $P(A_{v_1, v'} \cap A_{v_2, v'}) = P(A_{v_1, v'}) \cdot P(A_{v_2, v'}) = p^2$. Da V' genau $n - 2$ Elemente können wir nun einfach die Wahrscheinlichkeit errechnen, dass keines dieser möglichen Ergebnisse auftritt. Dies ist genau $P(\bigcap_{i=1}^{n-2} (A_i)^c \mid A_i = A_{v_1, v'_i} \cap A_{v_2, v'_i}) \stackrel{\text{unabh.}}{=} \prod_{i=1}^{n-2} (1 - p^2) = (1 - p^2)^{n-2}$ \square

Wir können also zeigen, wenn wir für p einfach eine Zahl einsetzen, dass wir bei größeren Graphen mehr Verbindungen haben. Dies scheint soweit logisch doch betrachten wir dieses im Extremen:

Korollar 3.3. *Die Wahrscheinlichkeit, dass zwei Knoten keine Verbindung vom Grad 2 haben konvergiert bei einem konstanten, von n unabhängigen $p > 0$ und einem n das gegen unendlich geht gegen 0.*

Beweis. $\lim_{n \rightarrow \infty} (1 - p^2)^{n-2} = 0$ Da wir wissen, dass $(1 - p^2) < 1$. □

Wir können hier den trivialen Fall von $p = 0$ ausschließen, da wir in diesem Fall den erzeugten Graphen absolut sicher als $G(V, \emptyset)$ identifizieren können. Dieses Corollar sagt uns nun, dass es bei immer größeren Graphen fast sicher ist, dass zwei Knoten eine Verbindung vom Grad 2 haben, beziehungsweise, dass es praktisch ausgeschlossen ist, dass diese nicht vorliegt. Das bedeutet, dass bei sehr großen Graphen jeder Knoten der "Nachbar eines Nachbarn" ist für jeden anderen Knoten. Dies erscheint nicht zweckdienlich für unsere Simulation, da sie mit zunehmender Größe nicht unbedingt enger zusammenwachsen sollte.

Bemerkung $G(n, \frac{c}{n})$

Um dieses Dilemma zu umgehen werden wir im Folgenden die Graphen als $G(n, \frac{c}{n})$ beschreiben. Hierbei spezifizieren wir also unser p , sodass es von n abhängig wird. Derweil sei c ein beliebiger, zumeist konstanter, positiver Wert mit dem wir unsere Wahrscheinlichkeit näher bestimmen.

Diese $G(n, \frac{c}{n})$ Graphen haben sogar einen sehr günstigen Vorteil:

Lemma 3.4. *Im Zufallsgraphen $G(n, \frac{c}{n})$ gilt für jeden Knoten e_i , dass die erwartete Anzahl an benachbarten Knoten gegen c konvergiert.*

Beweis. Sei $v \in V$ ein beliebiger Knoten. Wir definieren eine Familie von Ereignissen $A_i (A_i = A_{v, v_i}, v_i \in V \setminus v$ und den zugehörigen Indikatorfunktionen

$$\chi_{A_i}(\omega) = \begin{cases} 1, & \text{wenn } \omega \in A_i \\ 0, & \text{wenn } \omega \notin A_i \end{cases} .$$

Nach Lemma 2.13 gilt nun: $E(\chi_{A_i}) = \frac{c}{n}$ Da $\frac{c}{n} \geq 0$ können wir für alle Graphen mit $n < 1$ die σ -Additivität des Erwartungswerts nutzen und erhalten $E(\sum_{i=1}^{n-1} \chi_{A_i}) = \sum_{i=1}^{n-1} E(\chi_{A_i}) = \frac{c}{n} + \frac{c}{n} + \dots = \frac{c}{n}(n-1) \approx c$ □

Während wir für kleinere n eine Abweichung hinnehmen müssen, geht $\frac{n-1}{n}$ bei zunehmendem n gegen 1 womit wir eine recht genaue Abschätzung der Anzahl der Nachbarn in unserem Graphen haben. Die Graphen die wir ausgeklammert haben, solche mit $n = 1$ können wir direkt vorhersagen, da sie auf jeden Fall genau 0 Kanten haben werden.

Satz 3.5. Satz von Erdős und Renyi [a, Satz 6.1]

Sei $G(n, \frac{c}{n})$ ein Erdős Renyi Zufallsgraph. Sei weiterhin X_n die Anzahl der Elemente von G in der größten zusammenhängenden Menge von G und X'_n der Anteil der Elemente von G in der zweitgrößten verbundenen Menge. Dann gilt folgendes:

i) Es existiert eine Funktion $g(c)$ mit

$$P(\lim_{n \rightarrow \infty} X_n = g(c)) = 1 \text{ mit}$$

$$g(c) \begin{cases} = 0, & \text{für } c \leq 1 \\ > 0, & \text{für } c > 1 \end{cases} .$$

ii) $\lim_{n \rightarrow \infty} X'_n = 0$ unabhängig von c

Beweis. Sei $B(V, E) \subseteq G$ ein Baum, dessen Knoten genau denen in X_n entsprechen. Wir können dies annehmen, da wir nur an dem Zuwachs des Teilgraphens interessiert sind und nicht an seiner Verknüpfung. Nun brauchen wir einen Satz zur Kantenperlokation, der wie folgt lautet:

Satz 3.6 (a, Satz 5.9.). Sei $d \in \mathbb{N}, d \geq 2$. Sei T_d ein Baum, in dem ein Knoten (die Wurzel) d Nachbarn hat und jeder andere Knoten exakt $d + 1$ Nachbarn hat. Dieser Baum hat eine Kantenwahrscheinlichkeit von p . Die Wahrscheinlichkeit $\psi_{T_d}(p)$, dass eine unendliche zusammenhängende Komponente existiert ist gleich:

$$\psi(p) \begin{cases} 0, & \text{für } p \leq \frac{1}{d} \\ 1, & \text{für } p > \frac{1}{d} \end{cases} .$$

Dies entspricht in großen Teilen unserer Fragestellung. Wir haben einen Baum allerdings mit einer endlichen Anzahl an Knoten, hinzu kommt das Problem, dass jede Generation die Möglichkeiten der Verbindung, für die Nachfolgende einschränkt. Insofern

Für ii) können wir bedenken, dass für die ersten beiden Optionen das Gleiche gilt für die zweitgrößte Komponente gilt wie für die größte. Das heißt, sie wird mit zunehmendem n gegen 0 konvergieren. Für $c < 1$ nehmen wir an, dass für unser größtes Element X_n und unser zweitgrößtes Element X_m beide gilt $\lim_{n/m \rightarrow \infty} X_{n/m} > 0$

In dem Fall gilt, dass die Anzahl der möglichen Verbindungen zwischen X_n und X_m ebenfalls einen Grenzwert über 0 entwickeln muss. Was bedeuten würde, dass beide Mengen sich verbinden und zusammen die größte Menge sind. Alternativ gilt $\lim_{m \rightarrow \infty} X_m = 0$, was unsere Aussage ist. \square

Dies ist ein sogenanntes Schwellenwert Problem. Unser Graph verhält sich ab der Schwelle $\frac{1}{d}$ anders als vorher und ebenso ergeht es uns im nächsten Satz. In diesem geht es um isolierte Graphen und er stellt in gewisser Weise ein Gegenpol zu dem Satz 3.5 dar, indem er uns zeigt wie sich die kleinsten Elemente unseres Graphen, die isolierten Knoten verhalten. Doch zuerst brauchen wir die sogenannte Markov Ungleichung:

Lemma 3.7. [b, S.8] Sei X eine nicht negative Zufallsvariable, die nur ganzzahlige Werte annimmt. Dann gilt: $P(X > 0) = P(X \geq 1) \leq E(X)$

Die allgemeine Fall ist etwas umfassender für unseren folgenden Beweis reicht jedoch diese Aussage, damit widmen wir uns auch unserem Satz über isolierte Knoten im $G(n, p)$

Satz 3.8. [b, S.9ff] Sei $p(n) = \frac{1}{n}(\ln n + \gamma(n))$ und $\gamma(n) \rightarrow \infty$. Sei X die Anzahl der isolierten Knoten in $G(n, p(n))$, dann ist $P(X > 0) \rightarrow 0$ für $n \rightarrow \infty$

Beweis. Zuerst sei zu beachten, dass $X = \sum_{i=1}^n X_i$ ist. Mit

$$X_i = \begin{cases} 1, & \text{falls } i \text{ isoliert ist} \\ 0, & \text{falls } i \text{ nicht isoliert ist} \end{cases} .$$

Nach Corollar 2.13 gilt nun, dass der Erwartungswert von X_i gleich $P(X_i = 1)$ ist. In diesem speziellen Fall gilt:

$$E(X_i) = (1 - p)^{n-1} = \exp(-(n-1)\ln(1-p))$$

Da $0 < p < 1$ gilt können wir über die Taylorreihe abschätzen dass $\ln(1-p) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (-p)^k \cdot R_n(p)$ ist. Dabei geht $R_n(p)$ gegen 0 für $n \rightarrow \infty$ Wenn wir dies nun einsetzen erhalten wir:

$$= \exp(-n-1) \left(\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (-p)^k \right)$$

$= \exp(-n \left(\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (-p)^k \right) - \left(\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (-p)^k \right))$ Wir ziehen den ersten Summanden aus der ersten Summe heraus und erhalten:

$= \exp(-np + n \left(\sum_{k=2}^{\infty} \frac{(-1)^{k+1}}{k} (-p)^k \right) - \left(\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (-p)^k \right))$ Diese beiden Summen können wir durch ihren ersten Summanden abschätzen, da sie nicht schneller steigen werden und schreiben somit:

$$= \exp(-np + O(p + p^2n))$$

Nun setzen wir unser vorgegebenes $p(n) = \frac{1}{n}(\ln n + \gamma(n))$ ein und erhalten:

$$E(X) = \sum_{i=1}^n E(X_i) = n \cdot \exp(-np + O(p + p^2n))$$

$$= n \cdot \exp\left(-n \cdot \frac{1}{n}(\ln n + \gamma(n)) + O\left(\frac{1}{n}(\ln n + \gamma(n)) + \left(\frac{1}{n}(\ln n + \gamma(n))\right)^2 n\right)\right)$$

$$= n \cdot \exp\left(O\left(\frac{1}{n}(\ln n + \gamma(n)) + \left(\frac{1}{n}(\ln n + \gamma(n))\right)^2 n\right)\right) \cdot \exp(-(\ln n + \gamma(n)))$$

$$= n \cdot \exp\left(O\left(\frac{1}{n}(\ln n + \gamma(n)) + \left(\frac{1}{n}(\ln n + \gamma(n))\right)^2 n\right)\right) \cdot \frac{1}{n} \cdot e^{-\gamma(n)}$$

$$= \exp\left(O\left(\frac{1}{n}(\ln n + \gamma(n)) + \frac{1}{n^2}(\ln n + \gamma(n))^2 n\right)\right) \cdot e^{-\gamma(n)}$$

$= \exp\left(O\left(\frac{1}{n}(\ln n + \gamma(n)) + (\ln n + \gamma(n))^2\right)\right) \cdot e^{-\gamma(n)}$ Dies können wir nun abschätzen nach den Exponentialregeln. Da $e^{\frac{1}{n}}$ gegen 1 läuft können wir eine obere Schranke von 1 annehmen und erhalten so:

$$= (1 + o(1))e^{-\gamma}$$

Da $\gamma(n)$ gegen Unendlich geht erhalten wir also, dass $E(X) = (1 + o(1))e^{-\gamma}$ gegen 0 geht und mit der Markov Ungleichung können wir nun sagen:

$$P(X > 0) \leq E(X) \rightarrow 0 \quad \square$$

Wir können also, abhängig von unserem gewählten p , eine Aussage darüber treffen ob es isolierte Knoten in unserem Graphen geben wird, zumindest wenn er gegen unendlich strebt. Allerdings reicht das schon um bei größeren Graphen eine

gute Abschätzung zu erreichen. Diese Regel ist ein Spezialfall der “First Moment Method”, die im Bereich der Stochastik angewandt wird und einen Spezialfall der allgemeinen Markov Ungleichung darstellt. Mit diesem können wir sogar noch mehr Aussagen treffen als Beispiel folgende:

Lemma 3.9. [e, S 18] Sei $G(n, \frac{c}{n})$ ein Erdős Renyi Zufallsgraph mit $c = o(n^{\frac{1}{3}})$, dann gibt es mit großer Wahrscheinlichkeit, keinen K_4 Teilgraphen in $G(n, \frac{c}{n})$.

Beweis. Ein K_4 ist ein Graph mit 4 Knoten wobei jeder Knoten mit jeweils allen anderen verbunden ist. Demnach hat K_4 also 6 Kanten. Wir definieren zuerst $K = \{K_1, K_2, \dots, K_m\}$, $m \in \mathbb{N}$ die Menge aller 4 elementigen Teilmengen von

$$\chi_{K_i} = \begin{cases} 1, & \text{falls } K_i \text{ ein } K_4 \text{ ist} \\ 0, & \text{falls } K_i \text{ kein } K_4 \text{ ist} \end{cases} .$$

Da ein K_4 genau 6 Kanten enthält gilt $P(\chi_{K_i} = 1) = (\frac{c}{n})^6, \forall i$ Durch die Kombinatorik erhalten wir die Anzahl der K_i diese ist: $\binom{n}{4}$. Das heißt unser Erwartungswert über alle χ_{K_i} ist $E(\chi_K) = \binom{n}{4} \cdot (\frac{c}{n})^6$ Wenn wir nun $c = n^{\frac{1}{3}}$ setzen erhalten wir die Gleichung:

$$\frac{n^4 - 6n^3 + 11n^2 - 6n}{24} \cdot (\frac{n^{\frac{1}{3}}}{n})^6 = \frac{n^4 - 6n^3 + 11n^2 - 6n}{24} \cdot \frac{1}{n^4}$$

Da der erste Term mit Sicherheit langsamer wachsen wird als der zweite können wir, dank der Markov Ungleichung sagen, da $P(\chi_K \geq 1) \leq E(\chi_K)$ gilt und $E(\chi_K)$ bei wachsendem n gegen 0 korreliert. Dass für alle $c = o(n^{\frac{1}{3}})$ die Wahrscheinlichkeit für die Existenz eines K_4 ebenfalls gegen 0 geht. \square

Wir haben nun Erdős Renyi Graphen besprochen, diese haben die Form $G(n, p)$ beziehungsweise meistens auch $G(n, \frac{c}{n})$. Allerdings gibt es nicht nur andere Zufallsgraphen, sondern sogar andere Erdős Renyi Zufallsgraphen und zwar die Zufallsgraphen der Form $G(n, m)$:

Definition 3.10. Wir definieren die Erdős-Renyi Zufallsgraphen der Form $G(n, m)$ als Graphen mit n Knoten und m Kanten wobei für $A_G = G(n, m) = G(V, E) V = 1, \dots, n, E \subset \binom{n}{2}, \#E = m$, gilt $P(A_G) = p \forall G(V, E)$

Diese Art der Zufallsgraphen gibt somit jedem Graphen mit m Kanten den gleichen Zufallswert. Die Verbindung herzustellen zu unseren $G(n, p)$ Graphen ist an dieser Stelle auch recht simpel:

Bemerkung Ein Erdős-Renyi Zufallsgraph $G(n, p)$ verhält sich ähnlich zu einem Graphen $G(n, m)$ wenn gilt $m \approx \binom{n}{2} \cdot p$. Im Fall $p = \frac{c}{n}$ also $m \approx \binom{n}{2} \cdot \frac{c}{n} = \frac{c(n-1)}{2}$

Es gibt also einen Zusammenhang zwischen den beiden Typen von Zufallsgraphen allerdings werden wir hier nicht näher darauf eingehen, stattdessen kommen wir zum nächsten Kapite indem wir uns der technischen Realisierung unseres mathematischen Modells widmen werden.

Außerdem wird bei der Betrachtung von Zufallsgraphenmodellen und der Auswahl des besten Modells notwendigerweise die Transitivitätskennzahl des resultierenden Graphen berücksichtigt.

3.1.2 Das verallgemeinerte Erdős-Rényi-Modell und zufällige Abstandsgraphen

Gegeben sei die Zahl $n \in \mathbb{N}$ und die Menge $V = \{1, \dots, n\}$. Man nehme $N = C_n^2$ und bezeichne mit e_1, \dots, e_N die Kanten des vollständigen Graphen auf der Menge der Eckpunkte V .

Früher haben wir die Kanten im Zufallsgraphen nach dem Schema 2.15 von N Bernoulli-Tests ausgewählt, jeder von ihnen hat eine Erfolgswahrscheinlichkeit (eine Kante im Graphen zu nehmen) gleich eins und gleich $p \in [0, 1]$. Nun ziehen wir eine Kante zwischen den Knoten i und j mit der Wahrscheinlichkeit p_{ij} in Abhängigkeit von den genannten Knoten. Bezeichnen wir das beschriebene Modell mit $G(n, p_{ij})$. Es ist im Wesentlichen ein Wahrscheinlichkeitsraum

$$G(n, p_{ij}) = (\Omega_n, \mathfrak{S}_n, P_{n, p_{ij}}) \quad (13)$$

in dem

$$|\Omega_n| = 2^N, P_{n, p_{ij}}(G) = \prod_{(i,j) \in E} p_{ij} \cdot \prod_{(i,j) \notin E} (1 - p_{ij}) \quad (14)$$

Gegeben sei für jede natürliche Zahl n (oder für jedes n einer beliebigen unendlichen Folge von natürlichen Zahlen) ein Graph $G_n = (V_n, E_n)$, für den $V_n = V$, und $E_n \subseteq \{e_1, \dots, e_N\}$. Man betrachte $p = p(n) \in [0, 1]$ und setze $p_{ij} = p$, wenn $(i, j) \in E_n$, und $p_{ij} = 0$, wenn $(i, j) \notin E_n$.

Mit anderen Worten, die Kanten des Graphen G_n erscheinen im Zufallsgraphen unabhängig voneinander mit der gleichen Wahrscheinlichkeit, und die Kanten, die nicht im Graphen G_n erscheinen, erscheinen überhaupt nicht im Zufallsgraphen. Das neue Modell wird folgendermaßen definiert

$$G(G_n, p) = (\Omega_n, \mathfrak{S}_n, P_{n, p}) \quad (15)$$

Hier sind

$$|\Omega_n| = 2^{|E_n|}, P_{n, p}(G) = p^{|E_n|} (1 - p)^{|E| - |E_n|} \quad (16)$$

Offensichtlich kehren wir bei $E_n = \{e_1, \dots, e_N\}$ zum klassischen Erdős-Rényi-Modell zurück: $G(K_n, p) = G(n, p)$.

3.2 Der Graph von Watts und Strogatz

3.2.1 Die allgemeine Beschreibung

Im Jahr 1998 schlugen Watts und Strogatz [f] ein alternatives Kleine-Welt-Modell vor, das besser mit unserem alltäglichen Verständnis der Natur von sozialen Netzwerken übereinstimmt.

Ihr Vorschlag war, ein Modell zu konstruieren, das im Wesentlichen ein niedrigdimensionales reguläres Gitter ist, das einen gewissen Grad an Zufälligkeit aufweist, wie z. B. ein Zufallsgraph, um einen Kleine-Welt-Effekt zu erzeugen. Sie schlugen dazu das folgende spezielle Schema vor.

Nehmen wir ein eindimensionales Gitter aus 1.

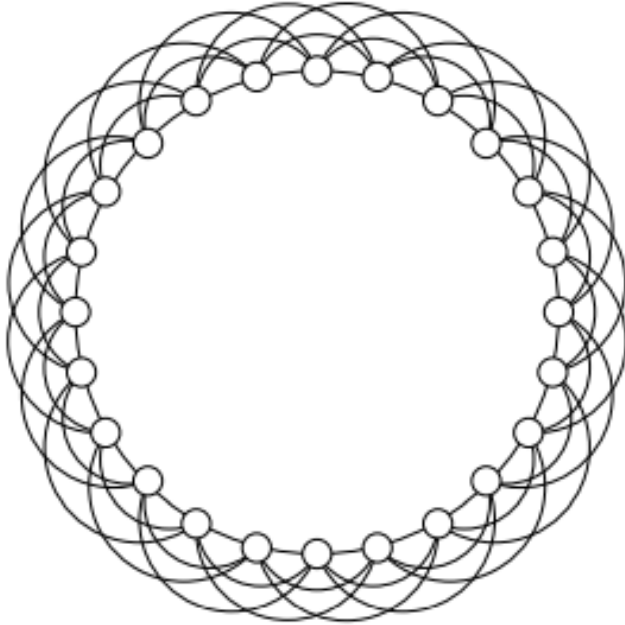


Abbildung 1: Ein eindimensionales Gitter

Gehen wir durch jedes der Glieder des Gitters und ordnen wir mit einer gewissen Wahrscheinlichkeit p zufällig ein Glied um, d. h. wir verschieben eines seiner Enden an eine neue Position, die zufällig aus dem Rest des Gitters ausgewählt wird.

Wenn p klein ist, erhalten wir einen Graphen, der immer noch größtenteils regulär ist, aber einige Verbindungen hat, die sich über lange Strecken durch das Gitter ziehen, wie in 2.

Die Koordinatenzahl des Gitters ist nach wie vor im Durchschnitt z , obwohl die Anzahl der Nachbarn eines bestimmten Knotens größer oder kleiner als z sein kann.

Auf sozialer Ebene können wir dieses Modell damit begründen, dass die meisten Menschen mit ihren unmittelbaren Nachbarn befreundet sind - den Nachbarn in ihrer Straße, den Menschen, mit denen sie arbeiten, den Menschen, denen sie von ihren Freunden vorgestellt werden -, dass aber einige Menschen auch mit ein oder zwei Menschen befreundet sind, die sozial weit entfernt sind - Menschen aus anderen Ländern, Menschen aus anderen Lebensbereichen, Bekannte aus früheren Lebensabschnitten usw. Diese entfernten Bekanntschaften werden im Modell von Watts und Strogatz durch entfernte Verbindungen dargestellt.

Offensichtlich werden die Werte des Clusterkoeffizienten $2.25 C$ für das Watts-Strogatz-Modell bei kleinen Werten von p in der Nähe der oben angegebenen Werte für das perfekt geordnete Gitter liegen, die für feste kleine d und große z zu $3/4$ tendieren. Watts und Strogatz haben auch mit numerischen Simulationen gezeigt, dass der durchschnittliche Abstand ℓ von Knotenpunkt zu Knotenpunkt mit dem eines echten Zufallsgraphen vergleichbar ist, sogar für ziemlich kleine Werte von p .

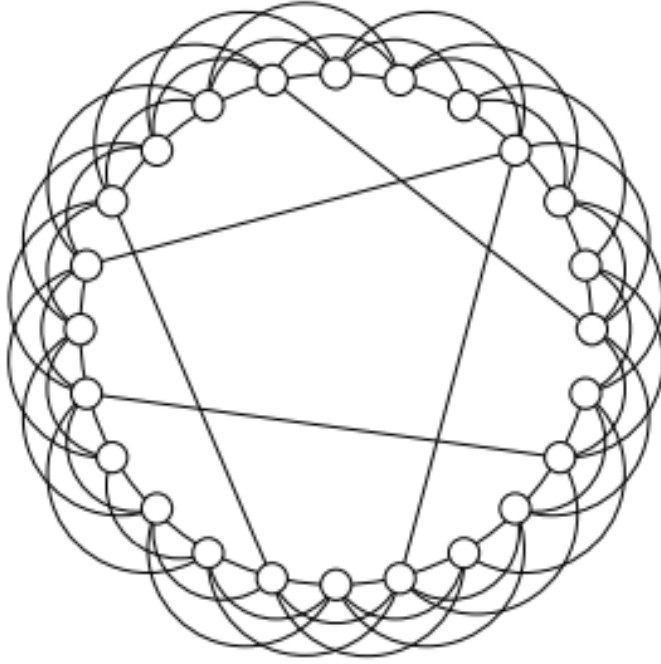


Abbildung 2: Die lange Strecken durch das Gitter

Zum Beispiel fanden sie für einen Zufallsgraphen mit $N = 1000$ und $z = 10$ heraus, dass der durchschnittliche Abstand zwischen zwei zufällig ausgewählten Knoten etwa $\ell = 3.2$ beträgt. Für ihr Rekonnexionsmodell war der durchschnittliche Abstand nur geringfügig größer, $\ell = 3.6$, und die Wahrscheinlichkeit der Rekonnexion $p = 1/4$, verglichen mit $\ell = 50$ für einen Graphen ohne jegliche Rekonnexionen.

Und selbst für $p = 1/64 = 0.0156$ fanden sie $\ell = 7.4$, etwas mehr als das Doppelte des Wertes für einen Zufallsgraphen.

Das Modell weist also sowohl Cluster- als auch Kleine-Welt-Eigenschaften auf. Dieses Ergebnis wurde sowohl durch weitere Simulationen als auch durch analytische Arbeiten an Kleine-Welt-Modellen bestätigt, die im nächsten Abschnitt beschrieben werden [g].

3.2.2 Analytische und numerische Ergebnisse für ein Kleine-Welt-Modell

Ein Großteil der neueren Arbeiten zu Kleine-Welt-Modellen wurde unter Verwendung einer Abwandlung des Watts-Strogatz-Modells durchgeführt, das von Newman und Watts 1999 [h, i] vorgeschlagen wurde

In dieser Version des Modells werden die Verbindungen zwischen den Objekten nicht mehr neu verbunden, wie in Abb. 2 werden zusätzliche Verbindungen, oft Labels genannt, zwischen zufällig ausgewählten Objektpaaren hinzugefügt, aber es werden keine Verbindungen aus dem Hauptgitter entfernt.

Dieses Modell ist etwas einfacher zu analysieren als das ursprüngliche Watts-Strogatz-Modell, da in ihm keine Region des Graphen vom Rest getrennt wird, während das im ursprünglichen Modell passieren kann. Mathematisch lässt sich ein divergenter Bereich des Graphen dadurch darstellen, dass der Abstand von einem beliebigen Knoten in diesem Bereich zu einem Knoten, der sich irgendwo im Rest des Graphen

befindet, unendlich ist. Das heißt jedoch, dass der durchschnittliche Knotenpunkt-
abstand ℓ im Modell, wenn er über alle möglichen Realisierungen des Graphen gemittelt
wird, für jeden endlichen Wert von p

Tatsächlich lässt sich zeigen, dass die Zerlegung der Reihe ℓ/L im Grad p in Bezug
auf $p = 0$ bis zur Ordnung p^{z-1} gut funktioniert, aber die Expansionskoeffizienten
für alle höheren Ordnungen unendlich sind.

Für die Version des Modells, in der keine Verbindungen entfernt werden, nehmen
die Expansionskoeffizienten bis zur Ordnung p^{z-1} , die gleichen Werte an, sind aber
auch für alle höheren Ordnungen endlich. Im Allgemeinen werden beide Versionen
des Modells als Kleine-Welt-Modelle oder manchmal auch als Kleine-Welt-Graphen
bezeichnet.

Viele Ergebnisse wurden für Kleine-Welt-Modelle erzielt, wobei viele weitere ihrer
Eigenschaften numerisch untersucht wurden.

Barthélemy und Amaral (1999) schlugen vor, dass der durchschnittliche Abstand von
einem Knoten zu einem Knoten ℓ der Gleichung $\ell = \xi G(L/\xi)$ unterliegt, wobei $G(x)$
die universelle Skalierungsfunktion ihres Arguments x und ξ die charakteristische
Längenskala für das Modell ist, von der angenommen wird, dass sie für kleine p
entsprechend $\xi \sim p^{-\tau}$ divergiert. Basierend auf numerischen Ergebnissen nahmen
Barthélemy und Amaral an, dass $\tau = 2/3$ ist. Barratt (1999) widerlegte diese zweite
Hypothese mit einem einfachen physikalischen Argument, das zeigte, dass τ nicht
kleiner als 1 sein kann, und behauptete, basierend auf numerischen Ergebnissen,
dass es in der Tat genau 1 ist.

Newman und Watts (1999) zeigten, dass das Kleine-Welt-Modell nur eine nicht-
triviale Längenskala außer dem Gitterabstand hat, die wir mit der obigen Variable
 ξ gleichsetzen können und die gegeben ist durch

$$\xi = \frac{1}{pz} \tag{17}$$

für das eindimensionale Modell, oder

$$\xi = \frac{1}{(pzd)^{1/d}} \tag{18}$$

im allgemeinen Fall τ sollte also tatsächlich gleich 1 für $d = 1$ sein, oder $\tau = 1/d$ für
ein beliebiges d , und da es keine anderen Längenskalen gibt, hätte ℓ die Form

$$\ell = \frac{L}{2dz} F(pzL^d) \tag{19}$$

wobei $F(x)$ eine weitere universelle Skalierungsfunktion ist. (Der Anfangsmultipli-
kator $(2d)^{-1}$ vor der Skalierungsfunktion ist beliebig. Sie ist so gewählt, dass sie F
einen einfachen Grenzwert für kleine Werte ihres Arguments liefert - siehe Gleichung.

Diese Form der Skalierung ist äquivalent zur Barthélemy- und Amaral-Form, indem
 $G(x) = xF(x)$ ersetzt wird, wenn $\tau = 1$. Es wurde weitgehend durch numerische

Simulationen bestätigt. (Newman and Watts 1999a, de Menezes et al. 2000).

Die Divergenz von ξ bei $p \rightarrow 0$ ergibt in diesem Grenzwert so etwas wie einen kritischen Punkt. (De Menezes und andere. (2000) argumentierte, dass wir diesen Punkt aus technischen Gründen als "kritischen Punkt erster Ordnung" bezeichnen sollten (Fisher und Berker 1982).)

Dies erlaubte Newman und Watts (1999a), die Renormgruppen-Transformation des reellen Raums auf das Modell in der Nachbarschaft dieses Punktes anzuwenden und zu beweisen, dass die obige Skalierungsform genau im Grenzbereich von kleinem p und großem L .

Die Gleichung (9) zeigt uns, dass die durchschnittliche Entfernung von Knoten zu Knoten in einem Kleine-Welt-Graphen zwar auf den ersten Blick eine Funktion von drei Parametern, p, z und L , zu sein scheint, aber in Wirklichkeit vollständig durch eine einzige skalare Funktion einer einzigen skalaren Variablen bestimmt wird. Wenn wir die Form dieser einzigen Funktion kennen, dann wissen wir alles. Tatsächlich ist diese Aussage nur wahr, wenn $\xi \gg 1$ wenn wir die andere Längenskala in der Problemstellung, den zugrunde liegenden Gitterparameter, gefahrlos ignorieren können. Daher ist zu erwarten, dass die Skalenform nur dann auftritt, wenn p klein ist, d. h. in einer Region, in der die meisten menschlichen Kontakte lokal sind und nur ein kleiner Teil über große Entfernungen erfolgt. (Der vierte Parameter d geht ebenfalls in die Gleichung ein, aber nicht gleichberechtigt mit den anderen, da die Funktionsform F mit d variiert und somit Gleichung (9) nicht aussagt, wie ℓ mit der Dimensionalität variiert).

Sowohl die Skalierungsfunktion $F(x)$, als auch die Skalierungsvariable $x \equiv pzL^d$ haben einfache physikalische Interpretationen. Die Variable x ist die zweifache durchschnittliche Anzahl von Reduktionen auf dem Graphen für einen gegebenen Wert von p , und $F(x)$ ist der durchschnittliche Bruchteil, um den der Abstand zwischen Knoten auf dem Graphen für einen gegebenen Wert von x reduziert wird. Aus den Ergebnissen in Abb. () ist zu erkennen, dass etwa $5\frac{1}{2}$ Reduktionen erforderlich sind, um den durchschnittlichen Abstand von Knoten zu Knoten zu halbieren, und dass 56 Reduktionen erforderlich sind, um um den Faktor 10 zu reduzieren.

Im Grenzwert von großem p wird das Kleine-Welt-Modell zu einem Zufallsgraphen oder annähernd zu einem solchen. Daher erwarten wir, dass der Wert von ℓ logarithmisch mit der Größe des Systems L , skaliert werden muss, wenn p groß ist, und auch, wie die Skalierungsform zeigt, wenn L groß ist. Andererseits erwarten wir, dass l linear mit L skaliert wird, wenn p oder L klein ist. Das impliziert, dass $F(x)$ diese Grenzformen aufweist

$$F(x) = \begin{cases} 1 & \text{für } x \ll 1 \\ \frac{\log x}{x} & \text{für } x \gg 1 \end{cases}$$

Theoretisch müsste es hier eine führende Konstante vor der Form des großen x , geben, aber, wie in Kürze besprochen wird, ist diese Konstante gleich eins. Der Übergang zwischen kleinen und großen x Moden muss in der Nachbarschaft von $L = \xi$, auftreten, da ξ die einzige verfügbare Längenskala ist, die diesen Punkt definiert.

Weder die tatsächliche Verteilung der Weglängen im Kleine-Welt-Modell noch die durchschnittliche Weglänge ℓ sind bisher genau berechnet worden; exakte analytische Berechnungen haben sich für das Modell als sehr schwierig erwiesen. Einige genaue Ergebnisse wurden von Kulkarni und anderen (2000) erzielt, die z. B. zeigten, dass der Wert von ℓ schlicht mit dem Mittelwert $\langle s \rangle$ und dem mittleren Quadrat s^2 der kürzesten Entfernung s zwischen zwei Punkten auf diametral gegenüberliegenden Seiten des Graphen zusammenhängt, gemäß

$$\frac{\ell}{L} = \frac{\langle s \rangle}{L-1} - \frac{\langle s^2 \rangle}{L(L-1)} \quad (20)$$

Leider ist die Berechnung des kürzesten Abstandes zwischen gegenüberliegenden Punkten genauso schwierig wie die direkte Berechnung von ℓ , analytisch oder numerisch.

Newman und andere (2000) berechneten die Form der Skalenfunktion $F(x)$ für $d = 1$ ine-Welt-Graphen mit Hilfe einer Mittelwertfeld-ähnlichen Näherung, die für kleine oder große Werte von x genau ist, aber nicht in dem Bereich, wo $x \simeq 1$.

Ihr Ergebnis lautet wie folgt

$$F(x) = \frac{4}{\sqrt{x^2 + 4x}} \tanh^{-1} \frac{x}{\sqrt{x^2 + 4x}} \quad (21)$$

Diese Form ist auch in Abb.6 eingezeichnet (gestrichelte Linie). Da dies für große x gilt, kann durch $1/x = 0$, zerlegt werden, um zu zeigen, dass die führende Konstante in der Form $F(x)$ für große x gleich 1 ist, Gleichung (21) wie oben angegeben.

Newman und andere haben auch die vollständige Längenverteilung zwischen den Knotenpunkten im Modell im Rahmen der Molekularfeldtheorie gelöst. Diese Verteilung kann verwendet werden, um ein einfaches Modell der Krankheitsausbreitung in einer kleinen Welt zu erstellen. Wenn eine Krankheit von einer Person irgendwo auf der Welt ausgeht und sich zuerst auf alle Nachbarn dieser Person ausbreitet, dann auf alle zweiten Nachbarn und so weiter, dann ist die Anzahl der Menschen n , die nach t Zeitschritten an dieser Krankheit erkranken, nichts anderes als die Anzahl der Menschen, die vom ursprünglichen Wirt durch die Entfernung t oder weniger getrennt sind.

Newman und Watts (1999) gaben zuvor eine approximative Differentialgleichung für $n(t)$ auf einem unendlichen Kleine-Welt-Graphen an, die sie für den eindimensionalen Fall gelöst haben; Moukarzel (1999) hat sie später für den Fall des allgemeinen d gelöst. Die Mittelfeldbetrachtung verallgemeinert die Lösung für $d = 1$ auf endliche Gitterdimensionen. (Ein ähnliches Ergebnis für das mittlere Feld wurde für ein etwas anderes Krankheitsausbreitungsmodell in Kleczkowski und Grenfell (1999) erzielt). Die resultierende Form für $n(t)$ ist in Abb. 2 dargestellt und hat eindeutig die richtige allgemeine Sigmoidform für die epidemische Ausbreitung. Tatsächlich ist diese Form von n auch typisch für standardmäßige logistische Wachstumsmodelle der Krankheitsausbreitung, die hauptsächlich auf Zufallsgraphen basieren (Sattenspiel und Simon 1988, Kretschmar und Morris 1996).

3.2.3 Sonderfälle der Berechnung der Modellkennzahlen nach Watts-Strogatz

Nehmen wir ein besonderes Beispiel für ein reguläres Gitter: ein reguläres eindimensionales Gitter mit Knoten vom Grad k [j].

Definition 3.11. Ein eindimensionales reguläres Gitter mit Knoten vom Grad k (mit geradem k) ist eine Kette von N Knoten, gekennzeichnet mit $i = 1, 2 \dots N$, wobei N gerade ist, und so beschaffen, dass

$$A_{ij} = \begin{cases} 1, & \text{wenn } i \neq j \text{ \& } \left| \frac{N}{2} - \left| \frac{N}{2} - |i - j| \right| \right| \leq \frac{k}{2} \\ 0 & \text{ansonsten} \end{cases}$$

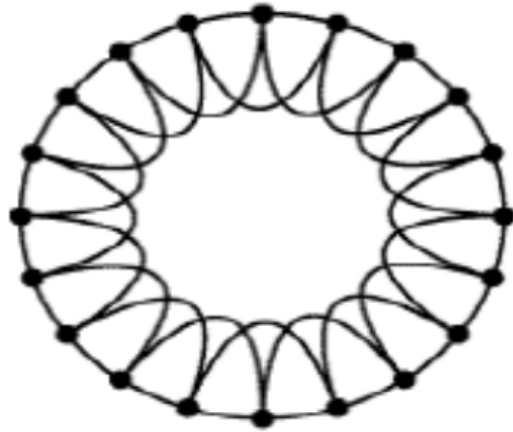


Abbildung 3: Ein regelmäßiges eindimensionales Gitter mit Knoten vom Grad $k = 4$

Ein Beispiel für ein regelmäßiges eindimensionales Gitter mit Knoten vom Grad $k = 4$ ist in Abbildung 3 dargestellt

Satz 3.12. Der Watts-Strogatz-Clusterkoeffizient C_{ws} in einem regelmäßigen eindimensionalen Gitter mit Knoten vom Grad k ist wie folgt gegeben

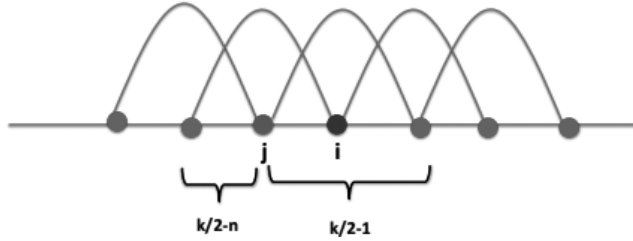
$$C_{ws} = \frac{3k - 2}{4k - 1} \quad (22)$$

Beweis. Die Anzahl der Dreiecke, die durch einen bestimmten Knoten des Gitters verlaufen, ist gegeben durch

$$\left(\frac{k}{2} - 1 \right) \frac{3}{4} k \quad (23)$$

Tatsächlich gibt es $(k/2 - 1) + (k/2 - n)$ Dreiecke, die durch den Knoten i und den Knoten j , verlaufen, die um $n \in [1, k/2]$ Schritte im Ring vom ursprünglichen Knoten i , entfernt wurden, wie in Abbildung 4 gezeigt. Wenn wir die um $n \in [1, k/2]$ nach rechts oder links vom ursprünglichen Knoten i , entfernten Knoten j aufsummieren,

Abbildung 4: Anzahl der Dreiecke, die durch zwei verbundene Knoten i und j im Abstand $n = \text{mod}(i - j, N/2)$ auf dem Ring verlaufen. Die Anzahl der geschlossenen Dreiecke ist gleich $(k/2 - 1) + (k/2 - n)$ für ein regelmäßiges eindimensionales Gitter mit einer Anzahl von Knoten vom Grad k . Die Abbildung zeigt die Fläche eines eindimensionalen Gitters mit Knoten vom Grad $k = 4$. Für unser spezielles Beispiel ist $n = 1$.



erhalten wir die Gesamtzahl der Dreiecke, die durch den Knoten i verlaufen, multipliziert mit zwei, da jedes Dreieck zweimal gezählt wird.

Daher ist die Anzahl der Dreiecke, die den Knoten durchlaufen, wie folgt gegeben

$$\sum_{k=1}^{n/2} k - 1 - n = (k - 1) \frac{k}{2} - \frac{k}{4} \left(\frac{k}{2} + 1 \right) \quad (24)$$

Daher ist der Clusterkoeffizient wie folgt definiert

$$C_i = \frac{\frac{3}{4}k \left(\frac{k}{2} - 1 \right)}{\frac{1}{2}k (k - 1)} = \frac{3k - 2}{4k - 1} \quad (25)$$

für jedes i . Daraus folgt, dass

$$C_{ws} = \frac{3k - 2}{4k - 1} \quad (26)$$

□

Daher ist der Clusterkoeffizient eine von der Netzwerkgröße unabhängige Konstante und somit ist C_{ws} im Grenzwert $N \rightarrow \infty$ endlich.

Satz 3.13. *Der Durchmesser eines regelmäßigen eindimensionalen Gitters mit Knoten vom Grad k hat einen Durchmesser, der im Grenzwert großer Netze $N \gg$ wie folgt ausfällt*

$$D \simeq \frac{N}{k} \quad (27)$$

daher besitzt das Netzwerk nicht die Eigenschaft der kleinen Welt.

Beweis. Ein Paar von Knoten, die voneinander entfernt sind, sind Knoten mit einem Abstand von $N/2$. Da die Knoten im Abstand $k/2$ direkt miteinander verbunden sind, können wir den Durchmesser des Netzwerks berechnen als

$$D \simeq \frac{N/2}{k/2} = \frac{N}{k} \quad (28)$$

ähnlicher Weise können wir die folgende Aussage für den durchschnittlichen Abstand eines regelmäßigen eindimensionalen Gitters mit Knoten vom Grad k zeigen. \square

Satz 3.14. *Der durchschnittliche Abstand eines regelmäßigen eindimensionalen Gitters mit Knoten vom Grad k steigt linear mit der Größe des Netzwerks, d. h.*

$$\ell \simeq \frac{N}{k} \quad (29)$$

3.3 Song and Wang Modell

3.3.1 Die allgemeine Beschreibung

Wie aus dem vorherigen Abschnitt ersichtlich, implementiert das Watts-Strogatz-Modell eine intuitive Darstellung von realen Netzwerken, deren Topologie weder komplett regulär noch komplett zufällig ist.

Ausgehend von einer regulären, lokal verbundenen Struktur, führt der Rekonnectionsprozess entfernte Kanten ein, bis er in einem vollständig randomisierten Zustand endet. Das resultierende System ist also etwas zwischen zwei gut untersuchten physikalischen Systemen: der Kristallstruktur und der durch Brownsche (Wiener) Prozesse erzeugten Unordnung.

Selbst für eine kleine Anzahl von wiederverbundenen Kontakten bleibt die Wahrscheinlichkeit, dass zwei Nachbarn miteinander verbunden sind (ein häufiger Fall in sozialen Netzwerken), praktisch unverändert, während kurze Pfade zwischen einzelnen Knoten fast unmittelbar auftauchen, was erklärt, wie soziale Netzwerke beides sein können: stark geclustert und so, dass sie es ermöglichen, dass ein Knoten durch eine kleine Anzahl von Knoten zu einem anderen Knoten gelangt.

Wie bereits erwähnt, ist ein wichtiges Merkmal des Modells, dass die durchschnittliche Anzahl der Knoten, mit denen ein Knoten verbunden ist, ein konstanter Wert ist, was ein Kontrollparameter erster Ordnung für mehrere dynamische Systeme ist, die z. B. auf Random Walks oder epidemischer Ausbreitung basieren. Die Annahme, dass der Grenzprozess der Rekonfiguration (Rekonnexion) einen Zufallsgraphen reproduziert, d. h. das Erdős-Rényi-Zufallsgraphenmodell $G(N, p)$, wobei N die Anzahl der Knoten und p die Wahrscheinlichkeit ist, dass zwei beliebige Knoten miteinander verbunden sind, wurde in dieser Studie aufgestellt.

Wie aus der Modelldefinition hervorgeht, ist das maximal randomisierte Watts-Strogatz-Modell jedoch nicht wirklich äquivalent zum Erdős-Rényi-Modell. Außerdem wird bei der Erwähnung des Grenzwertes der Unordnung (Zufälligkeit) nicht explizit auf das Erdős-Rényi-Modell als Grenzwert verwiesen. Außerdem hat der

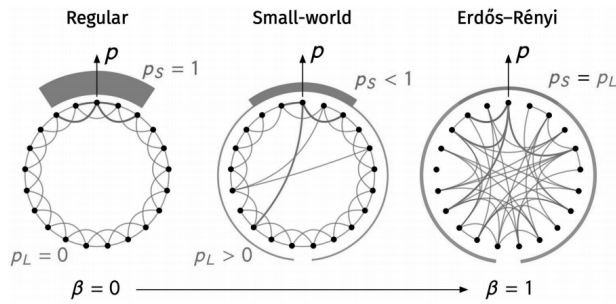


Abbildung 5: Schematische Darstellung des alternativen Kleine-Welt-Modells

Grenzzustand des Watts-Strogatz-Modells eine mehrdeutige Formulierung und kann daher schnell fehlinterpretiert werden.

Die zweite Schwierigkeit bei der praktischen Anwendung des Watts-Strogatz-Modells ist seine ursprüngliche Rekonnexionsprozedur, bei der jeder Knoten jede seiner Kanten mit der Wahrscheinlichkeit p_r mit seinen ganz rechts liegenden Nachbarn verbindet. Dieser Vorgang wird oft als einfaches Neuverbinden einer beliebigen Kante interpretiert.

Während Unterschiede oder geringfügige Variationen im Modell bei einigen dynamischen Systemen nicht signifikant sein mögen, können sie bei anderen deutliche Abweichungen von den erwarteten Ergebnissen verursachen, was zu Verwirrung oder fehlerhaften Interpretationen führen kann.

Ein Modell, das die oben skizzierten Probleme behandelt, wurde von Song und Wang [k] vorgestellt. Im Rahmen ihrer Studie zeigten sie, dass das Abtasten von Kanten aus der auf der Distanz basierenden Verbindungswahrscheinlichkeit die Schätzung von z. B. Gradverteilung und Clusterkoeffizient erleichtert. Betrachten wir dieses modifizierte Modell als eine der Möglichkeiten zur Modellierung sozialer Netzwerke. Ein weiterer wesentlicher Vorteil des betrachteten Modells ist, dass es zur analytischen Erklärung des Kleine-Welt-Phänomens herangezogen werden kann, indem der Clusterkoeffizient beim Erreichen der oberen Grenze der Nachrichtenübermittlungszeit verglichen wird.

3.3.2 Definition des Modells

Die Definition des Modells, der Einfachheit halber, bauen wir auf den Vergleich dieses Modells mit der klassischen Interpretation des Watts-Strogatz-Modells.

Im ursprünglichen Modell sind N Knoten auf dem Ring äquidistant und dann lokal verbunden, d. h. sie sind mit Knoten in ihrer Nachbarschaft mit maximalem Gitterabstand $d \leq k/2$ verbunden, wobei k eine gerade positive ganze Zahl ist.

In diesem Zustand hat jeder Knoten den Grad k , wobei sich "Grad" auf die Anzahl der Nachbarn des Knotens bezieht. Für den Prozess der Wiederverbindung (Reconnexion) verbindet jeder Knoten seine Verbindungen mit $k/2$ rechten Nachbarn zu jedem anderen Knoten im Netzwerk mit der Wahrscheinlichkeit p_r . Es ist leicht zu sehen, dass für die Randomisierungsgrenze $p = 1/r$ jeder Knoten den minimalen Grad $k/2$ hat.

Außerdem wurde die ursprüngliche Kante, die mit dem Knoten u verbunden ist, neu

verbunden und kann nur existieren, wenn sie durch ein anderes Rekonnectionsereignis auf der Basis des entsprechenden ganz rechten Nachbarn reproduziert wird. Das heißt, wenn $p_r = 1$ ist, existiert die ursprüngliche Kante mit der Wahrscheinlichkeit $1/(N - 1)$. Diese beiden Eigenschaften führen zu konzeptionellen Abweichungen vom Erdős -Rényi -Modell, in dem jede Kante mit der Wahrscheinlichkeit $P_{ER} = k/(N - 1)$ existiert und Knoten den Grad $< k/2$.

In einer Variante des modifizierten Modells von Song und Wang haben die Kanten eine intrinsische Existenzwahrscheinlichkeit, die bei nahen (S) und fernen (L) Kontakten variiert. Ein potenzieller Kontakt zwischen Knoten (i, j) wird als kurzreichweitig betrachtet, wenn ihr Abstand unter periodischen Randbedingungen $d(i, j) < k/2$ ist; und mit der Wahrscheinlichkeit p_S vorliegt.

Er gilt als weitreichend, wenn $d(i, j) > k/2$ ist und mit der Wahrscheinlichkeit p_L gegeben ist. Der Abstand wird berechnet als $d(i, j) = \min(\|j - i\|, N - \|j - i\|)$. Kurz gesagt, zwei Knoten mit Gitterabstand d sind verbunden mit der Wahrscheinlichkeit

$$p_d = \begin{cases} p_S, & \text{wenn } d \leq k/2 \\ p_L, & \text{andernfalls} \end{cases}$$

Wenn also $p_S = 1$ und $p_L = 0$ ist, erzeugt das Modell eine Struktur, die dem Ausgangspunkt des ursprünglichen Modells, dem eindimensionalen k -nächsten-Nachbarn-Gitter, entspricht. Wenn andererseits $p_S = p_L \equiv p$, dann existiert jede Kante mit der Wahrscheinlichkeit p und somit reproduziert das Modell einen Zufallsgraphen $G(N, p)$. Wir können den durchschnittlichen Grad bestimmen, indem wir feststellen, dass er aus dem Nahgrad $\langle k_S \rangle$ und dem Ferngrad $\langle k_L \rangle$ besteht. Jeder Knoten hat k potenzielle nahe Nachbarn und $N - 1 - k$ potenzielle ferne Nachbarn. Daher ist der erwartete Grad

$$p_S k + p_L (N - 1 - k) = \langle k_S \rangle + \langle k_L \rangle \quad (30)$$

Um den durchschnittlichen Grad konstant zu halten, führen wir einen Kontrollparameter β ein, der das Verhältnis der Verbindungswahrscheinlichkeit im Nah- und Fernmodus so steuert, dass $p_L = \beta p_S$. Man beachte, dass, wenn $\beta = 0$ ist, $p_L = 0$ ist, wodurch aus Gleichung (30) folgt, dass $p_S = 1$ ist, und wenn $\beta = 1$ ist, ist $p_L = p_S \equiv p$.

Damit der mittlere Grad konstant ist, ergibt die Gleichung (30) Wahrscheinlichkeiten basierend auf dem Abstand

$$p_S(\beta) = \frac{1}{1 + \beta (N - 1 - k) / k} p_L(\beta) = \frac{\beta}{1 + \beta (N - 1 - k) / k} = \beta p_S(\beta) \quad (31)$$

Der Grad des Nahbereichsknotens k_S entspricht der Binomialverteilung $\beta(k, p_S)$, und der Grad des Fernbereichsknotens k_L entspricht der Binomialverteilung $\beta(N - 1 - k, p_L)$, wobei $\beta(n, p)$ eine Wahrscheinlichkeitsmassenfunktion $f_k(n, p) = \binom{n}{k} (1 - p)^{n-k} p^k$ besitzt.

Eine schematische Erläuterung des Modells ist in Abb. () dargestellt. Ein einfacher Algorithmus zur Netzwerkgenerierung wird wie folgt beschrieben. Jeder Knoten $0 \leq u \leq N - 1$ ist mit jedem seiner k^2 extrem rechten nächsten Nachbarn mit der Wahrscheinlichkeit p_S verbunden. Anschließend werden die fernen Kanten m_L gebildet, wobei $m_L \beta(N(N - 1 - k)/2, p_L)$ entspricht. Für jede ferne Kante wird ein einheitlicher zufälliger Quellknoten u aus dem Bereich $[0, N - 1]$ gewählt.

Dieser Knoten ist dann mit seinem fernen Nachbarn $v = (u + k/2 + z) \bmod N$ verbunden, wobei die ganze Zahl z gleichmäßig zufällig aus dem Intervall $[1, N - k - 1]$ gewählt wird. Wenn eine bereits vorhandene Kante gewählt wurde, muss der Vorgang wiederholt werden. Dieser Algorithmus hat die Komplexität $O(Nk + \langle m_L \rangle)$ für dünnbesetzte Netzwerke.

Wie im ursprünglichen Modell beginnen wir mit N Knoten in gleichem Abstand zueinander auf einem Ring. Anstatt sich jedoch neu zu verbinden, verbindet sich jedes Paar von Knoten mit der Wahrscheinlichkeit p_d basierend auf dem Abstand, wobei d ihr minimaler Abstand im Ring ist. Innerhalb der Entfernung $d \leq k/2$ verbinden sich die Knoten mit der Wahrscheinlichkeit p_S über kurze Strecken. Bei größeren Abständen verbinden sich die Knoten mit der Wahrscheinlichkeit $p_L = \beta p_S$. Mit zunehmendem Verteilungsparameter $0 \leq \beta \leq 1$ wird die Verbindungswahrscheinlichkeit vom Nahbereichsmodus zum Fernbereichsmodus umverteilt, während der durchschnittliche Grad k konstant bleibt. Wenn also $\beta = 0$ ist, ist die Verbindungswahrscheinlichkeit bei kurzen Entfernungen gleich eins und bei langen Entfernungen gleich null, wodurch ein Gitter mit k - nächsten Nachbarn entsteht. Mit zunehmendem β werden die langreichweitigen "kurzen Wege" wahrscheinlicher, bis bei $\beta = 1$ beide Verbindungswahrscheinlichkeiten gleich werden und damit das Modell dem Erdős-Rényi-Modell entspricht.

Man beachte, dass für kleine β die Wahrscheinlichkeit einer kurzen Verbindung p_S ungefähr gleich der Wahrscheinlichkeit sein sollte, dass die Kante im ursprünglichen Modell nicht umverbunden wurde $1 - p_r$. Um sicherzustellen, dass $p_r = 1$ für $\beta = 1$ ist, setzen wir $p_r = [1 - p_S]/[1 - p_{ER}]$.

Um die beiden dynamischen Modelle zu vergleichen, berechnen wir die diskrete Random-Walk-Zeit wie folgt: Bei jedem diskreten Zeitschritt wählt der Walker, der sich am Knoten u befindet, mit der gleichen Wahrscheinlichkeit einen Übergang zu einem seiner Nachbarn, was sich unendlich wiederholt. Random Walks werden häufig zur Modellierung von Ausbreitungs- und Suchprozessen in der Physik, Biologie und Informatik verwendet.

In diesem Zusammenhang ist die durchschnittliche Erstdurchlaufzeit τ_{vu} - die erwartete Anzahl von Schritten, die ein zufälliger Walker bis zum Knoten v zurücklegen muss, wenn er vom Knoten u aus gestartet ist, was also als Obergrenze für einen beliebigen Suchprozess interpretiert werden kann.

Im Gegensatz dazu ist die Länge des kürzesten Weges zwischen zwei Knoten der kürzeste Weg zwischen zwei Knoten für den optimalen Suchprozess [1]. Basierend auf der durchschnittlichen Erstdurchlaufzeit dient die paargemittelte Erstdurchlaufzeit $\tau = (N(N - 1))^{-1} \sum_{\nu=1}^N \sum_{u \neq \nu}^N \tau_{u\nu}$ als grobe Schätzung, wie schnell ein zufälliger Suchprozess zwischen zwei beliebigen Knoten in einem bestimmten Netzwerk durchgeführt werden kann.

Wir berechnen eine paarweise durchschnittliche Erstdurchlaufzeit für Kleine-Welt-Netzwerke mit $N = 512$ Knoten und einem durchschnittlichen Grad $k = 10$. Die Steuerungsparameter sowohl des alternativen als auch des ursprünglichen Modells wurden entsprechend variiert β und $p_r = (1 - p_S(\beta))/(1 - p_{ER})$. Für jeden Wert von β haben wir den Durchschnitt, gemittelt über Paare von Erstdurchlaufzeiten, anhand der größten Komponente der 10.000 unabhängigen Netzwerkimplementierungen ermittelt.

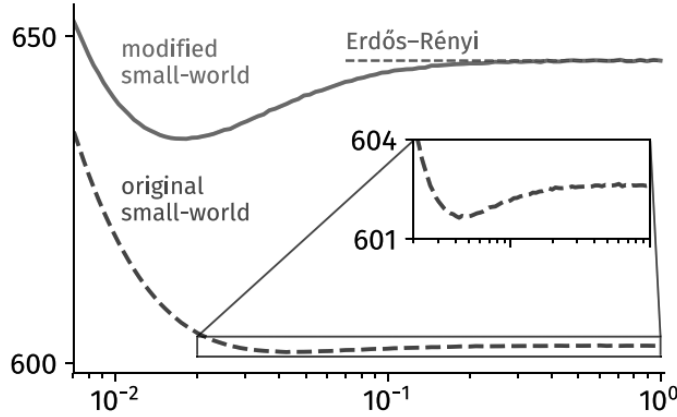


Abbildung 6: Paargemittelte Erstdurchlaufzeit (PAFPT pair-averaged first passage time) eines zeitdiskreten Random-Walk-Prozesses, ist ein Beispiel für ein beobachtetes Netzwerk, das sich von dem entsprechenden Ergebnis des Erdős-Rényi-Modells in der $\beta = 1$ Grenze unterscheidet.

Die in Abb. 7 gezeigten Ergebnisse bedeuten, dass der Unterschied zwischen den beiden Modellen in der Tat signifikant sein kann und relative Differenzwerte von bis zu $\approx 7\%$ im randomisierten Grenzbereich erreicht. Dieser Unterschied wird durch die Tatsache verursacht, dass im ursprünglichen Modell jeder Knoten mindestens $k/2$ Grad hat, während es im modifizierten Modell Knoten mit einem niedrigeren Grad geben kann. Das Auftauchen dieser Unterscheidung ist ein Beweis für die Relevanz des alternativen Modells für die Untersuchung des Einflusses der Kleine-Welt-Topologie auf die Ergebnisse der dynamischen Systeme.

3.3.3 Netzwerkeigenschaften des alternativen Modells

Wir beginnen unsere Untersuchung der Eigenschaften des Modells mit der Graddispersion, die wichtig ist, um die quantitative Heterogenität der Knoten im Netzwerk anhand ihrer Konnektivität zu bewerten. Auch die Gradvarianz spielt eine wichtige Rolle bei der Schätzung der mittleren Zeit von Random Walks $[m, n]$.

Da im alternativen Kleine-Welt-Modell der Grad eines Knotens als Überlagerung von Nah- und Ferngrad gegeben ist, kann die Gradvarianz einfach berechnet werden als

$$Var[k] = Var[k_s] + Var[k_L] = kp_S(1 - p_S) + (N - 1 - k)p_L(1 - p_L) \quad (32)$$

Mit zunehmendem β nehmen sowohl die Nah- als auch die Fernvarianz zu, so dass die Gradvarianz eine zunehmende Funktion von β ist, wie in Abb. 9 dargestellt [o]. Die komplette Gradverteilung kann berechnet werden, indem man beachtet, dass jeder Gradknoten $k_i = k_{S,i} + k_{L,i}$ so beschaffen ist, dass seine Verteilung durch Faltung gegeben ist.

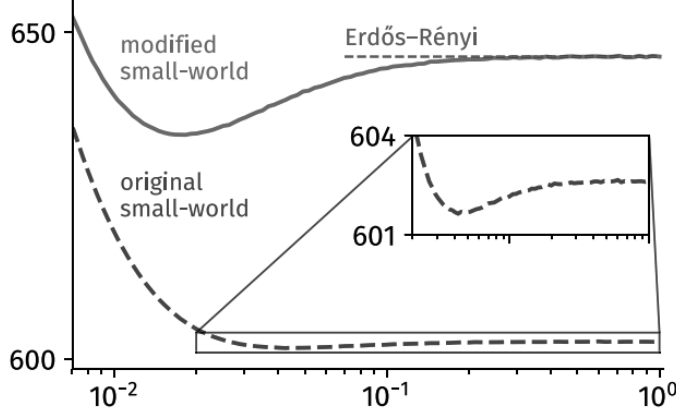


Abbildung 7: Paargemittelte Erstdurchlaufzeit (PAFPT pair-averaged first passage time) eines zeitdiskreten Random-Walk-Prozesses, ist ein Beispiel für ein beobachtetes Netzwerk, das sich von dem entsprechenden Ergebnis des Erdős-Rényi-Modells in der $\beta = 1$ Grenze unterscheidet.

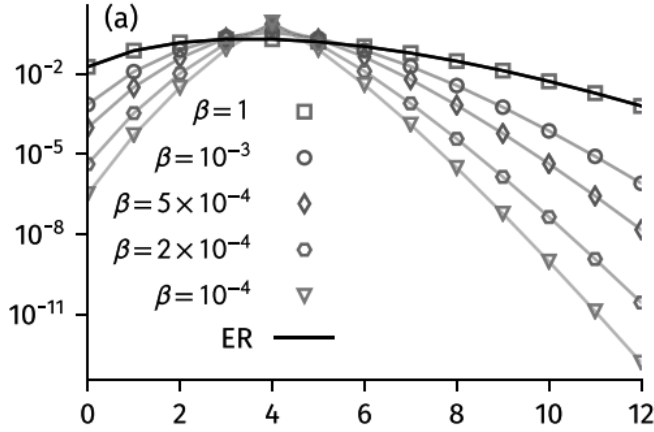
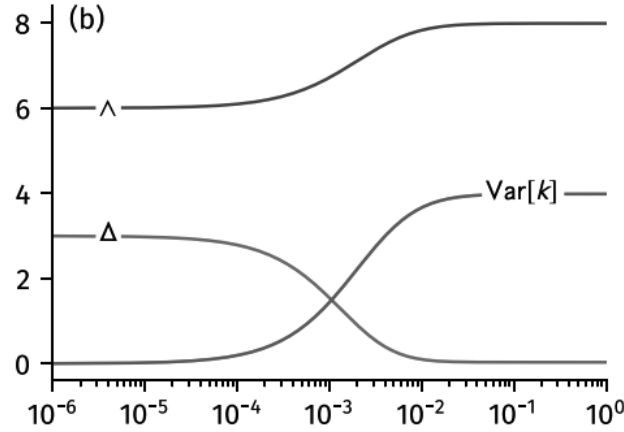


Abbildung 8: Analytische Ergebnisse für die Gradverteilung von Gleichung, die erwartete Anzahl von zwei Sternen pro Knoten \wedge und die erwartete Anzahl von Dreiecken pro Knoten Δ (Gleichung (Nr. 4)) sowie die Knotengradvarianz wie im Haupttext angegeben. Während die Gradstreuung und die Anzahl der Zwei-Stern-Knoten mit steigendem Fernumverteilungsparameter β zunimmt, sinkt die Anzahl der Dreiecke.



$$\begin{aligned}
 P_k' &= \sum_{k_S=0}^{\infty} \sum_{k_L=0}^{\infty} f_{k_S}(k, p_S) f_{k_L}(N-1-k, p_L) \delta_{k', (k_S+k_L)} \\
 &= \sum_{k_S=0}^{\min(k', k)} \binom{k}{k_S} \binom{N-1-k}{k'-k_S} (1-p_S)^{k-k_S} \\
 &\quad \times P_S^{k_S} (1-p_L)^{N-1-k'+k_S} p_L^{k'-k_S}
 \end{aligned} \tag{33}$$

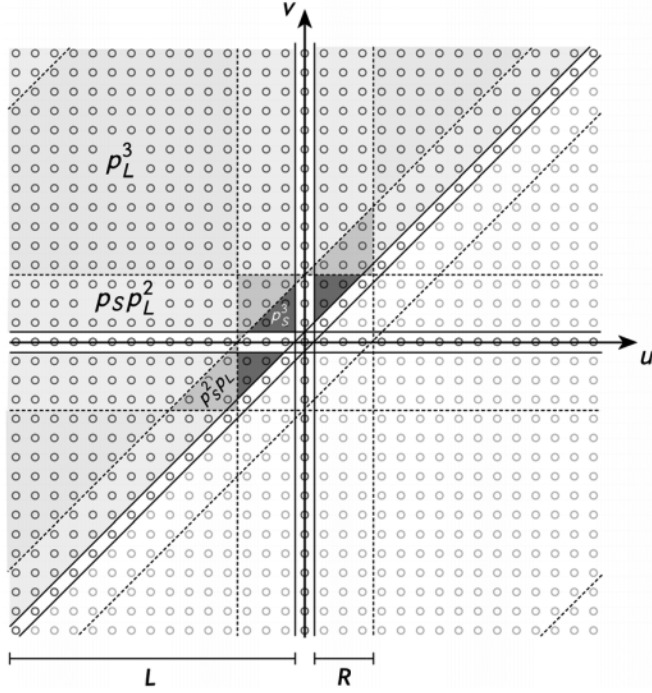


Abbildung 9: Abschätzung der Summationsbereiche zur Ermittlung der erwarteten Anzahl von Dreiecken pro Knoten Δ für eine ungerade Anzahl von Knoten N nach Gleichung (4). Die Summierung wurde zu $\sum_{u=-\frac{N-1}{2}+1}^{(N-1)/2} \sum_{v=u+1}^{(N-1)/2}$ geändert, so dass u und v gleich ihrem Gitterabstand zum zentralen Knoten 0 sind.

Es ist zu beachten, dass wir in der obigen Schlussfolgerung das Kronecker-Delta $\delta_{ij} = 0$ verwendet haben, wenn $i \neq j$, ansonsten $\delta_{ij} = 1$. Sowohl die Ergebnisse der Gradvarianz als auch die Ergebnisse der Gradverteilung unterstreichen die Einfachheit des alternativen Modells, das eine einfache analytische Auswertung im Vergleich zu den komplizierteren Berechnungen im ursprünglichen, auf Rekonexion basierenden Modell ermöglicht.

Obwohl es viele solcher Definitionen gibt, spiegelt der Clusterkoeffizient typischerweise die Wahrscheinlichkeit eines triadischen Abschlusses wider: Gegeben sei eine Struktur, in der Knoten i mit den Knoten v und u , verbunden ist, der Clusterkoeffizient ist die Wahrscheinlichkeit, dass u und v ebenfalls verbunden sind. Unter Verwendung der $(N \times N)$ -großen Netzwerk-Adjazenzmatrix $A_{ij} = 1$, wenn die Knoten i und j verbunden sind, und andernfalls $A_{ij} = 0$, definieren wir den globalen Clusterkoeffizienten als die bedingte Wahrscheinlichkeit

$$C = P[A_{iu}A_{uv}A_{vi} = 1 | A_{iu}A_{iv} = 1] = \frac{\langle A_{iu}A_{uv}A_{vi} \rangle}{\langle A_{iu}A_{iv} \rangle} \equiv \frac{\Delta}{\Lambda} \quad (34)$$

Wir leiten das Endergebnis mit einem eher geometrischen Ansatz wie folgt ab. Die Wahrscheinlichkeit $\Lambda = \langle A_{iu}A_{iv} \rangle$ ist die erwartete Anzahl von zwei Sternen pro Knoten (eine Struktur, in der Knoten i sowohl mit Knoten u als auch mit Knoten v verbunden ist). Um diesen Wert abzuschätzen, ist zu beachten, dass ein Knoten vom Grad k_v ein Teil $(1/2)k_v(k_v - 1)$ von zwei Sternen ist. Somit ist ihm $\Lambda = (\frac{1}{2})Var[k] + k(k - 1)$ zugeordnet. Er folgt damit qualitativ dem in Abb. 8 gezeigten Varianzgradverhalten.

Um die erwartete Anzahl von Dreiecken pro Knoten zu finden, nehmen wir an, dass jeder Knoten statistisch gleichwertig ist. Ohne Verlust der Allgemeingültigkeit berechnen wir daher die Anzahl der Dreiecke pro Knoten $i = 1$ als die Summe aller

möglichen verbleibenden Knotenpaare unter Berücksichtigung ihrer Verbindungswahrscheinlichkeit aufgrund der Entfernung als

$$\Delta = \sum_{u=2}^{N-1} \sum_{v=u+1}^N P_{d(u,1)} P_{d(v,1)} P_{d(u,v)} = F_{p_s}^3 + G_{p_s}^2 p_L^2 + H p_s P_L^2 + I_{p_L}^3 \quad (35)$$

Dabei sind F , G , H und I die in Abb. 9 hervorgehobenen Summenbereiche, wobei jeweils drei (grau), zwei (rosa), ein (grün) und kein (orange) Knotenpaar in geringem Abstand zueinander liegen.

Betrachtet man den Fall einer ungeraden Zahl N , so ist es möglich, die Summanden so zu verändern, dass sie vom Gitterabstand $-N/2$ zum Abstand $N/2$ um den zentralen Knoten bei $d = 0$ verlaufen, so dass sich das Auffinden der entsprechenden Regionen angesichts der Bedingungen für Verbindungen in kurzen Abständen auf eine geometrische Übung reduziert. Durch Bestimmung der Längen $L = (N - 1)/2$ und $R = k/2$, wie in Abb. 4 ist es zunächst möglich, die Nutzeneinheit des Kurzlängenbereichs in Form des Dreiecks $T = \frac{R^2 - R}{2} + R$ (in Abb. 4 rosa markiert) zu finden. Folglich sind die Summenbereiche gegeben als

$$\begin{aligned} F &= 3(T - R) = \frac{3k}{8}(k - 2) \\ G &= 3T = \frac{3k}{8}(k + 2) \\ H &= 2((L - R)R - T) + T + 2(L - R)R + 2((L - 1)R - T) = \\ &\quad (k/8)(12N - 26 - 11k) \\ I &= (L - R)^2 - 2((L - 1)R - T) - (L - R) + (L - R)^2 - T \\ &= (1/8)[5k^2 - k(12N - 26) + 4(N^2) - 3N + 2] \end{aligned} \quad (36)$$

Die erwartete Anzahl der Dreiecke von Gl. (35) nimmt erwartungsgemäß mit zunehmendem β ab, wie in Abb. 8 gezeigt. Angesichts von Gleichung (31) wird der genaue Wert des Clusterkoeffizienten wie folgt bestimmt

$$C(\beta) = p_s^3 \times \frac{F + G\beta + H\beta^2 + I\beta^3}{(1/2)\text{Var}[k] + k(k - 1)} \quad (37)$$

Im Rahmen der entsprechenden Grenzwerte gilt folgendes

$$C(\beta = 0) = \frac{3(k - 2)}{4(k - 1)} \quad (38)$$

$$C(\beta = 1) = \frac{\sum_{u=2}^{N-1} \sum_{v=u+1}^N p^3}{\sum_{u=2}^{N-1} \sum_{v=u+1}^N p^2} = p \quad (39)$$

was das erwartete Ergebnis sowohl für das k -Nächste-Nachbarn-Gitter als auch für den Erdős-Rényi-Graphen ist. Angesichts der Gleichungen (31) und (37) und unter

Berücksichtigung der Tatsache, dass $Var[k](\beta \rightarrow 0) = 0$ ist, können wir für den Grenzwert einer kleinen weiträumigen Umverteilung feststellen, dass

$$\frac{C(\beta \ll 1)}{C(0)} \approx p_S^3 = 1 - 3\beta \frac{N - k - 1}{k} + \mathcal{O}(\beta^2) \quad (40)$$

was für die Quantifizierung des Kleine-Welt-Effekts in den nachfolgenden Berechnungen wichtig sein wird.

3.3.4 Der Kleine-Welt-Effekt

Im ursprünglichen Modell wird der Kleine-Welt-Effekt durch die Ermittlung des Clusterkoeffizienten $\langle p \rangle$ und der durchschnittlichen Länge des kürzesten Pfades im Netzwerk dargestellt. Während zufällige Netzwerke kürzeste Pfadlängen haben, weisen sie eine geringe Clusterung auf, dagegen sind reguläre Netzwerke stark geclustert und die Knoten sind im Durchschnitt recht weit voneinander entfernt.

Indem nur eine kleine Anzahl von Kanten des geordneten Netzwerks neu verbunden wurde, konnte gezeigt werden, dass sofort kürzere Pfade auftreten, während eine hohe Clusterbildung bestehen bleibt, was den Kleine-Welt-Effekt erklärt. Ferner wird argumentiert, dass die algorithmische Suche, die lokale Informationen benötigt, notwendig ist, um diese kurzen Wege zu identifizieren.

In Situationen, in denen die Suche jedoch weniger zielgerichtet ist und einer eher diffusen Dynamik folgt, wie z. B. bei der epidemischen Ausbreitung in der Luft oder der Synchronisation in Oszillatoren, verliert die Rolle der durchschnittlichen kürzesten Weglänge an Bedeutung.

Vielmehr sind Random-Walk-Relaxation und Laufzeiten wichtige beobachtbare Parameter, die diese Dynamik charakterisieren, insbesondere für die Vorhersage des Zeitpunkts des Auftretens von Krankheiten oder der Wahrscheinlichkeit einer globalen Synchronisation. Daher verwenden wir im Folgenden einen Ansatz, der sich auf Random Walks konzentriert.

Eines der Ziele des ursprünglichen Watts-Strogatz-Modells war es, das Milgram-Experiment der “kleinen Welt” zu erklären, bei dem die Teilnehmer Briefe an Fremde schicken mussten, die sie an eine ihnen bekannte Person weiterleiteten und diese anweisen sollten, den Brief weiterzuleiten.

Veranschaulichen wir den “Kleine-Welt”-Effekt, indem wir zeigen, dass die obere Grenze für die Zustellzeit dieser Nachrichten viel schneller abnimmt als der Clusterkoeffizient, wenn die Wahrscheinlichkeit für weitreichende Kanten steigt. Da diese Obergrenze für die zufällige Suche auch die durchschnittliche Länge des kürzesten Weges einschränkt, was der Ankunftszeit bei der maximal informierten Suche entspricht, verallgemeinert das folgende Ergebnis den “Kleine-Welt”-Effekt auf die zufällige Dynamik.

Unter Berücksichtigung völlig uninformatierter Personen wird der Versandprozess als ein zufälliger Wanderprozess modelliert, bei dem zufällige Schritte den an die Empfänger zu versendenden Briefen entsprechen.

Zu jedem ganzzahligen Zeitschritt t befindet sich der Brief am Knoten u . Dann wird einer der Nachbarn u, v , zufällig als nächster Empfänger der Nachricht ausgewählt. Beim nächsten Zeitschritt $t + 1$ befindet sich der Brief am Knoten v . Dieser Prozess wiederholt sich unendlich und unterliegt der Hauptgleichung $\phi_\nu(t) = \sum_{u=1}^N \left(\frac{A_{\nu u}}{k_u} \right) \phi_u(t-1)$, $\phi_\nu(t)$ ist die Wahrscheinlichkeit, dass der Brief zum Zeitpunkt t am Knoten v ankommt und $W_{vu} = \frac{A_{vu}}{k_u}$ die Wahrscheinlichkeit, dass der Brief vom Knoten u zum Knoten v gesendet wird.

Anstatt Adjazenzmatrizen zu erzeugen und die Ergebnisse der entsprechenden Übergangsmatrizen zu mitteln, werden wir eine durchschnittliche Adjazenzmatrix berechnen, bei der jede Kante im Netzwerk durch die Wahrscheinlichkeit ersetzt wird, dass diese Kante so existiert, dass $W_{vu}^{avg} = \frac{\langle A_{vu} \rangle}{k} = p_{d(v,u)}/k$. Es kann nachgewiesen werden, dass die Zeitskala, mit der sich die Gleichgewichtsverteilung in diesem gemittelten Netzwerk nähert, durch die Übergangseigenwertlücke W_{vu}^{avg} als $t_{mix}^{-1} = 1 - \omega_1$, gegeben ist, wobei $\omega_0 = 1$ der größte Eigenwert und ω_1 der zweitgrößte Eigenwert ist. Die mittlere Übergangsmatrix der Umgebung W_{vu}^{avg} zirkuliert auf der Basis des Vektors

$$w = k^{-1} \left(\underbrace{0, p_S, \dots, p_S}_{k/2}, \underbrace{p_L, \dots, p_L}_{N-1-k}, \underbrace{p_S, \dots, p_S}_{k/2} \right) \quad (41)$$

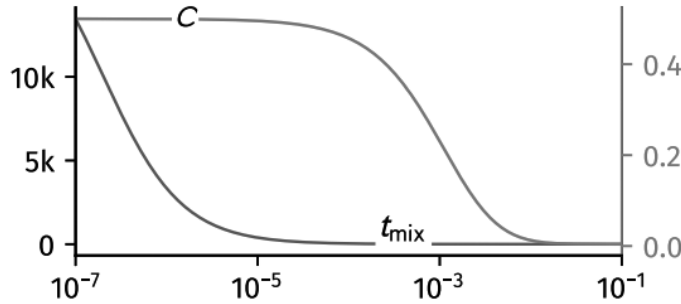


Abbildung 10: Der ‘Kleinen-Wel’-Effekt, veranschaulicht durch analytische Berechnungen.

In diesem Fall ist der j -te Eigenwert W_{vu}^{avg} gegeben als $\omega_j = \sum_{v=1}^N w_v \exp(i\pi v/N)$, so dass der zweitgrößte Eigenwert leicht berechnet werden kann als $\omega_1 = p_S \Gamma / k - p_L(1 + \Gamma) / k$, wobei $\Gamma = 2 \sum_{j=1}^{k/2} \cos(2\pi j/N) = k - (\pi/N)^2 k(k/2 + 1)(k + 1)/3 + \mathcal{O}(N^{-4})$, was die Überlagerungszeit aufzeigt

$$t_{mix}(\beta) = \left[1 - \frac{\Gamma - \beta(1 + \Gamma)}{k + \beta(N - 1 - k)} \right]^{-1} \quad (42)$$

In Abb. 5 wird gezeigt, wie der Clusterkoeffizient und die Überlagerungszeit abnehmen, wenn der Fernumverteilungsparameter β zunimmt. Im Grenzfalle finden wir die erwarteten Ergebnisse aus k -regulären Netzen und der gemittelten Approximation

des Erdős-Rényi-Graphen.

$$t_{mix}(\beta = 0) = \left[1 - \frac{\Gamma}{k}\right]^{-1} \xrightarrow{N \gg k/2} \frac{N^2}{\pi^2} \frac{3}{(k/2 + 1)(k + 1)} \quad (43)$$

$$t_{mix}(\beta = 1) = \left[1 - \frac{1}{N-1}\right]^{-1} = 1 - \frac{1}{N} \quad (44)$$

Das bedeutet, dass bei kleinen weiträumigen Umverteilungen die relative Überlagerungszeit abnimmt mit

$$\frac{t_{mix}(\beta)}{t_{mix}(0)} = 1 - \beta \left[\frac{3N^3}{\pi^2(k/2 + 1)(k + 1)k} - \frac{N}{k} + \frac{1}{k} \right] + \mathcal{O}(\beta^2) \quad (45)$$

Wenn man die Gleichungen (41) und (45) vergleicht [q], kann man leicht erkennen, dass für kleine β die Reduktionsrate der Überlagerungszeit die Ordnung N^3 besitzt, während die Reduktionsrate des Clusterkoeffizienten die Ordnung N besitzt, was ein Unterschied von zwei Größenordnungen ist. Dies zeigt, dass selbst für einen kleinen Wert der Fernverbindungswahrscheinlichkeit die Zustellzeit von zufällig übertragenen Nachrichten schnell abnimmt, während die Clusterbildung beibehalten wird. Da die optimale Suchstrategie den kürzesten Weg zwischen zwei Knoten vorgibt, und der ursprüngliche “Kleine Welt”- Effekt für diese kürzesten Wege nachgewiesen wurde, verallgemeinert dieses Ergebnis den “Kleine Welt”- Effekt auf zufällige Suchstrategien.

4 Mathematische Modellierung

4.1 Modellierung

Im praktischen Teil der Arbeit werden wir eine Software erstellen, die es erlaubt, die betrachteten Typen von Zufallsgraphen bei verschiedenen Werten der Anfangsparameter zu visualisieren, sowie einige Messwerte zu vergleichen: durchschnittliche Pfadlänge, Clustering, Transitivität und andere. Für die Erstellung verwenden wir UML 2.0 und für die Software-Implementierung Python 3.9 mit den notwendigen Visualisierungs- und Graphenbibliotheken. Die Sprache UML 2.0, wie auch ihre Vorgängerversionen, wurde entwickelt, um die folgende Aufgabe zu lösen: allen Anwendern eine leicht verständliche und ausdrucksstarke Schreibweise für die visuelle Modellierung zur Verfügung zu stellen, die speziell für die Entwicklung und Dokumentation von Modellen komplexer Systeme mit unterschiedlichen Zielsetzungen konzipiert wurde. Der Punkt ist, dass ein wichtiger Faktor für die weitere Entwicklung und breite Nutzung der UML-Methodik die intuitive Klarheit und Verständlichkeit der Grundkonstrukte der entsprechenden Modellierungssprache ist. Die Sprache UML 2.0 enthält nicht nur abstrakte Konstruktionen zur Darstellung von System-Metamodellen, sondern auch eine Reihe von konkreten Begriffen mit ganz bestimmter Semantik. Damit erreicht UML 2.0 gleichzeitig nicht nur die Einheitlichkeit der Darstellung von Modellen für verschiedene Anwendungen, sondern auch die Möglichkeit, ausreichend feine Details der Realisierung dieser Modelle in Bezug auf bestimmte Systeme zu beschreiben. Obwohl UML 2.0 eine formale Spezifikationsprache ist, unterscheidet sich die Formalität ihrer Beschreibung sowohl von der Syntax traditioneller formal-logischer Sprachen als auch bekannter Programmiersprachen. Die Entwickler der OMG gehen davon aus, dass die UML 2.0 wie keine andere Sprache für bestimmte Themenbereiche angepasst werden kann. Das ist deshalb möglich, weil es in der UML 2.0-Sprachbeschreibung selbst einen Mechanismus zum Erweitern von Basisbegriffen gibt, der ein eigenes Element ist und eine eigene Beschreibung in Form von Erweiterungsregeln hat. Angesichts einer recht einfachen linearen Benutzeroberfläche, die auf die Eingabe von Ausgangsdaten und die darauf basierende Generierung von Ergebnissen abzielt, beschränken wir uns auf das Anwendungsfalldiagramm und die Aktivitätsdiagramme. Anhand letzterer wollen wir die Algorithmen zur Erzeugung der für die Modellierung in Frage kommenden Graphen beschreiben.

Die Reihenfolge der Nutzung des erstellten Systems durch den Anwender ist im Diagramm der Anwendungsfälle dargestellt

Die Abbildungen 4.1, 4.1, 4.1 zeigen Aktivitätsdiagramme für jedes der Zufallsgraphenmodelle, die soziale Netzwerke simulieren, die im vorherigen Abschnitt besprochen wurden.

Unabhängig von der Art des verwendeten Modells wird jeder Graph durch die Anzahl seiner Eckpunkte und die Wahrscheinlichkeit/Wahrscheinlichkeitstupel für bestimmte Verbindungen oder Transformationen definiert. Für jeden der Graphen, die durch den Zufallsprozess erhalten wurden, werden wir den mittleren Abstand, den mittleren normalisierten Abstand und den Clusterkoeffizienten berechnen und die Abhängigkeit des Clusters und des mittleren Abstands von der in den Bedingun-

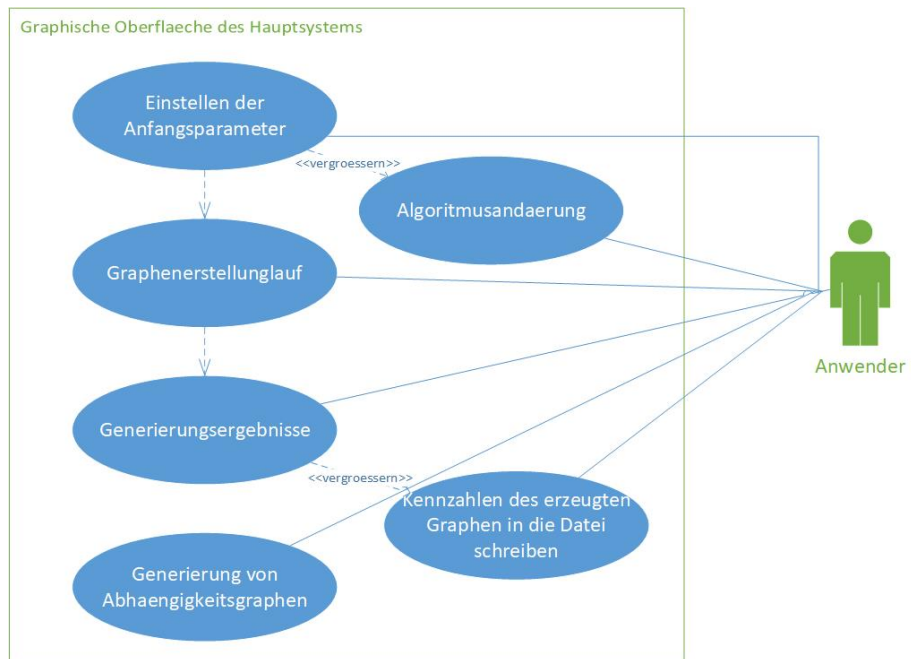


Abbildung 11: Das Anwendungsfalldiagramm

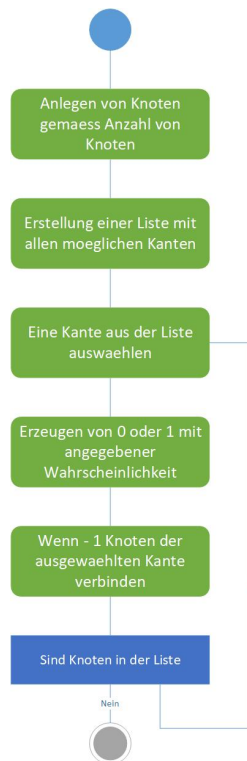


Abbildung 12: Das Anwendungsfalldiagramm

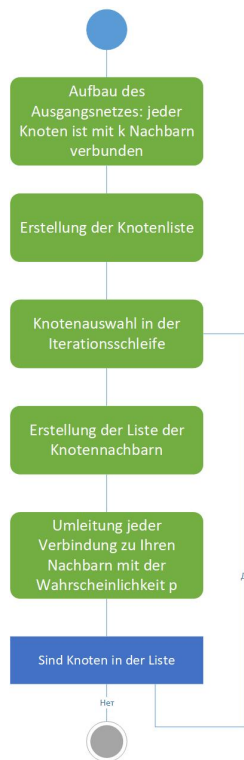


Abbildung 13: Das Anwendungsfalldiagramm

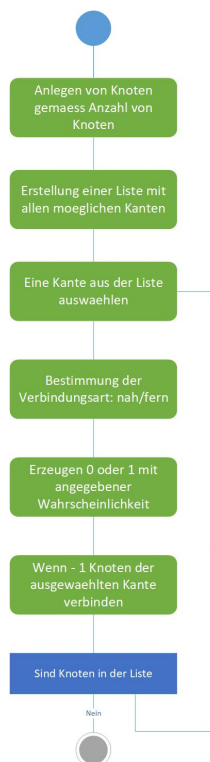


Abbildung 14: Das Anwendungsfalldiagramm

gen angegebenen Ereigniswahrscheinlichkeit darstellen. Für das Modell von Song und Wang erhält man ein dreidimensionales Diagramm, da das Modell durch zwei Wahrscheinlichkeiten gegeben ist: Kurzstrecken- und Langstreckenverbindung.

Das Computermodell ermöglicht es, die Abhängigkeit der Kennzahlen der Modelle von ihren Anfangsparametern zu untersuchen: numerisch und grafisch

Das Computermodell wird mit Python, Version 3.9, implementiert. Python wurde als eine der drei beliebtesten Programmiersprachen ausgewählt. Python ist eine Skriptsprache, die vor der Ausführung der Anwendung nicht kompiliert wird. Python-Skripte können auf fast allen Plattformen von iOS und Android bis hin zu Server-OS programmiert und verwendet werden.

In Verbindung mit Python wird der Paketmanager Anaconda verwendet, der auch zusätzliche Funktionen und Bibliotheken für das wissenschaftliche Arbeiten bereitstellt.

In einer separaten virtuellen Umgebung können nur die benötigten Bibliotheken installiert und verwendet werden. Dadurch wird sichergestellt, dass keine unnötigen, nicht berücksichtigten Abhängigkeiten in der Anwendung vorhanden sind. Das Vorhandensein solcher Abhängigkeiten führt zu Fehlern bei der Kompilierung von Paketen und der Verwendung von Skripten und Anwendungen auf anderen Computern als dem Computer des Anwendungsentwicklers.

Als Basisbibliothek für die Analyse und Visualisierung von Graphen wird das weit verbreitete Paket Networkx verwendet, mit dem sich die gestellten Aufgaben zuverlässig und recht einfach lösen lassen. Die Nachteile des Pakets sind die Geschwindigkeit für einige Algorithmen, was aber für die durchgeführte Modellierung nicht entscheidend ist. Der rechnerische Teil der Aufgabe wird mit Hilfe von NumPy und SciPy durchgeführt, ohne die fast jede mathematische Aufgabe in Python nicht gelöst werden kann.

Die Visualisierung der Lösung erfolgt mit einem Bundle aus PyQt5 und Matplotlib. Das erste ist ein C++-basiertes Framework zur Erstellung von grafischen Oberflächen, das zweite ist eine Bibliothek zur Erstellung von Graphen und anderen notwendigen Visualisierungen in Python.

4.2 Program

Betrachten wir einige Beispiele für die Modellierung von Kleine-Welt-Graphen unter Verwendung der erstellten Software. Der Hauptbildschirm des Programms ermöglicht die Einstellung der grundlegenden Parameter der Diagrammbildung 15.

Hier kann eines von drei Modellen gewählt werden: Erdős-Rényi, Watts-Strogatz und Song-Wang und dessen Parameter definiert werden. Jedes Modell wird durch die Anzahl der Knoten des erzeugten Graphen spezifiziert. Zusätzlich wird für das Erdős-Rényi-Modell die Konnektivität zweier beliebiger Eckpunkte definiert. Das Watts-Strogatz Modell definiert die Anzahl der verbundenen nächsten Nachbarn und die Wahrscheinlichkeit der Wiederverbindung ("rewiring") einer Kante, während das Song-Wang Modell durch Nah- und Fernverbindungswahrscheinlichkeiten zusätzlich

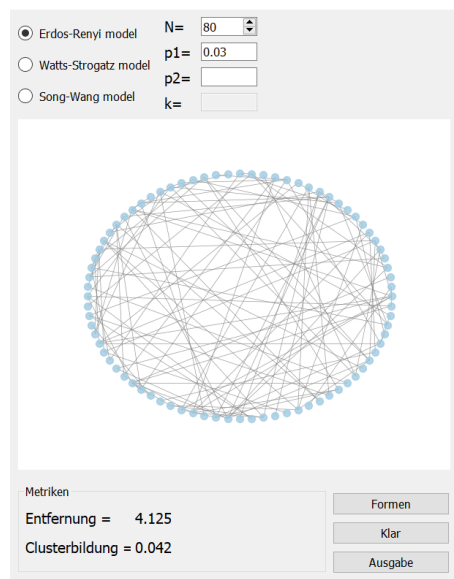


Abbildung 15: Gesamtansicht des Programms

zur Verbindung mit den nächsten Nachbarn definiert ist. Wie aus der Beschreibung der Modelle hervorgeht, muss k ausschließlich gerade sein.

Als Kennzahlen werden der mittlere Abstand und der Clusterkoeffizient berechnet. Bei der Berechnung der durchschnittlichen Distanz werden die im Erdős-Rényi-Modell vorhandenen isolierten Knoten aus dem Graphen entfernt. Der Clusterkoeffizient wird nach der Formel 2.25.

Die Bildung eines Zufallsgraphen nach dem Erdős-Rényi-Modell ist in Abbildung 16 dargestellt. Wie wir sehen, ist der Graph durch eine große Anzahl von Fernverbindungen und einen niedrigen Clusterkoeffizienten gekennzeichnet, was genau mit der theoretischen Ableitung übereinstimmt. Reduzieren wir die Anzahl der Knoten für eine bequemere Betrachtung der einzelnen Eigenschaften.

Erstens ist das Vorhandensein von isolierten Knoten zu beachten. Für ein echtes soziales Netzwerk ist es eher eine Ausnahme. Bei der Erzeugung eines Zufallsgraphen im Erdős-Rényi-Modell werden praktisch immer isolierte Knoten gebildet. Die Erhöhung des Clusterkoeffizienten lässt sich durch den kleinstmöglichen Abstand zwischen den Knoten im erhaltenen Graphen erklären. Aus diesem Grund hat sich auch der durchschnittliche Abstand verringert. In einer Reihe von Experimenten kann jedoch festgestellt werden, dass der durchschnittliche Abstand um nicht mehr als das 1,2-fache abnimmt, wenn die Anzahl der Knoten um den Faktor vier abnimmt.

Betrachten wir den Grenzfall des Erdős-Rényi-Modells.

Bei der Wahrscheinlichkeit der Knotenverbindung $p=0$ erhalten wir erwartungsgemäß die Menge der isolierten Knoten

Wenn wir die Wahrscheinlichkeit gleich eins setzen, erhalten wir einen vollständigen Graphen mit mittlerem Abstand und Clusterkoeffizienten gleich von 1.

Bei der Betrachtung des Watts-Strogatz Modells legen wir die Wahrscheinlichkeit

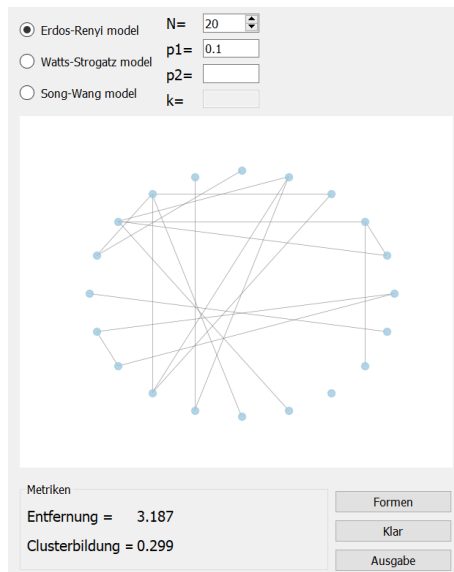


Abbildung 16: Erdős-Rényi-Modell

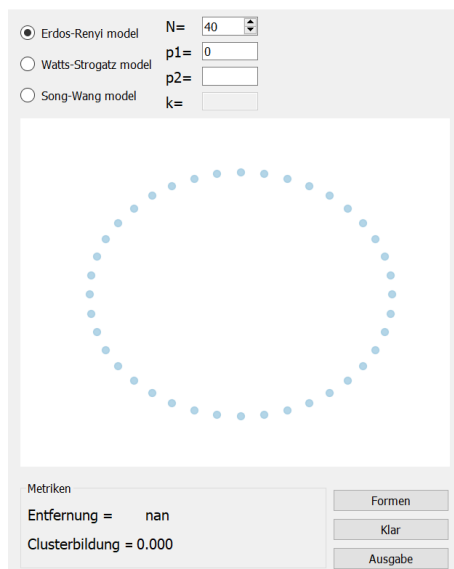


Abbildung 17: Erdős-Rényi-Modell $p = 0$

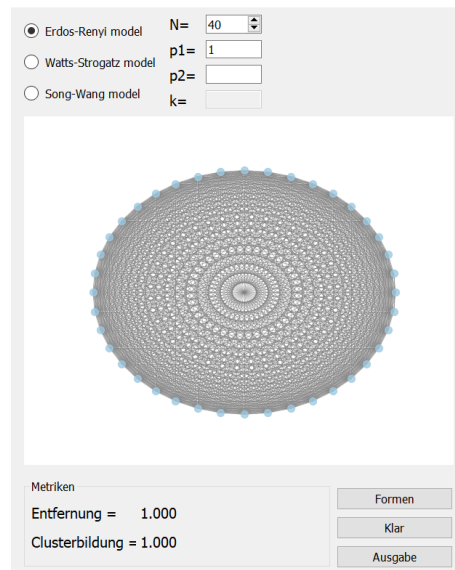


Abbildung 18: Erdős-Rényi-Modell $p = 1$

der Verbindung entfernter Knoten gleich $p = 0,1$ und die Anzahl der verbundenen nächsten Nachbarn gleich 4 fest (19).

Man beachte den deutlich erhöhten Clusterkoeffizienten im Vergleich zum Erdős-Rényi-Modell. Das wird durch die anfängliche Konnektivität der nahen Knoten erklärt.

Durch die Verringerung der Verbindungswahrscheinlichkeit von weit entfernten Knoten wird erwartet, dass sich der durchschnittliche Abstand im Graphen erhöht

So wird in einem Graphen mit der Anzahl der Knoten $N=300$ und der Wahrscheinlichkeit der Verbindung der entfernten Knoten $p=0.05$ der durchschnittliche Abstand $l=5,471$ sein, während der Clusterkoeffizient im Vergleich zum vorher betrachteten Fall unwesentlich zunimmt. (20)

Abschließend wenden wir uns dem Modell von Song und Wang zu

Zu beachten ist auch der hohe Clusterkoeffizient und die umgekehrte Abhängigkeit der durchschnittlichen Distanz von der Fernverbindungswahrscheinlichkeit.

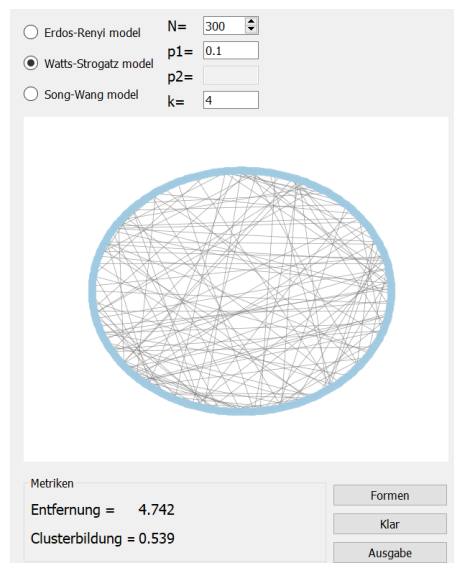


Abbildung 19: Watts-Strogatz-Modell $p = 0.1$

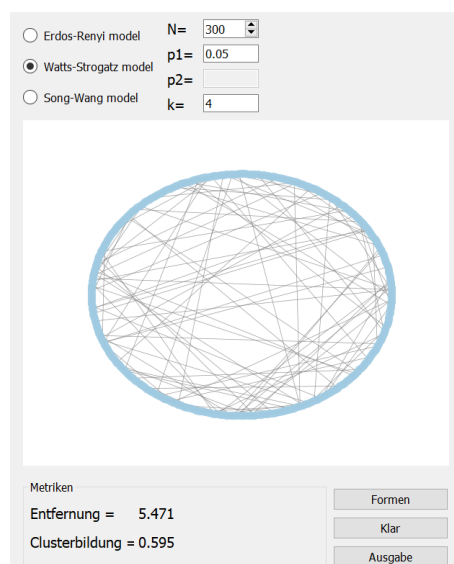


Abbildung 20: Watts-Strogatz-Modell $p = 0.05$

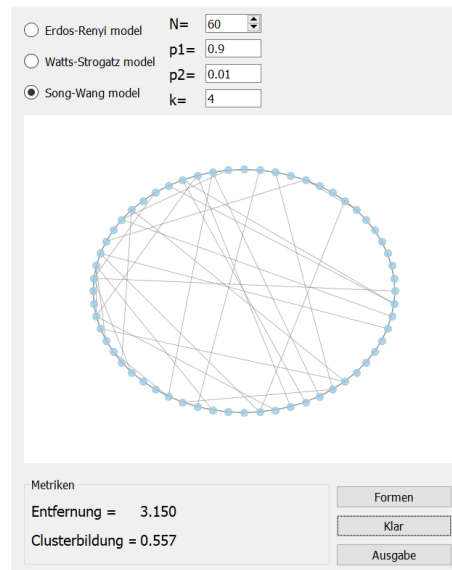


Abbildung 21: Song-Wang-Modell $p_n = 0.9, p_f = 0.01$

5 Fazit

In dieser Arbeit haben wir die von Stanley Milgram initiierte Modellierung der kleinen Welt mit Hilfe von Zufallsgraphen betrachtet. Das 1959 vorgeschlagene Modell des Zufallsgraphen von Erdős-Rényi wurde als Basismodell betrachtet, und eine Reihe seiner Eigenschaften wurde untersucht und eine Verallgemeinerung vorgeschlagen, die für die Lösung angewandter Probleme wesentlich bequemer ist.

Bei der Modellierung von sozialen Netzwerken haben das Erdős-Rényi-Modell und seine Modifikationen einen bedeutenden Nachteil in Form eines niedrigen Clusterkoeffizienten, der das Modell von den in der Praxis existierenden sozialen Netzwerken entfernt, in denen die Teilnehmer recht häufig die sogenannten "Triplets" bilden.

Um die Qualität der Darstellung eines sozialen Netzwerks durch einen Zufallsgraphen zu verbessern, werden Zufallsgraphenmodelle von Watts-Strogatz und Song und Wang eingeführt. Das erste der Modelle basiert auf der anfänglichen Verbindung mit k nächsten Nachbarn, die in der Regel symmetrisch ist, und der Rekonnexion von Knoten mit zufällig ausgewählten Knoten mit einer gegebenen und konstanten Wahrscheinlichkeit.

Die Realisierung der Rekonnexion erwies sich in der Praxis bei der Verwendung von Graphen mit einer großen Anzahl von Knoten als ziemlich kompliziert und erschwerte auch die Beschreibung einiger Eigenschaften des resultierenden Graphen. Um dieses Problem zu lösen, schlugen Song und Wang ein Modell vor, das durch ein Tupel aus zwei Wahrscheinlichkeiten beschrieben wird: die Wahrscheinlichkeit von Nahverbindungen und die Wahrscheinlichkeit von Fernverbindungen.

Im praktischen Teil der Arbeit haben wir mit Hilfe von Python Software zur Erzeugung und Visualisierung von Zufallsgraphen auf Basis der in der Arbeit betrachteten Modelle entworfen und implementiert.

Mit Hilfe des Visualisierungsprogramms wurden verschiedene im theoretischen Teil

der Arbeit betrachtete Grenzfälle untersucht. Die entwickelten Algorithmen zur Erzeugung von Zufallsgraphen basierend auf den gegebenen Anfangsparametern können ohne Änderungen des Codes für andere Forschungsaufgaben verwendet werden, bei denen Modelle basierend auf Zufallsgraphen und auch die Berechnung von Kennzahlen für diese Modelle benötigt werden.

6 Literatur

- [a] Olle Häggström: *Streifzüge durch die Wahrscheinlichkeitstheorie*, Springer 2006, ISBN 978-3-540-23050-2
- [b] Tomasz Luczak, Introduction to Random Graphs, unveröffentlichte Folien, 2015
- [c] R. Diestel: *Graph Theory*, Springer 2010, 5. Auflage, ISBN 978-3-662-53621-6
- [d] David Meintrup, Stefan Schäffler: *Stochastik*, Springer 2005, 1. Auflage, ISBN 3-540-21676-6
- [e] A. Frieze, M. Karonski: *Introduction to Random Graphs*, <https://www.math.cmu.edu/~af1p/BOOK.pdf>
- [f] Watts, D. J. & Strogatz, S. H. Collective dynamics of "small-world" networks. *Nature* 393, 440 (1998)
- [g] Kleinberg, J. The Small-world Phenomenon: An Algorithmic Perspective. In Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing, STOC-00, pp. 163-170 (ACM, New York, NY, USA, 2000)
- [h] Watts, D. J., Dodds, P. S. & Newman, M. E. J. Identity and Search in Social Networks. *Science* 296, 1302 (2002)
- [i] Newman, M. E. J. & Watts, D. J. Renormalization group analysis of the small-world network model. *Physics Letters A* 263, 341 (1999)
- [j] Barrat, A. & Weigt, M. On the properties of small-world network models. *Te European Physical Journal B* 13, 547 (2000)
- [k] Song, H. F. & Wang, X.-J. Simple, distance-dependent formulation of the Watts-Strogatz model for directed and undirected small-world networks. *Physical Review E* 90, 062801 (2014)
- [l] Maier, B. F., Huepe, C. & Brockmann, D. Modular hierarchical and power-law small-world networks bear structural optima for minimal first passage times and cover time. *Journal of Complex Networks*, <https://doi.org/10.1093/comnet/cnz010> (2019)
- [m] Benjamin F. Maier, Generalization of the small-world effect on a model approaching the Erdős-Rényi random graph. *Scientific Reports* volume 9, Article number:

9268 (2019), <https://www.nature.com/articles/s41598-019-45576-3>

[n] H. Lavicka, Simulations of agents on social networks, doctoral Thesis, Czech Technology University, Prague, 2010

[o] Duncan J. Watts, Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440-442, 1998

[p] Schank T., Wagner D. Approximating Clustering-Coefficient and Transitivity, University of Karlsruhe, Germany, (2004),
https://www.iti.kit.edu/_media/projects/spp1126/files/sw-acct-05.pdf

[q] Newman, M. E. J. Models of the Small World A Review. Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, <https://arxiv.org/abs/cond-mat/0001118>

Besonderer Dank an

7 Anhang

7.1 Code

```
from PyQt5 import QtWidgets as qw
from app.gui.main_window import Ui_MainWindow
import sys
import networkx as nx
import numpy as np
from random import choice, shuffle
from itertools import combinations
from random import randint
from networkx.algorithms.approximation.clustering_coefficient import average_c

options = {
    "node_color": "#AOCBE2",
    "width": 0.5,
    "with_labels": False,
    'node_size': 50,
    'alpha': 0.8,
    'edge_color': 'grey'
}

def prob_func(prob):
    p = int(prob * 100)
    arr = np.zeros(100)
    arr[:p] = 1
    arr = list(arr)
    shuffle(arr)
    return choice(arr)

class TheWindow(qw.QMainWindow):
    """
    Program's graphic interface
    """

    def __init__(self):
        """
        Main window constructor
        """
        # main window init
        super(TheWindow, self).__init__()
        self.ui = Ui_MainWindow()
        self.ui.setupUi(self)
```

```

# mathematics init
self.G = nx.Graph()
self.possible_edges = None

# set elements
self.ui.radioButton.setChecked(True)
self.ui.spinBox.setValue(20)
self.ui.prob1.setText("0.1")
self.ui.k.setEnabled(False)

# set handlers
self.ui.pushButton.clicked.connect(self.on_exit)
self.ui.pushButton_2.clicked.connect(self.on_clear)
self.ui.pushButton_3.clicked.connect(self.on_plot)

self.ui.radioButton.toggled.connect(self.onClicked)
self.ui.radioButton_2.toggled.connect(self.onClicked)
self.ui.radioButton_3.toggled.connect(self.onClicked)

def onClicked(self):
    rb = self.sender()
    if rb.isChecked():
        if rb.text() == 'Erdos-Renyi model':
            self.ui.k.setEnabled(False)
            self.ui.prob2.setEnabled(False)
            self.ui.k.setText("")
            self.ui.prob2.setText("")
        elif rb.text() == 'Watts-Strogatz model':
            self.ui.k.setEnabled(True)
            self.ui.prob2.setEnabled(False)
            self.ui.k.setText("4")
            self.ui.prob2.setText("")
        elif rb.text() == 'Song-Wang model':
            self.ui.k.setEnabled(True)
            self.ui.prob2.setEnabled(True)
            self.ui.k.setText("4")
            self.ui.prob1.setText("0.85")
            self.ui.prob2.setText("0.05")

    @staticmethod
    def make_n_k(k):
        if k % 2 != 0:
            raise ValueError("k shouldn't be odd")

        n_k = np.array(list(range(k + 1)))
        return n_k - int(k / 2)

def on_exit(self):

```

```

"""
Exit button method
:return: None
"""
try:
self.qtbd.close()
except:
pass
sys.exit(0)

def on_plot(self):
self.on_clear()

self.ui.pushButton_3.setDisabled(True)
self.G.add_nodes_from(list(range(self.ui.spinBox.value())))

if self.ui.radioButton.isChecked():
self.ER_edges()
elif self.ui.radioButton_2.isChecked():
self.WS_edges()
elif self.ui.radioButton_3.isChecked():
self.SW_edges()

# draw graph
layout = nx.circular_layout(self.G)
nx.draw(self.G, layout, **options, ax=self.ui.static_ax)

self.ui.cs.draw()

# average path length
avg_ls = list(filter(lambda x: x > 0.5,
[cur_path := nx.average_shortest_path_length(C) for C in (self.G.subgraph(c).copy()
for c in nx.connected_components(self.G))]))
self.ui.distance.setText(f'{np.mean(avg_ls):5.3f}')

# clustering coefficient
cl_koef = average_clustering(self.G, trials=10000)
self.ui.clustering.setText(f'{cl_koef:4.3f}')

self.ui.pushButton_3.setDisabled(False)

def ER_edges(self):
p = float(self.ui.prob1.text())
self.possible_edges = combinations(list(self.G.nodes), 2)
edge_list = []
for edge in self.possible_edges:
if prob_func(p) == 1:
edge_list.append(edge)

```

```

self.G.add_edges_from(edge_list)

def WS_edges(self):
    N = int(self.ui.spinBox.value())
    p = float(self.ui.prob1.text())

    # neighbours
    k = int(self.ui.k.text())
    n_k = self.make_n_k(k)

    # initial wiring
    near_edges = []
    for node in list(self.G.nodes):
        neighbours = n_k + node
        for one in neighbours:
            if one != node:
                if one < 0:
                    one += N
                elif one > N - 1:
                    one -= N
                near_edges.append((node, one))

    # rewiring
    final_edges = []
    for edge in near_edges:
        if prob_func(p) == 1:
            # make rewiring
            node = edge[0]
            node_left = node + 1 if (node + 1) < N else (node + 1 - N)
            node_right = node - 1 if (node - 1) > -1 else (node - 1 + N)
            while (new_node := randint(0, N)) in [node_right, node, node_left]:
                pass
            final_edges.append((edge[0], new_node))
        else:
            final_edges.append((edge[0], edge[1]))
    self.G.add_edges_from(final_edges)

def SW_edges(self):
    N = int(self.ui.spinBox.value())
    p1 = float(self.ui.prob1.text())
    p2 = float(self.ui.prob2.text())

    # neighbours
    k = int(self.ui.k.text())
    n_k = self.make_n_k(k)

    # wiring
    near_edges = []

```

```

far_edges = []
for node in list(self.G.nodes):
    neighbours = n_k + node
    # short distance
    for one in neighbours:
        if probab_func(p1) == 1:
            if one != node:
                if one < 0:
                    one += N
                elif one > N - 1:
                    one -= N
                near_edges.append((node, one))

    # long distance
    not_neighbours = [node for node in list(self.G.nodes) if node not in neighbours]
    for one in not_neighbours:
        if probab_func(p2) == 1:
            if one < 0:
                one += N
            elif one > N - 1:
                one -= N
            far_edges.append((node, one))

self.G.add_edges_from(near_edges)
self.G.add_edges_from(far_edges)

def on_clear(self):
    self.ui.static_ax.clear()
    self.ui.cs.draw()
    self.G = nx.Graph()

```


7.2 Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich, , die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Jedwede Erwähnung einer Quelle, ob im Sinn oder Wortlaut habe ich entsprechend kenntlich gemacht.

Unterschrift