

Skalenfreie Netze

Proseminar Algorithmen und Datenstrukturen
Wintersemester 2005/2006
Universität Trier

Eingereicht bei Prof. Dr. B. Walter
Fachbereich IV, Informatik
am 06.03.2006
von Ansgar Jonietz
mailto:ansgar@jonietz.de
<http://www.jonietz.de>
4. Fachsemester Informatik

Inhaltsverzeichnis

1. Einleitung	3
1.1 Milgram-Versuch	3
1.2 Kevin-Bacon-Orakel	4
1.3 Erdős-Zahlen	4
1.4 Weitere Beispiele	5
2. Ältere Modelle	6
2.1 Zufallsgraphen	6
2.2 Reguläre Graphen	7
2.3 Small World	7
3. Skalenfreie Netze	9
3.1 Struktur	9
3.2 Eigenschaften	10
3.3 Beispiele	10
3.4 Erzeugung	11
3.5 Deterministischer Algorithmus	12
4. Anwendungen	14
4.1 Angriffssicherheit	14
4.2 Virenausbreitung	14
4.3 Durchsuchbarkeit	15
5. Ausblick	17

„Fred Jones of Peoria, sitting in a sidewalk cafe in Tunis, and needing a light for his cigarette, asks the man at the next table for a match. They fall into conversation; the stranger is an Englishman who, it turns out, spent several months in Detroit studying the operation of an interchangeable-bottlecap-factory. ‚I know it’s a foolish question,‘ says Jones, ‚but did you ever by any chance run into a fellow named Ben Arkadian? He’s an old friend of mine, manages a chain of supermarkets in Detroit...‘

‚Arkadian, Arkadian,‘ the Englishman mutters. ‚Why, upon my soul, I believe I do! Small chap, very energetic, raised merry hell with the factory over a shipment of defective bottlecaps.‘

‚No kidding!‘ Jones exclaims in amazement.

‚Good lord, it’s a small world, isn’t it?‘¹

¹ Stanley Milgram, 1967 zit. nach [Kle00].

1. Einleitung

Die zunehmende digitale Speicherung von Informationen hat die Erforschung von Graphtopologien in den letzten Jahren erheblich vorangebracht. Immer mehr Daten liegen in elektronischer Form vor und können – als Graph interpretiert – in ihrer Struktur verhältnismäßig einfach untersucht werden. Diese Arbeit mit realen Daten ermöglicht einerseits die Verifizierung von Theorien, zum anderen wurden aber auch ganz neue Modelle über die Struktur großer Netze¹ entwickelt.

Daten, die sich als Graphen interpretieren lassen, finden sich neben der naheliegenden Informationstechnologien auch in vielen anderen Gebieten, etwa in der Biologie, der Soziologie oder den Wirtschaftswissenschaften. Viele dieser größtenteils selbstorganisierten Vernetzungen zeigen eine ähnliche Struktur, die insbesondere an einer sehr geringen mittleren Distanz und einem hohen lokalen Clustering-Koeffizient zu erkennen ist – später mehr hierzu.

1.1 Milgram-Versuch

Zur Erforschung dieser speziellen Graphtopologie dienten in erster Linie die sozialen Netze. Ein erster Versuch in diesem Bereich wurde 1967 von Stanley Milgram, einem Psychologen aus Harvard durchgeführt. Er übergab 160 Personen aus dem Süden der US-Westküste je ein Paket mit der Aufgabe, es an eine ihnen unbekannte Zielperson aus Massachusetts zu übermitteln; dabei sollten die Pakete nur an bekannte Personen weitergegeben werden. Bei der Auswertung dieses Versuchs kam Milgram zu dem für ihn überraschenden Ergebnis, dass im Mittel nur sechs Personen nötig waren, um ein Paket an die Zielperson zuzustellen. Er stellte daraufhin die unter dem Namen „Six degrees of Separation“ bekannte Theorie auf, dass alle Menschen der Welt maximal sechs Bekanntschaftsbeziehungen voneinander entfernt sind. [Zie05]

Inzwischen wurde diese vielzitierte Theorie allerdings angezweifelt, etwa in einer detaillierten Untersuchung von Judith Kleinfeld [Kle00]. Sie kritisiert unter anderem, dass die empirische Basis für Milgrams Versuche (im-

¹ Der Begriff „Netz“ steht hier nicht für einen Begriff aus der Graphentheorie, sondern als allgemeiner Oberbegriff für die Zusammenfassung von Objekten, die untereinander in bestimmten Beziehungen stehen.

merhin kamen nur 44 der 160 Pakete überhaupt am Ziel an) zu klein für seine Schlussfolgerung sei.

Vor einigen Jahren wurde der Milgram-Versuch von einer Arbeitsgruppe an der Columbia Universität wiederholt, allerdings wurden nun statt Paketen eMails verschickt und beobachtet. Auch hier waren nur 5 bis 7 Kontaktpersonen nötig, bis eine Zielperson erreicht wurde – die Wissenschaftler sahen in Ihrer Arbeit daher eine Bestätigung des Milgram-Versuchs. Doch ist die empirische Basis wieder vergleichsweise dünn: von circa 1,1 Millionen möglichen Ketten konnten die Soziologen 24163 festmachen, von denen tatsächlich aber nur 384 Ketten bis zum Ziel vordrangen. [Sch03]

1.2 Kevin-Bacon-Orakel

In den Neunziger Jahren des vergangenen Jahrhunderts konnten dann andere überraschenden Erkenntnisse ähnlich zur Theorie von Milgram gefunden werden. So wird etwa beim Kevin-Bacon-Orakel der ungerichtete Graph betrachtet, in dem ein Knoten für einen Schauspieler steht und eine Kante zwischen zwei Knoten eingetragen wird, wenn die zwei entsprechenden Schauspieler gemeinsam in einem Film aufgetreten sind. Die Bacon-Zahl gibt dabei die Anzahl der Schritte bis zum Knoten mit Kevin Bacon an; Bacon selbst hat also die Bacon-Zahl 0.

Erstaunlicherweise haben selbst viele deutsche Schauspieler eine Bacon-Zahl von etwa 3, die durchschnittliche Bacon-Zahl liegt knapp unter 3.² Das Kevin-Bacon-Orakel kann online unter [OoB] abgefragt werden. Möglich wurde dies erst dank der eingangs erwähnten Digitalisierung der Informationen; die Webseite zum Kevin-Bacon-Orakel nutzt z. B. die Informationen aus der Internet Movie Database imdb.com.

1.3 Erdős-Zahlen

Ein ähnlicher Ansatz beruht auf Paul Erdős (1913-1996), einem Mathematiker mit überdurchschnittlich vielen Veröffentlichungen. Betrachtet wird hier ein Graph, dessen Knoten Autoren darstellen, die genau dann mit einer Kante verbunden sind, wenn sie gemeinsam einen Text publiziert haben.³ Die

² Laut [OoB] ist die durchschnittliche Bacon-Zahl 2,946 (Stand: 29.06.2004) – betrachtet man andere Schauspieler als „Zentrum“ des Graphen läßt sich dieser Wert noch steigern. Sean Connery liefert etwa eine durchschnittliche Connery-Zahl von 2,731. Insgesamt gibt es 1048 „bessere Zentren“ als Kevin Bacon.

³ In der verbeiteteren Variante ist es unwichtig, wieviele Autoren an der Veröffentlichung beteiligt waren, ein Artikel mit 6 Autoren erzeugt damit 15 Kanten im Graph. Alternativ können nur Veröffentlichungen mit genau zwei Autoren betrachtet werden, die dann entstehenden Erdős-Zahlen werden als „Erdős numbers of the second kind“ bezeichnet. [ENP]

Erdős-Zahl gibt dann für jeden Autor die Länge des kürzesten Pfades bis zu Paul Erdős an: Erdős selbst hat die Zahl 0, seine Koautoren die Erdős-Zahl 1. Gibt es für einen Autor keinen Pfad zu Erdős wird seine Erdős-Zahl auf unendlich gesetzt. Unter [ENP] findet sich das „Erdős Number Project“ mit weiteren Informationen; die Daten stammen aus verschiedenen Literaturdatenbanken wie dem DBLP.

Auffallend ist wieder die mit 4,65 recht niedrige durchschnittliche Erdős-Zahl; auch derzeitige deutsche Doktoranden haben häufig eine Erdős-Zahl im Bereich von 5 [Zie05, S. 189]. Die maximale endliche Erdős-Zahl, die bisher gefunden wurde, ist 13.

1.4 Weitere Beispiele

Neben diesen relativ bekannten Beispielen finden sich Netze mit ähnlichen Eigenschaften (geringe mittlere Distanz, hoher lokaler Clustering-Koeffizient) in vielen weiteren Gebieten. Dazu zählen verschiedene Energie- und Datennetze, etwa das amerikanische Stromnetz – Knoten sind hier die Umspannstationen, Kanten die Hochspannungsleitungen. Auch das Internet hat eine ähnliche Struktur, hierbei repräsentieren die Knoten die einzelnen Router.

Die einzelnen Dienste im Internet können ebenfalls als Graph untersucht werden, z. B. das WWW: Knoten sind einzelne HTML-Seiten, Kanten entstehen durch Verweise in den Dokumenten. Als Teilmenge aller Webseiten kann man beispielsweise die Wikipedia-Artikel untersuchen: für die englische Wikipedia gibt es unter dem Namen „Six degrees of Wikipedia“ [SDoW] die Möglichkeit, von einem beliebigen Artikel den kürzesten Pfad durch den Verlinkungsgraph zu einem anderen Artikel zu finden.⁴

An all diesen Netzen verwundert immer wieder die auffallend kurze Distanz zwischen zwei beliebigen Knoten. Im folgenden Kapitel werden nun verschiedene Verfahren vorgestellt, um solche Netze zu modellieren. Dies führt zu den Begriffen „Small World“ und schließlich zu den „skalenfremen Netzen“.

⁴ Da die Zitations- und Verlinkungseigenschaft nicht symmetrisch ist, werden zur Modellierung gerichtete Graphen eingesetzt.

2. Ältere Modelle

2.1 Zufallsgraphen

Erste Modellierungsansätze zur Beschreibung großer Netze basierten auf den von Erdős und Rényi erklärten Zufallsgraphen [BA99] – Graphen, bei denen die Knoten einer vorgegebenen N -elementigen Knotenmenge zufällig mit Kanten verknüpft werden. Die Wahrscheinlichkeit p , dass zwei Knoten mit einer Kante verbunden werden, ist dabei fest und unabhängig von anderen Kanten.

In einem so erzeugten Graph ist die mittlere Distanz zwischen zwei beliebigen Knoten gering. Lokales Clustering kommt bei Zufallsgraphen im Gegensatz zu natürlichen Netzen allerdings nicht vor.

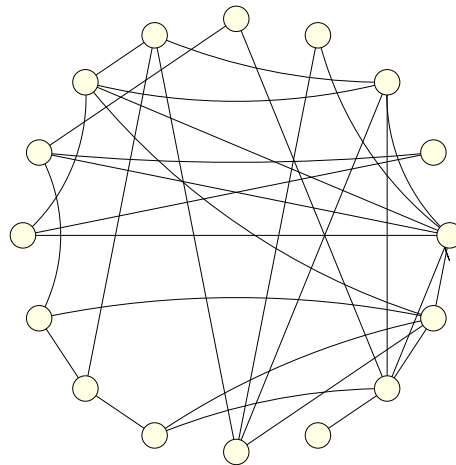


Abbildung 2.1. Zufallsgraph

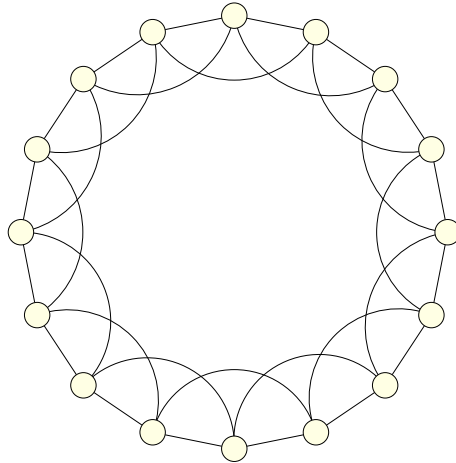


Abbildung 2.2. Regulärer Graph

2.2 Reguläre Graphen

Reguläre Graphen werden erzeugt, indem – ausgehend von einer auf einer Ebene angeordneten Knotenmenge – zwei Knoten genau dann mit einer Kante verbunden werden, wenn ihr Abstand in der Ebene kleiner ist als ein bestimmter fester Betrag. Anschaulich kann man etwa eine Knotenmenge gleichmäßig auf einem Kreis anordnen und jeden Knoten mit den beiden nächsten und übernächsten verbinden (siehe Abbildung 2.2).

Erkennen läßt sich hier ein hoher lokaler Clustering-Koeffizient, da die Nachbarn eines Knotens mit hoher Wahrscheinlichkeit untereinander verbunden sind. Die mittlere Distanz zwischen zwei beliebigen Knoten ist allerdings sehr hoch, im Beispiel mit kreisförmig angeordneten Knoten ist im schlimmsten Fall ein halber Durchlauf des Ringes nötig um einen Knoten zu erreichen [Zie05].

Diese beiden Modelle stellen jeweils nur einen Aspekt realer Netze dar: reguläre Graphen modellieren gut die lokalen Verbindungen (hoher Clustering-Koeffizient), während Zufallsgraphen mit ihrer geringen mittleren Distanz die globalen Verbindungen darstellen.

2.3 Small World

Eine Kombination dieser Modelle wurde 1998 von Duncan Watts und Steven Strogatz unter dem Namen „Small-World-Modell“ vorgestellt. Sie starten mit einem regulären Graph, trennen dann aber einige wenige Kanten auf und verbinden sie mit zufälligen anderen Knoten. Auf diese Weise entstehen Abkürzungen, die ein schnelles Durchschreiten des Graphen ermöglichen

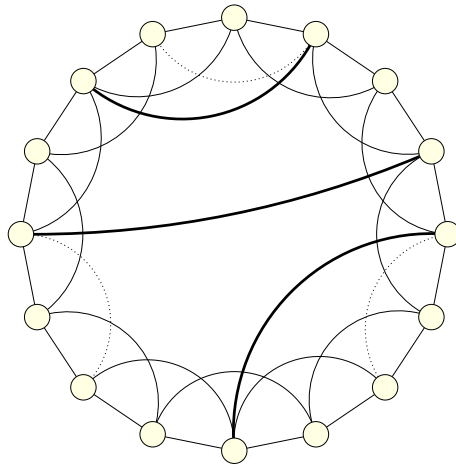


Abbildung 2.3. Small World

und für eine geringe mittlere Distanz sorgen. Da die meisten Kanten des ursprünglichen regulären Graphen nicht verändert werden, bleibt ein hohes lokales Clustering.¹

Abbildung 2.3 zeigt eine solche „Small World“: ausgehend von einem regulären Graphen wurden die gepunktet angedeuteten Kanten aufgetrennt und mit einem zufällig gewählten Knoten verbunden, so dass die hervorgehobenen Abkürzungen entstehen.

Mit der Arbeit von Watts und Strogatz lassen sich die in der Einleitung vorgestellten Phänomene zwar erklären, es zeigen sich aber dennoch einige Schwächen bei der Simulation realer Netze:

- Das Small-World-Modell geht – wie auch die anderen vorgestellten Modelle – von einem statischen Graphen aus. Reale Netze werden aber beständig durch neue Knoten ergänzt. Dieses Wachstum muss daher auch in der Simulation berücksichtigt werden.
- Da die Anzahl der Kanten pro Knoten im ganzen Graphen konstant ist, bezeichnet man die von diesem Modell erzeugten Graphen als egalitär – auch dies entspricht in den meisten Fällen nicht der Realität.²

¹ Diese Strategie lässt sich leicht an einem sozialen Netz nachvollziehen: die meisten Freunde befinden sich im lokalen Umfeld (und sind meist auch gegenseitig bekannt), jeder hat aber ein paar weiter entfernte Bekannte – sie bilden die Abkürzungen oder das „Backbone“ [Zie05].

² In sozialen Netzen gibt es stärker oder schwächer eingebundene Menschen, ein Wissenschaftler kann überdurchschnittlich viele Koautoren haben usw.

3. Skalenfreie Netze

3.1 Struktur

Bei ihrer Arbeit mit verschiedenen komplexen Netzen untersuchten die Physiker Albert-László Barabási und Réka Albert auch die Gradverteilung, also die Wahrscheinlichkeit dass ein Knoten mit einer bestimmten Anzahl von Kanten verbunden ist. Dabei fanden Sie allerdings nicht die bei Zufallsgraphen und Small Worlds entstehende Gauß'sche Glockenkurve, sondern eine Verteilung die einem Potenzgesetz folgt. [Zie05]

Die Wahrscheinlichkeit, dass ein Knoten k Kanten hat, verhält sich nach diesem Potenzgesetz wie folgt:

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

Der Parameter γ liegt bei realen Netzen meist zwischen 2 und 4 – genauere Werte werden im Abschnitt 3.3 aufgeführt.

Die Verteilung wird nach dem italienischen Ingenieur, Soziologen und Ökonomen Vilfredo Pareto „Pareto-Verteilung“ genannt.¹ Sie unterscheidet sich von der Gauß-Verteilung in erster Linie dadurch, dass auch Werte weit ab vom Durchschnitt vorkommen. Auf Graphen übertragen finden wir also auch einige Knoten, die überdurchschnittlich viele Verbindungen haben – weit mehr als bei einer Gauß-Verteilung möglich wären.

In der Literatur ist diese Verteilung meist unter der englischen Bezeichnung „Power Law“ zu finden.

Graphen, deren Verlinkungsgrad Pareto-verteilt ist, weisen also keine typische Anzahl von Verbindungen pro Knoten auf. Da ihr Verlinkungsgrad keiner Skala folgt, bezeichnet man sie als „skaleninvariant“ oder „skalenfrei“.

Die im vorigen Abschnitt vorgestellten Small Worlds können ebenfalls skalenfrei sein, Voraussetzung ist eine entsprechende Verteilung der Kanten. Umgekehrt haben auch viele skalenfreie Netze ähnliche Eigenschaften wie die Small Worlds – bei bestimmten Eigenschaften wie dem Clustering kann es allerdings zu Abweichungen zwischen den Modellen kommen.

¹ Pareto fand bei Untersuchungen zur Vermögensverteilung in Italien heraus, dass 80% des Vermögens bei 20% der Familien konzentriert war. Dieser Effekt, dass eine kleine Anzahl von hoch bewerteten Elementen sehr viel zum Gesamtwert einer Menge beitragen, ist auch unter dem Begriff „80:20-Regel“ bekannt. [Par]

3.2 Eigenschaften

Durch die Pareto-Verteilung entstehen bestimmte Knoten mit überdurchschnittlich vielen Verbindungen; sie werden als „Hubs“ bezeichnet. Diese Hubs fungieren als Backbone und erlauben kurze Wege zwischen zwei beliebigen Knoten.

Cliquenbildung ist keine direkte Eigenschaft der skalenfreien Netze, da im Modell keine Aussagen zum Clustering getroffen werden. Wie der im Abschnitt 3.4 beschriebene simple Algorithmus demonstriert, können auch Pareto-verteilte Graphen erzeugt werden, die über kein lokales Clustering verfügen.

Um reale Netze zu modellieren werden aber meist Graphen behandelt, die ein hohes lokales Clustering aufweisen. In diesen Netzen entstehen dann abgegrenzte Cliques, die durch einzelne Kanten mit den Hubs verbunden sind – diese Verbindungskanten entsprechen den neu verbundenen Kanten im Small-World-Modell [Zie05]. Die Verteilung folgt weiter dem obigen Potenzgesetz, so dass es sich auch mit Clustering um ein skalenfreies Netz handelt.

3.3 Beispiele

Barabási und Albert haben ihr Modell an einigen realen Netzen verifiziert und den jeweiligen Parameter γ bestimmt. Beispielsweise läßt sich das in Kapitel 1.2 auf Seite 4 vorgestellte Netz von Schauspielern („Kevin-Bacon-Orakel“) als skalenfreier Graph mit $\gamma = 2,3 \pm 0,1$ interpretieren.

Um die Topologie des WWW zu untersuchen wurde ein Robot entwickelt, der – ausgehend von einer Webseite – rekursiv allen Links folgt und die URL in eine Datenbank einträgt. Aus den so ermittelten Daten wurde die Wahrscheinlichkeit bestimmt, dass eine Webseite k ein- bzw. ausgehende Verweise hat; die entsprechenden Parameter für die Pareto-Verteilung wurden mit $\gamma_{in} = 2,1$ und $\gamma_{out} = 2,45$ berechnet. Zur Überprüfung wurde der Robot mit verschiedenen Webseiten gestartet, die Ergebnisse zeigten aber kaum Abweichungen. [BAJ00]

Selbst bei relativ kleinen Netzen mit knapp 5000 Knoten (elektrisches Hochspannungsnetz der westlichen USA) konnten Barabási und Albert ihr Modell anwenden und einen Parameter γ von etwa 4 berechnen. [BA99]

Auf Basis von Daten der amerikanischen Telefongesellschaft AT&T konnte ein Graph aufgestellt werden, der den Telefonverkehr eines Tages repräsentiert – jeder Knoten entspricht einer Person, die Kanten Anrufe. Unabhängig davon ob man ausgehende oder eingehende Gespräche betrachtet ergibt sich immer eine Pareto-Verteilung mit $\gamma \approx 2,1$ [ALPH01].

3.4 Erzeugung

Mit einem simplen Algorithmus ist es möglich, skalenfreie Netze zu generieren. Der Algorithmus startet mit einer kleinen Zahl m_0 von Knoten. In jedem Zeitabschnitt wird nun ein neuer Knoten hinzugefügt und durch m Kanten mit existierenden Knoten verbunden. Bei der Auswahl der existierenden Knoten werden die mit vielen Verbindungen bevorzugt; die Wahrscheinlichkeit ist proportional zur Anzahl von Kanten, die ein Knoten bereits besitzt.

Auf diese Weise entsteht ein skalenfreies Netz mit Parameter $\gamma = 2,9 \pm 0,1$. Dabei ist die skalenfreie Verteilung unabhängig von der Laufzeit des Algorithmus und damit der Größe des gesamten Graphen – er organisiert sich selbst trotz seines beständigen Wachstums in einem skalenfreien Zustand. [BA99]

Die Bevorzugung von Knoten mit vielen Verbindungen bei der Erzeugung von Kanten ist eine elementare Eigenschaft von skalenfreien Netzen, die man auch als „preferential attachment“ bezeichnet. Dies entspricht in vielen Fällen der Realität, etwa werden neue Webseiten bevorzugt Verweise auf sehr bekannte Seiten setzen, Autoren zitieren überwiegend weit verbreitete Werke, junge Schauspieler werden zunächst in Nebenrollen mit erfahrenen Kollegen auftreten usw. Allgemein entspricht dies dem sogenannten „Rich get richer“-Prinzip.

Um zu untersuchen, ob tatsächlich sowohl das Wachstum als auch die bevorzugte Knotenauswahl berücksichtigt werden müssen, um einen skalenfreien Graphen zu erzeugen, wurden von Barabási und Albert noch zwei alternative Modelle entwickelt. Das erste enthält das Anwachsen des Graphen, wählt für die Auswahl eines Knotens aber eine konstante Wahrscheinlichkeit, unabhängig von den bereits bestehenden Verbindungen der Knoten. Das zweite Modell startet mit N unverbundenen Knoten, von denen in jedem Schritt einer zufällig gewählt und mit einem anderen Knoten verbunden wird, die Wahrscheinlichkeit bei Wahl dieses zweiten Knotens ist dabei abhängig von der Anzahl der bereits bestehenden Kanten – im Vergleich zum ursprünglichen Modell ist also das Wachstum der Knotenmenge entfallen.

Aus der Tatsache dass beide Modelle zu einer Verteilung führen, die nicht skalenfrei ist,² läßt sich schließen dass ein Modell zur Erzeugung skalenfreier Netze Wachstum und preferential attachment beinhalten muss. [BA99, S. 6]

Der von Barabási beschriebene Algorithmus vermittelt zwar gut ein Verständnis vom Aufbau skalenfreier Netze, ist in seiner praktischen Anwendung aber eher beschränkt. Wie die Beispiele in Abschnitt 3.3 zeigen, variiert der Parameter der Pareto-Verteilung je nach Anwendung; obiger Algorithmus generiert aber immer Graphen mit $\gamma \approx 3$.

² Das zweite Modell entwickelt anfangs eine Pareto-Verteilung, führt aber bald zu einem vollständigen Graphen.

Abhilfe verschaffen universellere Algorithmen: so wurde von D. Volchenkov und Ph. Blanchard an der Universität Bielefeld ein Algorithmus entwickelt, der für $\gamma \in (1, \infty)$ skalenfreie Netze erzeugt. Zudem kann ein getrenntes γ_{in} und γ_{out} festgelegt werden, wie es z. B. bei Verlinkungs- und Zitationsgraphen notwendig ist. [VB05]

3.5 Deterministischer Algorithmus

Die meisten Algorithmen zur Generierung skalenfreier Netze arbeiten stochastisch, die Knoten werden also zufällig miteinander verbunden. Auch wenn der Zufall bei realen Netzen eine wesentliche Eigenschaft ist, haben deterministisch erzeugte Netze in bestimmten (theoretischen) Bereichen Vorteile.

In [BRV03] wird ein solcher Algorithmus vorgestellt. Er arbeitet wie folgt:

0. Gestartet wird mit einem einzelnen Knoten, der Wurzel.
1. Es werden zwei Knoten hinzugefügt, die beide mit der Wurzel verbunden werden.
2. Nun werden zwei Komponenten aus je drei Knoten eingefügt, so dass beide Komponenten identisch zu dem Graph aus dem vorherigen Schritt sind. Die vier unteren Knoten der beiden Komponenten werden mit der Wurzel verbunden.

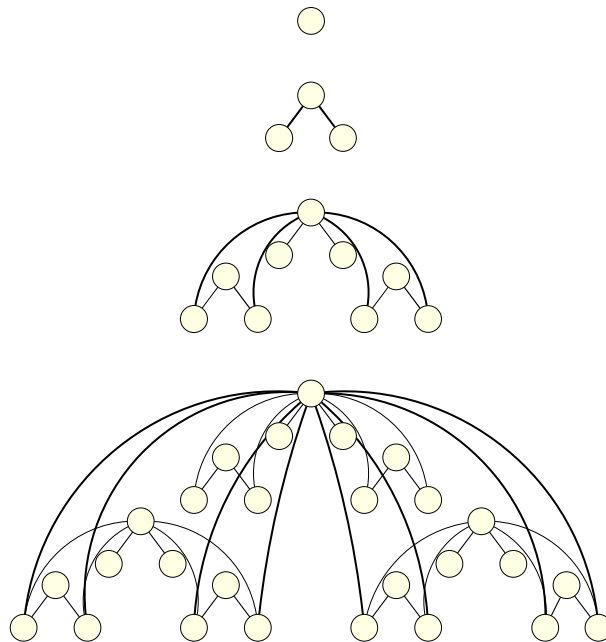


Abbildung 3.1. Nullter bis dritter Schritt des deterministischen Algorithmus zur Erzeugung skalenfreier Netze

3. Es werden zwei Komponenten mit je neun Knoten hinzugefügt, beide identisch zu dem Graph aus dem vorigen Schritt. Die acht unteren Knoten werden mit der Wurzel verbunden.

Allgemein führt der Algorithmus in Schritt n folgende Operationen aus:

- n . Es werden zwei Komponenten mit je 3^{n-1} Knoten eingefügt, die beide identisch zum Graph in Schritt $n - 1$ sind. Jeder der 2^n unteren Knoten wird mit der Wurzel verbunden.

Bildlich gesprochen wird also der existierende Graph in jeder Iteration zweimal kopiert, um anschließend die unteren Knoten der beiden Kopien mit der Wurzel zu verbinden.

Offensichtlich erzeugt der Algorithmus verschieden stark verbundenen Hubs, wobei die Wurzel durch Kanten zu $\frac{2}{3}$ aller Knoten der größte Hub ist. An Hand der Verbindungen der Hubs läßt sich zeigen, dass die Anzahl der Kanten pro Knoten Pareto-verteilt ist mit dem Parameter

$$\gamma = 1 + \frac{\ln 3}{\ln 2}.$$

Wenn die Anzahl der Knoten, die in jedem Schritt mit der Wurzel verbunden werden variiert wird, kann dieser Parameter einfach geändert werden.

Da in dem so erzeugten Graph keine Querverbindungen vorkommen, ist der Clustering-Koeffizient 0. Auch das läßt sich mit kleinen Anpassungen im Algorithmus abändern, etwa indem bei jeder Iteration die Hubs der beiden Komponenten mit einer Kante verbunden werden.

4. Anwendungen

Die Beschäftigung mit Netztopologien ist nicht nur von theoretischem Interesse; dieser Abschnitt soll exemplarisch einige praktische Anwendungen der Kenntnisse über skalenfreie Netze vorstellen.

4.1 Angriffssicherheit

Wie bereits erläutert sind viele Kommunikationsnetze skalenfrei, eine wichtige Fragestellung ist dabei die Robustheit dieser Netze. Unterschieden werden zwei Arten von Störungen, zum einen die zufälligen Ausfälle, zum anderen gezielte Attacken, die darauf ausgelegt sind die Kommunikation in einem möglichst großen Teil des Netzes zu unterbrechen.

Während bei Zufallsgraphen nur ein marginaler Unterschied zwischen diesen beiden Angriffsformen besteht, ist er bei skalenfreien Graphen erheblich ausgeprägter. Diese zeigen sich als „hochgradig robust gegen unkoordinierte Angriffe“ [Zie05, S. 192], werden doch in den meisten Fällen Knoten getroffen, deren Ausfall keine Folgen für die Kommunikation im Graphen hat.

Ganz anders sieht es hingegen bei gezielten Angriffen aus: kennt der Angreifer die Struktur des Netzes, konzentriert er seine Attacken auf die Hubs und kann mit geringem Aufwand das Netz stark schädigen.

In skalenfreien Datennetzen wie dem Internet wird versucht diese Gefährdung der Hubs durch eine hohe Redundanz auszugleichen, bisher konnte die damit erreichte Robustheit auch überzeugen:

„Als beispielsweise im Oktober 2005 der größte deutsche Internet-Austauschpunkt DE-CIX in Frankfurt teilweise ausfiel, nahm kaum jemand Notiz davon.“ [DP06, S. 161]

4.2 Virenausbreitung

Kenntnisse über die Netztopologie sind wichtig, um Voraussagen über die Ausbreitung von ansteckenden Krankheiten treffen zu können. Nur dann können auch geeignete Maßnahmen getroffen werden, um die Krankheit einzudämmen. Neben den biologischen Viren in sozialen Netzen (etwa AIDS)

findet die gleiche Theorie auch bei Computerviren in Kommunikationsnetzen Anwendung.

Zur Untersuchung der Virenausbreitung dient das „Susceptible-Infected-Susceptible“-Modell (SIS). In diesem Modell kann jeder Knoten entweder gesund oder infiziert sein, eine Ansteckung eines Knotens ist nur möglich wenn durch eine Kante eine direkte Verbindung zu einem infizierten Knoten besteht. In jedem Zeitabschnitt wird ein gesunder Knoten mit der Wahrscheinlichkeit ν infiziert, vorausgesetzt er ist mit mindestens einem kranken Knoten verbunden. In der gleichen Zeit wird ein infizierter Knoten mit der Wahrscheinlichkeit δ wieder gesund – als effektive Ausbreitungsgeschwindigkeit ergibt sich also $\lambda = \frac{\nu}{\delta}$. [DB03]

Bei Zufallsgraphen und regulären Graphen läßt sich ein Schwellwert λ_c mit folgender Bedeutung ermitteln: ist die effektive Ausbreitungsgeschwindigkeit kleiner als λ_c , wird der Virus bald ausgelöscht. Ist umgekehrt $\lambda > \lambda_c$ wird die Krankheit weiter bestehen.

Für skalenfrie Netze (mit $\gamma \leq 3$) wurde jedoch gezeigt, dass $\lambda_c = 0$ ist – selbst Viren mit einer sehr niedrigen Ausbreitungsgeschwindigkeit können also bestehen und sich weiter ausbreiten. Ursache dafür sind die Hubs, da sie zu sehr vielen Knoten Verbindung haben sind sie auch leicht zu infizieren und geben den Virus dann an viele Knoten weiter.

Die zur Eindämmung von Viren eingesetzten Gegenmittel (z. B. Antibiotika bei biologischen Viren) sind in vielen Fällen nur begrenzt verfügbar oder sehr teuer und sollen daher möglichst effektiv eingesetzt werden. In skalenfrie Netzen ist es dazu notwendig, die Hubs zu identifizieren: werden die Gegenmittel gezielt an den Hubs eingesetzt, wird der Schwellwert λ_c wieder größer als Null und eine Eindämmung des Virus ist damit möglich. [DB03]

Da im Moment die Hubs in den sozialen Netzen weitgehend unbekannt sind, muss ein großer Teil der Bevölkerung mit Gegenmitteln versorgt werden, um einen Virus einzudämmen. Beispielsweise müssen zur erfolgreichen Bekämpfung von Masern etwa 90% der Menschen geimpft werden; erst dann kann davon ausgegangen werden dass alle Hubs getroffen wurden [Zie05, S. 192].

4.3 Durchsuchbarkeit

Die Pareto-Verteilung ermöglicht unter bestimmten Voraussetzungen eine effiziente Suche im Graphen, ohne ihn komplett durchschreiten zu müssen. Als Beispiel sollen hier Peer-to-Peer-Netze dienen, die zum direkten Dateiaustausch zwischen Computern entworfen wurden – eine wesentliche Aufgabe ist es dabei, zu ermitteln, auf welchem Computer eine gewünschte Datei zu finden ist. Während zunächst (etwa bei der Musiktauschbörse Napster) zentrale Server Listen mit den Dateien auf jedem Computer führten, sind bei neueren Tauschbörsen wie Gnutella alle Teilnehmer gleichberechtigt – auf zentrale Komponenten wurde bewusst verzichtet.

Der klassische Suchalgorithmus von Gnutella nutzt die Skalenfreiheit des Netzes ($\gamma \approx 2$) allerdings nicht aus: ein Computer schickt die Suchanfrage an alle Nachbarn in einer bestimmten Umgebung. Kann einer dieser Nachbarn die gewünschte Datei anbieten, sendet er eine positive Antwort an den Anfragenden. Mit dieser Methode können Dateien schnell gefunden werden, doch die dazu eingesetzten Broadcasts beanspruchen eine recht hohe Bandbreite. Problematisch sind auch langsam angebundene Computer im Netz, die die eintreffenden Suchanfragen nicht schnell genug weiterreichen können.

In [ALPH01] wird ein alternativer Algorithmus vorgestellt, der die Pareto-Verteilung nutzt. Im Unterschied zum obigen Vorgehen wird die Suchanfrage immer nur an einen Nachbarn weitergereicht – bevorzugt werden dabei stärker verbundene Knoten. Voraussetzung ist, dass jeder Computer weitere Daten seiner Umgebung speichert, etwa Listen der Dateien, die seine Nachbarn anbieten. Durch verschiedene Zusatzbedingungen wird versucht, auch bei wenig lokalen Informationen einen möglichst kurzen Weg durch das Netz zu erreichen.

5. Ausblick

Seit den Forschungen der Arbeitsgruppe um Albert-László Barabási zu skalenfreien Netzen hat die Beschäftigung mit Netztopologien erheblich an Popularität gewonnen – die beeindruckende Zahl von Veröffentlichungen zeigt dies. Wissenschaftler aus den verschiedensten Bereichen konnten die Modelle der Small Worlds und Skalenfreien Netzen anwenden; die Bandbreite der untersuchten Gebiete reicht dabei von Unternehmensallianzen bis hin zu chemischen Interaktionen zwischen Enzymen und Molekülen in Zellen [Zie05].

Wurden früher bevorzugt Zufallsgraphen zur Modellierung großer Netze eingesetzt, so ermöglichen die skalenfreien Netze nun ein weit realistischeres Modell. Doch läßt sich auch dieses Modell natürlich noch weiter optimieren und von seiner relativ allgemeinen Beschreibung auf einige speziellere Anwendungen anpassen. Ein Ansatz zur Erweiterung ist etwa die Berücksichtigung des Alters der Knoten, denn in einigen realen Netzen läßt sich beobachten, dass mit zunehmendem Alter seltener Verbindungen zu einem Knoten aufgenommen werden.

Auch Nichtwissenschaftler kommen seit einiger Zeit vermehrt mit skalenfreien Netzen in Berührung, etwa durch Dienste, die ein soziales Netz im Internet abbilden. Nach Registrierung bei einem dieser Anbieter werden neben persönlichen Details auch die Bekanntschaftsbeziehungen zu anderen Mitgliedern eingegeben.¹ Zu jedem fremden Mitglied kann dann der Bekanntschaftsgraph angezeigt werden – eine deutliche Demonstration der niedrigen mittleren Distanz in skalenfreien Netzen. Während einige Dienste wie der Open Business Club openbc.com in erster Linie die Vermittlung von Geschäftskontakten zum Ziel haben, setzen andere den Fokus eher auf private Bekanntschaften, z. B. Google mit orkut.com.

Doch ungeachtet dieser praktischen Anwendungen üben die in der Einleitung vorgestellten Projekte wie die Erdős-Zahlen oder das Kevin-Bacon-Orakel² eine gewisse Faszination aus. Das jüngste Projekt dieser Art wurde kürzlich im Rahmen des Kulturprogramms zur FIFA Weltmeisterschaft 2006 vorgestellt: auf netz2006.de sollen alle Fußballspieler angeben, mit welchen Mitspielern sie schon gemeinsam auf dem Platz gestanden haben. So soll ein großes Netz aller Spieler entstehen – vermutlich skalenfrei.

¹ Bei Orkut können diese Kanten im sozialen Netz sogar gewichtet werden: anzugeben ist dazu wie „eng“ die Beziehung zu der anderen Person ist.

² Laut [Zie05] gehörte die Webseite zum Kevin-Bacon-Orakel [OoB] zeitweise zu den meistbesuchten Webseiten überhaupt.

Literatur

- [ALPH01] Adamic, Lada A.; Lukose, Rajan M.; Puniyani, Amit R.; Huberman, Bernardo A.: *Search in power-law networks*. HP Labs, Palo Alto; Department of Physics, Stanford University, 2001.
- [BA99] Barabási, Albert-László; Albert, Réka: *Emergence of Scaling in Random Networks*. Department of Physics, University of Notre-Dame, 1999.
- [BAJ00] Barabási, Albert-László; Albert, Réka; Jeong, Hawoong: *Scale-free characteristics of random networks: the topology of the world-wide web*. Department of Physics, University of Notre-Dame, 2000.
- [BRV03] Barabási, Albert-László; Ravasz, Erzsébet; Vicsek, Tamás: *Deterministic Scale-Free Networks*. Department of Physics, University of Notre-Dame, 2003.
- [DB03] Deszö, Zoltán; Barabási, Albert-László: *Halting viruses in scale-free networks*. Department of Physics, University of Notre-Dame, 2003.
- [DP06] Dierichs, Stefan, Pohlmann, Prof. Norbert: *Geordnetes Chaos. Wie Routing dem Internet seine Selbstheilungskräfte verleiht*. In: c't 03/2006, Heise Zeitschriften Verlag Hannover, 2006. S. 160-165.
- [Kle00] Kleinfeld, Judith: *Could It Be a Big World After All? What the Milgram Papers in the Yale Archives Reveal About the Original Small World Study*. University of Alaska Fairbanks, 2000.
http://www.columbia.edu/itc/sociology/watts/w3233/client_edit/big_world.html
- [Sch03] Schreiber, Cecilia: *Die Welt des Internet ist klein: 24163 Email-Ketten können nicht irren. Oder doch?* 04.09.2003.
<http://www.heise.de/tp/r4/artikel/15/15557/1.html>
- [VB05] Volchenkov, D.; Blanchard, Ph.: *An algorithm generating scale free graphs*. BiBoS, Universität Bielefeld, 2005.
- [Zie05] Ziegler, Dr. Cai: *Schöne kleine Welt. Vom Wesen natürlicher Vernetzung*. In: c't 24/2005, Heise Zeitschriften Verlag Hannover, 2005. S. 188-192.
- [Par] *Pareto-Verteilung*. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 10. Dezember 2005, 22:40 UTC.
<http://de.wikipedia.org/w/index.php?title=Pareto-Verteilung&oldid=11516717>
- [SDoW] *Six degrees of Wikipedia*.
<http://tools.wikimedia.de/sixdeg/index.jsp>
- [ENP] *The Erdős Number Project*.
<http://www.oakland.edu/enp/>
- [OoB] *The Oracle of Bacon at Virginia*. University of Virginia.
<http://www.cs.virginia.edu/oracle/>

Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Ausarbeitung selbständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Mir ist bekannt, dass ein Verstoß hiergegen als Betrugsversuch angesehen wird. Dies führt dazu, dass ich für diese Veranstaltung keinen Leistungsnachweis mehr erhalten kann.

Weiterhin bin ich mir darüber bewußt, dass dieser Betrugsversuch dem Hochschulprüfungsamt gemeldet werden kann, was zu einem entsprechenden Eintrag in meiner Studentenakte führt, oder, im Wiederholungsfall, auch zur Exmatrikulation führen kann.

Trier, den 06.03.2006