Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Felix Eberhard

Grundlagen für das Mapping des Internet-Graphen

Student Thesis SA-2003.30
Summer Term 2003

Tutor: Prof. Dr. Thomas Erlebach

Supervisor:
Prof. Dr. Thomas Erlebach

31.7.2003

Institut für Technische Informatik und Kommunikationsnetze
Computer Engineering and Networks Laboratory

Zusammenfassung

Wir betrachten den Internet-Graphen auf Ebene der Autonomous Systems (AS). Um die Topologie des Internet in Erfahrung zu bringen, können Messungen an einzelnen Knoten des Netzwerks durchgeführt werden. In dieser Arbeit wird untersucht, wie viele Messungen dazu nötig sind und an welchen Knoten diese am besten gemacht werden.

Wir gehen vom idealisierten Fall aus, dass durch eine Messung an einem Knoten alle kürzesten Pfade, die an diesem Knoten beginnen, gefunden werden. In dieser Arbeit untersuchen wir zuerst, wie viele Messungen für einfache Topologien nötig sind, um den Graphen komplett zu rekonstruieren. Für allgemeine Graphen geben wir eine untere Schranke an. Auf Grundlage dieser Resultate entwickeln wir zwei Online-Algorithmen, welche eine Menge von Messknoten bestimmen. Diese Algorithmen vergleichen wir mit einer zufälligen Wahl von Messknoten und den hergeleiteten theoretischen Aussagen. Als Testgraphen verwenden wir die zuvor untersuchten einfachen Topologien, Zufallsgraphen nach Erdös-Rényi und Waxman, als auch Power-law-Topologien. Letztere sind in ihrer Struktur dem wirklichen Internet-Graphen sehr ähnlich.

Abkürzungen

AS	Autonomous System
BA	Barabási-Albert
BGP	Border Gateway Protocol
GLP	Generalized Linear Preference
ISP	Internet Service Provider
LT	Schichtengraph (Layered Tree)
MST	Minimaler Spannbaum
O.B.d.A	Ohne Beschränkung der Allgemeinheit

Inhaltsverzeichnis

Zusammenfassung **i**

Abkürzungen **iii**

Inhaltsverzeichnis **v**

Abbildungsverzeichnis **vii**

1 Einleitung **1**

2 Aufgabenstellung **3**

3 Notation **7**

4 Theoretische Ergebnisse **9**
4.1 Kreis 9
4.2 Radgraph 10
4.3 Baum 12
4.4 Allgemeine Graphen 19

5 Algorithmen **23**

6 Versuche **31**
6.1 Versuchsanordnung 31
6.2 Resultate 33

7 Schlussfolgerung **39**

8 Ausblick **41**

9 Schlussbemerkung **43**

Literaturverzeichnis **45**

Abbildungsverzeichnis

3.1 Beispiel für einen Schichtengraphen 8

4.1 Messschema für den Radgraphen 10
4.2 Fälle mit unentdeckten Kanten beim Radgraphen 11
4.3 Beispiel für Teilbäume 13
4.4 Beispiel für optimale Messungen in einem Baum 15
4.5 Induktionsschritt vom Baum B' nach B 16
4.6 Graph mit konstantem Grad 3 und $|S| = \frac{|V|}{4}$ für $|V| \geq 8$ . . . 19
4.7 Vergleich der Anzahl Messungen im Graphen und im MST . . 20

5.1 Prozedur reconstructed 24
5.2 Algorithmus MinDegree 24
5.3 Algorithmus MinMaxEdges 26
5.4 Prozedur discovered des Algorithmus MinMaxEdges 27
5.5 Beispiel zum Algorithmus MinMaxEdges 27
5.6 Beispiel für unterschiedliche Anzahl entdeckter Kanten und Nichtkanten 28

6.1 Parameter von BRITE 31
6.2 Parameter der Modelle Waxman, BA und GLP 32
6.3 Anzahl Messungen bei Baum 33
6.4 Anzahl Messungen beim Radgraphen, Erdös-Rényi ($p = 0.05$) und Waxman 34
6.5 Anzahl Messungen bei BA und GLP 35
6.6 Prozentsatz erkannter Kanten und Nichtkanten in Abhängigkeit der Messung 36
6.7 Variationskoeffizienten von Erdös-Renyi ($p = 0.1$) und BA . . 37

Kapitel 1

Einleitung

Das Internet kann auf zwei verschiedenen Ebenen als Graph aufgefasst werden: Einerseits auf Router-Ebene und andererseits auf der Ebene der Autonomous Systems (AS). Ein AS besteht aus einem oder mehreren Netzwerken, welche typischerweise von einer Institution (ISP, Unternehmung, Universität, ...) administriert werden. Benachbarte AS benutzen das Border Gateway Protocol (BGP), um Informationen über Pfade auszutauschen [4].

Die Router resp. AS bilden die Knotenmenge des Internet-Graphen. Direkte Kommunikation zwischen zwei Routern / AS wird durch eine Kante modelliert.

Die genaue Topologie des Internet ist sowohl auf AS- wie auch auf Router-Ebene nicht bekannt. Es wird vermutet, dass auf beiden Ebenen der Grad und die Frequenz der Knoten einem Potenzgesetz (power law) gehorchen [6] [3]. Das heisst, einigen wenigen Knoten mit sehr hohem Grad stehen sehr viele mit geringem Grad gegenüber.

Um die Topologie des Internet in Erfahrung zu bringen werden häufig an verschiedenen Knoten des Netzwerkes Messungen durchgeführt. Auf Router-Ebene wird typischerweise *traceroute* verwendet. Dadurch erhält man alle Knoten und Kanten auf dem Weg von einem Quell- zu einem Zielknoten. Auf AS-Ebene kann ebenfalls *traceroute* benutzt werden, es ist aber auch möglich, einzelne BGP Routing Tabellen auszulesen. Dadurch erhält man die Pfadinformation von einem AS.

Die Messpunkte können im allgemeinen aber nicht frei gewählt werden. Mit *traceroute* werden typischerweise Messungen von einer geringen Anzahl Quellknoten zu einer grossen Anzahl Zielknoten durchgeführt [3]. Um BGP Routing Tabellen auslesen zu können, ist die Kooperation der Administration des entsprechenden AS nötig. Die Universität von Oregon stellt beispielsweise mit dem Projekt *Route Views* [7] Routinginformation zur Verfügung.

In dieser Arbeit betrachten wir Messungen auf AS-Ebene. Wir gehen von einem idealen Szenario aus, in dem Messknoten frei wählbar sind und eine Messung alle kürzesten Pfade ausgehend von einem Messknoten zurückliefert.

Die Bestimmung einer möglichst geringen Anzahl Messpunkte, welche die Topologie des Graphen aufdeckt, ist ein Online-Problem. Beim entsprechenden Offline-Problem ist die Topologie bereits zu Beginn bekannt, und es soll wiederum eine minimale Anzahl Messpunkte bestimmt werden, so dass der Graph komplett rekonstruiert wird.

Im Kapitel 4 werden wir für einfache Topologien die Anzahl der benötigten Messungen für das Offline-Problem herleiten, sowie für allgemeine Graphen Schranken beweisen. Danach werden in Kapitel 6 verschiedene Online-Algorithmen vorgeschlagen, die jeweils eine Menge von Messknoten bestimmen. Diese Algorithmen werden dann auf verschiedenen Topologien getestet.

Kapitel 2

Aufgabenstellung

Aufgabenstellung für die Semesterarbeit

„Grundlagen für das Mapping des Internet-Graphen“

Institut TIK, D-ITET, ETH Zürich
Sommersemester 2003

Student:	Felix Eberhard
Verantwortlicher Professor:	Thomas Erlebach
Betreuung:	Thomas Erlebach

Thematischer Hintergrund

Aus verschiedenen Gründen besteht grosses Interesse daran, die genaue Topologie des Internet in Erfahrung zu bringen und als Graph zu repräsentieren. Eine häufig angewandte Methode dazu besteht darin, von einzelnen Netzwerk-Knoten aus Messungen zu machen, die die Links entlang der Wege von diesem Knoten zu allen anderen Knoten feststellen. Eine wichtige Frage dabei ist, wie viele solche Messungen nötig sind, um ein möglichst wirklichkeitsgetreues Abbild des realen Netzes zu bekommen. Die Erkundung des Internet kann dabei auf zwei Ebenen erfolgen: auf der Router-Ebene, wobei traceroute-basierte Tools verwendet werden, und auf der Ebene der *Autonomous Systems*, wobei die Routing-Tabellen mehrerer zentraler BGP-Router ausgelesen werden.

Problemstellung

In dieser Arbeit soll die Fragestellung der Erkundung einer unbekannten Netzwerktopologie durch Messungen systematisch untersucht werden. Dabei sind einerseits theoretische Überlegungen zur Machbarkeit solcher Untersuchungen und zur Entwicklung geeigneter Algorithmen erwünscht, anderer-

seits sind Implementierungen und Experimente mit realen und synthetischen Internet-Graphen gefragt. Es geht dabei nicht um tatsächliche Messungen des Internets, sondern um Verfahren, mit denen Messungen geplant und ausgewertet werden können.

Wir betrachten das Netz als ungerichteten Graphen. Durch Messungen bei einem Knoten v können die Pfade von dem Knoten zu allen anderen Knoten in Erfahrung gebracht werden, entweder die Pfade in einem Kürzeste-Wege-Baum oder sogar alle Pfade auf kürzesten Wegen, die bei dem Knoten beginnen. In beiden Modellen können die folgenden Probleme betrachtet werden:

- Wie viele Messungen sind bei verschiedenen Graph-Typen nötig, um den Graph gut zu rekonstruieren?
- Wie hängt die Anzahl gefundener Kanten bzw. Nicht-Kanten von der Anzahl der Messungen ab?
- Wie sehen der Worst-Case und der Best-Case aus?
- Nach welchen Kriterien sollten die Knoten ausgewählt werden, um die Messungen zu machen?
- Wie vergleicht sich eine optimale Wahl der Knoten mit einer zufälligen Wahl?
- Welche Beobachtungen lassen sich bei zufällig generierten Internet-artigen Graphen machen?
- Wenn bereits einige Messungen gemacht wurden, wie kann dann für einen verbleibenden Knoten abgeschätzt werden, wie viele Kanten oder Nicht-Kanten man bei einer Messung bei diesem Knoten mindestens / höchstens entdecken wird?
- Wie verändert sich die Situation, wenn nicht beliebige Pfade zugelassen sind, sondern nur Pfade, die aufgrund der Geschäftsbeziehungen zwischen Autonomous Systems erlaubt sind?

Teilaufgaben

Im Rahmen der Semesterarbeit sind im wesentlichen folgende Teilaufgaben zu bewerkstelligen:

1. Arbeiten Sie sich in die Thematik ein.
2. Betrachten Sie die oben erwähnten Problemstellungen und erarbeiten Sie theoretische Aussagen und Vorschläge für Algorithmen.

3. Implementieren Sie geeignete Algorithmen und führen Sie Experimente mit verschiedenen Graphtypen (reale Internet-Graphen, synthetische AS-Graphen, Zufallsgraphen, etc.) durch.
4. Werten Sie die experimentellen Ergebnisse aus.
5. Erstellen Sie einen schriftlichen Bericht über Ihre Arbeit, der gut strukturiert und gut lesbar ist.
6. Präsentieren Sie die Ergebnisse Ihrer Arbeit in einem Vortrag in der letzten Semesterwoche.

Viel Erfolg!

Literaturverzeichnis

[1] Barford-Betavros-Byers-Crovella/01 P. Barford, A. Bestavros, J. Byers, and M. Crovella. On the marginal utility of deploying measurement infrastructure. In *Proceedings of the ACM SIGCOMM Internet Measurement Workshop 2001*, November 2001.

[2] Lakhina-Byers-Crovella-Xie/03 A. Lakhina, J. Byers, M. Crovella, and P. Xie. Sampling biases in IP topology measurements. In *Proceedings of INFOCOM'03*, 2003.

[3] Shankar/02 L. S. Ram. Tree-based graph reconstruction, 2002. CGC Pre-Doc Project, ETH Zürich.

[4] Subramanian-Agarwal-Rexford-Katz/02 L. Subramanian, S. Agarwal, J. Rexford, and R. Katz. Characterizing the Internet hierarchy from multiple vantage points. In *Proceedings of INFOCOM'02*, June 2002.

Kapitel 3

Notation

Wir betrachten ein Netzwerk als ungerichteten, ungewichteten und zusammenhängenden Graphen $G = (V, E)$.

Definition 3.1 $\deg(v)$ *bezeichnet den Grad des Knotens v, d.h. die Anzahl aller zu v inzidenten Kanten.*

Definition 3.2 *$dist(u, v)$ bezeichnet die Länge des kürzesten Weges zwischen den Knoten u und v im Graph G. Wenn mit mehreren Graphen argumentiert wird, so ist $dist_G(u, v)$ die Distanz von u nach v in G. $dist(u, v)$ werden wir auch den Abstand von u zu v nennen.*

Definition 3.3 *Der Durchmesser $diam(G)$ eines Graphen bezeichnet den längsten kürzesten Pfad zwischen zwei beliebigen Knoten $u, v \in G$.*

Definition 3.4 *Ein Schichtengraph (layered Tree) mit Wurzel v, abgekürzt $LT(v)$, bezeichnet den Kürzeste-Wege-Graphen, der alle kürzesten Wege von v zu allen anderen Knoten enthält.*

Es sei darauf hingewiesen, dass der Schichtengraph im Normalfall kein Baum ist. Eine Kante $e = \{t, u\} \notin LT(v)$ genau dann wenn t und u den selben Abstand von v haben. Wir sagen auch, dass sich t und u auf dem selben Level befinden.

In Abbildung 3.1 ist die Kante $\{a, c\} \notin LT(b)$, da a und c von b den Abstand eins haben.

Definition 3.5 *Eine Messung an einem Knoten v entdeckt eine Kante e genau dann wenn $e \in LT(v)$. v wird Messknoten genannt. Die Menge aller Messungen bezeichnen wir als S.*

Messknoten werden in Grafiken durch einen Rahmen gekennzeichnet, wie beim Knoten b in Abbildung 3.1 ersichtlich ist.

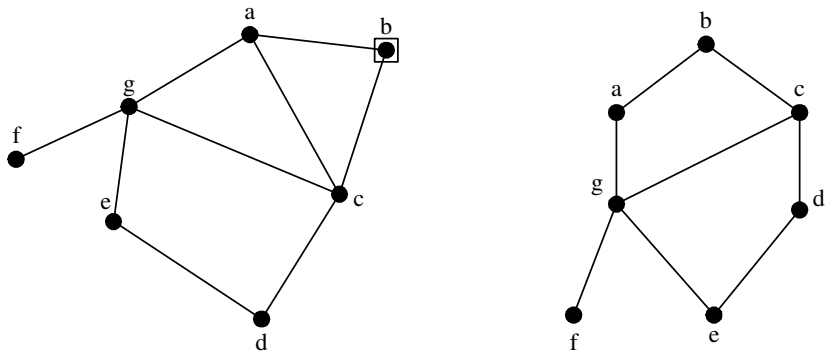


Abbildung 3.1: Graph G (links) und der Schichtengraph LT(b)

Definition 3.6 *Eine Nichtkante zwischen zwei Knoten v und w bezeichnet die Tatsache, dass v und w nicht durch eine Kante verbunden sein können.*

Die Information, dass eine Kante nicht vorhanden sein kann, ist genauso viel wert, wie die Information über das Vorhandensein einer Kante.

Wir werden sagen, dass eine Kante von einer Messung nicht entdeckt wird, auch wenn nicht klar ist, ob es sich um eine Kante oder um eine Nichtkante handelt.

In Abbildung 3.1 gibt es beispielsweise eine Nichtkante zwischen a und d, die bei der Messung an b entdeckt wird. Denn es gilt $dist(b, a) \neq dist(b, d)$ und zwischen a und d ist keine Kante vorhanden.

Kapitel 4

Theoretische Ergebnisse

In diesem Kapitel betrachten wir zuerst die einfachen Netztopologien Kreis, Radgraph und Baum. Wir geben jeweils die genaue Anzahl der Messungen an, die nötig sind um jede Kante resp. Nichtkante zu entdecken. Danach werden Schranken für allgemeine Graphen hergeleitet. In diesem Kapitel beschäftigen wir uns nur mit der vollständigen Rekonstruktion eines Graphen und nicht mit Approximationen.

Die Anzahl Messungen für den vollständigen Graphen wird in [1] bewiesen und beträgt $n-1$ für K_n.

4.1 Kreis

Als erstes betrachten wir Kreise. Es stellt sich heraus, dass das Problem für diesen Fall sehr einfach zu lösen ist.

Satz 4.1 *Sei der unterliegende Graph G ein Kreis mit $|V| \geq 3$. Es werden genau zwei Messungen benötigt, um alle Kanten und Nichtkanten zu entdecken.*

Beweis. Weniger als zwei Messungen sind nicht ausreichend: Nach einer Messung am Knoten $v \in V$ beträgt im $LT(v)$ der Abstand von v zu seinen beiden Nachbarn jeweils 1. Eine Kante zwischen diesen zwei Knoten wird somit nicht erkannt.

Zwei Messungen sind ausreichend: Wir betrachten zunächst den Fall, dass G eine ungerade Anzahl Kanten enthält. Für beliebige zwei Knoten t und u, $t \neq u$, gibt es nur einen Knoten v, mit $dist(t,v) = dist(u,v)$. Finden die zwei Messungen an verschiedenen Knoten statt, so können sich t und u höchstens einmal im gleichen Abstand vom Messknoten befinden. Betrachten wir nun den Fall, dass G eine gerade Anzahl Kanten enthält und eine Messung am Knoten v durchgeführt wurde. Es existiert ein Knoten v', $v \neq v'$, so dass für zwei beliebige Knoten t und u, $t \neq u$, mit $dist(t,v) = dist(u,v)$

gilt, dass $dist(t, v') = dist(u, v')$. v' ist der Knoten mit dem grösstem Abstand zu v. Wird die zweite Messung an $w \in V \backslash \{v, v'\}$ durchgeführt, so ist offensichtlich, dass es keine zwei Knoten geben kann, die von v und von w den gleichen Abstand haben. ■

4.2 Radgraph

Der Radgraph ist folgendermassen definiert: Sei $n = |V|$ und $G = (\{a\} \cup \{1, 2, \ldots, (n-1)\}, \{(i, i \bmod (n-1)+1) : i = 1, 2, \ldots, (n-1)\} \cup \{(a, i) : i = 1, 2, \ldots, (n-1)\})$. Oder informell formuliert: a ist ein Zentrumsknoten, der mit allen anderen Knoten verbunden ist. Diese äusseren Knoten werden von 1 bis $(n-1)$ nummeriert und jeweils mit den beiden Nachbarn verbunden.

Ein Radgraph mit $|V| = 4$ entspricht dem kompletten Graphen. Daraus folgt, dass die Anzahl Messungen drei beträgt. Die Graphen mit $|V| \in \{5, 6, 7\}$ sind Spezialfälle, auf die hier nicht weiter eingegangen wird. Die Anzahl der benötigten Messungen beträgt zwei für $|V| \in \{5, 6\}$, sowie drei für $|V| = 7$. Wir schauen uns nun Radgraphen mit mindestens acht Knoten genauer an.

Satz 4.2 *Sei der unterliegende Graph G ein Radgraph mit $|V| \geq 8$. Die Anzahl der benötigten Messungen beträgt* $\left\lceil \frac{2}{5}(|V| - 2) \right\rceil$.

Beweis. Die Menge der Messungen ist $S = \{1, 5, 7, 10, 12, 15, \ldots\}$. Falls gilt $n \bmod 5 = 0$ muss noch zusätzlich am Knoten $n-1$ gemessen werden. Es sei darauf hingewiesen, dass sich zwischen den ersten beiden Messknoten drei Knoten befinden, an denen keine Messung stattfindet. Sonst befinden sich immer abwechslungsweise ein resp. zwei Knoten ohne Messung zwischen zwei Messknoten.

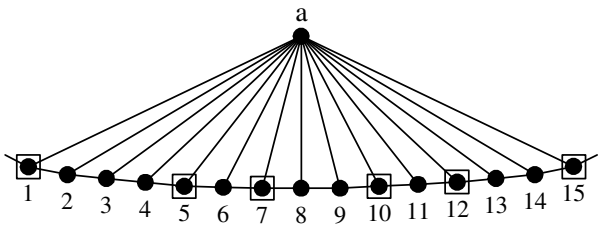


Abbildung 4.1: Messschema für den Radgraphen

Als erstes werden wir nun beweisen, dass weniger Messungen nicht ausreichend sind. Bei den folgenden drei Fällen (vgl. Abbildung 4.2) wird jeweils mindestens eine Kante nicht erkannt. Diese Fälle dürfen deshalb nicht auftreten.

- Es existieren vier paarweise disjunkte Knoten $v_1, v_2, v_3, v_4 \in \{1, \ldots, n-1\}$ an denen keine Messung durchgeführt wird. Und es existieren drei Kanten e_1, e_2, e_3 mit $e_1 = \{v_1, v_2\}, e_2 = \{v_2, v_3\}$ und $e_3 = \{v_3, v_4\}$. Die Kante e_2 wird von keiner Messung erkannt, da v_2 und v_3 zum Zentrumsknoten a adjazent sind und von allen anderen möglichen Messknoten den Abstand zwei haben.

- Es existieren fünf paarweise disjunkte Knoten $v_1, v_2, v_3, v_4, v_5 \in \{1, \ldots, n-1\}$, wobei nur am Knoten v_3 eine Messung durchgeführt wird. Und es existieren vier Kanten e_1, e_2, e_3, e_4 mit $e_1 = \{v_1, v_2\}, e_2 = \{v_2, v_3\}, e_3 = \{v_3, v_4\}$ und $e_4 = \{v_4, v_5\}$. Eine Kante zwischen v_2 und v_4 wird nicht erkannt, da v_2 und v_4 zu v_3 und zu a adjazent sind und von allen anderen möglichen Messknoten den Abstand zwei haben.

- Es existieren sechs paarweise disjunkte Knoten $v_1, v_2, v_3, w_1, w_2, w_3 \in \{1, \ldots, n-1\}$ an denen keine Messung durchgeführt wird. Und es existieren vier Kanten e_1, e_2, e_3, e_4 mit $e_1 = \{v_1, v_2\}, e_2 = \{v_2, v_3\}, e_3 = \{w_1, w_2\}$ und $e_4 = \{w_2, w_3\}$. Eine Kante zwischen v_2 und w_2 wird nicht erkannt, da v_2 und w_2 zu a adjazent sind und von allen anderen möglichen Messknoten den Abstand zwei haben.

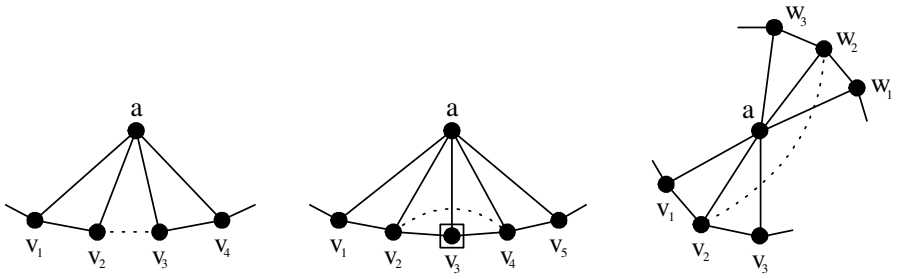


Abbildung 4.2: Fälle mit unentdeckten Kanten beim Radgraphen

Das vorgeschlagene Messschema ist die Anordnung mit den wenigsten Messknoten, so dass keiner dieser drei Fälle auftritt. Daraus folgt, dass wenn weniger Messungen durchgeführt werden, nicht alle Kanten erkannt werden können.

Als nächstes werden wir beweisen, dass diese Menge S von $\left\lceil \frac{2}{5}(|V| - 2) \right\rceil$ Messungen auch ausreichen, um alle Kanten und Nichtkanten zu erkennen. Dazu werden die zwei folgenden Lemmata benötigt:

Lemma 4.1 *Für einen Knoten $v \in V \backslash \{a\} \backslash S$ existiert ein Messknoten $m \in S$, so dass gilt $dist(m, v) = 2$.*

Beweis. Der Knoten a ist adjazent zu allen anderen Knoten, und somit auch zu allen Messknoten. Aus Satz 4.2 folgt, dass die Anzahl der Messungen für jeden Radgraphen mit $|V| \geq 8$ mindestens drei beträgt. v ist adjazent zu genau drei Knoten, einer davon ist a, welcher kein Messknoten ist. Also ist v zu mindestens einem Messknoten nicht adjazent. ■

Lemma 4.2 *Für jeden Knoten $v \in V \backslash \{3\} \backslash S$ existiert ein Messknoten $m \in S$, so dass v zu m adjazent ist.*

Beweis. Die Korrektheit dieses Lemmas folgt direkt aus der Anordnung der Messpunkte. ■

Nun sind wir bereit um den Beweis fortzusetzen: Da Kanten und Nichtkanten inzident zu einem Messknoten immer erkannt werden, genügt es zu beweisen, dass eine Kante zwischen zwei beliebigen Knoten v und w, an denen keine Messung durchgeführt wird, erkannt wird. Wir unterscheiden die Fälle $v = a$, $v = 3$ und $v \in V \backslash \{a, 3\} \backslash S$.

- $v = a, w \in V \backslash \{v\} \backslash S$. Aus Lemma 4.1 folgt, dass ein Messknoten $m \in S$ existiert, so dass gilt $dist(v, m) = 1$ und $dist(w, m) = 2$. Dadurch wird eine Kante zwischen v und w erkannt.

- $v = 3, w \in V \backslash \{v\} \backslash S$. Aus Lemma 4.2 folgt, dass ein Messknoten $m \in S$ existiert, so dass gilt $dist(v, m) = 2$ und $dist(w, m) = 1$. Dadurch wird eine Kante zwischen v und w erkannt.

- $v \in V \backslash \{a, 3\} \backslash S, w \in V \backslash \{a, 3, v\} \backslash S$. Aus Lemma 4.2 folgt, dass es einen zu v adjazenten Messknoten m gibt. m wiederum ist adjazent zu a und einem Knoten $u \neq v$. Falls $w \neq u$ ist, so gilt dass $dist(m, v) = 1$ und $dist(m, w) = 2$, da m zu w nicht adjazent sein kann. Falls $w = u$ ist folgt aus der Anordnung der Messungen dass ein Messknoten $m' \neq m$ existiert, der entweder zu v oder zu w adjazent ist. Daraus folgt, dass $dist(m', v) \neq dist(m', w)$.

Damit ist Satz 4.2 vollständig bewiesen. ■

4.3 Baum

Als nächstes betrachten wir den Fall, dass die Topologie des Netzwerkes ein Baum ist. Um die Anzahl der Messungen anzugeben, führen wir zuerst die Notation eines Teilbaumes ein.

Definition 4.3 *Gegeben sei ein Baum $G = (V, E)$. Ein Teilbaum $T = (V', E')$ ist ein Subgraph von G mit $V' \subseteq V$ und $E' \subseteq E$. Der Teilbaum $T(v, \{v, w\})$ mit $w, v \in V, w$ adjazent zu v, bezeichnet den Baum mit Wurzel*

v, den man erhält, wenn man im Baum G alle zu v inzidenten Kanten bis auf $\{v, w\}$ weglässt. Die Menge aller Teilbäume von v, d.h. mit v als Wurzel, wird $T(v)$ genannt. Es gilt $|T(v)| = \deg(v)$.

Definition 4.4 *Die Menge aller Teilbäume von v, die einen Knoten u mit $\deg(u) \geq 3$ enthalten, heisst $T_{\geq 3}(v)$. Analog gilt $T_{<3}(v) = T(v) - T_{\geq 3}(v)$.*

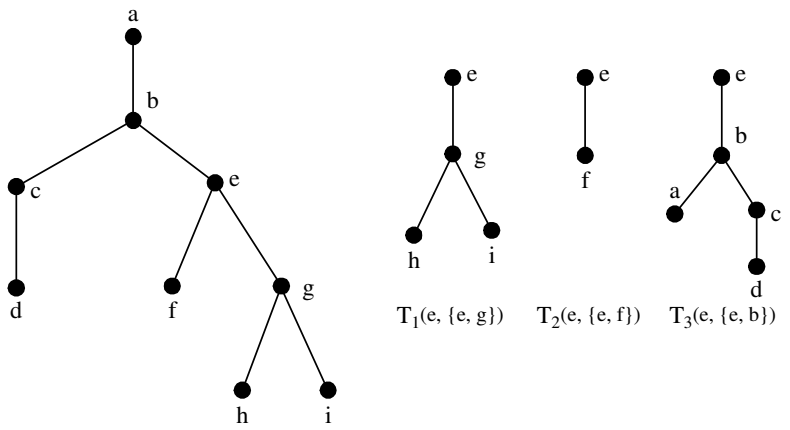


Abbildung 4.3: Beispielbaum (links) und die drei Teilbäume vom Knoten e (rechts)

In der Abbildung 4.3 besteht $T(e)$ aus T_1, T_2 und T_3. $T_{\geq 3}(e) = \{T_1, T_3\}$ und $T_{<3}(e) = \{T_2\}$, da es in T_2 keinen Knoten vom Grad ≥ 3 gibt. Als nächstes benötigen wir mehrere einfache Aussagen:

Lemma 4.5 *Eine Messung an einem Knoten $u \in T(v, \{v, w\})$ entdeckt eine Kante zwischen s und t, $s \neq t$ und $s, t \in V \backslash T(v, \{v, w\})$ genau dann, wenn die Kante von einer Messung an v entdeckt wird.*

Beweis. Der kürzeste Weg von u zu s muss über v führen, da v die einzige Verbindung des Teilbaumes $T(v, \{v, w\})$ zum Rest des Graphen ist. Dasselbe gilt für den kürzesten Weg von u zu t. Daraus ergibt sich

$$\begin{aligned} |dist(u, s) - dist(u, t)| &= |(dist(u, v) + dist(v, s)) - (dist(u, v) + dist(v, t))| \\ &= |dist(v, s) - dist(v, t)| \end{aligned}$$

■

Lemma 4.6 *Gegeben sei eine Menge von Knoten $v_1, v_2, \ldots, v_i \in V$, wobei jeweils v_k und $v_{k+1}, k = 1, \ldots, i-1$ durch eine Kante verbunden sind. Sei $\deg(v_1) = 1$, $\deg(v_i) \geq 1$ und $\deg(v_k) = 2, k = 2, \ldots, i-1$. Eine Messung an v_1 erkennt alle Kanten und Nichtkanten von $v_1, v_2, \ldots, v_i$.*

Beweis. Die Distanz von v_1 zu v_j beträgt $j-1$. Daraus folgt $dist(v_1, v_j) \neq dist(v_1, v_k)$ für $j \neq k$. Für einen Knoten $w \in V \backslash \{v_1, \ldots, v_i\}$ gilt $dist(v_1, v_i) < dist(v_1, w)$. ■

Lemma 4.6 besagt, dass in einer Kette mit mindestens einem Blatt mit einer Messung alle Kanten und Nichtkanten von Knoten in dieser Kette erkannt werden.

Lemma 4.7 *Für einen Knoten $v \in V$ werden* $\deg(v)-1$ *Messungen benötigt, um jede Kante zwischen zwei beliebigen zu v adjazenten Knoten zu erkennen.*

Beweis. Werden weniger Messungen durchgeführt, so gibt es mindestens zwei Teilbäume $T_1(v, \{v, w_1\})$ und $T_2(v, \{v, w_2\})$, die keinen Messknoten $m \neq v$ enthalten. Aus Lemma 4.5 folgt nun, dass für jeden Messknoten m' gilt, dass $dist(m', w_1) = dist(m', w_2)$. Eine Kante zwischen w_1 und w_2 wird somit nicht erkannt.

Um zu zeigen, dass $\deg(v) - 1$ Messungen genügen, werden die Messungen so ausgeführt, dass bis auf einen alle Teilbäume von v einen Messknoten $m \neq v$ enthalten. Für beliebige zwei zu v adjazente Knoten w_1 und w_2, $w_1 \neq w_2$ gilt nun, dass mindestens einer der Teilbäume $T_1(v, \{v, w_1\})$ und $T_2(v, \{v, w_2\})$ einen Messknoten $m' \neq v$ enthält. O.B.d.A enthalte T_1 den Messknoten m'. Die Distanz von m' zu w_1 ist offensichtlich kleiner als die Distanz von m' zu w_2. Das Lemma folgt. ■

Aus Lemma 4.7 folgt damit, dass für einen Stern mit Zentrumsknoten x $\deg(x) - 1$ Messungen nötig sind um Sicherheit über alle Kanten und Nichtkanten zu erlangen. Der maximale Grad $\deg_{max}$ in einem Graphen liefert eine erste untere Schranke für die Anzahl Messungen in einem Baum. Es müssen mindestens $\deg_{max} - 1$ Messungen durchgeführt werden. Im folgenden Satz wird die Anzahl Messungen für einen Baum exakt bestimmt.

Satz 4.3 *Sei der unterliegende Graph G ein Baum. Falls kein Knoten vom Grad ≥ 3 existiert genügt eine Messung, ansonsten beträgt die Anzahl der benötigten Messungen*

$$\sum_{v:\deg(v)\geq 3} \max(0, |T_{<3}(v)| - 1)$$

Beweis. Die Messungen werden so durchgeführt, dass für jeden Knoten v, $\deg(v) \geq 3$, gilt, dass nur in einem seiner $\deg(v)$ Teilbäume kein Messknoten vorhanden ist. Dazu wird an $|T_{<3}(v)| - 1$ Blättern von Teilbäumen $\in T_{<3}(v)$ gemessen. Es sei darauf hingewiesen, dass es sich bei einem Teilbaum $\in T_{<3}(v)$ um eine Kette handelt. Abbildung 4.4 zeigt an einem Beispiel, wie eine optimale Messung aussehen kann.

Für den Fall, dass kein Knoten vom Grad ≥ 3 existiert, ist es einfach zu sehen, dass mit der ersten Messung alle Kanten und Nichtkanten erkannt

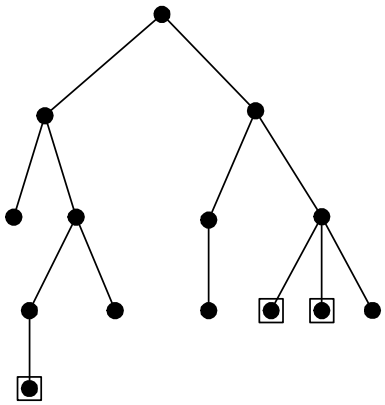


Abbildung 4.4: Beispiel für optimale Messungen in einem Baum

werden. Der Baum ist dann nämlich eine Kette und wenn an einem der beiden Blätter gemessen wird, so können keine zwei unterschiedlichen Knoten den gleichen Abstand zum Messknoten haben.

Wir beweisen für den Fall, dass es mindestens einen Knoten vom Grad ≥ 3 gibt, dass die Anzahl der Messungen ausreichend ist. Der Beweis erfolgt per Induktion über die Anzahl der Knoten vom Grad ≥ 3. Sei n die Anzahl Knoten vom Grad ≥ 3.

- Induktionsverankerung: $n = 1$. Es gibt nur einen Knoten x mit $\deg(x) \geq 3$. Da keiner der $\deg(x)$ Teilbäume von x einen Knoten vom Grad mindestens drei enthält, folgt daraus

  $$\sum_{v:\deg(v)\geq 3} \max(0, |T_{<3}(v)| - 1) = \deg(x) - 1$$

  Die Teilbäume $T_i(x, \{x, w_i\})$ sind Ketten mit Wurzel x und je einem Blatt b_i, für $i = 1, \ldots, \deg(x)$. Es wird nun an jedem ausser einem dieser Blätter gemessen, insgesamt also $\deg(x) - 1$ Messungen. Aus Lemma 4.6 folgt, dass eine Messung an b_i alle Kanten und Nichtkanten von Knoten im Teilbaum T_i erkennt. Somit bleibt nur noch zu zeigen, dass alle Kanten und Nichtkanten zwischen zwei unterschiedlichen Knoten $t, u \in T_j \backslash \{x\}$ ebenfalls erkannt werden, wobei T_j der Teilbaum ist, an dessen Blatt b_j keine Messung durchgeführt wurde. Dies ist aber offensichtlich, da $dist(m, t) \neq dist(m, u)$ für jeden Messknoten m gilt.

- Induktionsschritt: $n \to n+1$. Gegeben sei ein Baum B mit $n+1$ Knoten vom Grad ≥ 3. Sei x ein Knoten mit $\deg(x) \geq 3$ und $|T_{\geq 3}(x)| = 1$. Das bedeutet, dass jeder Weg von x zu einem anderen Knoten vom Grad ≥ 3 über dieselbe zu x inzidente Kante führt. Sei y der Knoten mit minimaler Distanz zu x und $\deg(y) \geq 3$. Sei B' der einzige

Teilbaum $\in T_{\geq 3}(x)$. Für einen beliebigen Knoten $v \in B$ bezeichne v' den entsprechenden Knoten in B'. Abbildung 4.5 verdeutlicht die Anordnung.

Aus der Induktionsannahme folgt, dass für B' eine Messanordnung $S(B')$ mit

$$\sum_{v' \in B' : \deg(v') \geq 3} \max(0, |T_{<3}(v')| - 1)$$

Messungen existiert.

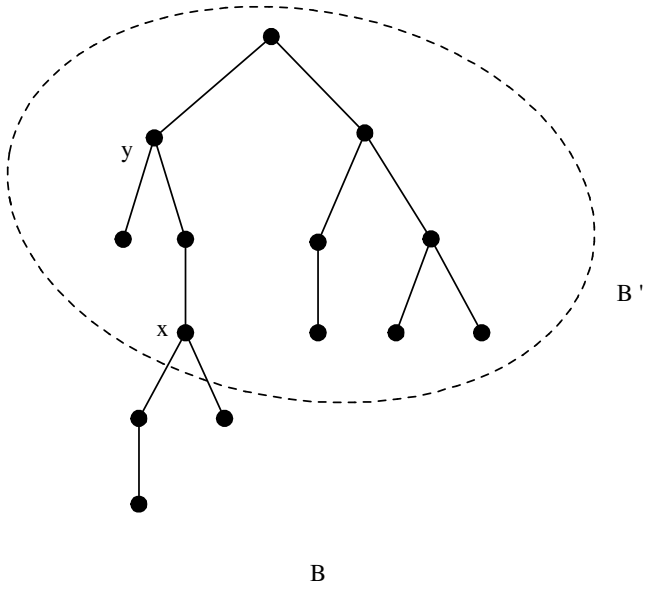


Abbildung 4.5: Induktionsschritt vom Baum B' nach B

Es gelten nun folgende Gleichungen:

$$|T_{<3}(y')| = |T_{<3}(y)| + 1 \tag{4.1}$$

$$|T_{<3}(v')| = |T_{<3}(v)| \text{ für } v \neq x, v \neq y \text{ und } \deg(v) \geq 3 \tag{4.2}$$

Gleichung 4.1 gilt, da sich x' in einem Teilbaum von y ohne Knoten vom Grad ≥ 3 befindet und der Grad von x per Definition ≥ 3 ist.

Für alle Knoten $v \neq x, v \neq y$ und $\deg(v) \geq 3$ gilt Gleichung 4.2, da sich x und y im gleichen Teilbaum T von v befinden. T enthält mit y einen Knoten vom Grad grösser oder gleich drei. Beim Übergang von

B' nach B ändert sich deshalb nichts an der Tatsache dass $T \in T_{\geq 3}(v)$ ist.

Wir unterscheiden zwei Fälle:

1. $|T_{<3}(y')| - 1 > 0$. In der Messanordnung für B' wird mindestens eine Messung an einem Teilbaum T von y' mit $T \in T_{<3}(y')$ durchgeführt. $x' \in B'$ ist ein Blatt und es kann o.B.d.A angenommen werden, dass eine Messung an x' stattfindet. In B werden die gleichen Messungen wie in B' durchgeführt, mit Ausnahme der Messung an x', welche weggelassen wird. Stattdessen wird an $|T_{<3}(x)| - 1 = \deg(x) - 2$ Blättern in Teilbäumen $\in T_{<3}(x)$ gemessen. Für die Messanordnung $S(B)$ gilt also:

$$\begin{aligned}
|S(B)| &= |S(B')| - 1 + \deg(x) - 2 \\
&= \left[\sum_{v':\deg(v')\geq 3} \max(0, |T_{<3}(v')| - 1) \right] - 1 + \deg(x) - 2 \\
&= \left[\sum_{v':\deg(v')\geq 3, v'\neq y'} \max(0, |T_{<3}(v')| - 1) \right] + \\
&\quad + \underbrace{|T_{<3}(y')| - 1 - 1}_{|T_{<3}(y)|-1} + \underbrace{\deg(x) - 2}_{|T_{<3}(x)|-1} \\
&= \sum_{v:\deg(v)\geq 3} \max(0, |T_{<3}(v)| - 1)
\end{aligned}$$

   Die letzte Umformung gilt dank den Gleichungen 4.1 und 4.2 und der Vorbedingung $|T_{<3}(y')| - 1 > 0$.

2. $|T_{<3}(y')| - 1 = 0$. Aus der Vorbedingung folgt, dass y' genau einen Teilbaum ohne Knoten vom Grad mindestens drei enthält. Und zwar ist dies der Teilbaum, in dem auch x zu finden ist. Wie im vorherigen Fall muss zusätzlich an $|T_{<3}(x)| - 1 = \deg(x) - 2$ Blättern in Teilbäumen $\in T_{<3}(x)$ gemessen werden. Es gilt

$$\begin{aligned}
|S(B)| &= |S(B')| + \deg(x) - 2 \\
&= \left[\sum_{v':\deg(v')\geq 3} \max(0, |T_{<3}(v')| - 1) \right] + \deg(x) - 2 \\
&= \left[\sum_{v':\deg(v')\geq 3, v'\neq y'} \max(0, |T_{<3}(v')| - 1) \right] + \\
&\quad + \underbrace{|T_{<3}(y')| - 1}_{0} + \underbrace{\deg(x) - 2}_{|T_{<3}(x)|-1}
\end{aligned}$$

$$= \sum_{v:\deg(v)\geq 3} \max(0, |T_{<3}(v)| - 1)$$

Die Überlegungen sind analog zum vorherigen Fall, mit dem Unterschied, dass $|T_{<3}(y')| - 1 = 0$ ist. Mit Gleichung 4.1 folgt dann dass $|T_{<3}(y)| - 1 = -1$ ist.

Dass auch tatsächlich alle Kanten entdeckt werden, kann man sich wie folgt klar machen: Wegen der Induktionsannahme gilt, dass alle Kanten und Nichtkanten zwischen Knoten im Baum B' erkannt wurden. Die verbleibenden Knoten gehören zu Teilbäumen $T_i(x, \{x, w_i\}), i = 1, \ldots, |\deg(x) - 1|$. Diese Teilbäume sind Ketten und es werden Messungen an allen ausser einem Teilbaum T_j durchgeführt. mit Lemma 4.6 folgt somit dass für einen Knoten $s \notin T_j$ alle Kanten und Nichtkanten erkannt werden. Für einen Knoten $t \in T_j$ kann es also auch keine unentdeckten Kanten zu s geben. Die Nichtkanten von t nach $u \in T_j, t \neq u$ werden ebenfalls entdeckt, somit bleibt nur noch zu zeigen, dass eine Kante zwischen $v \in B'$ und t erkannt wird. Es gibt mindestens einen Messknoten $m_1 \in T_i, i \neq j$. Zudem gibt es in B' mindestens einen Messknoten m_2. Damit eine Kante zwischen v und t nicht erkannt wird muss gelten $dist(m_1, t) = dist(m_1, v)$. Die Wege von m_1 nach t und von m_1 nach v trennen sich im Knoten x, daraus folgt $dist(x, t) = dist(x, v)$. Für den zweiten Messknoten gilt: Der Weg von m_2 nach v ist kürzer als der Weg von m_2 über x nach v, da x in B' ein Blatt ist:

$$\begin{aligned}
dist(m_2, v) &< dist(m_2, x) + dist(x, v) \\
dist(m_2, v) &< dist(m_2, x) + dist(x, t) \text{ , da } dist(x, t) = dist(x, v) \\
dist(m_2, v) &< dist(m_2, t)
\end{aligned}$$

Die letzte Ungleichung gilt, da der Weg von m_2 nach t immer über x führen muss. Eine Kante zwischen v und t wird folglich erkannt.

Nachdem wir nun gezeigt haben wie viele Messungen ausreichen, wollen wir als nächstes beweisen, dass mit weniger Messungen der Graph nicht vollständig erkannt werden kann. Dazu benötigen wir folgendes Lemma:

Lemma 4.8 *Seien v und v' zwei Knoten mit Grad mindestens drei. Für einen Knoten $u \in T_{<3}(v, \{v, w\}), u \neq v$ gilt: $u \in T_{<3}(v', \{v', w'\})$ genau dann wenn $v' = v$ und $w' = w$.*

Beweis. Angenommen obige Behauptung ist falsch und es gilt $v' = v$ und $w' \neq w$. Es folgt daraus dass es von u nach w und nach w' je einen Weg gibt, der nicht über v führt. Da aber w und w' adjazent zu v sind gibt es von u nach v zwei unterschiedliche Wege und somit einen Kreis. Widerspruch.

Angenommen es gilt $v' \neq v$. Es gibt einen Weg von u nach v und einen Weg von u nach v'. Somit gibt es auch einen Weg von v nach v'. Per Definition wird ein Teilbaum durch Weglassen aller ausser einer zum Wurzelknoten inzidenten Kanten erzeugt. Somit hat entweder v oder v' noch den ursprünglichen Grad, nämlich mindestens drei. Dies steht im Widerspruch dazu dass wir Teilbäume mit Knoten nur vom Grad kleiner drei betrachten. ■

Die behauptete Anzahl Messungen ist so gewählt, dass jeder Knoten x mit $\deg(x) \geq 3$ höchstens einen Teilbaum $\in T_{<3}(x)$ besitzt, an dem keine Messung durchgeführt wird. Lemma 4.8 besagt, dass ein Knoten vom Grad kleiner drei sich in maximal einem Teilbaum mit Knoten nur vom Grad kleiner drei befinden kann (vorausgesetzt es werden nur Teilbäume von Knoten mit Grad mindestens drei betrachtet). Werden nun weniger Messungen durchgeführt, so folgt daraus dass es mindestens einen Knoten y gibt, mit $\deg(y) \geq 3$ für den gilt: Es gibt zwei Teilbäume $T_1(y, \{y, z\}), T_2(y, \{y, z'\}) \in T_{<3}(y), z \neq z'$ an denen entweder gar nicht oder nur am Knoten y gemessen wird. Mit Lemma 4.5 und der Tatsache, dass z und z' zu y adjazent sind folgt, dass eine Kante zwischen z und z' nicht erkannt wird. ■

4.4 Allgemeine Graphen

Gute Schranken für allgemeine Graphen zu finden ist nicht einfach. Aus dem maximalen Grad $\deg_{max}$ ergibt sich nicht mehr eine einfache obere Schranke wie dies beim Baum der Fall war. Dies kann man sich zum Beispiel mit Hilfe des Radgraphen verdeutlichen. Wie im Abschnitt über Radgraphen gezeigt genügen weniger als $\deg_{max} - 1$ Messungen um alle Kanten und Nichtkanten zu erkennen. Bei den bisherigen Betrachtungen blieb bei konstanten Grad auch die Anzahl der Messungen konstant (z.B. beim Kreis). Im allgemeinen Fall gilt aber auch diese Beobachtung nicht wie beispielsweise Abbildug 4.6 zeigt.

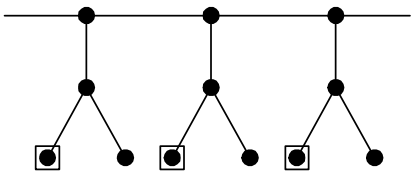


Abbildung 4.6: Graph mit konstantem Grad 3 und $|S| = \frac{|V|}{4}$ für $|V| \geq 8$

Auch ein minimaler Spannbaum (MST) liefert weder eine untere noch eine obere Schranke, wie in Abbildung 4.7 dargestellt. Ob ein MST mehr oder weniger Messungen benötigt ist abhängig vom Graphen.

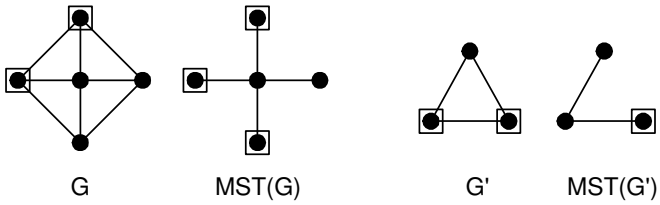


Abbildung 4.7: Beispiele für mehr (links) und weniger (rechts) Messungen im MST als im ursprünglichen Graphen

Im folgenden Satz wird nun eine untere Schranke für allgemeine Graphen vorgestellt. Die Güte der Schranke ist dabei stark von der Wahl des Subgraphen abhängig.

Satz 4.4 *Gegeben sei ein Graph $G = (V, E)$. Für einen Subgraphen $H = (V', E')$ mit $V' \subseteq V$, $E' \subseteq E$ und Durchmesser $diam(H)$ gilt: Es werden mindestens $\log_{(diam(H)+1)}(|V'|)$ Messungen an Knoten von G benötigt um alle Kanten und Nichtkanten von G zu entdecken.*

Beweis. Sei $n' = |V'|$ und sei $m_i \in V$ der Knoten an dem die i-te Messung stattfindet. Für $m_i, i \geq 1$ und zwei Knoten $v, w \in H$ gilt $|dist(m_i, v) - dist(m_i, w)| \leq diam(H)$, denn es gibt immer einen Weg von v nach w der Länge maximal $diam(H)$. Daraus folgt aber, dass alle Knoten des Teilgraphen H sich auf höchstens $diam(H) + 1$ Levels im $LT(m_i)$ befinden. Nach der ersten Messung gibt es mindestens $\left\lceil \frac{n'}{diam(H)+1} \right\rceil$ Knoten, die den selben Abstand von m_1 haben. Es gilt also:

$$\exists L_1 \subseteq V(H) \text{ mit } \forall a, b \in L_1 : dist(m_1, a) = dist(m_1, b)$$
$$|L_1| = \left\lceil \frac{n'}{diam(H) + 1} \right\rceil$$

Bei der zweiten Messung können sich diese Knoten wiederum auf maximal $diam(H) + 1$ Levels aufteilen. Daraus folgt:

$$\exists L_2 \subseteq L_1 \text{ mit } \forall a, b \in L_2 : dist(m_2, a) = dist(m_2, b)$$
$$|L_2| = \left\lceil \frac{n'}{(diam(H) + 1)^2} \right\rceil$$

Im allgemeinen Fall gilt:

$$\exists L_i \subseteq L_{i-1}, i \geq 2, \text{ mit } \forall a, b \in L_i : dist(m_i, a) = dist(m_i, b)$$
$$|L_i| = \left\lceil \frac{n'}{(diam(H) + 1)^i} \right\rceil$$

Das heisst, dass es nach i Messungen $|L_i|$ Knoten gibt, die von jedem Messpunkt die gleiche Entfernung haben. Kanten zwischen diesen Knoten werden aber nicht erkannt. Um alle Kanten und Nichtkanten zu erkennen muss deshalb für die letzte Messung am Knoten m_k gelten, dass $|L_k| \leq 1$ ist.

$$\begin{aligned} |L_k| = \left\lceil \frac{n'}{(diam(H)+1)^k} \right\rceil &\leq 1 \\ n' &\leq (diam(H)+1)^k \\ \log_{(diam(H)+1)}(n') &\leq k \end{aligned}$$

■

Natürlich ist es schwierig einen geeigneten Subgraphen zu finden, so dass die Schranke aus Satz 4.4 möglichst gut ist. Die folgenden Korollare zeigen deshalb wie diese Schranke für einfache Subgraphen aussieht.

Korollar 4.9 *In einem Graphen G mit maximalem Grad* $\deg_{max}$ *beträgt die Anzahl Messungen mindestens* $\log_3(\deg_{max}+1)$.

Beweis. Der Beweis folgt direkt aus Satz 4.4. Die Knotenmenge des Teilgraphen $H = (V', E')$ besteht aus dem Knoten v mit maximalem Grad und allen zu v adjazenten Knoten. Die Kanten inzident zu v bilden die Kantenmenge. ■

Korollar 4.10 *Es werden mindestens* $\log_{(diam(G)+1)}(|V|)$ *Messungen benötigt um alle Kanten und Nichtkanten des Graphen $G = (V, E)$ zu entdecken.*

Beweis. Dieses Korollar beschreibt den trivialen Fall $G = H$ des Satzes 4.4. ■

Korollar 4.11 *Für einen Graphen mit einer Clique der Grösse k sind mindestens* $\log_2(k)$ *Messungen nötig um den Graphen vollständig zu erkunden.*

Beweis. Der Beweis des Korollares folgt wiederum direkt aus Satz 4.4, wenn H der Clique entspricht. ■

Die Schranke aus Korollar 4.11 wird auch in [1] gezeigt. Eine Clique maximaler Grösse in einem Graphen zu finden ist aber ein $\mathcal{NP}$-vollständiges Problem, deshalb wird diese Schranke im Normalfall kaum Anwendung finden.

Kapitel 5

Algorithmen

In diesem Kapitel beschäftigen wir uns mit Algorithmen, welche eine Menge von Messpunkten bestimmen. Alle Algorithmen funktionieren im Wesentlichen nach dem selben Prinzip: Der erste Messknoten wird zufällig gewählt. Dann wird solange jeweils ein Messknoten bestimmt und eine Messung durchgeführt, bis der Graph gut genug rekonstruiert wurde. Die Prozedur *reconstructed* (Abbildung 5.1) bestimmt, wann dies der Fall ist.

Random

Der erste Algorithmus wählt den Messknoten jeweils zufällig aus der Menge der Knoten. Es wird verhindert, dass ein Knoten gewählt wird, an dem bereits eine Messung durchgeführt wurde.

MinDegree

Die Idee hinter dem Algorithmus MinDegree (siehe Abbildung 5.2) ist folgende: Messungen an Knoten mit hohem Grad entdecken meistens weniger Kanten als Messungen an Knoten mit kleinem Grad. Ein Knoten mit Grad x entdeckt z.B. keine Kante zwischen zwei seiner Nachbarsknoten. Somit werden bereits $\binom{x}{2}$ Kanten nicht erkannt. Beim Baum beispielsweise wurde gezeigt, dass mit Messungen nur an Knoten vom Grad eins das Optimum erreicht werden kann.

MinDegree wählt deshalb immer einen Knoten von minimalem Grad als nächsten Messpunkt. Natürlich kann sich aber der Grad, der aufgrund bisheriger Messungen gesehen wird, vom tatsächlichen Grad stark unterscheiden.

MinMaxEdges

Der dritte Algorithmus (siehe Abbildung 5.3) berechnet für jeden möglichen Messknoten, wie viele Kanten im besten und wie viele im schlechtesten Fall erkannt werden. Diese Anzahl wird min respektive max genannt. Es wird

Eingabe: H: Bisher erkannter Graph
P: Menge der potentiellen Kanten
Ausgabe: Ob der Graph genug gut rekonstruiert wurde

procedure reconstructed(H, P)

```
begin
    return (P = ∅);
end
```

(a)

```
begin
    return (|P| < (|V(H)| choose 2) * 5%);
end
```

(b)

Abbildung 5.1: Prozedur reconstructed für komplette Rekonstruktion (a) sowie Erkennen von mehr als 95% aller Kanten und Nichtkanten (b)

Eingabe: $G = (V, E)$
Ausgabe: Menge der Messungen S
Variablen: H: Bisher erkannter Graph
P: Menge der potentiellen Kanten

begin
  $S :=$ {zufälliger Knoten $v \in V(G)$};
  Messe an v;
  $H :=$ (entdeckte Knoten, entdeckte Kanten);
  $P := \binom{V(H)}{2} \setminus E(H)$;
  while (**not** reconstructed(H, P)) **do**
    $S := S \cup$ {Knoten $v \in V(H) \setminus S$ mit minimalem Grad};
    Messe an v;
    $E(H) := E(H) \cup$ neu entdeckte Kanten;
    $P := P \setminus$ neu entdeckte Kanten und Nichtkanten;
  end while
end

Abbildung 5.2: Algorithmus MinDegree

dann jeweils an dem Knoten gemessen, der bezüglich seines min- und max-Wertes bestmöglich ist.

Die Funktion f berechnet, ob ein (min, max)-Paar besser als das bisher beste Paar ist. Die einfachste Möglichkeit ist min respektive max als primäres Kriterium zu nehmen und den anderen Wert nur bei Gleichheit des primären Kriteriums zu beachten. Wenn aber beispielsweise min das erste Kriterium ist, so kann es zu einem solchen Fall kommen:

$$v_1 : (min, max) = (10, 20)$$
$$v_2 : (min, max) = (9, 200)$$

Knoten v_1 wird gewählt, obwohl der min-Wert nur um 1 grösser ist als bei v_2. Auch im besten Fall erkennt v_1 nur 20 Kanten, hingegen könnte v_2 bis zu 200 Kanten entdecken. Analoge Fälle sind möglich wenn max als primäres Kriterium gewählt wird. Es ist deshalb sinnvoll, die Funktion f so zu wählen, dass beide Werte berücksichtigt werden.

Die Prozedur discovered (Abbildung 5.4) überprüft, ob eine Kante $e = \{u, t\}$ bei einer Messung an einem Knoten v erkannt wird. Sei G der unterliegende und zu rekonstruierende Graph. Der Graph H, der sich aus den Erkenntnissen bisheriger Messungen zusammensetzt, besitzt die gleiche Knotenmenge, aber weniger oder gleich viele Kanten wie G. H' sei nun der Graph den man erhält, wenn man bei H alle potentiellen Kanten hinzufügt, d.h. Kanten von denen man noch nicht weiss, ob sie vorhanden sind oder nicht. Die Kantenmenge von H' ist grösser oder gleich der Kantenmenge von G. Es gilt also:

$$V(H) = V(G) = V(H')$$
$$E(H) \subseteq E(G) \subseteq E(H')$$

Die Distanz zwischen zwei Knoten $a, b \in G$ lässt sich durch die Distanz von a nach b in H und H' abschätzen. Bezeichne $dist_G(a, b)$ die Distanz von a nach b im Graphen G. Es gelten folgende Schranken für die Länge des kürzesten Weges von a nach b:

$$dist_{H'}(a, b) \leq dist_G(a, b) \leq dist_H(a, b)$$

Die Prozedur discovered berechnet nun für einen möglichen Messpunkt v und eine potentielle Kante $e = \{u, t\}$, die obere und untere Schranke für $dist(v, u)$ und $dist(v, t)$. Falls die obere Schranke eines Knotens kleiner ist als die untere Schranke des anderen Knotens, so ist $dist(v, u)$ auf jeden Fall ungleich $dist(v, t)$. Daraus folgt, dass die Kante e bei einer Messung an v erkannt wird. Wenn alle vier Schranken denselben Wert haben, so

```
Eingabe: G = (V, E)
Ausgabe: Menge der Messungen S
Variablen: H: Bisher erkannter Graph
           P: Menge der potentiellen Kanten
           m: nächster Messknoten
           m_min: Anzahl Kanten, die mindestens entdeckt werden
           m_max: Anzahl Kanten, die höchstens entdeckt werden
begin
    m:= nil;
    m_min := 0;
    m_max := 0;
    S := {zufälliger Knoten v ∈ V(G)};
    Messe an v;
    H := (entdeckte Knoten, entdeckte Kanten);
    P := ( V(H) 2 ) \ E(H);
    while (not reconstructed(H, P)) do
        for (alle Knoten v ∈ V(H) \ S) do
            _min := 0;
            _max := 0;
            for (alle Kanten e = {t, u} ∈ P) do
                if (discovered(v, e, H, H ∪ P) = true) then
                    _min + +;
                    _max + +;
                else if (discovered(v, e, H, H ∪ P) = maybe) then
                    _max + +;
                end if
            end for
            if (f(_min, _max, m_min, m_max)) then
                m_min := _min;
                m_max := _max;
                m := v;
            end if
        end for
        S := S ∪ {m};
        Messe an m;
        E(H) := E(H) ∪ neu entdeckte Kanten;
        P := P \ neu entdeckte Kanten und Nichtkanten;
    end while
end
```

Abbildung 5.3: Algorithmus MinMaxEdges

gilt $dist(v, u) = dist(v, t)$. Die Kante e wird somit nicht entdeckt. In allen anderen Fällen ist es sowohl möglich, dass die beiden Distanzen gleich sind, aber auch dass sie ungleich sind. Es ist deshalb unklar, ob e beim Messen erkannt wird oder nicht.

Eingabe: v: möglicher Messknoten
e: potentielle Kante
H: bisher erkannter Graph
H': bisher erkannter Graph vereinigt mit allen potentiellen Kanten
Ausgabe: Ob die Kante e bei einer Messung an v erkannt wird

procedure discovered(v, e, H, H')
begin
  $//e = \{t, u\}$
  $vt_min := dist_{H'}(v, t)$;
  $vt_max := dist_{H}(v, t)$;
  $vu_min := dist_{H'}(v, u)$;
  $vu_max := dist_{H}(v, u)$;
  if $((vt_max < vu_min)$ **or** $(vu_max < vt_min))$ **then**
    return true;
  else if $(vt_min = vt_max = vu_min = vu_max)$ **then**
    return false;
  else
    return maybe;
  end if
end

Abbildung 5.4: Prozedur discovered des Algorithmus MinMaxEdges

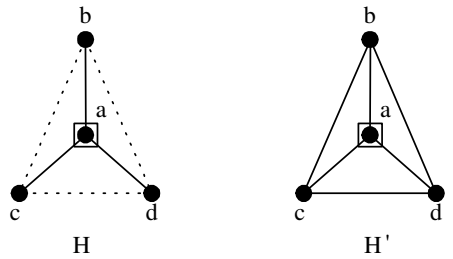


Abbildung 5.5: Beispiel zum Algorithmus MinMaxEdges

Betrachten wir den Beispielgraphen in Abbildung 5.5. Am Knoten a wurde die erste Messung durchgeführt. Die Menge der potentiellen Kanten ist: $\{\{b, c\}, \{b, d\}, \{c, d\}\}$. Für den möglichen Messknoten b ergibt sich folgendes

Bild:

Kante $\{b, c\}$: $dist_G(b, b) \geq dist_{H'}(b, b) = 0$
$dist_G(b, b) \leq dist_H(b, b) = 0$
$dist_G(b, c) \geq dist_{H'}(b, c) = 1$
$dist_G(b, c) \leq dist_H(b, c) = 2$
Die Kante $\{b, c\}$ wird erkannt.

Kante $\{b, d\}$: aus Symmetriegründen analog zu $\{b, c\}$.

Kante $\{c, d\}$: $dist_G(b, c) \geq dist_{H'}(b, c) = 1$
$dist_G(b, c) \leq dist_H(b, c) = 2$
$dist_G(b, d) \geq dist_{H'}(b, d) = 1$
$dist_G(b, d) \leq dist_H(b, d) = 2$
Ob die Kante $\{c, d\}$ erkannt wird hängt von der Topologie von G ab.

Für die zwei anderen potentiellen Messknoten c und d berechnen sich die Werte analog. Aus Symmetriegründen ergibt sich auch für c und d, dass zwei Kanten erkannt und eine Kante vielleicht erkannt wird.

Ob bei einer Messung an b die Kante $\{c, d\}$ erkannt wird oder nicht, hängt von der Topologie des unterliegenden Graphen ab. In Abbildung 5.6 sind die Fälle dargestellt, dass die Kante entdeckt wird (links) und dass sie nicht entdeckt wird (rechts).

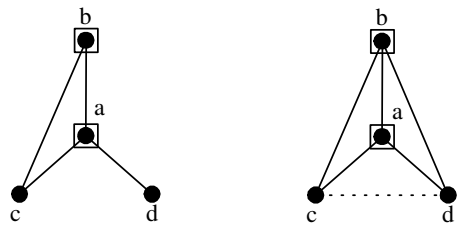


Abbildung 5.6: Beispiel für unterschiedliche Anzahl entdeckter Kanten und Nichtkanten, je nachdem ob die Kante $\{b, d\}$ vorhanden ist oder nicht

Das Offline-Problem

Die beschriebenen Algorithmen suchen online eine möglichst kleine Menge von Messpunkten. Im Offline-Fall, wenn die Topologie also bereits zu Beginn bekannt ist, kann man das Problem auf Minimum Set Cover reduzieren.

Das Problem des Minimum Set Covers ist das folgende: Gegeben ist eine Menge $U = \{a_1, a_2, \ldots, a_N\}$ und eine Menge $F = \{S_1, S_2, \ldots, S_t\}$. S_i, $i = 1, \ldots, t$, ist eine Teilmenge von U und hat Kosten $c(S_i) \in \mathbf{R}$. Gesucht ist $C \subseteq F$ mit minimalen Kosten, so dass U von C abgedeckt wird.

Für die Reduktion sei U die Menge aller Kanten und Nichtkanten des Graphen G. F besteht aus $|V|$ Teilmengen von U. Eine Teilmenge enthält die Menge der Kanten und Nichtkanten, die bei einer Messung an einem Knoten $v \in G$ entdeckt werden. Die Kosten aller Teilmengen sind identisch. Diese Reduktion ist analog zum Fall beschrieben in [1]. Es existiert ein Algorithmus für Minimum Set Cover mit Approximationsfaktor $O(\log(N))$. Das Problem kann im Offline-Fall also mit Approximationsfaktor $O(\log(|V|))$ gelöst werden, da $N = \frac{|V|*(|V|-1)}{2}$.

Kapitel 6

Versuche

Als nächstes werden nun die drei im vorhergehenden Kapitel besprochenen Algorithmen Random, MinDegree und MinMaxEdges auf ihre Tauglichkeit getestet. Zuerst beschreiben wir, wie die Versuche durchgeführt wurden, und anschliessend wie die Resultate aussehen.

6.1 Versuchsanordnung

Die drei Algorithmen wurden auf Graphen mit 10 bis 1000 Knoten getestet. Neben den Topologien Baum und Radgraph verwendeten wir Zufallsgraphen nach dem Modell von Erdös-Rényi, sowie von BRITE [5] erzeugte Topologien nach den Modellen Waxman, Barabási-Albert (BA) und Generalized Linear Preference (GLP). BA und GLP erzeugen Graphen, die einem Potenzgesetz gehorchen.

Die Bäume wurden zufällig erzeugt. Ausgehend von einem Knoten wurde dem Graphen jeweils ein neuer Knoten und eine neue Kante hinzugefügt. Die Kante war dabei inzident zum neuen Knoten und einem zufälligen Knoten des bisherigen Graphen.

Für die Graphen nach Erdös-Rényi verwendeten wir drei verschiedene Werte für die Wahrscheinlichkeit p, dass eine Kante vorhanden ist: $p = 0.05, p = 0.1$ und $p = 0.2$.

Parameter	Wert
HS	1000
LS	100
Node Placement	Random
Growth Type	Incremental
Pref. Conn	On
Bandwidth Distr.	Constant

Abbildung 6.1: Parameter von BRITE

In Abbildung 6.1 sind die allgemeinen Parameter dargestellt, welche für die mit BRITE erzeugten Graphen verwendet wurden. Abbildung 6.2 zeigt die Parameter für die einzelnen Modelle.

Modell	Parameter
Waxman	$\alpha = 0.15$ $\beta = 0.2$ $m = 2$
BA	$m = 2$
GLP	$p(add) = 0.45$ $\beta = 0.64$ $m = 2$

Abbildung 6.2: Parameter der Modelle Waxman, BA und GLP

Auf den Graphen bis zu 100 Knoten liessen wir die Algorithmen auf jedem Graphen zehnmal laufen. Für eine feste Anzahl Knoten verwendeten wir einen Graphen, ausser bei Bäumen und Zufallsgraphen nach Erdös-Rényi, wo jeweils Tests auf 10 verschiedenen Graphen durchgeführt wurden. Auf den Graphen mit mehr als 100 Knoten liessen wir die Algorithmen wegen der langen Laufzeit nur einmal laufen.

Es wurden zwei Abbruchkriterien für die Algorithmen getestet: Vollständige Rekonstruktion des Graphen, sowie Erkennen von mehr als 95% aller Kanten und Nichtkanten. Dies entspricht den beiden Varianten der Prozedur reconstructed in Abbildung 5.1. Wir nennen diese beiden Versuchsreihen *Rekonstruktion* resp. 95%-*Approximation*. Rekonstruktion und 95%-Approximation wurden in verschiedenen Durchläufen gemessen, aber - bis auf die Topologien Baum und Erdös-Rényi - mit dem gleichen Set von Graphen.

Der Algorithmus MinMaxEdges bietet einen Freiheitsgrad an bezüglich der Bewertung des *min*- und *max*-Wertes eines Knotens (durch die Funktion f). In unseren Tests wurde einmal *min* und einmal *max* maximiert. In den Resultaten erscheint der Algorithmus deshalb zweimal, einmal als MinMaxEdges (min) und einmal als MinMaxEdges (max).

Der Kreis und der vollständige Graph eignen sich nicht für Versuche, da bereits eine zufällige Wahl der Messknoten das Optimum liefert.

Ausser bei Bäumen und Radgraphen steht kein optimaler Wert als Vergleich zur Verfügung. Deshalb wurde eine untere Schranke aus dem minimalen Wert von Korollar 4.10 und Korollar 4.9 berechnet. Diese Schranke darf aber keinesfalls mit dem Optimum verwechselt werden, denn es handelt sich dabei nur um eine grobe Abschätzung. In vielen Fällen sind mehr Messungen nötig als die Schranke angibt.

6.2 Resultate

In diesem Abschnitt betrachten wir die Resultate der Testläufe. Als erstes gehen wir auf das wichtigste Kriterium ein: Die Anzahl Messungen, die mit den verschiedenen Algorithmen benötigt wurden.

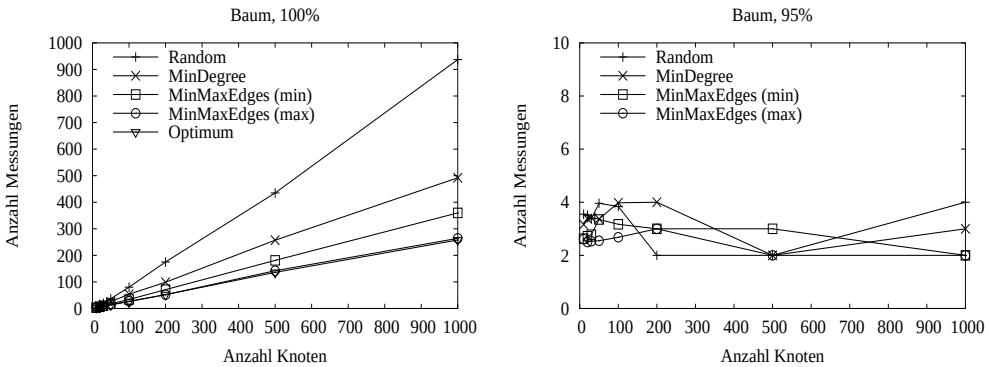
Anzahl Messungen

Bei der Rekonstruktion beim Baum zeigt sich, dass MinMaxEdges (max) praktisch dem Optimum entspricht. Auch MinMaxEdges (min) bringt sehr gute Resultate (Abbildung 6.3). Beim Radgraphen sind beide Varianten von MinMaxEdges sehr nahe am Optimum (Abbildung 6.4). MinDegree ist beim Baum deutlich besser als eine zufällige Wahl der Messpunkte, beim Radgraphen jedoch vergleichbar schlecht.

Bei der 95%-Approximation sieht es beim Radgraphen ähnlich aus wie bei der Rekonstruktion. Ganz anders jedoch beim Baum: Es sind viel weniger Messungen nötig als für die Rekonstruktion und alle Algorithmen liegen sehr nahe zusammen. Keiner ist den anderen überlegen.

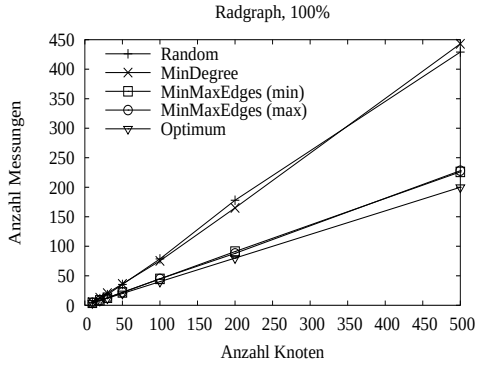
Bei den übrigen Graphen ist es so, dass bei der Rekonstruktion die beiden Varianten von MinMaxEdges durchwegs bessere Resultate als MinDegree und Random liefern. Sehr deutlich zeigt sich dies bei den Power-law-Toplogien BA und GLP (Abbildung 6.5). MinDegree ist im Normalfall kaum besser als Random.

Bei der 95%-Approximation werden allgemein nur sehr wenige Messungen benötigt. Oft genügen bereits vier oder fünf Messungen um die 95%-Grenze zu erreichen, bei den Zufallsgraphen nach Erdös-Rényi sind etwas mehr Messungen nötig. Es sind keine eindeutigen Tendenzen bezüglich der Güte der einzelnen Algorithmen auszumachen. Selbst eine zufällige Wahl der Messknoten bringt vergleichbare Resultate wie die anderen Algorithmen.

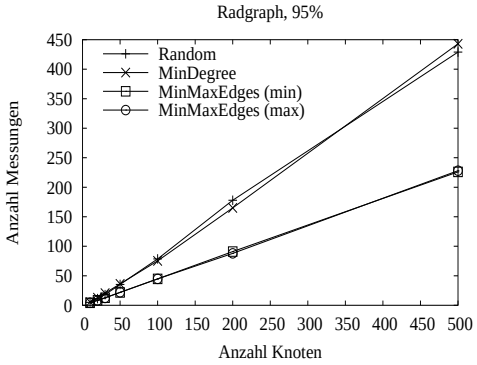


(a) Rekonstruktion (b) 95%-Approximation

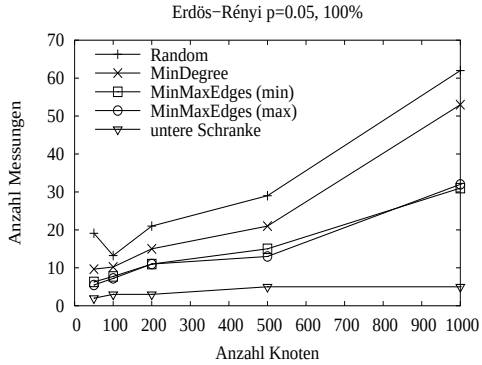
Abbildung 6.3: Anzahl Messungen bei Baum



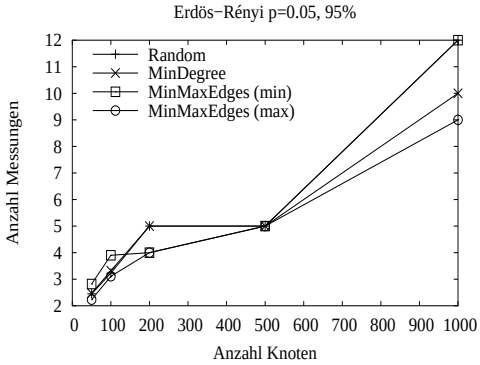
(a) Rekonstruktion



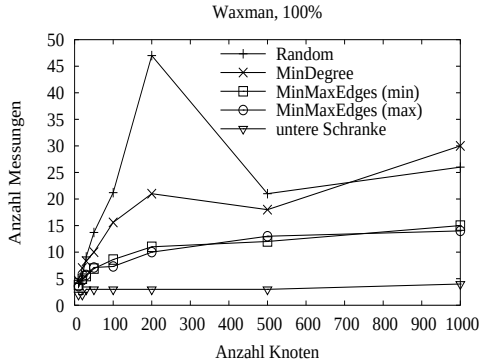
(b) 95%-Approximation



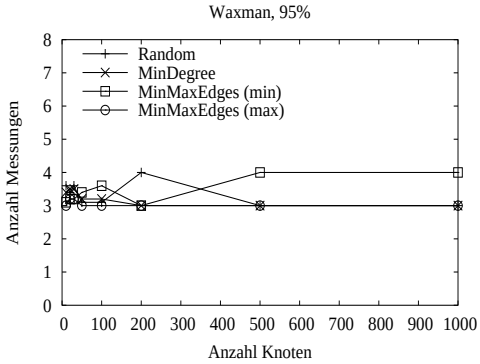
(c) Rekonstruktion



(d) 95%-Approximation

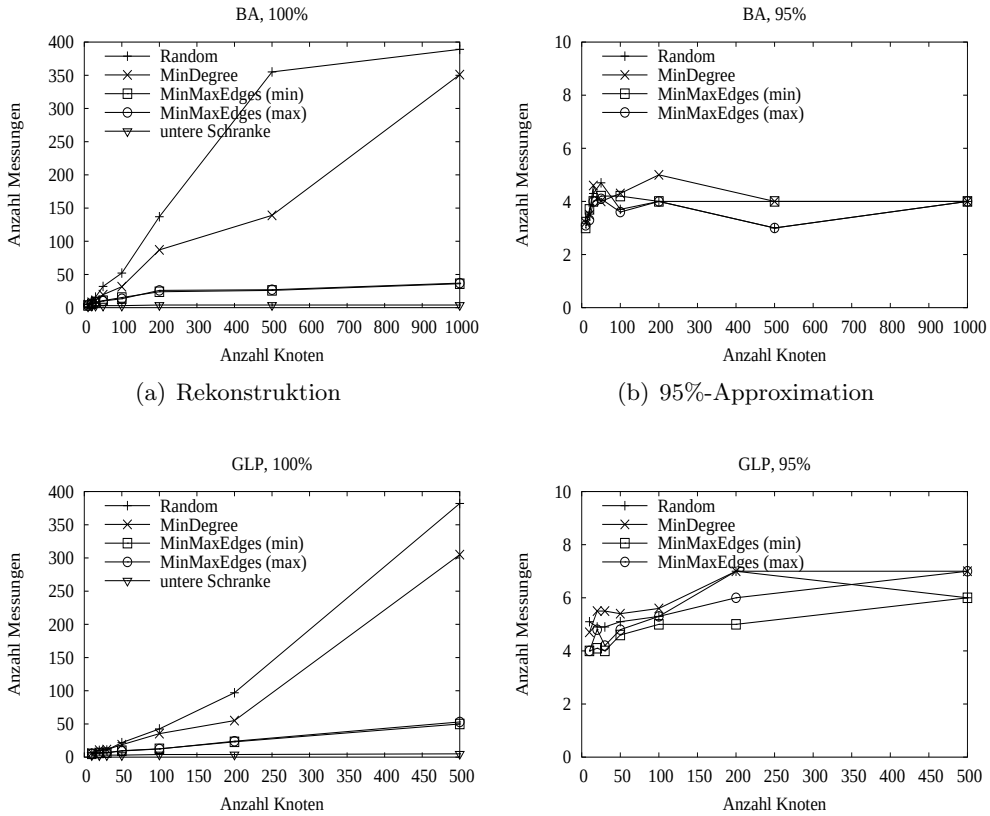


(e) Rekonstruktion



(f) 95%-Approximation

Abbildung 6.4: Anzahl Messungen beim Radgraphen, Erdös-Rényi ($p = 0.05$) und Waxman



(a) Rekonstruktion (b) 95%-Approximation

(c) Rekonstruktion (d) 95%-Approximation

Abbildung 6.5: Anzahl Messungen bei BA und GLP

Prozentsatz erkannter Kanten und Nichtkanten

Wir betrachten nun den Prozentsatz der erkannten Kanten und Nichtkanten in Abhängigkeit der Messungen. Die Resultate für Rekonstruktion und 95%-Approximation sind hier dieselben.

Allgemein zeigt sich, dass die erste Messung bereits den grössten Teil aller Kanten resp. Nichtkanten erkennt. Beim Baum mit 1000 Knoten lag dieser Anteil sogar bei ungefähr 90% (Abbildung 6.6a). Der Informationsgewinn nimmt danach exponentiell mit den Messungen ab. Ein Sonderfall ist der Radgraph (Abbildung 6.6b). Hier wächst der Prozentsatz linear mit der Anzahl Messungen. Dies ist auf die spezielle Struktur des Graphen zurückzuführen. Als weitere Beispiele sind die Graphen Erdös-Rényi mit $p = 0.2$ und Waxman (Abbildungen 6.6c,d) aufgeführt. Es zeigt sich, dass alle Algorithmen mit den ersten Messungen einen vergleichbaren Anteil des Graphen entdecken.

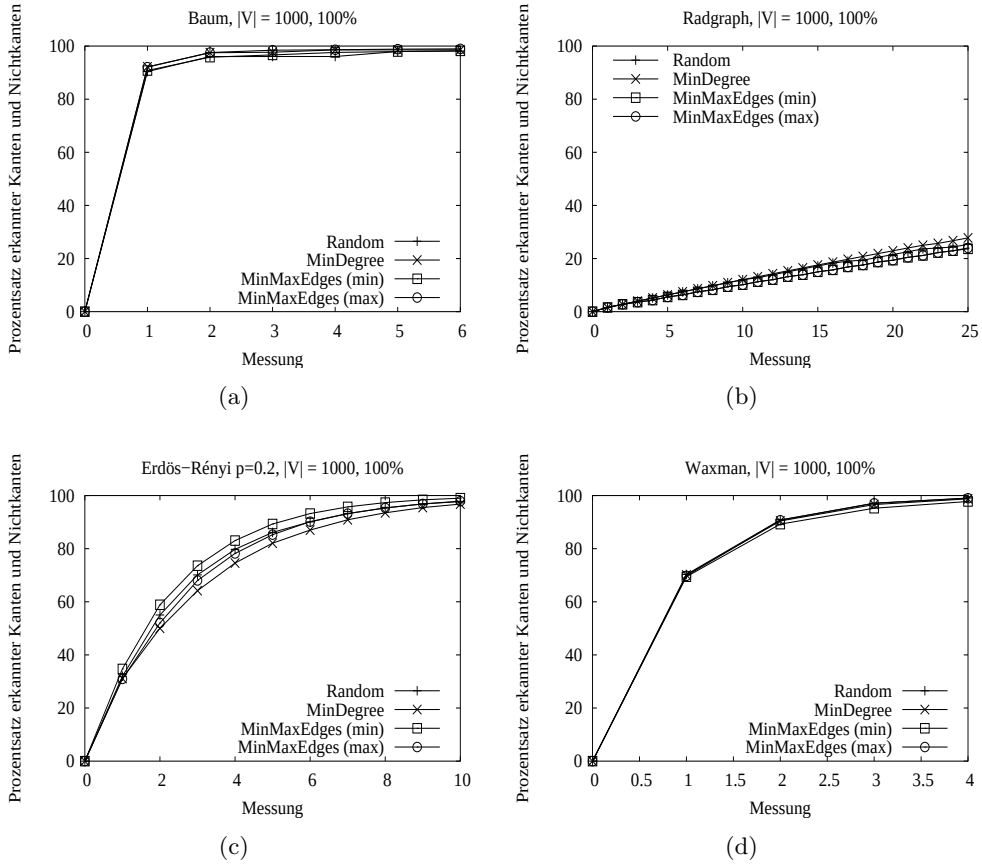


Abbildung 6.6: Prozentsatz erkannter Kanten und Nichtkanten in Abhängigkeit der Messung

Streuung

Für Graphen mit bis zu 100 Knoten haben wir je 10 Testläufe durchgeführt. Es lässt sich also bestimmen, wie stark die Streuung der einzelnen Durchläufe ist. Der Variationskoeffizient v berechnet sich folgendermassen:

$$v = \frac{\sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}}{\overline{x}}$$

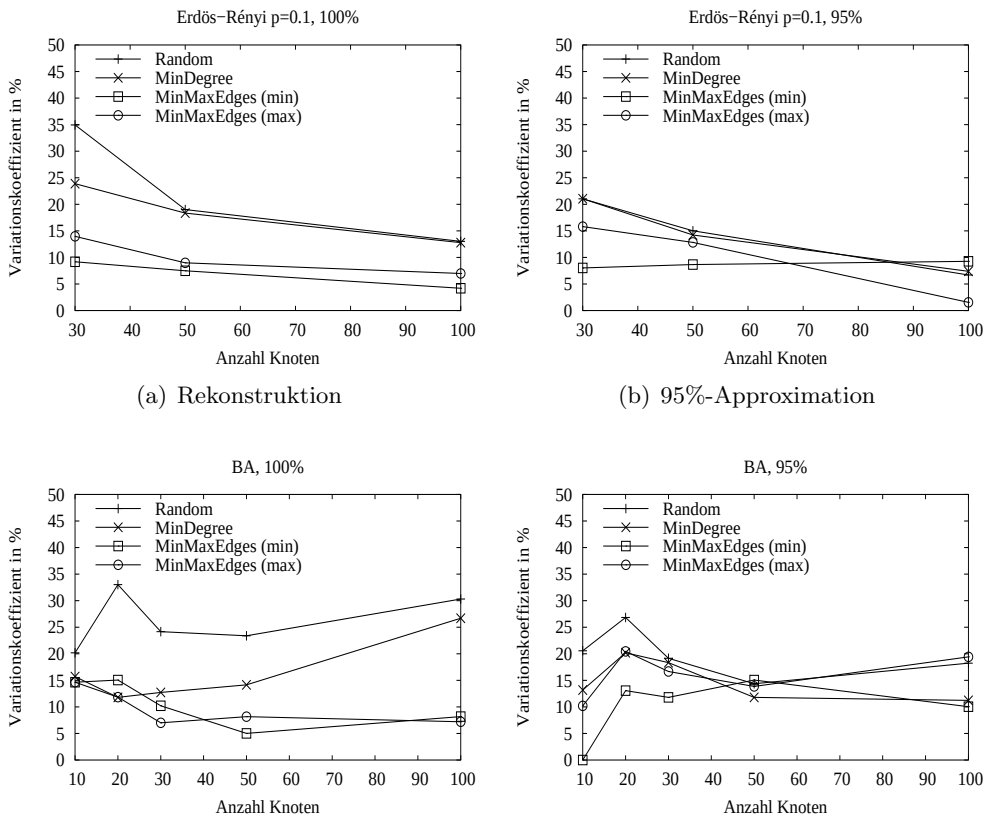
Dabei bezeichnet x_i die Anzahl Messungen, die im i-ten Durchlauf benötigt wurde und $\overline{x}$ die durchschnittliche Anzahl Messungen. In unserem Fall ist $n = 10$.

Bei der Rekonstruktion ist im allgemeinen der Variationskoeffizient bei MinMaxEdges viel kleiner als bei den Algorithmen Random und MinDegree. Dies entspricht auch den Erwartungen, da bei MinMaxEdges der Zufall nur bei der ersten Messung und wenn mehrere Messknoten die gleiche Güte auf-

weisen, zum Zuge kommt. Random und MinDegree dagegen sind mehr vom Zufall abhängig und somit besteht auch eher die Gefahr von sehr schlechten Messungen.

Bei der 95%-Approximation sind die Resultate weniger deutlich, tendenziell ist der Variationskoeffizient bei MinMaxEdges ebenfalls kleiner als bei Random und MinDegree. Abbildung 6.7 zeigt die Variationskoeffizienten der Graphen BA und Erdös-Rényi mit $p = 0.1$.

Es ist noch zu erwähnen, dass beim Algorithmus MinMaxEdges der Unterschied zwischen dem Durchlauf mit der minimalen Anzahl Messungen und demjenigen mit der maximalen Anzahl nur gering ist. Hingegen ist bei MinDegree und insbesondere bei Random der Unterschied zwischen dem besten und dem schlechtesten Fall manchmal erheblich. Dies gilt vor allem bei der Rekonstruktion.



(a) Rekonstruktion (b) 95%-Approximation

(c) Rekonstruktion (d) 95%-Approximation

Abbildung 6.7: Variationskoeffizienten von Erdös-Renyi ($p = 0.1$) und BA

Kapitel 7

Schlussfolgerung

Unsere Untersuchung hat gezeigt, dass die Anzahl Messungen, um einen Graphen vollständig zu rekonstruieren resp. zu approximieren, stark von der Topologie abhängig ist. Eine grössere Anzahl Kanten bedeutet nicht - wie intuitiv zu erwarten ist - dass auch mehr Messungen benötigt werden. Es gibt zwar Graphen mit vielen Kanten, die viele Messungen benötigen, z.B. der vollständige Graph. Ebenso gibt es aber auch Graphen mit wenigen Kanten, die sehr aufwändig sind, z.B. gilt dies für den Stern. Es ist nicht einfach, für einen allgemeinen Graphen anzugeben, wie viele Messungen benötigt werden, und wir haben in unserer Betrachtung nur eine grobe untere Schranke gezeigt.

Der Algorithmus MinMaxEdges benötigt für die komplette Rekonstruktion eines Graphen deutlich weniger Messungen als eine zufällige Wahl. Für den Algorithmus MinDegree gilt, dass er im Normalfall kaum besser als eine zufällige Wahl ist, teilweise sogar schlechter. Die Stärke von MinMaxEdges zeigt sich erst wenn bereits einige Messungen durchgeführt wurden, und es dann darum geht, die letzten Unsicherheiten zu beseitigen. Wenn der Graph nur approximiert werden soll, so bietet eine zufällige Wahl der Messknoten unter Umständen vergleichbare Resultate wie MinMaxEdges. Dies gilt besonders bei niedriger Approximationsgüte. Der Grund dafür ist, dass Messungen in unserem Modell sehr mächtig sind, da sie alle kürzesten Pfade von einem Knoten zurückliefern. Häufig entdeckt bereits die erste Messung weit über die Hälfte aller Kanten und Nichtkanten, die überhaupt im unterliegenden Graphen vorhanden sind. Der Einsatz von MinMaxEdges lohnt sich umso mehr, desto besser der unterliegende Graph approximiert werden soll.

Kapitel 8

Ausblick

In diesem Kapitel gehen wir auf einige Aspekte ein, die bei unseren Betrachtungen vernachlässigt wurden.

Im Kapitel 1 wurde bereits erwähnt, dass die Messpunkte nicht immer frei gewählt werden können. Unter Umständen ist es also gar nicht möglich, am gewünschten Knoten eine Messung durchzuführen.

Weiter wurde nicht beachtet, dass gewisse Pfade gar nicht zugelassen sind. Nehmen wir beispielsweise ein Autonomous System B, welches mit zwei Providern AS A und AS C verbunden ist. Obwohl es eine Verbindung von A über B nach C gibt wird B keinen Verkehr zwischen A und C weiterleiten [4].

Eine komplette Erkundung des Graphen ist in der Realität unmöglich, deshalb sind gute Approximationen gefragt. In unseren Betrachtungen verwendeten wir als Kriterium für eine Approximation, dass 95% aller Kanten und Nichtkanten entdeckt wurden. Es wurde aber nicht überprüft, wie gut der approximierte Graph anderen Kriterien genügt. Wie verhält sich beispielsweise die Grad-Verteilung der Knoten im gemessenen Graphen zum unterliegenden Graphen? Wie gross ist bei beiden Graphen der Durchmesser oder die durchschnittliche Pfadlänge?

Interessant wären Testläufe auf Graphen mit vergleichbarer Anzahl Knoten und Kanten wie der tatsächliche Internet-Graph. Wegen der langen Laufzeiten des Algorithmus MinMaxEdges beschränkten wir uns auf Graphen mit bis zu 1000 Knoten. Im Hinblick auf echte Messungen waren unsere Testgraphen aber zu klein und es ist nicht klar, wie sich die Algorithmen auf sehr grossen Graphen verhalten.

In unserer Arbeit haben wir keine theoretischen Aussagen für die Approximation von Graphen gemacht, sondern nur für die komplette Rekonstruktion. Ausserdem haben wir keine Analyse der präsentierten Algorithmen durchgeführt. Insbesondere ist offen, welche Approximationsrate erreicht wird und wie jeweils ein Worst-Case-Graph aussieht.

Kapitel 9

Schlussbemerkung

An dieser Stelle möchte ich Professor Thomas Erlebach danken, der mir diese Semesterarbeit ermöglicht hat. Er hat sich sehr viel Zeit genommen, um diese Arbeit zu betreuen und hat mir bei Schwierigkeiten jeweils sofort geholfen.

Alles in allem hat mir die Arbeit an diesem Thema gut gefallen. Einzig das Erarbeiten der theoretischen Ergebnisse war manchmal etwas frustrierend, da viele Ansätze nötig waren, bis einer zum Ziel führte.

Ich bestätige, dass ich die Arbeit selbständig und nur unter Verwendung der erwähnten Hilfsmittel geschrieben habe.

Felix Eberhard

Literaturverzeichnis

[1] L. S. Ram. Tree-based graph reconstruction, 2002. CGC Pre-Doc Project, ETH Zürich.

[2] P. Barford, A. Bestavros, J. Byers, and M. Crovella. On the marginal utility of deploying measurement infrastructure. In *Proceedings of the ACM SIGCOMM Internet Measurement Workshop 2001*, November 2001.

[3] A. Lakhina, J. Byers, M. Crovella, and P. Xie. Sampling biases in IP topology measurements. In *Proceedings of INFOCOM'03*, 2003.

[4] L. Subramanian, S. Agarwal, J. Rexford, and R. Katz. Characterizing the Internet hierarchy from multiple vantage points. In *Proceedings of INFOCOM'02*, June 2002.

[5] A. Medina, A. Lakhina, I. Matta, J. Byers. BRITE: Universal Topology Generation from a User's Perspective. Technical Report BUCS-TR2001 -003, Boston University, 2001.

[6] T. Bu, D. Towsley. On distinguishing between Internet power law topology generators. In *Proceedings of INFOCOM'02*.

[7] University of Oregon RouteViews project. http://www.routeviews.org.