

# Kleine Welten und dezentrale Suche

Martin Mader\*  
Universität Konstanz  
Seminar „Methoden der Netzwerkanalyse“  
Winter Semester 2005/06

12. Mai 2006

## **Inhaltsverzeichnis**

<b>1</b>	<b>Einführung</b>	<b>1</b>
<b>2</b>	<b>Grundlagen</b>	<b>1</b>
<b>3</b>	<b>Dezentrale Suche</b>	<b>6</b>
<b>4</b>	<b>Das Modell von Kleinberg</b>	<b>8</b>
<b>5</b>	<b>Zusammenfassung</b>	<b>12</b>

---

\*e-mail: mader@inf.uni-konstanz.de

# 1 Einführung

Das Phänomen der kleinen Welt, also dass alle Menschen über kleine Ketten von Bekannten verbunden sind, gehört zu unserer alltäglichen Erfahrung. Erste quantitative Studien dieses Phänomens wurden innerhalb der Sozialwissenschaften durchgeführt. Ein Experiment von Stanley Milgram et al. [7] in den 1960'ern ist wohl das bekannteste und meist zitierte innerhalb dieses Gebiets. Darin wurden Versuchspersonen aufgefordert, eine Nachricht an eine (unbekannte) Zielperson zu übermitteln, wobei die Nachricht aber nur an eine bekannte Person weitergegeben werden durfte. Das Ergebnis der Studie war dann, dass die Länge der erfolgreichen „Ketten“ zwischen fünf und sechs lag, was zum berühmten Prinzip der „*six degrees of separation*“ führte. Natürlich sind die Ergebnisse mit Vorsicht zu genießen, viele der Nachrichten kamen überhaupt nicht an, und die Zielperson hatte einen relativ hohen sozialen Status.

Dennoch lieferte die Studie eine qualitative Aussage über soziale Netzwerke im Allgemeinen: Sie weisen meist kurze Pfade zwischen Individuen auf. Später wurden dieselbe Eigenschaft auch für gänzlich andere Netzwerke, wie z.B. im neuronalen Netz des Fadenwurms *C. elegans*, dem Stromnetz der westlichen USA, oder dem Hyperlink-Netzwerk des Internets festgestellt. So beschäftigten sich seit Milgrams Experiment viele Wissenschaftler mit der Modellierung kleiner Welten. Der erste Teil dieser Arbeit behandelt kurz die Grundlagen der Modellierung kleiner Welten und führt auf das Modell von Watts und Strogatz [8] hin.

Eine wesentliche Erkenntnis des Experiments von Milgram ist aber zudem, dass kurze Wege zwischen Individuen nicht nur existieren, sondern dass die Individuen auch in der Lage sind, diese ohne eine globale Kenntnis des Netzwerks zu finden. Dies führt zum Konzept der dezentralen Suche in kleinen Welten, die in Abschnitt 3 behandelt wird. Es wird dabei gezeigt, dass es keinem dezentral arbeitenden Algorithmus möglich ist, die kurzen Wege im Modell von Watts und Strogatz in optimaler Zeit zu finden.

Jon Kleinberg veröffentlichte deshalb in [5] ein Modell, das im Wesentlichen das Modell von Watts und Strogatz erweitert und die dezentrale Suche in optimaler Zeit ermöglicht. Dieses Modell und die zentralen Aussagen dieser Arbeit werden in Abschnitt 4 behandelt.

# 2 Grundlagen

Wie in der Einführung beschrieben, treten kleine Welten in vielen sozialen und technologischen Netzwerken auf. Um diese komplexen Netzwerke zu verstehen, wird versucht, einfache Modelle zu bilden, welche wesentliche empirische Eigenschaften dieser Netzwerke repräsentieren. Oft führt die Analyse dieser Modelle dann zu neuen Fragestellungen, die wiederum empirisch überprüft werden können. In dieser Arbeit liegt der Fokus auf sozialen Netzwerken, alle Aussagen sind aber generell übertragbar.

Es stellt sich also die Frage, was kleine Welten auszeichnet. Empirische Beobachtungen legen folgende grundlegenden Eigenschaften nahe:

1. Beschränkte Anzahl von Bekannten pro Person.
2. Bekannte einer Person kennen sich untereinander.
3. Hohe Wahrscheinlichkeit für Existenz kurzer Wege zwischen allen Personen

Für gewöhnlich werden soziale Netzwerke als Graphen modelliert, in denen die Knoten die Personen darstellen. Eine Kante zwischen zwei Knoten bedeutet dann, dass diese

Personen sich „kennen“. Diesbezüglich können die obigen Eigenschaften umformuliert werden:

1. kleiner durchschnittlicher Knotengrad ohne Ausreißer.
2. lokal dichte Teilgraphen.
3. kleiner durchschnittlicher Abstand bzw. kleiner Durchmesser.

Um die obenstehenden Eigenschaften in Graphen zu formalisieren und überprüfen zu können, werden noch Kenngrößen für diese Eigenschaften benötigt. Wir betrachten dabei schlichte, gerichtete Graphen; für ungerichtete Graphen beziehen sich die folgenden Größen auf den entsprechenden gerichteten Graphen. Die Kenngröße für die erste Eigenschaft ist der durchschnittliche Knotengrad:

**Definition 1 (Durchschnittlicher Knotengrad)** Sei  $G = (V, E)$  ein (Multi-)graph und bezeichne  $d(v)$  den Knotengrad von einem Knoten  $v \in V$ . Dann heißt

$$d(G) = \frac{1}{n} \cdot \sum_{v \in V} d(v)$$

der *durchschnittliche Knotengrad* von  $G$ .

**Definition 2 (Dichte)** Die *Dichte*  $\mathbb{D}(G)$  eines (Multi-)Graphen  $G = (V, E)$  ist definiert durch

$$\mathbb{D}(G) = \frac{|\{(u, w) \in V \times V : (u, w) \in E, u \neq w\}|}{n(n-1)}$$

wobei  $n = |V|$ .

Die Dichte gibt also an, wieviele der prinzipiell in einem Graph möglichen Kanten auch tatsächlich vorhanden sind. Über die Dichte kann nun folgende Kenngröße definiert werden:

**Definition 3 (Cluster-Koeffizient)** Sei  $G = (V, E)$  ein Multigraph und  $G[V'] \subseteq G$  der von einer Teilmenge  $V' \subseteq V$  induzierte Teilgraph.  $N(v)$  bezeichne die Nachbarschaft eines Knoten  $v \in V$ . Die Dichte

$$C(v) = \mathbb{D}(G[N(v) \setminus \{v\}])$$

der Nachbarschaft eines Knoten  $v \in V$  heißt *Cluster-Koeffizient*  $C(v)$  von  $v$ .

Der Cluster-Koeffizient  $C(G)$  eines Graphen  $G$  ist der mittlere Cluster-Koeffizient seiner Knoten.

Der Cluster-Koeffizient eines Graphen gibt also an, wie stark die Nachbarn eines Knoten im Mittel untereinander verbunden sind. Somit ist er ein Maß für die lokale Dichte eines Graphen. Die dritte Eigenschaft, also der durchschnittliche Abstand zwischen Knoten, wird als charakteristische Weglänge definiert:

**Definition 4 (Charakteristische Weglänge)** Die *charakteristische Weglänge* eines Multigraphen  $G = (V, E)$  ist der mittlere Abstand seiner Knoten,

$$L(G) = \frac{1}{n(n-1)} \cdot \sum_{s \neq t \in V} d_G(s, t)$$

wobei  $d_G(s, t)$  den Abstand zwischen zwei Knoten  $s$  und  $t$  bezeichnet.

Im Folgenden sollen einige Graphenmodelle auf ihre Tauglichkeit überprüft werden, kleine Welten zu modellieren.

Betrachten wir zuerst den *vollständigen Graphen*  $K_n$  mit  $n$  Knoten. Hierbei ist  $C(G) = 1$  und  $L(G) = 1$ , allerdings ist der durchschnittliche Knotengrad  $d(G) = n - 1$ , und somit nicht klein. Klarerweise wird die charakteristische Weglänge mit zunehmender Beschränkung des Knotengrades steigen, insbesondere wird kein Modell für kleine Welten eine konstant beschränkte charakteristische Weglänge aufweisen. Es stellt sich also die Frage, welche Schranke für diese Kenngröße im Allgemeinen zu erwarten ist.

Deshalb betrachten wir als nächstes die Klasse  $\mathcal{G}(n, r\text{-reg})$  der Zufallsgraphen mit konstantem Knotengrad  $r$ , die sogenannten *zufälligen  $r$ -regulären Graphen*. Gilt  $r \ll n$ , so ist die erste Eigenschaft natürlich erfüllt.

**Theorem 1 (Bollobás, de la Vega 1982, [1])** Sei  $r \geq 3$  und  $\varepsilon > 0$  fest. Sei  $d = d(n)$  die kleinste natürliche Zahl, die

$$(r - 1)^{d-1} \geq (2 + \varepsilon) r n \log n$$

erfüllt. Dann haben fast alle  $r$ -regulären Graphen der Ordnung  $n$  einen Durchmesser von höchstens  $d$ .

**Korollar 1** Sei  $r \geq 3$  konstant. Wird zufällig ein Graph aus der Menge aller Graphen mit  $n$  Knoten gewählt, in dem jeder Knoten exakt Grad  $r$  hat, so ist jedes Knotenpaar mit hoher Wahrscheinlichkeit durch einen Weg der Länge  $O(\log n)$  verbunden.

Das bedeutet, dass in den meisten Zufallsgraphen (mit obigen Beschränkungen) zwischen allen oder zumindest einem Großteil der Knoten ein Weg existiert, dessen Länge logarithmisch in  $n$ , und damit exponentiell kleiner als die Anzahl der Knoten, ist. Eine logarithmisch – oder allgemeiner polylogarithmisch – beschränkte charakteristische Weglänge stellt also auch im Bezug auf Modelle für kleine Welten das erreichbare Optimum dar.

Somit weisen  $r$ -reguläre Graphen zwar einen beschränkten Knotengrad und eine kleine charakteristische Weglänge auf, sie erfüllen aber nicht den Anspruch an lokal dichte Teilgraphen, was den empirischen Beobachtungen zufolge eine wesentliche Eigenschaft von kleinen Welten ist. Ein  $r$ -regulärer Zufallsgraph  $G_{\text{random}}$  ist in der Regel lokal dünn besetzt. Nach [8] gilt  $C(G_{\text{random}}) \sim \frac{r}{n} \ll 1$ , so dass die Nachbarn eines Knotens mit hoher Wahrscheinlichkeit selbst keine Nachbarn sind.

Betrachten wir nun also ein Graphenmodell, in dem garantiert lokal dichte Teilgraphen auftreten, den *Torus*. Ein Torus ist im Wesentlichen ein  $d$ -dimensionales Gitter, in dem alle Knoten innerhalb eines „Radius“  $r$  (bezüglich der Manhattan-Distanz) verbunden sind. Dabei werden Knoten auch über die „Grenzen“ einer Dimension verbunden.

**Definition 5 (Torus)** Seien  $d, k, r \in \mathbb{N}$ ,  $V_k = \{0, \dots, k - 1\}$  und  $V_k^d = \underbrace{V_k \times \dots \times V_k}_{d \text{ mal}}$ .

Der Graph  $G = (V_k^d, E)$  mit

$$\{(v_1, \dots, v_d), (w_1, \dots, w_d)\} \in E \Leftrightarrow 1 \leq \sum_{i=1}^d \min\{v_i - w_i, w_i - v_i \bmod k\} \leq r$$

heißt  $(k, d, r)$ -Torus.

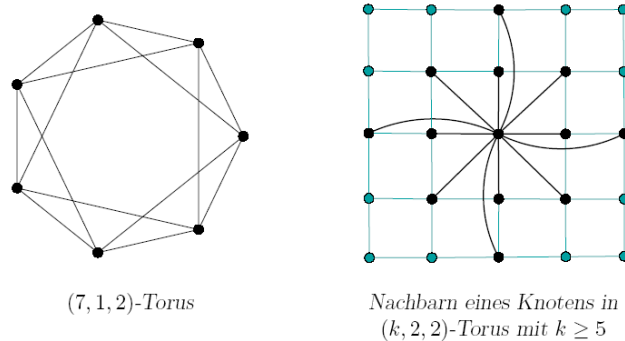


Abbildung 1: Beispiele für Tori, aus [3]

Im einfachsten Fall ( $d = 1$ ) besteht ein Torus aus einem Ring, bei dem ein Knoten  $v$  zusätzlich zu allen anderen Knoten innerhalb des Abstandes  $r$  auf dem Ring verbunden ist, siehe Abbildung 1. Für einen solchen  $(n, 1, r)$ -Torus  $G = (V, E)$  mit  $r \leq \frac{n-1}{2}$  können folgende Aussagen gezeigt werden [3]:

- Durchschnittlicher Knotengrad  $d(G) = 2r$
- $C(G) = \frac{3(r-1)}{4(r-\frac{1}{2})}$
- $L(G) \geq \lfloor \frac{n-1}{4r} \rfloor$

Für höhere Dimensionen können analoge Aussagen gezeigt werden.

Wie oben angedeutet besitzt dieses Modells also lokal dichte Teilgraphen, der Cluster-Koeffizient ist konstant und relativ hoch. Auch der Knotengrad ist beschränkt. Jedoch ist die charakteristische Weglänge linear in  $n$ , und somit exponentiell zur erwarteten polylogarithmisch beschränkten charakteristischen Weglänge.

Beide vorhergehenden Modelle beschreiben nur zwei der drei Eigenschaften von kleinen Welten. 1998 veröffentlichten Watts und Strogatz ein Modell für kleine Welten, das beide vorangegangenen Konzepte, also ein in hohem Maße geclustertes, reguläres und ein zufälliges Netzwerk, kombiniert [8]. Ein  $(k, d, r)$ -Torus bildet die Grundlage des Netzwerkmodells. Darauf aufbauend werden die Kanten mit einer festen Wahrscheinlichkeit  $p$  umgeleitet und erhalten einen neuen, zufällig gewählten Endknoten. Dabei wird uniform aus der Knotenmenge gewählt. Das Watts-Strogatz-Modell ist folgendermaßen definiert [3]:

**Definition 6 (Kleine Welt, Watts und Strogatz 1998)** Zu  $k, d, r \in \mathbb{N}$  und  $0 \leq p \leq 1$  heißt das Graphenmodell

$$\mathcal{S}(k, d, r, p) = \left( \mathcal{G}(k^d), P \right)$$

*kleine Welt* der Dimension  $d$ , falls  $P$  die Wahrscheinlichkeit angibt, dass der Graph aus einem  $(k, d, r)$ -Torus entsandt ist, indem jede Kante unabhängig mit der Wahrscheinlichkeit  $p$  genau einen neuen Endknoten erhalten hat.

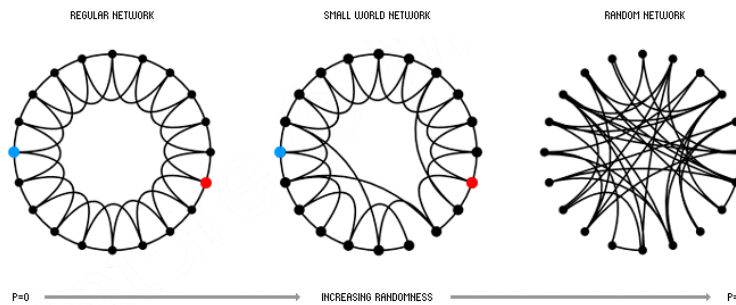
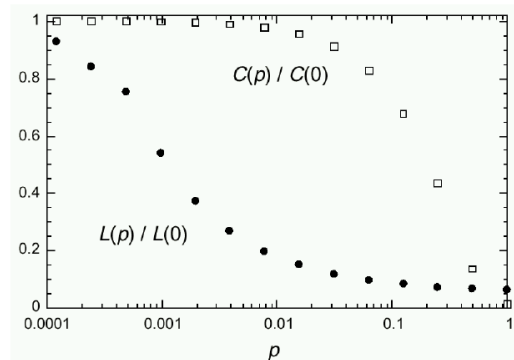


Abbildung 2: Entwicklung kleiner Welten nach dem Watts-Strogatz-Modell, aus [8]

Abbildung 3:  $C(G)$  und  $L(G)$  in Abhängigkeit der Wahrscheinlichkeit  $p$ , aus [8]

Wie in Abbildung 2 zu sehen, entstehen durch die zufällige Umleitung von Kanten „Abkürzungen“ im Netzwerk, so dass die charakteristische Weglänge sinkt. Allerdings sinkt an diesen Stellen mit hoher Wahrscheinlichkeit auch der Cluster-Koeffizient. Experimentelle Ergebnisse zeigen aber, dass mit Erhöhung der Wahrscheinlichkeit  $p$  im Bereich  $0 \leq p \leq 0.1$  die charakteristische Weglänge rapide sinkt und der Cluster-Koeffizient des Netzwerks dennoch hoch bleibt, wie Abbildung 3 zeigt. Die Daten sind Durchschnittswerte aus 20 Instanzen des Modells von Watts und Strogatz, wobei  $k = 1000$ ,  $d = 1$  und  $r = 5$ , und wurden mit den Werten  $L(0)$  und  $C(0)$  des regulären Gitters normalisiert.

Eine analytische Bestätigung dieser Ergebnisse gibt es bisher nicht. Allerdings wird in [2] gezeigt, dass der Durchmesser, und somit die charakteristische Weglänge, eines sehr ähnlichen Modells polylogarithmisch beschränkt ist.

**Theorem 2 (Bollobás, Chung 1988, [2])** Sei  $G$  ein Graph, der entsteht, wenn zu einem Kreis mit  $n$  Knoten ein zufälliges Matching hinzugefügt wird. Mit einer Wahrscheinlichkeit, die gegen 1 geht, wenn  $n \rightarrow \infty$ , erfüllt der Durchmesser  $D(G)$  von  $G$

$$\log n - c < D(G) \leq \log n + \log \log n + c$$

wobei  $c$  eine kleine Konstante ist.

Für das Watts-Strogatz Modell ist also eine ähnliche Aussage zu vermuten, und es erfüllt somit die gewünschten Eigenschaften eines Modells für kleine Welten: Der Knotengrad ist beschränkt und für geeignete  $p$  ist die lokale Dichte hoch und die charakteristische Weglänge klein.

### 3 Dezentrale Suche

In dem vorhergehenden Abschnitt wurde auf das Modell von Watts und Strogatz hingeführt, das die drei anfangs genannten strukturellen Eigenschaften erfüllt. Betrachtet man aber noch einmal das grundlegende Experiment von Milgram, oder auch andere Netzwerke, die kleine Welten darstellen, so zeigt sich eine erstaunliche „algorithmische“ Eigenschaft in diesen Netzwerken: Es existieren nicht nur kurze Wege zwischen beliebigen Knoten, die Individuen können diese auch finden, und das nur unter Benutzung *lokaler* Informationen!

Im Bezug auf Netzwerkmodelle für kleine Welten ergibt sich daraus folgende Frage:

Welche Eigenschaften muss ein Netzwerkmodell besitzen, damit ein *dezentral* arbeitender Algorithmus in kleinen Welten Wege zwischen beliebigen Knoten in polylogarithmisch beschränkter Zeit findet ?

Diese Frage führt nun zum Konzept der dezentralen Suche in kleinen Welten, ähnlich der Herangehensweise in Milgrams Experiment. Ein dezentraler Algorithmus versucht, eine virtuelle „Nachricht“ von einem beliebigen Startknoten  $s$  an einen beliebigen Zielknoten  $t$  in möglichst kurzer Zeit, also über einen möglichst kurzen Weg, zu übermitteln. Die Nachricht wird sequentiell von einem Knoten zu einem seiner bekannten, d.h. adjazenten, Knoten weitergegeben. Dabei arbeitet der Algorithmus ausschließlich auf lokaler Information; der Knoten, bei dem sich die Nachricht befindet, kennt nur

- Die Menge seiner Bekannten und deren Ort bzw. Koordinaten bezüglich der unterliegenden Struktur des Netzwerks (z.B. Torus im Modell von Watts und Strogatz) und
- Den Ort bzw. die Koordinaten des Zielknotens.

Im Folgenden wird stets ein einfacher Greedy-Algorithmus  $\mathcal{A}$  betrachtet: Sei  $u$  der Knoten, bei dem sich die Nachricht momentan befindet. Als Empfänger der Nachricht wird derjenige Knoten  $v$  unter allen Nachbarn von  $u$  ausgewählt, der die Nachricht dem Zielknoten  $t$  bezüglich des Abstands  $d(v,t)$  auf der unterliegenden Struktur am nächsten bringt.

Dabei wird die Nachricht entweder an einen Knoten weitergegeben, der durch die unterliegende Struktur benachbart ist; diese Art von Nachbarn wird als *lokaler Kontakt* bezeichnet. Oder die Nachricht wird an einen Knoten übermittelt, der durch das zufällige Umleiten oder Neueinfügen von Kanten benachbart ist, und somit potentiell eine Abkürzung auf dem Weg zum Zielknoten darstellt. Diese Art von Nachbarknoten wird als *entfernter Kontakt* bezeichnet.

Um nun zu beurteilen, inwieweit die „Individuen“ eines Modells für kleine Welten in der Lage sind, eine Nachricht nur mit Hilfe lokaler Informationen zu übermitteln, benötigt man ein Maß für die Güte eines Netzwerkmodells im Bezug auf dezentrale Suche:

**Definition 7 (Erwartete Lieferzeit)** Sei  $\mathcal{S}$  ein Netzwerkmodell für kleine Welten. Die *erwartete Lieferzeit*  $EX$  eines dezentralen Algorithmus  $\mathcal{A}$  ist die erwartete Anzahl von Schritten, die  $\mathcal{A}$  benötigt, um eine Nachricht zwischen zwei *zufällig* gewählten Knoten  $s$  und  $t$  aus  $\mathcal{S}$  zu übermitteln.

Zurück zum Modell von Watts und Strogatz: Wie oben festgestellt wurde, existieren in diesem Modell mit hoher Wahrscheinlichkeit kurze, also polylogarithmisch beschränkte, Wege. Aber kann ein dezentraler Algorithmus die Wege zwischen beliebigen Knoten auch in polylogarithmisch beschränkter Zeit finden? Man betrachte eine zweidimensionale Instanz des Modells:

**Theorem 3** Sei  $\mathcal{S}(k, 2, r, p)$  ein Netzwerkmodell nach Watts und Strogatz der Dimension 2 mit  $n = k^2$  Knoten.

Es existiert eine Konstante  $c_0$ , abhängig von  $r$  und  $p$ , aber unabhängig von  $n$ , so dass die erwartete Lieferzeit  $EX$  eines beliebigen dezentralen Algorithmus mindestens  $c_0 \cdot k^{2/3} = c_0 \cdot n^{1/3}$  beträgt.

**BEWEIS** Betrachte die Menge  $U$  der Knoten mit Abstand kleiner  $k^{2/3}$  zum Zielknoten  $t$ . Somit liegt der Startknoten  $s$  mit hoher Wahrscheinlichkeit außerhalb von  $U$ . Sei also  $s \notin U$ :

- Wird die Nachricht nie an einen entfernten Kontakt in  $U$  weitergegeben, so ist die benötigte Anzahl an Schritten proportional zu  $k^{2/3}$ .
- Andererseits ist die Wahrscheinlichkeit, dass irgendein Knoten einen entfernten Kontakt in  $U$  hat,  $p \cdot \frac{1}{n} \cdot |U|$ , da der Endknoten einer umgeleiteten Kante nach der uniformen Wahrscheinlichkeitsverteilung ausgewählt wird. Es gilt :

$$\begin{aligned} p \cdot \frac{1}{n} \cdot |U| &= p \frac{1}{n} \cdot \sum_{i=1}^{k^{2/3}} 4i \\ &= 4p \cdot \frac{1}{n} \cdot \frac{1}{2} \cdot k^{2/3} (k^{2/3} + 1) \\ &\leq 4p \cdot \frac{k^{4/3}}{k^2} \\ &\sim k^{-2/3} \end{aligned}$$

Somit ist die erwartete Anzahl an Schritten, um einen solchen Kontakt zu finden, ebenfalls proportional zu  $k^{2/3}$ .

□

Das Theorem zeigt, dass  $EX$  im Modell von Watts und Strogatz in  $\Omega(n^{1/3})$ , also polynomial in  $n$ , ist. Dies ist sogar exponentiell in der vermuteten polylogarithmisch beschränkten charakteristischen Weglänge.

Warum aber scheitert ein dezentraler Algorithmus im Watts-Strogatz Modell? Anscheinend gibt es für die Individuen in diesen Netzwerken keine Hinweise darüber, ob ein entfernter Kontakt die Nachricht dem Ziel wirklich näher bringt, also ob die Nachricht über diesen Kontakt andere Knoten erreichen kann, von denen eine Abkürzung zum Ziel existiert. Die entfernten Kontakte werden „zu zufällig“ ausgebildet, es fehlt an Korrelation zwischen der Bildung dieser Kontakte und der Geometrie der unterliegenden Struktur.

In [5] beschreibt Jon Kleinberg eine Familie von Netzwerkmodellen, die im Wesentlichen das Modell von Watts und Strogatz erweitert. Hier wird die Wahrscheinlichkeit für die Auswahl der Knoten, zu denen eine zufällige Kante ausgebildet wird, an die Geometrie der unterliegenden Struktur des Netzwerks angepasst. Im Folgenden wird diese Modellfamilie vorgestellt, und gezeigt, dass für ein Modell der Familie die erwartete Lieferzeit polylogarithmisch beschränkt ist.

## 4 Das Modell von Kleinberg

Auch im Modell von Kleinberg entstehen die Eigenschaften einer kleinen Welt durch die Kombination eines regulären Grundgerüsts mit einem zufälligen Netzwerk. Im Gegensatz zu dem Modell von Watts und Strogatz arbeitet Kleinberg mit einem gerichteten Graphen, dessen Grundlage ein  $d$ -dimensionales Gitter ist. Auch hier werden Kanten zu allen Knoten innerhalb eines bestimmten Abstands  $r$  ausgebildet; diese Knoten sind dann die lokalen Kontakte. Die Kanten zu entfernten Kontakten entstehen hier aber nicht durch die zufällige Umleitung bereits vorhandener Kanten; stattdessen wird für jeden Knoten eine fest vorgegebene Anzahl an zufälligen Kanten zu entfernten Kontakten hinzugefügt. Dabei wird der entfernte Kontakt nicht uniform aus der Knotenmenge gewählt, sondern gemäß der inversen Potenzverteilung. Im Folgenden wird die Analyse auf 2-dimensionale Modelle beschränkt:

**Definition 8** Sei  $n \in \mathbb{N}$  und  $G = (V, E = E_l \cup E_e)$  ein gerichteter Graph mit

$$V = \{(i, j) : i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, n\}\}$$

Wir nennen

$$d((i, j), (k, l)) = |k - i| + |l - j|$$

den Gitterabstand zwischen zwei Knoten  $(i, j)$  und  $(k, l)$ .

Sei  $r \in \mathbb{N}$  konstant und  $u, v \in V$ . Die Menge  $E_l$  von Kanten zu lokalen Kontakten sei

$$E_l = \{(u, v) : d(u, v) \leq r\}$$

Sei  $q \in \mathbb{N}$  konstant,  $a \in \mathbb{R}_{\geq 0}$  und  $u, v \in V$ .

Für jeden Knoten  $u$  werden der Menge  $E_e$  von Kanten zu entfernten Kontakten genau  $q$  zufällige Kanten  $(u, v)$  hinzugefügt, wobei

$$Pr((u, v)) = [d(u, v)]^{-a} / \sum_{w \in V} [d(u, w)]^{-a}$$

Diese Wahrscheinlichkeitsverteilung heißt „inverse  $a$ -te Potenzverteilung“.

Abbildung 4 zeigt beispielhaft den Aufbau eines Graphen nach dem Modell von Kleinberg. Betrachtet man  $q$  und  $r$  als konstant, so ergibt sich eine Familie von Netzwerkmodellen mit Parameter  $a$ . Es ergeben sich folgende Beobachtungen:

- Für  $a = 0$  entspricht die inverse Potenzverteilung der uniformen Wahrscheinlichkeitsverteilung, wie im Modell von Watts und Strogatz.
- Mit steigendem  $a$  werden Kanten zu entfernten Kontakten immer wahrscheinlicher in der näheren Umgebung eines Knotens ausgebildet.
- Mit steigendem  $a$  werden also die Abkürzungen, die durch die entfernten Kontakte entstehen, ineffektiver.

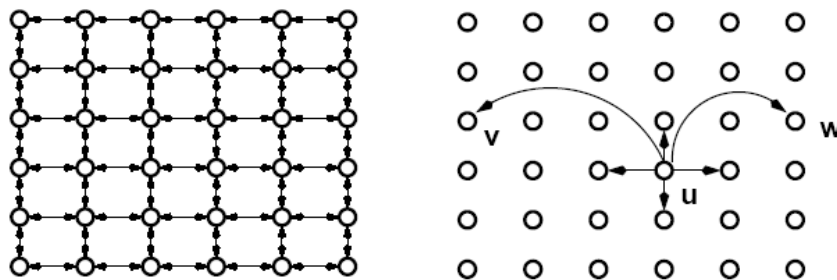
(a) Grundgerüst mit  $n = 6$  und  $r = 1$ (b) Kontakte eines Knoten  $u$  mit  $r = 1$  und  $q = 2$ 

Abbildung 4: Modell von Kleinberg, aus [5]

Es ist zu bemerken, dass sowohl das Modell von Watts und Strogatz als auch das Modell von Kleinberg für  $r = 1$  keine hohe lokale Dichte aufweisen ( $C(G) = 0$ ). Allerdings kann die Eigenschaft der lokal hohen Dichte auch so interpretiert werden, dass Individuen vor allem Kontakte in ihrem geographisch lokalen Umfeld haben.

Kommen wir nun zur entscheidenden Frage: Kann ein dezentraler Algorithmus in diesem Modell die Wege zwischen beliebigen Knoten in polylogarithmisch beschränkter Zeit finden?

**Theorem 4** Seien  $a = 2$  und  $r = q = 1$ .

Es gibt einen dezentralen Algorithmus  $\mathcal{A}$  und eine von  $n$  unabhängige Konstante  $c_2$ , so dass die erwartete Lieferzeit  $EX$  von  $\mathcal{A}$  maximal  $c_2 (\log n)^2$  beträgt.

**BEWEIS** Wir betrachten wir hier wieder den Algorithmus  $\mathcal{A}$ , der in jedem Schritt denjenigen Kontakt als Empfänger der Nachricht auswählt, der die Nachricht dem Zielknoten bezüglich des Gitterabstands am nächsten bringt.

Zunächst ist festzustellen, dass der maximale Abstand zwischen Startknoten  $s$  und Zielknoten  $t$   $2n - 2$  beträgt. Daher können wir die Ausführung des Algorithmus in Phasen unterteilen.  $\mathcal{A}$  befinde sich in Phase  $j$ , wenn  $2^j \leq d(u, t) \leq 2^{j+1}$ , der Algorithmus endet also nach höchstens  $1 + \log n$  Phasen. Da bei der Ausführung von  $\mathcal{A}$  der Abstand des Knotens, der die Nachricht momentan besitzt, zum Zielknoten streng monoton sinkt, kann angenommen werden, dass der entfernte Kontakt eines Knotens erst dann erzeugt wird, wenn die Nachricht den Knoten erreicht.

Im Folgenden wird gezeigt, dass für jede Phase  $j$  die erwartete Anzahl an Schritten  $EX_j$  in Phase  $j$  ist logarithmisch beschränkt. Da es höchstens  $1 + \log n$  Phasen gibt, folgt die Aussage.

Sei nun  $\log(\log n) \leq j < \log n$  und  $u$  der aktuelle Knoten, der die Nachricht besitzt. Es stellt sich die Frage, mit welcher Wahrscheinlichkeit Phase  $j$  im aktuellen Schritt endet. Dazu muss die Nachricht einen Knoten  $v \in B_j$  erreichen, wobei

$$B_j = \{w \in V : d(w, t) \leq 2^j\}$$

Die Wahrscheinlichkeit, dass der erzeugte entfernte Kontakt  $u$  mit einem Knoten  $v \in B_j$

verbindet, und somit Phase  $j$  beendet, ist

$$Pr((u, v) : v \in B_j) = |B_j| \cdot d(u, v)^{-2} / \sum_{u \neq w \in V} d(u, w)^{-2}$$

Wir wollen diese Wahrscheinlichkeit nun nach unten abschätzen.

- Abschätzung des Normalisierungsfaktors

$$\begin{aligned} \sum_{u \neq w \in V} d(u, w)^{-2} &\leq \sum_{i=1}^{2n-2} (4i) (i^{-2}) \\ &= 4 \sum_{i=1}^{2n-2} i^{-1} \\ &\stackrel{\text{harmonische Reihe}}{\leq} 4 + 4 \ln(2n-2) \\ &\leq 4 \ln(6n) \end{aligned}$$

- Abschätzung von  $|B_j|$

$$\begin{aligned} |B_j| &\geq 1 + \sum_{i=1}^{2^j} (i+1) \\ &> \sum_{i=1}^{2^j} i \\ &= 1/2 \cdot 2^j (2^j + 1) \\ &> 2^{2j-1} \end{aligned}$$

- Abschätzung von  $\frac{1}{d(u, v)^2} : v \in B_j$

$$\begin{aligned} d(u, v) &\leq d(u, t) + d(t, v) \\ &\leq \underbrace{2^{j+1}}_{\text{Phase } j} + \underbrace{2^j}_{v \in B_j} \\ &< 2^{j+2} \end{aligned}$$

Und somit

$$\frac{1}{d(u, v)^2} \geq \frac{1}{2^{2j+4}}$$

Für die Wahrscheinlichkeit, dass Phase  $j$  im aktuellen Schritt endet, gilt also insgesamt:

$$Pr((u, v) : v \in B_j) \geq \frac{2^{2j-1}}{4 \ln(6n) \cdot 2^{2j+4}} = \frac{1}{128 \ln(6n)}$$

Bezeichne nun  $X_j$  die Anzahl Schritte in Phase  $j$ . Die Wahrscheinlichkeit, dass  $X_j \geq i$  ist, ist die Wahrscheinlichkeit, dass die Nachricht  $i-1$  mal hintereinander nicht  $B_j$  erreicht, also  $[1 - Pr((u, v) : v \in B_j)]^{i-1}$ . Für die erwartete Anzahl Schritte  $EX_j$  in Phase

$j$  gilt somit:

$$\begin{aligned}
 EX_j &= \sum_{i=1}^{\infty} Pr[X_j \geq i] \\
 &\leq \sum_{i=1}^{\infty} \left(1 - \frac{1}{128 \ln(6n)}\right)^{i-1} \\
 &\stackrel{\text{geometrische Reihe}}{=} 128 \ln(6n)
 \end{aligned}$$

Analog kann gezeigt werden, dass die Aussage  $EX_j \leq 128 \ln(6n)$  auch für  $j = \log n$  erfüllt ist. Für  $j < \log(\log n)$  gilt trivialerweise  $EX_j \leq 128 \ln(6n)$ , da in diesem Fall höchstens  $\log n$  Schritte gebraucht werden, um  $t$  zu erreichen, selbst wenn nur lokale Kontakte benutzt werden.

Sei  $X$  die gesamte Anzahl aller Schritte des Algorithmus  $\mathcal{A}$ . Da es maximal  $1 + \log n$  Phasen gibt, gilt:

$$X \leq \sum_{j=0}^{\log n} X_j$$

und mit der Linearität von Erwartungswerten:

$$EX \leq (1 + \log n) (128 \ln(6n)) \leq c_2 (\log n)^2$$

für eine geeignet gewählte Konstante  $c_2$ .

□

Der einfache dezentrale Algorithmus  $\mathcal{A}$  ist also in der Lage, die Wege zwischen zwei Knoten in polylogarithmisch beschränkter Zeit zu finden. Die Erklärung liegt darin, dass die inverse Potenzverteilung für  $a = 2$  eine bestimmte Eigenschaft erfüllt: Es werden nicht entfernte Kontakte ausgebildet, die uniform über die Knotenmenge, sondern uniform in den Größenordnungen des Abstands verteilt sind. D.h. ein Knoten bildet mit gleicher Wahrscheinlichkeit Kanten zu Kontakten mit Abstand 2-4, 4-8, 8-16, usw. aus, wie Abbildung 5 schematisch darstellt. Aufgrund dieser Eigenschaft ist die Wahrscheinlichkeit hoch, dass die Nachricht dem Zielknoten in wenigen Schritten mindestens eine Größenordnung näher kommt, und zwar unabhängig vom tatsächlichen Abstand zu diesem.

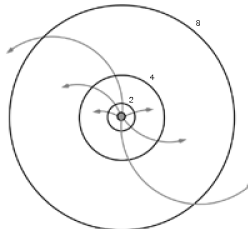


Abbildung 5: Uniforme Verteilung der entfernten Kontakte über Größenordnungen, aus [6]

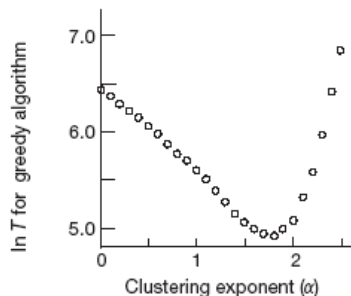


Abbildung 6: Experimentelle Laufzeit von  $\mathcal{A}$  in Abhängigkeit des Modellparameters  $a$ , aus [4]

Da diese uniforme Verteilung in den Größenordnungen nur bei  $a = 2$  auftritt, ist dies formal sogar das einzige Modell der Familie, das einem dezentralen Algorithmus ermöglicht, die Wege in polylogarithmisch beschränkter Zeit zu finden. Für höhere Dimensionen  $d$  muss nach [5] dazu  $a \stackrel{!}{=} d$  gelten. Experimentelle Ergebnisse zeigen aber, dass der dezentrale Algorithmus auch in einem kleinen Bereich um  $a = d$  effizient arbeitet. Abbildung 6 zeigt das Resultat einer Simulation von  $\mathcal{A}$  auf einem 2-dimensionalen Gitter mit  $n = 20000$ . Dabei ist jeder Datenpunkt der Durchschnitt von 1000 Durchläufen.

## 5 Zusammenfassung

Viele reale Netzwerke, insbesondere soziale, weisen das Phänomen der kleinen Welt auf - Es existiert eine hohe Wahrscheinlichkeit, dass kurze Wege zwischen zwei beliebigen Knoten existieren. Das Modell von Watts und Strogatz erfasst die strukturellen Eigenschaften von kleinen Welten, indem ein regulärer, lokal dichter Graph mit einem zufälligen Graphen kombiniert wird, so dass Abkürzungen innerhalb des Netzwerks entstehen.

Neben dieser strukturellen Eigenschaft besitzen diese Netzwerke aber auch die „algorithmische“ Eigenschaft, dass Individuen in kleinen Welten in der Lage sind, diese kurzen Wege nur mit lokalen Informationen schnell zu finden. Es wurde gezeigt, dass dezentrale Suchalgorithmen in bisherigen Modellen für kleine Welten, wie dem Modell von Watts und Strogatz, dazu nicht fähig sind. Das Modell von Kleinberg löst dieses Problem, indem die erzeugten Kontakte der Individuen mit der dem Modell unterliegenden Geometrie korreliert sind: Die Endknoten für Abkürzungen werden nicht uniform ausgewählt, sondern nach der inversen Potenzverteilung. Bei geeigneter Wahl der Parameter werden die entfernten Kontakte uniform über die Größenordnungen des Abstands gewählt. So ist ein einfacher, dezentraler Greedy-Algorithmus, in der Lage, die kurzen Wege des Netzwerks in optimaler Zeit zu finden.

## Literatur

- [1] Béla Bollobás. *Random Graphs*. Academic Press, 1985.
- [2] Béla Bollobás and Fan R. K. Chung. The diameter of a cycle plus a random matching. *SIAM J. Discrete Math*, 1(3), 1988.
- [3] Ulrik Brandes. *Skriptum zur Vorlesung Methoden der Netzwerkanalyse*. 2005.
- [4] Jon Kleinberg. Navigation in a small world. *Nature*, 406, 2000.
- [5] Jon Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proceedings of STOC*, 2000.
- [6] Jon Kleinberg. The small-world phenomenon and decentralized search. *SIAM News*, 37(3), 2004.
- [7] Stanley Milgram. The small world problem. *Psychology today*, 1(61), 1967.
- [8] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393, 1998.